2019-07-01

Isolation, Characterization, and Genomic Comparison of Bacteriophages of Enterobacteriales Order

Ruchira Sharma
Brigham Young University

Isolation, Characterization, and Genomic Comparison of Bacteriophages of *Enterobacteriales* Order

Ruchira Sharma

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Julianne H. Grose, Chair
Bradley D. Geary
Steven M. Johnson
Perry G. Ridge
K. Scott Weber

Department of Microbiology and Molecular Biology

Brigham Young University

ABSTRACT

Isolation, Characterization and Genomic Comparison of Bacteriophages of *Enterobacteriales* Order

Ruchira Sharma
Department of Microbiology and Molecular Biology, BYU
Doctor of Philosophy

According to CDC, every year at least 2 million people are affected and 23,000 dies as a result of antibiotic resistance in U.S. It is considered one of the biggest threats to global health. More and more bacterial infections are becoming harder to treat. One such infection is fire blight, one of the most destructive disease of apple and pear trees. It is caused by bacteria *Erwinia amylovora* and its outbreaks have been known to destroy entire orchards in a single season. The conventional method of treatments includes use of antibiotics like streptomycin and oxytetracycline but the incidences like presence of multi-drug resistant bacteria in the mammals grazing in the fields have raised concerns. Phage therapy is considered one of the few ways available to combat bacterial resistance and prevent fire blight. In this method, a cocktail of highly lytic bacteriophages is prepared and sprayed on the trees at different time intervals. Bacteriophages are an "intelligent" drug. They multiply at the site of the infection until there are no more bacteria and then they are excreted back into the nature. These phenomena make them more efficient than an antibiotic, which kills all kind of bacteria including good bacteria and can be maintained in the environment for long periods of time. These qualities of bacteriophage have resulted in many commercially available phage therapies.

The initial part of this research focuses on isolation, characterization and genomic comparison of bacteriophages that infect a plant pathogen *E.amylovora* of *Erwiniaceae* family of *Enterobacteriales* order. In this study, 28 novel bacteriophages were isolated, fully sequenced, characterized and grouped into seven families based on phage homology. To take this further, we characterized a novel jumbo family of bacteriophages that has a small burst size of 4.6-4.9 and are most similar to bacteriophages that infect *Pseudomonas* and *Ralstonia* rather than *Enterobacteriales* bacteria by protein similarity. These bacteriophages are shown to infect *Erwinia* and *Pantoea* bacterial strains, but no infection of 9 other bacterial strains tested, was seen, under laboratory conditions. The results of this work provide an insight on special characteristics that makes bacteriophage so unique and adaptable.

The final part of this research explores the enormous diversity of bacteriophages. In 2014 Grose and Casjens grouped 337 fully sequenced tailed phages into 56 diverse clusters (32 lytic and 24 temperate). We further expanded our current understanding of these clusters by performing the comprehensive analysis of genomes and proteomes of 1037 tailed bacteriophages, posted on GenBank. The results of this work provide insights into diversity and relatedness of bacteriophages and the data is posted on GenBank.

Keywords: *E. amylovora*, bacteriophage, *Enterobacteriaceae,* phage clusters, fire blight, *Pantoea*, phage therapy

ACKNOWLEDGEMENTS

This work is dedicated to my husband Yomesh, and kids Avani and Aniruddh. It would have been impossible without their sacrifice and support. I am indebted to my parents, siblings and in-laws for their trust and unconditional love.

I would like to begin with thanking my pillar of support and phabulous mentor Dr. Julianne Grose. Her creative thinking, caring attitude and prompt replies to my dumb questions even at 2 am in the morning, made this a fruitful journey. I would also like to thank my committee Dr. Weber, Dr. Johnson, Dr. Geary and Dr. Ridge for their valuable advice and encouragement. I am grateful to the graduate committee and Dr. Griffitts for admitting me into the program at the last minute and for enhancing the quality of my research with his wonderful insight. A big thank you goes to our academic advisors Marcia and Kevin, business managers Amanda, Becky and Heidi and graduate secretaries for their hard work, time and support. I would also like to thank Sue Pratley for fulfilling my last-minute room-scheduling requests and for reviewing this dissertation.

I am lucky to have built everlasting friendships in and out of our lab and greatly appreciate their affection and much needed constructive criticism. Thank you, Jenny, KaiLi, Trever, Elise, Daniel T, Daniel A, Claudia, Deborah, Khin, Israel, Anna, Nidhi, Jeralyn, Evita, Desi, Ian, Jordan and Garrett for mentoring and guiding me and lending your shoulders to cry on. I would also like to thank all the phage hunters and high school interns, who I had the honor of training, for making me a better scientist.

I appreciate the patience and kindness of my friends outside of school and everyone who has helped in making this possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: Isolation, Characterization and Genomic Comparison of Bacteriophages of *Enterobacteriales* Order

1.1 Background

Bacteriophages were likely first reported in 1896, when Ernest Hanbury Hankin discovered antibacterial activity against cholera in the waters of two large rivers in India, the Ganges and Yamuna (1). In the early 1900s two scientists, a bacteriologist Frederick Twort (2, 3) and a microbiologist Félix d'Herelle (4), independently discovered "something" that infected and killed bacteria (5). They found that filtrate obtained from sewage could dissolve cultures of some intestinal bacteria, and theorized that this was due to lysing of the bacterial cells caused by a virus (later known as bacteriophages) (4).

1.2 Introduction to bacteriophages

Bacteriophages have a structure composed of a protein coat encapsulating DNA or RNA. They are obligate intracellular parasites meaning that they depend on a host for their replication, making nucleic acid and protein from host resources. During the infection process, bacteriophages can transfer foreign DNA to their host (including virulence factors), integrate into the host genome, and/or kill their host through cell lysis (6). The sheer number of bacteriophages combined with their clear evolutionary influence makes them an important target for understanding the ecology and evolution of bacteria, including pathogenic strains (7, 8). In addition, their specificity, genomic plasticity, and rapid multiplication rates make them a potential weapon to treat bacterial infections (9, 10). Bacteriophages are obligate intracellular parasites meaning that they depend on a host for their replication, making nucleic acid and protein from host resources. There are many bacteriophages and very little is known about them.

Their specificity, genomic plasticity, and rapid multiplication rates make them a suitable drug for curing bacterial infections. In Europe, they have been used for over 90 years as an alternative to antibiotics and are seen as a possible therapy against multi-drug-resistant strains of many bacteria (11-13).

1.3 Structure and morphology of Caudovirales bacteriophages

The Caudovirales or tailed bacteriophages (14) are divided into three families based on morphology: *Myoviridae* (with long contractile tail and sheath)(15), *Podoviridae* (with small non contractile tail) (16) and *Siphoviridae* (with long non contractile tail) (17).The bacteriophages can either be lytic or temperate (prophages) with visible differences in plaque morphology. Lytic bacteriophages make clear plaques and temperate makes plaque with bullseye or cloudy appearance.

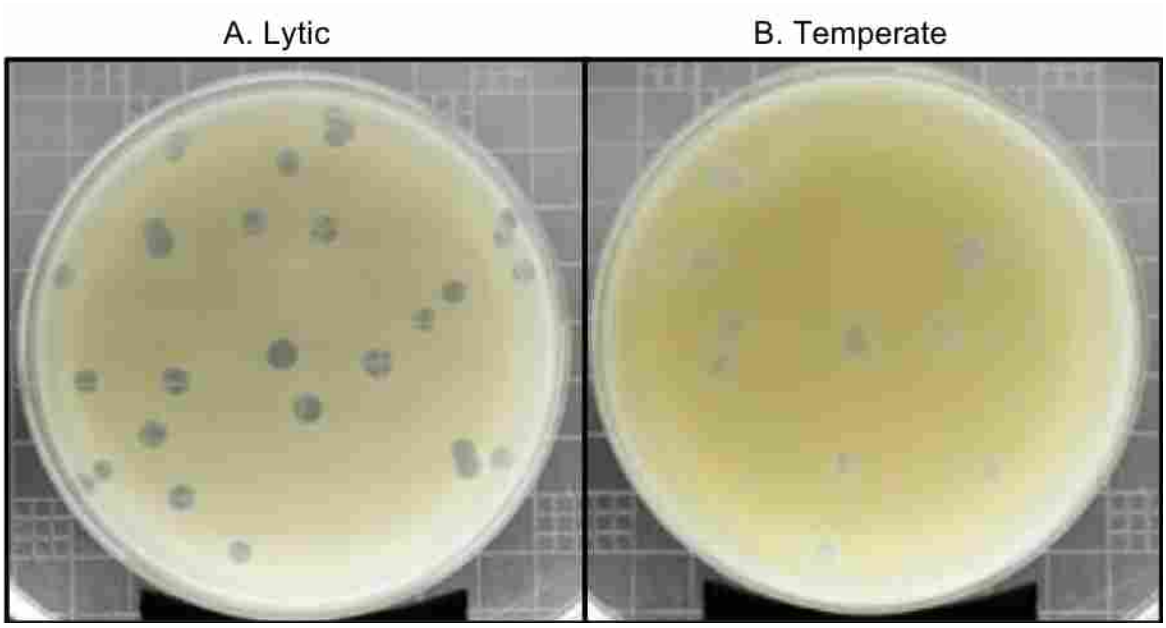


Figure 1.1 Two lifecycles of bacteriophages (A) lytic and (B) temperate distinguished on plate

1.4 The *Enterobacteriales* order

Enterobacteriales is an order of class *Gammaproteobacteria* and is home to eight families of bacteria including *Thorselliaceae* (18), and seven families proposed by Adelou *et.al.* (19) in 2017: *Budviciaceae*, *Enterobacteriaceae*, *Erwiniaceae*, *Moganellaceae*, *Pectobactericeae*, *Yersiniaceae*, and *Hafniaceae*. It also hosts some unclassified *Enterobacteriales*, and untested environmental sample. All bacterial species of *Enterobacteriales* order are Gram-negative, rod shaped, facultative anaerobic, non-spore forming bacteria (19). Many known animal and plant pathogens like *Salmonella*, *E.coli*, *Klebsiella*, *Serratia, Erwinia*, *Pantoea* etc are members of *Enterobacteriales* order and have shown increasing resistance to antibiotics (20-25). The immediate need of fighting antibiotic resistance have paved way to further investigations into phage therapy (26, 27).

1.5 *Erwinia amylovora*

Erwinia amylovora is a Gram negative, rod shaped bacterium of *Erwiniaceae* family of *Enterobacteriales* order. As a member of *Enterobacteriales*, *E. amylovora* is a close relative to opportunistic plant pathogens like *P. vagans* (28) and *P. agglomerans* (29) and other animal pathogens like *E. coli* (30) and *Salmonella* (20). The bacterium can range anywhere from 1.1 to 1.6 µm x 0.6 to 0.9 µm. It is a casual plant pathogen and causative agent of the contagious disease fire blight and is responsible for millions of dollars of loss in agriculture in US alone, annually (31). The conventional antibacterial treatment for fire blight is the use of copper or antibiotics like Streptomycin (32). Increase in number of reported cases of Streptomycin resistance and discolored fruits due to copper has drawn attention to research more effective

modes of treatment (33, 34). Chapter 3 in this study focuses on characterization and genomic comparison of a new genus (35) *Agrican357virus* of bacteriophages that infects *E. amylovora*

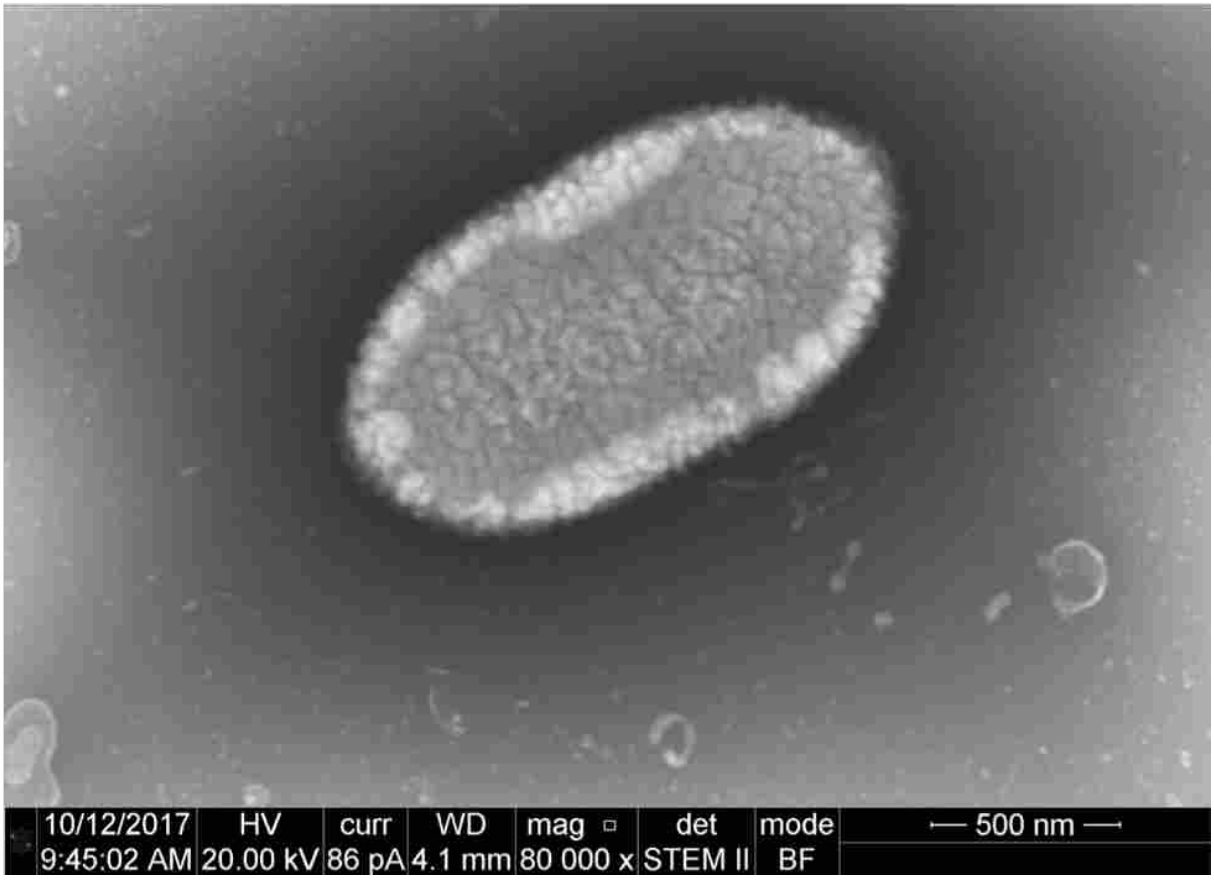


Figure 1.2 TEM image of *Erwinia amylovora*

1.6 Fire blight

Fire blight mainly affects pears, apples, and ornamental plants of the *Rosaceae* family. During cold season *E. amylovora* survives by residing in cankers. Owing to optimal moisture and temperature conditions, *E. amylovora* targets blossoms early spring and then travel through the entire tree infecting shoots, stem and roots. Progressive infection leads to the wilting, oozing and death of blossoms, shoots and branches. (36) Small droplets of sticky bacterial ooze (Figure

1.3A), blighted appearance and darkened wood (Figure 1.3B), as if scorched on fire, are characteristics of fire blight (31).

Figure 1.3 Fire blight infection as seen on (A) pear and (B) apple tree

The conventional treatment for fire blight includes use of antibiotics like streptomycin three times a year. This has led to increasing resistance to streptomycin (37). In 2013 Scherer *et.al.* (38) found multidrug-resistant bacteria in the nasal cavity and feces of sheep grazing through farms where streptomycin was sprayed for pant diseases. To avoid such conditions alternative methods have been researched to fight fire blight. One such method is using phage therapy to combat bacterial resistance and control of fire blight. Chapter 2 and Appendix I talks more about bacteriophages that infect *E. amylovora*, causative agent of fire blight.

1.7 The interplay between bacteria-bacteriophage

Bacteriophages have shown promise as an alternative treatment for bacterial infection where antibiotic resistance has become a major concern. (39). The bacteriophages attach themselves to bacteria, hack their machinery to make more bacteriophage progeny, and then exit by lysing the wall and killing bacteria. (40). To attach they look for receptors on bacterial surface which a bacterium can alter to become resistant to the bacteriophage. (41) To overcome this a cocktail of highly lytic bacteriophages is used in the preventative treatment that makes it very hard for bacteria to become resistant to. Prophages (temperate bacteriophage) on the other hand, can equip bacteria with necessary information and machinery through horizontal gene transfer or specialized transduction, to become resistant (42). Understanding this host-phage interplay is crucial to our study of bacteria and their evolution. By studying the bacteriophages, both lytic and temperate, we can enhance our understanding of their co-evolution. Chapter 4 and Appendix II in this study expands more on the enormous diversity of bacteriophages of *Enterobacteriales* order

1.8 Bacteriophage isolation and basic characterization

Bacteriophages used in chapter 2, 3 and Appendix I were isolated by collecting samples from infected trees like dead leaves, soil, twigs and ooze. In late spring to early summer, local farms were contacted for samples when bacteria are ready to spread their highly contagious infection to trees. These samples are then enriched in the lab by growing with *E. amylovora* bacteria and looked for bacteriophage through infection assays for the formation of a plaque, a place where a bacteriophage has landed and killed the bacteria. The bacteriophage is then isolated through subsequent rounds of plaque purification, the DNA is isolated and sequenced, and the phages are imaged using electron microscopy (EM). The sequence data is assembled, annotated and compared with other bacteriophage genomes. This characterization may identify genes that contribute to the evolution and virulence of *E. amylovora* strains.

1.9 Summary of research chapters

Chapter 1 begins with an introduction to the bacteriophages and dives into the importance and industrial application of research presented in following chapters.

Chapter 2 (43) is a published genome announcement of nine *E. amylovora* bacteriophages that fall into five distinct clusters based on genome similarity. Eight of these bacteriophages are *Myoviridae* whereas the ninth one is a *Podoviridae*. All nine of these bacteriophages were isolated on ATCC 29780 from various locations of Wasatch Front, Utah. In this announcement we studied the morphology, genomic similarity and mode of packaging of these bacteriophages. We found that based on genomic similarity these phages fall into five distinct clusters with vB_EamM_Bosolaphorus, vB_EamM_Desertfox, vB_EamM_MadMel and vB_EamM_Mortimer in *Agrican357virus*, vB_EamM_Asesino and vB_EamM_Wellington in

SPN3US/CR5 like, vB_EamM_Alexandra in Y2- like, vB_EamM_SunLIRen in PhiEa21-4 like and vB_EamPPavtok in Pep14 like cluster.

Chapter 3 is published research titled, "A novel, highly-related jumbo family of bacteriophages that were isolated against *Erwinia",*(44) and discusses in detail the characterization of a highly conserved family of eight jumbo bacteriophages which were recently added a genus *Agrican357virus* by ICTV along with another bacteriophage Ea35-70 found in Ontario, Canada. The bacteriophages of this family are myoviruses with genome size ranging from 271-275Kb. They are comparatively more similar to bacteriophages that infect *Pseudomonas* and *Ralstonia* than to other *Erwinia* which can be seen through their genomic and proteomic analysis. They have a broad host range, small burst size and harbor genes that may be helpful in the survival of this family in unfavorable conditions. In this chapter we discuss the unique characteristics of these jumbo bacteriophages.

In chapter 4 we shift our focus to a broader community of bacteriophages. Herein we performed the classification of more than 1000 lytic bacteriophages that infect bacterial hosts of *Enterobacteriales* order. The bacteriophages were put into clusters based on major capsid protein similarity and dot plot analysis. This research is a built upon previous research done by Grose and Casjens in 2014 (14) and follow methods designed by Dr. Hatfull and his coworkers (14, 45, 46). We were able to categorize these bacteriophages into 49 supercluster and 90 clusters (greater than 50% genomic and 33% proteomic similarity). This research contributes to our current understanding of bacteriophage relationships and confirms the previous analysis of grouping bacteriophages into clusters based on major capsid protein.

1.10 Summary of appendices

Appendix I contains a published Genome Announcement by the American Society of Microbiology titled, “Genome Sequences of 19 *Erwinia amylovora* bacteriophages (47). In this announcement we present the characterization and genomic comparison of 19 novel bacteriophages. Three podoviridae and 16 myoviridae phages were identified using TEM and were grouped into families based on their genomic similarity.

Appendix II contains a published manuscript on “Genomic comparison of 60 completely sequenced bacteriophages that infect *Erwinia* and/or *Pantoea* bacteria” (48). In this study we divided 60 bacteriophages from *Erwiniaceae* into 20 groups or clusters based on their nucleotide and protein homology.

Appendix III lists oral and poster presentations exhibited during this PhD.

REFERENCES

1. Abedon, S. T., C. Thomas-Abedon, A. Thomas, and H. Mazure. "Bacteriophage Prehistory: Is or Is Not Hankin, 1896, a Phage Reference?" [In eng]. *Bacteriophage* 1, no. 3 (May 2011): 174-78.
2. Twort, F. W. "Further Investigations on the Nature of Ultra-Microscopic Viruses and Their Cultivation." [In eng]. *J Hyg (Lond)* 36, no. 2 (Jun 1936): 204-35.
3. Twort, F.W., L.R.C.P. LOND., M.R.C.S. "An Investigation on the Nature of Ultra-Microscopic Viruses." *The Lancet* 186, no. 4814 (1915): 1241-43.
4. D'Herelle, F. "On an Invisible Microbe Antagonistic toward Dysenteric Bacilli: Brief Note by Mr. F. D'herelle, Presented by Mr. Roux. 1917." [In eng]. *Res Microbiol* 158, no. 7 (Sep 2007): 553-4.
5. Keen, E. C. "A Century of Phage Research: Bacteriophages and the Shaping of Modern Biology." [In eng]. *Bioessays* 37, no. 1 (Jan 2015): 6-9.
6. Chen, J., and R. P. Novick. "Phage-Mediated Intergeneric Transfer of Toxin Genes." [In eng]. *Science* 323, no. 5910 (Jan 2 2009): 139-41.
7. Boyd, E. F. "Bacteriophage-Encoded Bacterial Virulence Factors and Phage-Pathogenicity Island Interactions." [In eng]. *Adv Virus Res* 82 (2012): 91-118.
8. Bollback, J. P., and J. P. Huelsenbeck. "Parallel Genetic Evolution within and between Bacteriophage Species of Varying Degrees of Divergence." [In eng]. *Genetics* 181, no. 1 (Jan 2009): 225-34.
9. Shivaswamy, V. C., S. B. Kalasuramath, C. K. Sadanand, A. K. Basavaraju, V. Ginnavaram, S. Bille, S. S. Ukken, and U. N. Pushparaj. "Ability of Bacteriophage in Resolving Wound Infection Caused by Multidrug-Resistant Acinetobacter Baumannii in Uncontrolled Diabetic Rats." [In eng]. *Microb Drug Resist* 21, no. 2 (Apr 2015): 171-7.
10. Sulakvelidze, A. "Phage Therapy: An Attractive Option for Dealing with Antibiotic-Resistant Bacterial Infections." [In eng]. *Drug Discov Today* 10, no. 12 (Jun 15 2005): 807-9.
11. Zwet TVD, Keil HL. Fireblight: A Bacterial Disease of Rosaceous Plants. 1979.
12. Forster, H., G. C. McGhee, G. W. Sundin, and J. E. Adaskaveg. "Characterization of Streptomycin Resistance in Isolates of Erwinia Amylovora in California." [In eng]. *Phytopathology* 105, no. 10 (Oct 2015): 1302-10.
13. Tancos, K. A., S. Villani, S. Kuehne, E. Borejsza-Wysocka, D. Breth, J. Carol, H. S. Aldwinckle, and K. D. Cox. "Prevalence of Streptomycin-Resistant Erwinia Amylovora in New York Apple Orchards." [In eng]. *Plant Dis* 100, no. 4 (Apr 2016): 802-09.
14. Grose, J. H., and S. R. Casjens. "Understanding the Enormous Diversity of Bacteriophages: The Tailed Phages That Infect the Bacterial Family Enterobacteriaceae." [In eng]. *Virology* 468-470 (Nov 2014): 421-43.

15. Yap, M. L., and M. G. Rossmann. "Structure and Function of Bacteriophage T4." [In eng]. *Future Microbiol* 9, no. 12 (2014): 1319-27.

16. Aksyuk, A. A., V. D. Bowman, B. Kaufmann, C. Fields, T. Klose, H. A. Holdaway, V. A. Fischetti, and M. G. Rossmann. "Structural Investigations of a Podoviridae Streptococcus Phage C1, Implications for the Mechanism of Viral Entry." [In eng]. *Proc Natl Acad Sci U S A* 109, no. 35 (Aug 28 2012): 14001-6.

17. Friedman, D. I., and D. L. Court. "Bacteriophage Lambda: Alive and Well and Still Doing Its Thing." [In eng]. *Curr Opin Microbiol* 4, no. 2 (Apr 2001): 201-7.

18. Kampfer, P., S. P. Glaeser, L. K. Nilsson, T. Eberhard, S. Hakansson, L. Guy, S. Roos, H. J. Busse, and O. Terenius. "Proposal of Thorsellia Kenyensis Sp. Nov. And Thorsellia Kandunguensis Sp. Nov., Isolated from Larvae of Anopheles Arabiensis, as Members of the Family Thorselliaceae Fam. Nov." [In eng]. *Int J Syst Evol Microbiol* 65, no. Pt 2 (Feb 2015): 444-51.

19. Adeolu, M., S. Alnajar, S. Naushad, and S. Gupta R. "Genome-Based Phylogeny and Taxonomy of the 'Enterobacteriales': Proposal for Enterobacterales Ord. Nov. Divided into the Families Enterobacteriaceae, Erwiniaceae Fam. Nov., Pectobacteriaceae Fam. Nov., Yersiniaceae Fam. Nov., Hafniaceae Fam. Nov., Morganellaceae Fam. Nov., and Budviciaceae Fam. Nov." [In eng]. *Int J Syst Evol Microbiol* 66, no. 12 (Dec 2016): 5575-99.

20. D, V. T. Nair, K. Venkitanarayanan, and A. Kollanoor Johny. "Antibiotic-Resistant Salmonella in the Food Supply and the Potential Role of Antibiotic Alternatives for Control." [In eng]. *Foods* 7, no. 10 (Oct 11 2018).

21. Rasheed, M. U., N. Thajuddin, P. Ahamed, Z. Teklemariam, and K. Jamil. "Antimicrobial Drug Resistance in Strains of Escherichia Coli Isolated from Food Sources." [In eng]. *Rev Inst Med Trop Sao Paulo* 56, no. 4 (Jul-Aug 2014): 341-6.

22. Adamo, R., and I. Margarit. "Fighting Antibiotic-Resistant Klebsiella Pneumoniae with "Sweet" Immune Targets." [In eng]. *MBio* 9, no. 3 (May 15 2018).

23. Gajdacs, M., and E. Urban. "Resistance Trends and Epidemiology of Citrobacter-Enterobacter-Serratia in Urinary Tract Infections of Inpatients and Outpatients (Recesuti): A 10-Year Survey." [In eng]. *Medicina (Kaunas)* 55, no. 6 (Jun 18 2019).

24. Ge, Y., J. H. Lee, B. Hu, and Y. Zhao. "Loss-of-Function Mutations in the Dpp and Opp Permeases Render Erwinia Amylovora Resistant to Kasugamycin and Blasticidin S." [In eng]. *Mol Plant Microbe Interact* 31, no. 8 (Aug 2018): 823-32.

25. Mardaneh, J., and M. M. Dallal. "Isolation, Identification and Antimicrobial Susceptibility of Pantoea (Enterobacter) Agglomerans Isolated from Consumed Powdered Infant Formula Milk (Pif) in Nicu Ward: First Report from Iran." [In eng]. *Iran J Microbiol* 5, no. 3 (Sep 2013): 263-7.

26. Abedon, S. T., S. J. Kuhl, B. G. Blasdel, and E. M. Kutter. "Phage Treatment of Human Infections." [In eng]. *Bacteriophage* 1, no. 2 (Mar 2011): 66-85.

27. Alisky, J., K. Iczkowski, A. Rapoport, and N. Troitsky. "Bacteriophages Show Promise as Antimicrobial Agents." [In eng]. *J Infect* 36, no. 1 (Jan 1998): 5-15.

28. Smits, T. H., F. Rezzonico, T. Kamber, A. Goesmann, C. A. Ishimaru, V. O. Stockwell, J. E. Frey, and B. Duffy. "Genome Sequence of the Biocontrol Agent Pantoea Vagans Strain C9-1." [In eng]. *J Bacteriol* 192, no. 24 (Dec 2010): 6486-7.

29. Deletoile, A., D. Decre, S. Courant, V. Passet, J. Audo, P. Grimont, G. Arlet, and S. Brisse. "Phylogeny and Identification of Pantoea Species and Typing of Pantoea Agglomerans Strains by Multilocus Gene Sequencing." [In eng]. *J Clin Microbiol* 47, no. 2 (Feb 2009): 300-10.

30. Kolodner, R., S. D. Hall, and C. Luisi-DeLuca. "Homologous Pairing Proteins Encoded by the Escherichia Coli Rece and Rect Genes." [In eng]. *Mol Microbiol* 11, no. 1 (Jan 1994): 23-30.

31. Zwet TVD, Keil HL. Fireblight: A Bacterial Disease of Rosaceous Plants. 1979.

32. Forster, H., G. C. McGhee, G. W. Sundin, and J. E. Adaskaveg. "Characterization of Streptomycin Resistance in Isolates of Erwinia Amylovora in California." [In eng]. *Phytopathology* 105, no. 10 (Oct 2015): 1302-10.

33. Tancos, K. A., S. Villani, S. Kuehne, E. Borejsza-Wysocka, D. Breth, J. Carol, H. S. Aldwinckle, and K. D. Cox. "Prevalence of Streptomycin-Resistant Erwinia Amylovora in New York Apple Orchards." [In eng]. *Plant Dis* 100, no. 4 (Apr 2016): 802-09.

34. Tancos, K. A., and K. D. Cox. "Exploring Diversity and Origins of Streptomycin-Resistant Erwinia Amylovora Isolates in New York through Crispr Spacer Arrays." [In eng]. *Plant Dis* 100, no. 7 (Jul 2016): 1307-13.

35. Adams, M. J., E. J. Lefkowitz, A. M. Q. King, B. Harrach, R. L. Harrison, N. J. Knowles, A. M. Kropinski*, et al.* "Changes to Taxonomy and the International Code of Virus Classification and Nomenclature Ratified by the International Committee on Taxonomy of Viruses (2017)." [In eng]. *Arch Virol* 162, no. 8 (Aug 2017): 2505-38

36. SV, Thomson. *Epidemiology of Fire Blight.* Fire Blight; the Disease and Its Causative Agent, Erwinia Amylovora. Edited by Vanneste JL. Wallingford, United Kingdom: CABI Publishing, 2000.

37. Smits, T. H., V. M. Guerrero-Prieto, G. Hernandez-Escarcega, J. Blom, A. Goesmann, F. Rezzonico, B. Duffy, and V. O. Stockwell. "Whole-Genome Sequencing of Erwinia Amylovora Strains from Mexico Detects Single Nucleotide Polymorphisms in Rpsl Conferring Streptomycin Resistance and in the Avrrpt2 Effector Altering Host Interactions." [In eng]. *Genome Announc* 2, no. 1 (Jan 23 2014).

38. Scherer, A., H. R. Vogt, E. M. Vilei, J. Frey, and V. Perreten. "Enhanced Antibiotic Multi-Resistance in Nasal and Faecal Bacteria after Agricultural Use of Streptomycin." [In eng]. *Environ Microbiol* 15, no. 1 (Jan 2013): 297-304.

39. Svircev, A., D. Roach, and A. Castle. "Framing the Future with Bacteriophages in Agriculture." [In eng]. *Viruses* 10, no. 5 (Apr 25 2018).

40. Arens, D. K., T. S. Brady, J. L. Carter, J. A. Pape, D. M. Robinson, K. A. Russell, L. A. Staley*, et al.* "Characterization of Two Related Erwinia Myoviruses That Are Distant Relatives of the Phikz-Like Jumbo Phages." [In eng]. *PLoS One* 13, no. 7 (2018): e0200202.

41. Samson, J. E., A. H. Magadan, M. Sabri, and S. Moineau. "Revenge of the Phages: Defeating Bacterial Defences." [In eng]. *Nat Rev Microbiol* 11, no. 10 (Oct 2013): 675-87.

42. Penades, J. R., J. Chen, N. Quiles-Puchalt, N. Carpena, and R. P. Novick. "Bacteriophage-Mediated Spread of Bacterial Virulence Genes." [In eng]. *Curr Opin Microbiol* 23 (Feb 2015): 171-8.

43. Sharma R, Berg JA, Beatty NJ, Choi MC, Cowger AE, Cozzens BJR, Duncan SG, Fajardo CP, Ferguson HP, Galbraith T, Herring JA, Hoj TR, Durrant JL, Hyde JR, Jensen GL, Ke SY, Killpack S, Kruger JL, Lawrence EEK, Nwosu IO, Tam TC, Thompson DW, Tueller JA, Ward MEH, Webb CJ, Wood ME, Yeates EL, Baltrus DA, Breakwell DP, Hope S, Grose JH. "Genome Sequences of Nine Erwinia Amylovora Bacteriophages." *Microbiology Resource Announcements* 7 (October 11, 2018 2018).

44. Sharma, R, Pielstick, B.A, Bell, K.A.,Nieman, T.B., Stubbs, O.A., Yeates, E.L., Baltrus, D.A., Grose, J.H. " A Novel, Highly-related Jumbo Family of Bacteriophages that were Isolated against *Erwinia."*(In production), *Frontiers in Microbiology*, 2019 DOI: 10.3389/fmicb.2019.01533

45. Hatfull G. F. (2014). Molecular Genetics of Mycobacteriophages. *Microbiology spectrum*, *2*(2), 1-36.

46. Hatfull, G.F., Jacobs-Sera, D., Lawrence, J.G., Pope, W.H., Russell, D.A., Ko, C.C., Weber, R.J., Patel, M.C., Germane, K.L., Edgar, R.H., Hoyte, N.N., Bowman, C.A., Tantoco, A.T., Paladin, E.C., Myers, M.S., Smith, A.L., Grace, M.S., Pham, T.T., O'Brien, M.B., Vogelsberger, A.M., Hryckowian, A.J., Wynalek, J.L., Donis-Keller, H., Bogel, M. W., Peebles, C.L., Cresawn, S.G., Hendrix, R.W. (2010). Comparative genomic analysis of 60 Mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. *J. Mol. Biol*ogy, 397: 119–143.

47. Esplin, I. N. D., J. A. Berg, R. Sharma, R. C. Allen, D. K. Arens, C. R. Ashcroft, S. R. Bairett*, et al.* "Genome Sequences of 19 Novel Erwinia Amylovora Bacteriophages." [In eng]. *Genome Announc* 5, no. 46 (Nov 16 2017).

48. Thompson, D.T., Casjens, S.R., Sharma, R and Grose, J.H."Genomic comparison of 60 completely sequenced bacteriophages that infect *Erwinia* and/or *Pantoea* bacteria. [In Press], *Virology*, 2019.

CHAPTER 2: Genome Sequences of Nine *Erwinia amylovora* Bacteriophages

The following chapter is taken from an article submitted to Microbial Resource Announcements in American society of Microbiology Journal. Some unpublished sections, which were integral to obtaining results for this chapter were reinserted.

2.1 Abstract

Erwinia amylovora is a plant pathogen belonging to the *Enterobacteriaceae* family, a family containing many plant and animal pathogens. Herein, we announce nine genome sequences of *E. amylovora* bacteriophages isolated from infected apple trees along the Wasatch Front in Utah.

2.2 Introduction

At an estimated total number of 10^{31}, phages are by far the most abundant biological entity on the planet (1–7). They dramatically influence the evolution of bacteria by their ability to infect and kill their hosts and to transfer genetic material. *Erwinia amylovora* is a rod-shaped facultative anaerobic member of the Enterobacteriaceae bacterial family, which includes many well-characterized Gram-negative plant and animal pathogens, such as *Salmonella* spp., *Escherichia coli*, and *Klebsiella* spp. As the causative agent of fire blight, *Erwinia amylovora* infects members of the *Rosaceae* plant family, causing diseased areas to appear burnt (8–10). The isolation and characterization of phages that infect *E. amylovora* may aid in our understanding of these bacteria and provide potential treatment for this devastating agricultural disease.

2.3 Materials and methods

2.3.1 Bacteriophage isolation and genome sequencing

Environmental samples of leaves, branches and soil surrounding infected trees were collected from around the state of Utah, USA and used to create enrichment cultures with the host *E. amylovora*. To test for the presence of amplified bacteriophages, the enrichment cultures were spun at 4000 rpm and 4°C for 20 minutes and the supernatant was removed and used without filtering. 50µL of this supernatant was incubated at room temperature with 500µL of *E. amylovora* ATCC 29780 bacteria for 30-45 minutes, mixed with 5ml NBDYE top agar (at half concentration agar), plated on NBSYE agar plate, and incubated at 25°C overnight. Plaque presence on the plates was the primary indicator of bacteriophage presence in the environmental sample. Using a sterile needle or pipette tip, we picked a plaque from the initial bacteriophage identification plate and performed three rounds of plaque purification. All eight isolated bacteriophages were able to infect *E. amylovora,* ATCC 29780 (11). Bacteriophage DNA was extracted using the Phage DNA isolation kit (Norgen Biotek Corporation), and was sequenced, assembled and annotated as previously described (11).

2.3.2 Electron microscopy

Electron microscopy was performed at Brigham Young University in the Life Sciences Microscopy Lab using a FEI Helios NATOCAB 600i DualBeam FIB/SEM with STEM detector. The samples for SEM analysis were prepared by placing 15µL of high-titer bacteriophage lysate on a 200-mesh copper carbon type-B electron microscope grid for one-two minutes. The lysate was wicked away and the grids were stained for two minutes using 15µL of 2% phosphotungstic acid (pH = 7). Residual liquid was wicked away using Kimtech wipes and the grid was allowed

to dry before being imaged. Bacteriophage structures in electron micrographs were measured using ImageJ (12). The average and standard deviation for each measurement was calculated from a minimum of four separate measurements.

2.3.3 Computational analysis

These sequences were then used in Gepard (13) to generate the dot plots of nucleic acid. The Average Nucleotide Identity (ANI) percentages comparing each of the *E. amylovora* bacteriophage genomes were calculated using MAFFT (14) plugin in Geneious R8.1 (15)

2.4 Results and discussion

Herein, we announce the genome sequences of nine *E. amylovora* bacteriophages, vB_EamM_Asesino, vB_EamM_Alexandra, vB_EamM_Bosolaphorus, vB_EamM_Desertfox, vB_EamM_MadMel, vB_EamM_Mortimer, vB_EamP_Pavtok, vB_EamM_SunLIRen, and vB_EamM_Wellington. Phages were isolated from apple trees along the Wasatch Front in Utah that appeared to harbor fire blight infection. Phages were plaque purified through a minimum of three passages after amplification via enrichment culture (11). All nine phages reported here infect the *Erwinia amylovora* ATCC 29780 strain, as indicated by plaque assays, and their characteristics are summarized in Table 2.1. Genomic DNA was extracted (Phage DNA isolation kit; Norgen Biotek), a library was made using the Illumina TruSeq DNA Nano kit, and sample genomes were sequenced by Illumina HiSeq 2500 sequencing (250-bp paired end) and assembled with Geneious (15) version 8.1 using de novo assembly with medium-low sensitivity and various percentages of data. All phages circularized upon assembly and were annotated using DNA Master (http://cobamide2.bio.pitt.edu/computer.htm) , giving preference for calls that

Table 2.1 Properties of nine *Erwinia amylovora* bacteriophage genomes ORFs, open reading frames based on current annotation

Name	GenBank accession no.	SRA accession no.	Total no. of reads	No. of reads used	Assembly fold coverage (range [mean])	Length (bp)	No. of ORFs	No. of tRNAs	G+C content (%)
vB_EamP_Pavtok	MH426726	SRX4597602	1,301,332	386,192	492–2,086 [1,069]	61,401	62	0	36.9
vB_EamM_SunLIRen	MH426725	SRX4597606	1,301,332	386,192	8,249–42,422 [13,566]	84,559	141	22	36.3
vB_EamM_Wellington	MH426724	SRX4597603	626,048	372,488	133–514 [329.7]	244,950	295	8	50.3
vB_EamM_Asesino	KX397364	SRX4597609	2,222,038	1,022,382	512–1,378 [1,037.7]	246,290	289	12	51.2
vB_EamM_Alexandra	MH248138	SRX4597608	381,540	200,005	63–516 [166.3]	266,532	349	0	49.8
vB_EamM_Bosolaphorus	MG655267	SRX4597604	778,168	326,344	83–555 [248.4	272,228	321	1	49.4
vB_EamM_Desertfox	MG655268	SRX4597605	1,930,470	1,138,933	115–612 [352.9]	272,458	320	0	49.6
vB_EamM_Mortimer	MG655270	SRX4616109	2,581,160	287,396	47–207 [129.4]	273,914	325	1	49.5
vB_EamM_MadMel	MG655269	SRX4597607	1,604,720	1,443,568	567–1,577 [1,213.9]	275,000	321	0	49.4

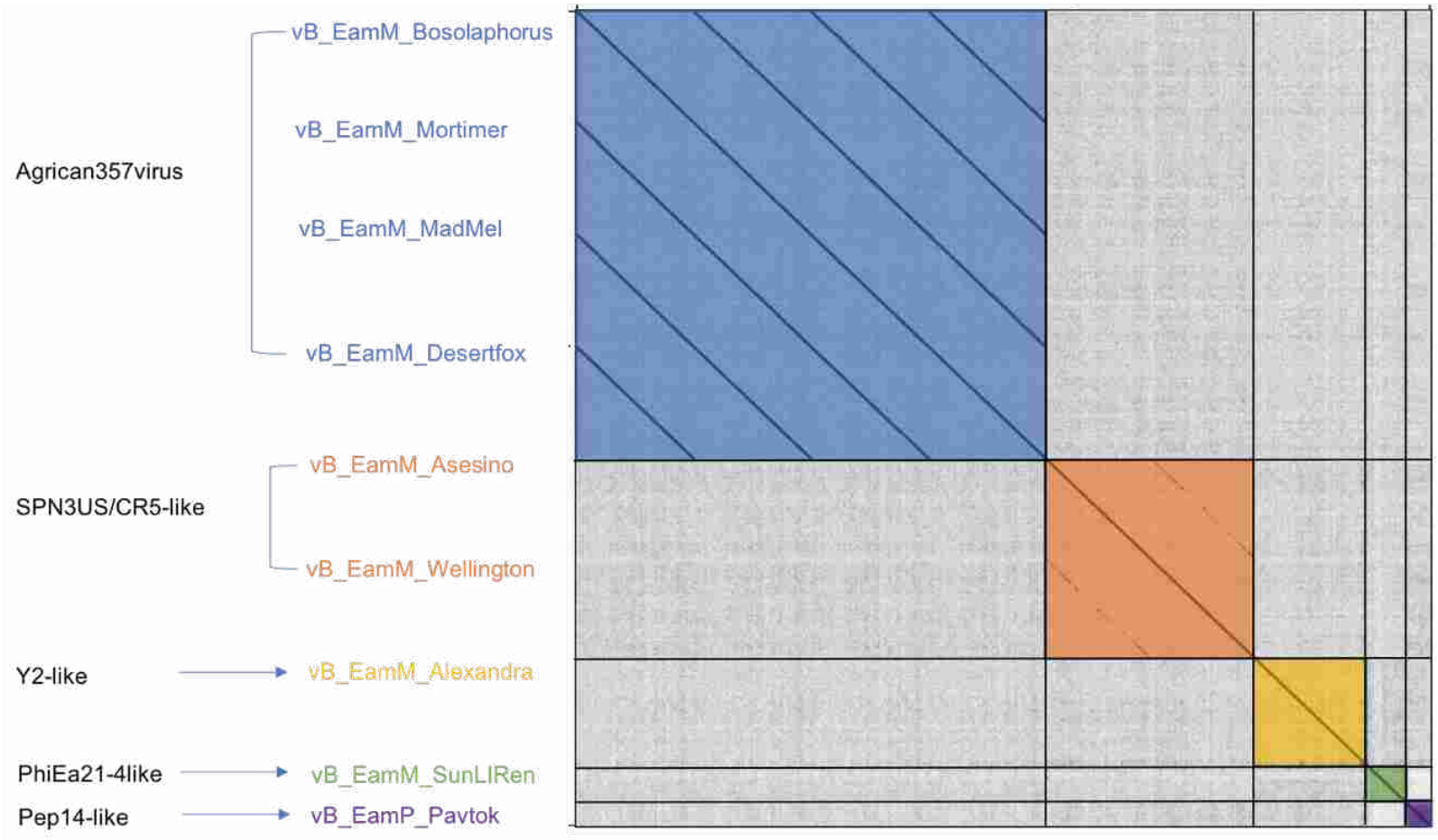


Figure 2.1 Whole genome dot plot of nine phages displays distinct clusters. Whole genome nucleotide dot plot constructed using Gepard was used to group phages in different clusters based on their nucleotide similarity. In Gepard dot plots greater than 50% similarity is represented with darker line at the word size of 10. Lack of line dark line is indicative of no significant similarity between phages of different clusters.

Table 2.2 Average Nucleotide Identity (ANI) of nine phages K-align was used to investigate the nucleotide identity of the phages. The phages with more than 50% average nucleotide similarity were grouped in same clusters. Less than 30% similarity is considered insignificant Shading indicates level of similarity from dark grey (>50% similar) to light grey.

	Bosolaphorus	Mortimer	MadMel	Desertfox	Wellington	Asesino	Alexandra	SunLIRen	Pavtok
Bosolaphorus	100	97.109	95.63	96.164	29.905	29.975	28.388	17.324	9.035
Mortimer	97.109	100	95.396	96.075	29.888	29.847	28.426	17.213	8.984
MadMel	95.63	95.396	100	96.325	29.773	29.743	28.291	17.143	8.943
Desertfox	96.164	96.075	96.325	100	30.009	29.998	28.552	17.298	9.014
Wellington	29.905	29.888	29.773	30.009	100	51.084	28.012	12.08	9.963
Asesino	29.975	29.847	29.743	29.998	51.08	100	28.407	12.007	10.062
Alexandra	28.388	28.426	28.291	28.552	28.012	28.407	100	10.682	7.872
SunLIRen	17.324	17.213	17.143	17.298	12.08	12.007	10.682	100	17.989
Pavtok	9.035	8.984	8.943	9.014	9.963	10.062	7.872	17.989	100

gave full coding potential coverage. The nine phages were grouped into five distinct clusters by genomic dot plot (Figure 2.1) and average nucleotide identity analyses (Table 2.2), as previously described (11, 14), with the first three groups containing jumbo *Myoviridae.* The first jumbo group included four myoviruses, vB_EamM_Bosolaphorus, vB_EamM_Desertfox, vB_EamM_MadMel, and vB_EamM_Mortimer, which are similar to previously published *Erwinia* phage Ea35-70 (17), as well as other phages we have isolated (11). The second group included two jumbo myoviruses, vB_EamM_Asesino and vB_EamM_Wellington, with similarity to the well characterized *Salmonella* SPN3US phage (18) and related phages. The third is a single jumbo myovirus, vB_EamM_Alexandra, which has similarity to previously published *Erwinia* phages vB_EamM_Yoloswag (11) and vB_EamM_Y3 (19). Podovirus vB_EamP_Pavtok and myovirus vB_EamM_SunLIRen are similar to *Erwinia* phages PEp14 and phiEa21-4 (20), respectively.

The three jumbo myovirus groups package DNA by headful packaging based on homology to phage phiKZ terminase (21), and their bp 1 was chosen by alignment to their phage family. PhageTerm (22) was used to determine the packaging strategy of SunLIRen and Pavtok. SunLIRen appeared to have headful packaging, and its bp 1 was assigned based on homology alignment to *Erwinia* phage phiEa21-4, while the packaging strategy of Pavtok is unknown, and its bp 1 was assigned due to homology to PEp14.

2.5 Acknowledgements

We thank the Howard Hughes Medical Institute Science Education Alliance–Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) for phage analysis training. In addition, we thank Ed Wilcox (BYU DNA Sequencing Center) and Michael Standing

(BYU Microscopy Lab). This work was graciously funded by a USDA grant (to D.A.B., University of Arizona) and the Department of Microbiology and Molecular Biology and the College of Life Sciences at Brigham Young University, as well as a private donor.

REFERENCES

1. Bergh DA, Borsheim KY, Bratbak G, Heldal M. 1989. High abundance of viruses found in aquatic environments. Nature 340, 467-468.
2. Woomack KE, Colwell RR. 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. 64, 69-114.
3. Brussow H, Hendrix, RW. 2002. Phage genomics: small is beautiful. Cell 108, 13-16.
4. Wilhelm SW, Jeffrey WH, Suttle CA, Mitchell DL. 2002. Estimation of biologically damaging UV levels in marine surface waters with DNA and viral dosimeters. Photochem Photobiol. 76(3), 268-73.
5. Hendrix RW. 2003. Bacteriophage genomics. Curr. Opin. Microbiol. 6, 506-511.
6. Hambly E, Suttle CA. 2005. The viriosphere, diversity, and genetic exchange within phage communities. Curr. Opin. Microbiol. 8, 444-450.
7. Suttle CA. 2005 Viruses in the sea. Nature 437, 356-361.
8. Khan MA, Zhao Y, Korban SS. 2012. Molecular Mechanisms of Pathogenesis and Resistance to the Bacterial Pathogen *Erwinia amylovora*, Causal Agent of Fire Blight Disease in Rosaceae. Plant Molecular Biology Reporter 30:247-260.
9. MA K, Y Z, S K. 2012. Molecular Mechanisms of Pathogenesis and Resistance to the Bacterial Pathogen *Erwinia amylovora,* Causal Agent of Fire Blight Disease in Rosaceae. Plant Molecular Biology Reporter 30:247-260.
10. Schroth M, Thomson S, Hildebrand D, Moller W. 1974. Epidemiology and control of fire blight. Annual Review of Phytopathology 12:389-412.
11. Esplin IND, et al. 2017. Genome Sequences of 19 Novel *Erwinia amylovora* Bacteriophages. Genome Announc. 5(46).
12. Abramoff, M.D., P.J. Magalhaes, and S.J. Ram. "Image Processing with Imagej." no. Biophotonics International (2004).
13. Krumsiek, J., R. Arnold, and T. Rattei. "Gepard: A Rapid and Sensitive Tool for Creating Dotplots on Genome Scale." [In eng]. *Bioinformatics* 23, no. 8 (Apr 15 2007): 1026-8.
14. Katoh, K., K. Misawa, K. Kuma, and T. Miyata. "Mafft: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform." [In eng]. *Nucleic Acids Res* 30, no. 14 (Jul 15 2002): 3059-66.
15. Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton*, et al.* "Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data." [In eng]. *Bioinformatics* 28, no. 12 (Jun 15 2012): 1647-9.

16. Thomson SV. 2000. Epidemiology of Fire Blight 2, p 9-36. Fire Blight: the disease and its causative agent, *Erwinia amylovora*. CABI Publishing, Wallingford, United Kingdom.

17. Yagubi AI, Castle AJ, Kropinski AM, Banks TW, Svircev AM. 2014. Complete Genome Sequence of *Erwinia amylovora* Bacteriophage vB_EamM_Ea35-70. Genome Announc 2(4):e00413-14.

18. Lee JH, Shin H, Kim H, Ryu S. 2011. Complete genome sequence of *Salmonella* bacteriophage SPN3US. J Virol 85:13470-1.

19. Buttimer C, Born Y, Lucid A, Loessner MJ, Fieseler L, Coffey A. 2018. *Erwnia* amylovora phage vB_EamM_Y3 represents another lineage of hairy Myoviridae. S0923-2508(18)30062-7.

20. Lehman SM, Kropinski AM, Castle AJ, Svircev AM. Complete genome of the broad-host-range *Erwinia amylovora* phage phiEa21-4 and its relationship to *Salmonella* phage felix O1. Applied and environmental microbiology. 2009;75(7):2139-47.

21. Mesyanzhinov VV, Robben J, Grymonprez B, Kostyuchenko VA, Bourkaltseva MV, Sykilinda NN, Krylov VN, Volckaer G. 2002. The genome of bacteriophage phiKZ of *Pseudomonas aeruginosa*. J Mol Biol 317:1-19.

22. Garneau JR, Depardieu F, Fortier LC, Bikard D, Monot M. PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. Sci Rep. 7(1), 8292.

CHAPTER 3: A Novel, Highly Related Jumbo Family of Bacteriophages That Were Isolated Against *Erwinia*

The following chapter is taken from an article published in Frontiers in Microbiology Journal. All content and figures have been formatted for this dissertation, but it is otherwise unchanged.

3.1 Abstract

Erwinia amylovora is a plant pathogen from the *Erwiniaceae* family and a causative agent of the devastating agricultural disease fire blight. Here we characterize eight lytic bacteriophages of *E. amylovora* that we isolated from the Wasatch front (Utah, USA) that are highly similar to vB_EamM_Ea35-70 which was isolated in Ontario, Canada. With the genome size ranging from 271-275 kb, this is a novel jumbo family of bacteriophages. These jumbo bacteriophages were further characterized through genomic and proteomic comparison, mass spectrometry, host range and burst size. Their proteomes are highly unstudied, with over 200 putative proteins with no known homologs. The production of 27 of these putative proteins was confirmed by mass spectrometry analysis. These bacteriophages appear to be most similar to bacteriophages that infect *Pseudomonas* and *Ralstonia* rather than *Enterobacteriales* bacteria by protein similarity, however we were only able to detect infection of *Erwinia* and the closely related strains of *Pantoea*.

3.2 Introduction

In 1998, Whitman *et al*.(1) estimated that there are approximately 5 x10^{30} bacteria on earth, which is more than the number of plants and animals combined. Most, or likely all, bacteria are subject to infection by one or more viruses or "bacteriophages", making bacteriophages the most common and diverse biological entity at an estimated 10^{32} (2-4). Bacteriophages were likely first reported in 1896, when Ernest Hanbury Hankin discovered

antibacterial activity against cholera in the waters of two large rivers in India, the Ganges and Yamuna (5). They were independently characterized and named in the 1900s by bacteriologist Frederick Twort (6,7) and microbiologist Félix d'Herelle (8,9). During the infection process, bacteriophages can transfer foreign DNA to their host (including virulence factors), integrate into the host genome, and/or kill their host through cell lysis (10). The sheer number of bacteriophages combined with their clear evolutionary influence makes them an important target for understanding the ecology and evolution of bacteria, including pathogenic strains (11,12). In addition, their specificity, genomic plasticity, and rapid multiplication rates make them a potential weapon to treat bacterial infections (13, 14).

One such bacterial infection caused by a phytopathogen *Erwinia amylovora* (15) is called fire blight that mainly affects ornamental plants of the *Rosaceae* family. The symptoms of the infected tissues include wilting, ooze production and death of blossoms, shoots branches and entire trees (16). We have recently isolated and characterized twenty-eight bacteriophages that infect *E. amylovora* (17, 18). Out of these 28*,* there is a distinct group of eight highly related bacteriophages: vB_EamM_Special G (Special G), vB_EamM_Simmy50 (Simmy50), vB_EamM_RAY (RAY), vB_EamM_Deimos-Minion (Deimos-Minion or DM), vB_EamM_Bosolaphorus (Bosolaphorus), vB_EamM_Desertfox (Desertfox), vB_EamM_MadMel (MadMel) and vB_EamM_Mortimer (Mortimer) very similar to *Erwinia* bacteriophage Ea35-70 which was isolated in Ontario, Canada (19). These nine bacteriophages were recently added as the *Agrican357virus* genus of bacteriophages by the ICTV (20) and are considered jumbo bacteriophages due to their large genome (>200 kb) and particle size (21).

As reviewed in 2017, jumbo bacteriophages have diverse genome sizes (ranging from 208-497 kb) as well as diverse virion morphology and complex virion structure (21). They often encode greater than 60 structural proteins with some displaying complex head structures composed of more than five proteins (22) or long, wavy, curly tail fibers (23). Jumbo bacteriophages were also found to be highly diverse, with over 11 clusters and five singleton bacteriophages suggested from 52 complete jumbo bacteriophage genomes analyzed in 2017, many of which are uncharacterized (21). Only a few jumbo bacteriophage families have been characterized beyond sequence analysis and EM, including the phiKZ-like bacteriophages 201phi2-1 (24), KTN4 (25), phiPA3 (26), phiRSL2 (27), phiRSF1(27), OBP (28), EL (29) and phiKZ (30), related bacteriophages phiRSL1 (31) and PaBG (32), C*ronobacter* bacteriophage CR5 (33), *Prochlorococcus* bacteriophage P-SSM2 (34), related bacteriophages KVP40 (35) and Aeh1 (36), *Aeromonas* bacteriophage phiAS5 (37), *Pectobacterium* bacteriophage CBB (38), *Caulobacter* bacteriophage phiCbK (39), related *Erwinia* bacteriophages Joad and RisingSun (40), related bacteriophage RaK2 (41) and GAP32 (42), *Bacillus* bacteriophage 0305phi8-36 (43), related B*acillus* bacteriophages BpSp (23) and AR9 (44). Herein we further analyze the genome, proteome, and host range of our eight *Agrican357virus* jumbo bacteriophages. Their lytic nature and plethora of novel genes makes them a unique entity to be studied further and analyzed. As a close relative of the animal pathogens *Escherichia coli* and *Salmonella* (45), viruses that infect *E. amylovora* may help us understand the evolution of pathogenic strains in this family.

3.3 Materials and methods

3.3.1 Bacteriophage isolation, electron microscopy and genome sequencing

Environmental samples of leaves, branches and soil surrounding infected trees were collected from around the state of Utah (USA) and used to create enrichment cultures with the host *E. amylovora*. To test the presence of amplified bacteriophages, the enrichment cultures were spun at 4000 rpm and 4°C for 20 minutes and the supernatant was removed and used without filtering. 50µL of this supernatant was incubated at room temperature with 500µL of *E. amylovora* ATCC 29780 bacteria for 30-45 minutes, mixed with 5ml NBDYE top agar (at half concentration agar), plated on NBSYE agar plate, and incubated at 25°C overnight. Plaque presence on the plates was the primary indicator of bacteriophage presence in the environmental sample. Using a sterile needle or pipette tip, we picked a plaque from the initial bacteriophage identification plate and performed three rounds of plaque purification. All eight isolated bacteriophages: Special G (KU886222), Simmy50 (KU886223), RAY (KU886224), Deimos-Minion (KU886225), Bosolaphorus (MG655267), Desertfox (MG655268), MadMel (MG655269) and Mortimer (MG655270) were able to infect *E. amylovora* ATCC 29780 (17, 18). Bacteriophage DNA was extracted using the Phage DNA isolation kit (Norgen Biotek Corporation), and was sequenced, assembled and annotated as previously described (17, 18).

3.3.2 Electron microscopy

Electron microscopy was performed at Brigham Young University in the Life Sciences Microscopy Lab using a FEI Helios NATOCAB 600i DualBeam FIB/SEM with STEM detector. The samples for SEM analysis were prepared by placing 15µL of high-titer bacteriophage lysate on a 200-mesh copper carbon type-B electron microscope grid for one-two minutes. The lysate

was wicked away and the grids were stained for two minutes using 15µL of 2% phosphotungstic acid (pH = 7). Residual liquid was wicked away using Kimtech wipes and the grid was allowed to dry before being imaged. Bacteriophage structures in electron micrographs were measured using ImageJ (46). The average and standard deviation for each measurement was calculated from a minimum of four separate measurements.

3.3.3 Burst size

Burst size was calculated by performing single-infection assay as described by M. Delbruck (47). The bacteria-bacteriophage mixture was allowed to adsorb for 10 minutes at a multiplicity of infection (MOI) of 100. The lysate was then removed at different time-intervals ranging from 1-6 hours and diluted to avoid secondary infection. Soft agar plaque method was used to determine titers and a graph of 10 separate readings was plotted with their average titers and time.

3.3.4 Host range

Host range of all eight bacteriophages was determined using the soft agar plaque method (48). For this, 50µL of bacteriophage lysate dilutions were incubated with 500µL of bacteria grown overnight for 30 min before plating in top agar. The plates were incubated with the top agar facing up at 25°C overnight for this assay. Seventeen bacterial strains including *E. amylovora* ATCC 29780 (49) as control were used including five other *E. amylovora* strains (Ea110 (49), GH9 (50), EaBH (50), RB02 (50), Ea273 (51)), *Pantoea agglomerans* E325 (52), *Pantoea vagans* C-91 (53, 54) , *E. coli* K-12 BW 25113 (54), *Salmonella enterica* LT2 (generously donated by John Roth lab), *Yersinia pestis* KIM6 (56, 57), *Enterobacter cloacae* ATCC 13047 (58), *Klebsiella pneumoniae* ATCC 10031 (59), *Bacillus subtilis* ATCC 6033 (60),

Cronobacter sakazakii ATCC 29544 (61, 62), standard clinical isolate *Pseudomonas aeruginosa* PA100 (63) and *Pseudomonas chlororaphis* ATCC 13985 (64). An average of two readings was taken to obtain bacteriophage titers post infection.

3.3.5 Computational analysis and genomic comparison

Bacteriophages with any similarities to *Agrican357virus* genus were identified using a blastx analysis of their putative major capsid and terminase proteins, and the corresponding bacteriophage for all retrieved hits with a cutoff e-value of less than 1.00E-04 and 33% similarity were downloaded from GenBank (65-67). In addition, any bacteriophages that showed up in at least three qblast hits while annotating were also retrieved. These sequences were then used in Gepard (68) to generate the dot plots of nucleic acid and protein sequences. PhamDB, a web interface (69) was used for creating databases and Phamerator, (70) an open-source program was used to compare bacteriophage genes and genomes. PhamDB uses kClust (71) to cluster large protein sequence databases. The default settings of PhamDB were used in this comparison.of Splitstree (72) protein analysis was produced from the exported pham table of conserved proteins converted to a Nexus file using Janus (http://cobamide2.bio.pitt.edu). The Average Nucleotide Identity (ANI) percentages comparing each of the *E. amylovora* bacteriophage genomes were calculated using MAFFT (73) plugin in Geneious R8.1 (74). The genome sequences of all eight bacteriophages were compared against one phage from each potential cluster formed in whole genome dot plot analysis.

The evolutionary history was inferred by using the Maximum Likelihood method and Poisson correction model (75). The bootstrap consensus tree inferred from 100 replicates (76) is taken to represent the evolutionary history of the taxa analyzed (76). Branches corresponding to

partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches (76). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. This analysis involved 59 amino acid sequences. There were a total of 1302 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (77).

3.3.6 Mass spectrometry

Sample preparation was performed (78) by diluting crude lysates of RAY and Deimos-Minion in TNE (50mM Tris pH 8.0, 100mM NaCl, 1mM EDTA) buffer and adding RapiGest SF reagent (Waters Corp.) to a final concentration of 0.1%. Samples were then boiled for 5 min followed by addition of 1mM (final concentration) of TCEP (Tris (2-carboxyethyl) phosphine) and incubated at 37°C for 30 min. Afterwards, carboxymethylation of samples was done with 0.5 mg/ml of iodoacetamide for 30 min at 37°C followed by neutralization with 2mM TCEP (final concentration). Trypsin (trypsin: protein ratio - 1:50) was used overnight at 37°C to digest the crude lysates prepared as above. The samples were treated with 250mM HCl at 37°C for 1h followed by centrifugation at 14000 rpm for 30 min at 4°C to degrade and remove RapiGest. The soluble fraction was then added to a new tube and Aspire RP30 desalting columns (Thermofisher Scientific) were used for extraction and desalting of the peptides.

High pressure liquid chromatography (HPLC) coupled with tandem mass spectroscopy (LC-MS/MS) using nano-spray ionization was used to analyze Trypsin-digested peptides (79). A TripleT of 5600 hybrid mass spectrometer (ABSCIEX) interfaced with nano-scale reversed-

phase HPLC (Tempo) using a 10 cm-100-micron ID glass capillary packed with 5-µm C18 Zorbax™ beads (Agilent Technologies, Santa Clara, CA) was used to perform the nano-spray ionization experiments. By using a linear gradient (5–60%) of ACN (Acetonitrile) at a flow rate of 250µl/min for 1h, peptides were eluted from the C18 column into the mass spectrometer The ACN gradient was created using these buffers: buffer A (98% H_2O, 2% ACN, 0.2% formic acid, and 0.005% TFA) and buffer B (100% ACN, 0.2% formic acid, and 0.005% TFA). In a data-dependent manner MS/MS data were acquired in which the MS1 data was acquired for 250 ms at m/z of 400 to 1250 Da and the MS/MS data was acquired from m/z of 50 to 2,000 Da. For Independent data acquisition (IDA) parameters MS1-TOF 250 milliseconds, followed by 50 MS2 events of 25 milliseconds each. The IDA criteria; over 200 counts threshold, charge state of plus 2-4 with 4 seconds exclusion window. Finally, MASCOT® (Matrix Sciences) was used to analyze the collected data and Protein Pilot 4.0 (ABSCIEX) was used for peptide identifications.

3.3.7 Extracellular polymeric substance (EPS) depolymerase mediated biofilm degradation assay

Soft agar plaque method (48), as described previously in host range method, was used to detect the presence of halo zone on *P. vagans* strain C9-1 and *E. amylovora* ATCC 29780. The putative EPS-depolymerase from bacteriophage RAY was PCR amplified from lysate using primers designed to amplify the full length gp76. It was cloned by digesting with enzymes NdeI/SalI into a similarly digested pET15b. The resulting plasmid (JG1700) was amplified by transforming into *E. coli* DH5α and plated on LB-amp. Resulting colonies were PCR checked and were used to start overnight cultures and DH5α without plasmid pJG1700 was grown as a control. The protein was induced using IPTG and extracted by lysing cells via sonication. Post

sonication, cell debris was removed from both cultures by centrifuging at 12000 rpm and 4°C for 2x20 minutes. 10µl of resulting supernatant was spotted on bacterial lawns of *P. vagans* strain C9-1and *E. amylovora* ATCC 29780 embedded in top agar after plating for 2 hours.

3.3.8 Motif identification and analysis

MEME (80) and FIMO (81) tools at public phage galaxy (https://cpt.tamu.edu/galaxy-pub/) were used to scan bacteriophage genome of *Agrican357virus* for statistically significant motifs. Motifs found by MEME (80) with e-value less than 1e-002 were selected by FIMO (81) to be searched for their coordinates and iterations in their respective genomes. User defined cutoff values (P-value < 1e-3, Q-value < 0.05), as described in Berg *et al* (82) were used to maximize the results. The location of the motifs within bacteriophage genomes was determined from the annotated GenBank files (17, 18).

3.4 Results and discussion

3.4.1 Isolation and characterization of eight closely related large bacteriophages infecting *E. amylovora*

Eight novel bacteriophages (Deimos-Minion, Special G, RAY, Simmy50, Bosolaphorus, Desertfox, Mortimer and MadMel) that infect *E. amylovora* were plaque isolated and their genomes were subsequently sequenced and annotated as previously described (17, 18). All eight bacteriophages have relatively large genomes with genome sizes of 271 to 275 kb (Table 3.1), which are comparable to the related bacteriophage Ea35-70 (271084 bp). These bacteriophages have correspondingly large putative proteomes, with 317 to 324 predicted ORFs. A search for tRNA's using tRNA ScanSE (83) suggests that RAY, Simmy50, Bosolaphorus and Mortimer have 1 tRNA each coding for Asparagine, whereas no tRNA's were detected for DM,

Table 3.1 General characteristics of nine related bacteriophage Deimos-Minion (DM), RAY, Special G, Desertfox, MadMel, Mortimer, Bosolaphorus, Simmy50, and Ea35-70 that infect E. amylovora ATCC 29780. Sample type is as reported by collectors, no sample type was recorded for Mortimer. Due to the high conservation of this family, differences in encoded genes is also provided with missing genes numbered with respect to Deimos-Minion.

Phage Name	**GenBank Accession**	**Genome length (bp)**	**Sample type**	**Conserved Domains**	**ORFs (tRNAs)**	**Gene Differences compared to DM**	
						extra genes	**Missing**
Deimos-Minion **(DM)**	KU886225	273,501	fruit	39	324		
RAY	KU886224	271,182	leaves, stem	39	317 (1)	0	gp49, gp50, gp90, gp91, gp166, gp234
Special G	KU886222	273,224	branches, blossoms	41	321	gp63, gp203, gp231	gp90, gp91, gp111, gp166, gp234
Desertfox	MG655268	272,458	soil	39	320	gp106, gp231, gp256,	gp48, gp50, gp90, gp91, gp111, gp234
Madmel	MG655269	275,000	soil	41	321	gp62, gp202, gp230	gp90, gp91, gp111, gp252
Mortimer	MG655270	273,914	–	40	324 (1)	gp62, gp110, gp238, gp261	gp48, gp117, gp234
Bosolaphorus	MG655267	272,228	orchard dirt	39	320 (1)	gp223	gp48, gp90, gp91, gp234
Simmy50	KU886223	271,088	bark	39	322 (1)	gp8, gp63, gp209, gp210	gp51, gp90, gp91, gp166, gp234
Ea35-70	KF806589	271,084	soil	36	318 (1)	gp61, gp115, gp224	gp86, gp93, gp120, gp166, gp232, gp234, gp252

Special G, MadMel and Desertfox. No lysogeny related genes were identified (including integrase, excisionase or repressors). Their clear plaque morphology and ease in obtaining higher titers (~10^8-10^{10} pfu/ml) suggest they may be lytic bacteriophages, however rigorous testing for bacterial lysogeny has not been performed.

3.4.2 Electron microscopy reveals myovirus structure eight *E. amylovora* bacteriophages

Deimos-Minion, Special G, RAY, Simmy50, Mortimer, MadMel, Desertfox, and Bosolaphorus were all found to be similarly sized Myoviridae (Figure 3.1), having contractile tails (average size 159 nm +- 11.4 nm), a tail sheath (average size 78.5 nm+- 9.28 nm), visible tail fibers, and large capsids (average size 128 nm +- 5.96 nm). This morphology is supported by their protein-based relationships to other jumbo *Myoviridae* discussed below. Due to apparent similarity within these bacteriophages, only RAY's morphological calculations are listed but all eight of these bacteriophages were imaged extensively.

3.4.3 Host range and burst size

Bacteriophages of the *Agrican357virus* family were tested for activity against seventeen different bacterial strains (Table 3.2). Out of these, fifteen were from the *Enterobacteriales- P. agglomerans* E325 (52), *P. vagans* C-91 (53, 54), *E. coli* K-12 BW 25113 (55), *S. enterica* (generous donation by roth lab), *Y. pestis* KIM6 (56, 57), *E. cloacae* ATCC 13047 (58), *K. pneumoniae* ATCC 10031 (59), *B. subtilis* ATCC 6033 (60), *C. sakazakii* ATCC 29544 (61, 62), *E. amylovora* Ea110 (49), *E. amylovora* GH9 (50), *E. amylovora* EaBH (50), *E. amylovora* RB02 (50), *E. amylovora* Ea273 (51), *E. amylovora* ATCC 29780 (control) (49) and two from *Pseudomonadaceae- P. aeruginosa* PA100 (63) and *P. chlororaphis* ATCC 13985 (64) *Enterobacteriales* strains were chosen due to being members of the same bacterial order as

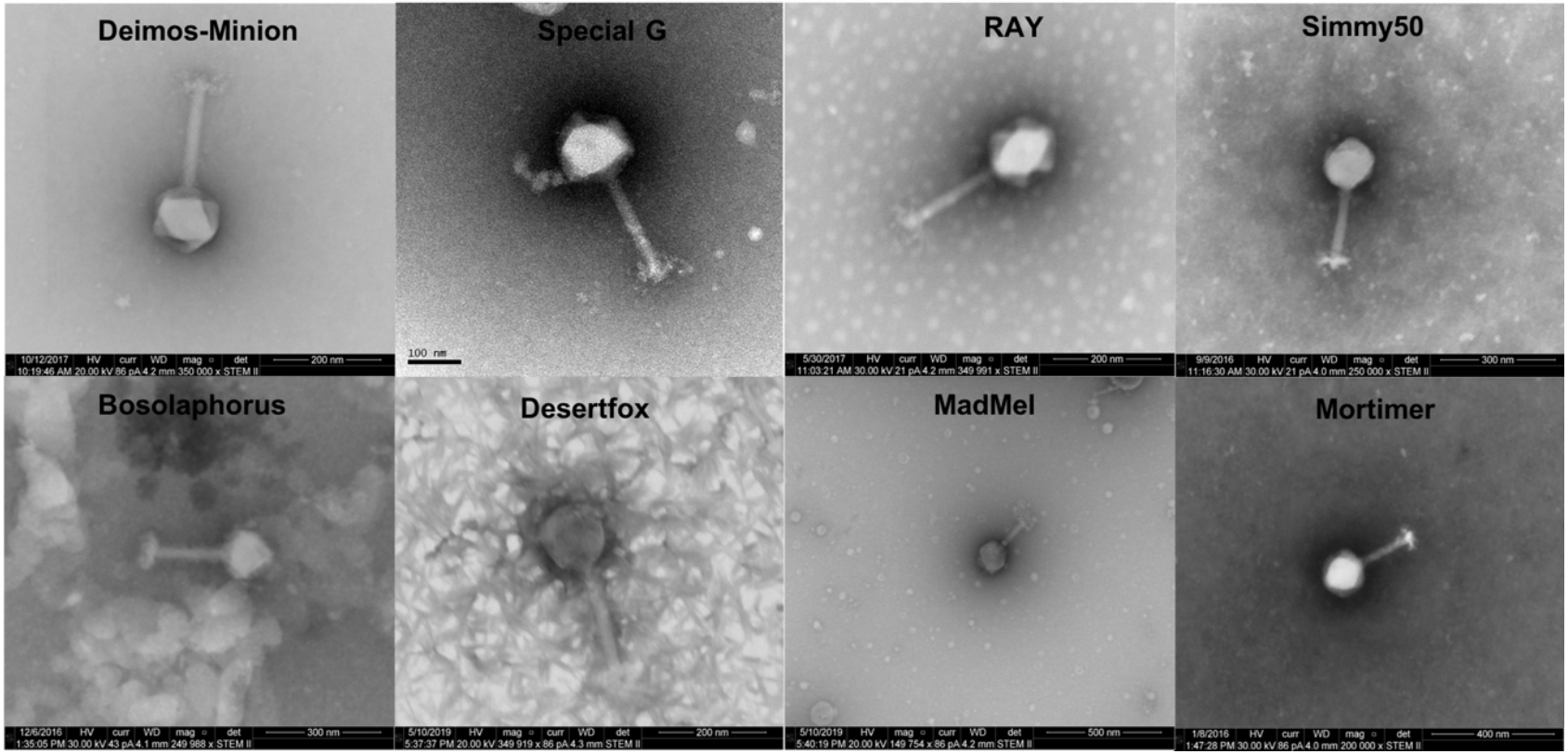


Figure 3.1 Electron microscopy STEM images of Deimos-Minion, Special G, RAY, Simmy50, Bosolaphorus, Desertfox, MadMel, and Mortimer revealed *Myoviruses* with long contractile tails

Erwinia, whereas *Pseudomonadaceae* strains were the hosts of bacteriophages related to the *Agrican357virus* bacteriophages based on protein BLAST.

Our current analyses displayed that *Agrican357virus* bacteriophages infect all *Erwinia* strains (with the exception of Special G and Mortimer that failed to infect GH9 and EaBH, respectively) as well as closely related genera also commonly found on plants– *P. agglomeran*s (84) and *P. vagans* (85*)* (Table 3.2). Owing to the large nature of *Agrican357virus* bacteriophages, we investigated the burst size of bacteriophage Deimos-Minion on *E. amylovora* strain ATCC 29780. Burst size studies suggested that when infected at MOI of 100 Deimos-Minion has burst size of 4.6-4.9 with latent period of 3-4 hours before the first burst (Figure 3.2) under the laboratory growth conditions used herein, consistent with their large size. As seen in Figure 3.2, a second burst is appearing at the end of this six hours period. Owing to the large nature of *Agrican357virus* bacteriophages, we investigated the burst size of bacteriophage Deimos-Minion on *E. amylovora* strain ATCC 29780. Burst size studies suggested that when infected at MOI of 100 Deimos-Minion has burst size of 4.6-4.9 with latent period of 3-4 hours before the first burst (Figure 3.2) under the laboratory growth conditions used herein, consistent with their large size. As seen in Figure 3.2, a second burst is appearing at the end of this six hours period. The observed burst size (~5) was confirmed with phage RAY (data not shown) and is consistent with other large *Myoviridae* in that *Pseudomonas aeruginosa* bacteriophage KTN4has a reported burst of 6-8 and may be due to the need to build internal cellular structures for the Jumbo viruses to be built (25), or due to sub-optimal assay conditions for proliferation.

Table 3.2 Host range analysis of eight *Agrican357virus* bacteriophages. Host range tests on *Agrican357virus* displays infection of *E. amylovora* strains ATCC 29780 (control), GH9, Ea110, EaBH, RBO2, Ea273, *P. agglomerans* (E325) and *P. vagans* (C9-1) only. Bacteriophages Special G and Mortimer failed to infect strain EaGH9 and EaBH respectively. All other bacterial strains remained uninfected. Plaque forming units (pfu) should be compared to the ATCC strain, because the same amount of the same lysate was used to infect each strain

Bacterial strains (strain number)	**Bacteriophages**							
	Deimos-Minion	**RAY**	**Special G**	**Desertfox**	**MadMel**	**Mortimer**	**Bosolaphorus**	**Simmy50**
E. amylovora (ATCC 29780)	5.20E+09	7.80E+09	3.40E+09	2.56E+07	5.42E+08	2.87E+06	3.29E+04	4.33E+08
E. amylovora GH9	3.03E+10	3.90E+09	–	1.77E+07	5.00E+06	3.49E+05	5.09E+04	4.15E+08
E. amylovora EA110	6.60E+09	5.20E+09	4.52E+09	9.00E+06	3.63E+08	5.26E+06	5.65E+04	8.97E+08
E. amylovora EaBH	5.70E+09	4.40E+09	2.60E+09	1.06E+07	5.78E+08	–	6.04E+04	3.00E+08
E. amylovora RB02	3.25E+09	4.05E+09	1.84E+08	3.65E+07	4.47E+08	1.06E+06	5.03E+04	2.42E+08
E. amylovora 273	1.04E+10	9.75E+09	1.45E+07	2.36E+07	4.37E+08	5.39E+06	6.15E+03	2.61E+08
P. vagans (C9-1)	3.14E+10	2.64E+10	1.00E+11	5.01E+07	2.05E+09	6.39E+06	4.05E+03	4.95E+09
P. agglomerans (E325)	3.10E+10	9.30E+09	2.60E+10	5.80E+06	2.67E+09	2.90E+06	2.79E+04	4.48E+09
P. chlororaphis (ATCC 13985)	–	–	–	–	–	–	–	–
E. coli k-12 (BW 25113)	–	–	–	–	–	–	–	–
B. subtilis (ATCC 6033)	–	–	–	–	–	–	–	–
C. sakazakii (ATCC 29544)	–	–	–	–	–	–	–	–
K. pneumoniae (ATCC 10031)	–	–	–	–	–	–	–	–
S. enterica (Roth lab)	–	–	–	–	–	–	–	–
E. cloacae (ATCC13047)	–	–	–	–	–	–	–	–
P. aeruginosa (PA100)	–	–	–	–	–	–	–	–
Y. pestis (KIM6)	–	–	–	–	–	–	–	–

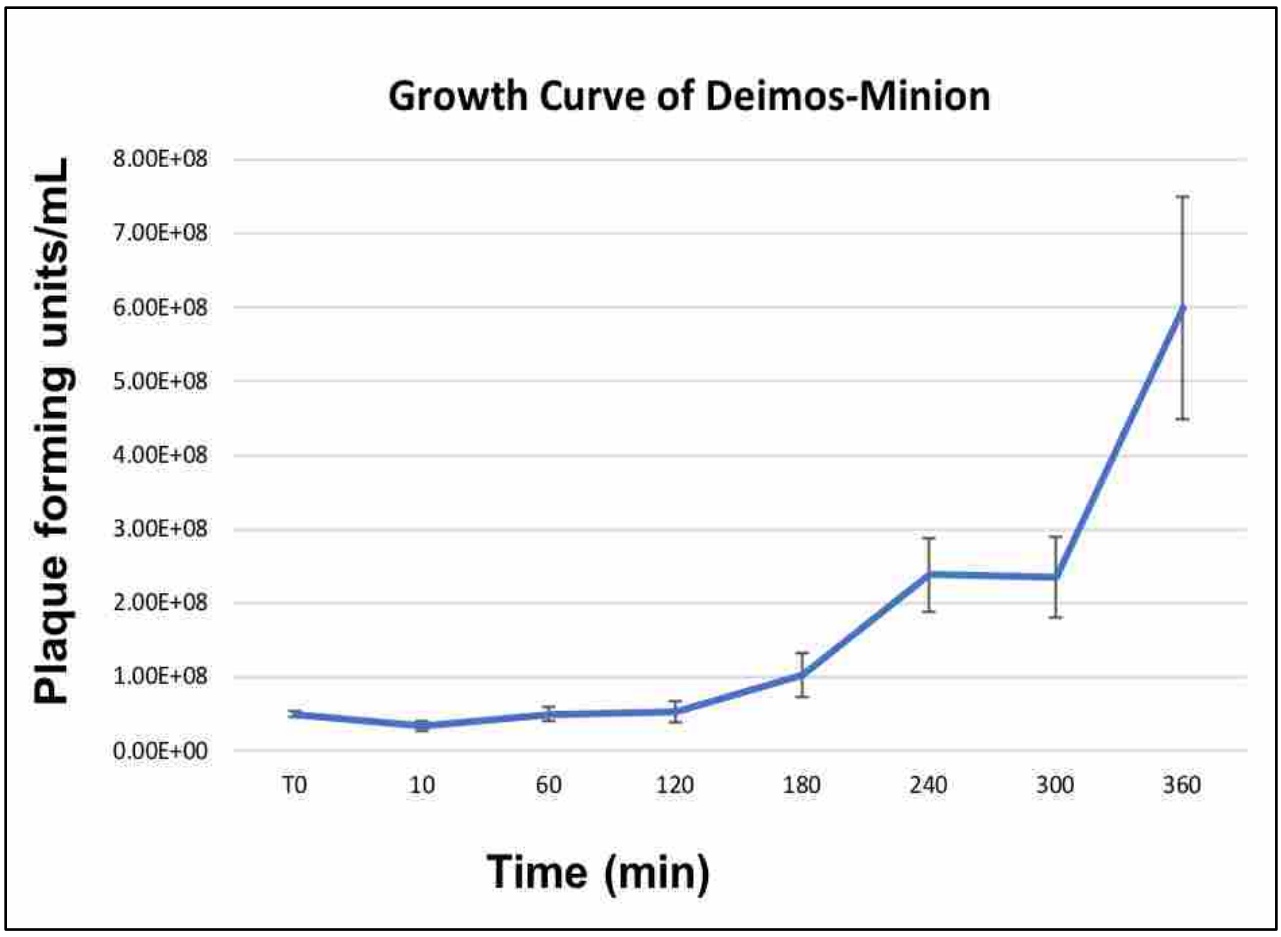


Figure 3.2 Growth curve for Deimos-Minion with host *Erwinia amylovora* ATCC 29780 by plaque assays shows first burst at ~4 hours and second burst at ~6 hours.

3.4.4 Genomic and evolutionary characteristics

To determine the overall genomic and proteomic similarity of our eight novel bacteriophages to available bacteriophages in GenBank, related bacteriophages were identified by BLAST (qblast) using each of the putative gene products encoded by RAY. The bacteriophages with e-values below 1.00E-04 and above 33% identity that were identified in three or more BLAST searches were then compared using Gepard dot plot (68) average nucleotide identity (ANI analysis) (86), and BLAST alignment (65). Dot plots were constructed using whole genome sequences, major capsid protein amino acid, and terminase amino acid

sequences (Figure 3.3 A-C respectively). While looking at the results of the whole genome dot plot, all eight of our bacteriophages show no similarity to any other bacteriophages used in the dot plot except for very close similarity to Ea35-70 (KF806589) (19), an *Erwinia* bacteriophage isolated in Canada in 2014 (see Figure 3.3A). In addition, their average nucleotide identity (ANI) using Geneious (74) was remarkably high >94% (see supplementary table 3.S1). These results indicate that these eight bacteriophages Deimos-Minion, Simmy50, RAY, Special G, Bosolaphorus, Desertfox, MadMel and Mortimer along with Ea35-70 make a distinct family of bacteriophages, consistent with the International committee on taxonomy of viruses' classification as new species of a new genus *Agrican357virus* in the family *Myoviridae* of order *Caudovirales* (20).

The major capsid protein (MCP) and terminase proteins are two of the most conserved proteins in bacteriophage genomes and have been used to group bacteriophages in families by single gene analysis (87). In order to identify distant bacteriophage relatives, a proteomic comparison of these bacteriophages was performed using terminases (see Figure 3.3B) and MCPs (see Figure 3.3C) by Gepard dot plot (68). The same bacteriophage order from the whole genome dot plot (Figure 3.3A) was used in these dot plots.

Whole genome and terminase dot plots both displayed limited synteny between *Agrican357virus* bacteriophages and *Erwinia* bacteriophage phiEaH1 (4.00E-155 from blastp of terminase) indicating this bacteriophage as the closest known relative from *Erwiniaceae*. In contrast, little similarity to *Pseudomonas* bacteriophages phiKZ (8.00E-156), KTN4 (8.00E-156), phiPA3 (2.00E-149), 201phi2-1 (5.00E-140), OBP (2.00E-101), EL (3.00E-77) and *Ralstonia* bacteriophages RSF1 (9.00E-122) and RSL2 (3.00E-120)

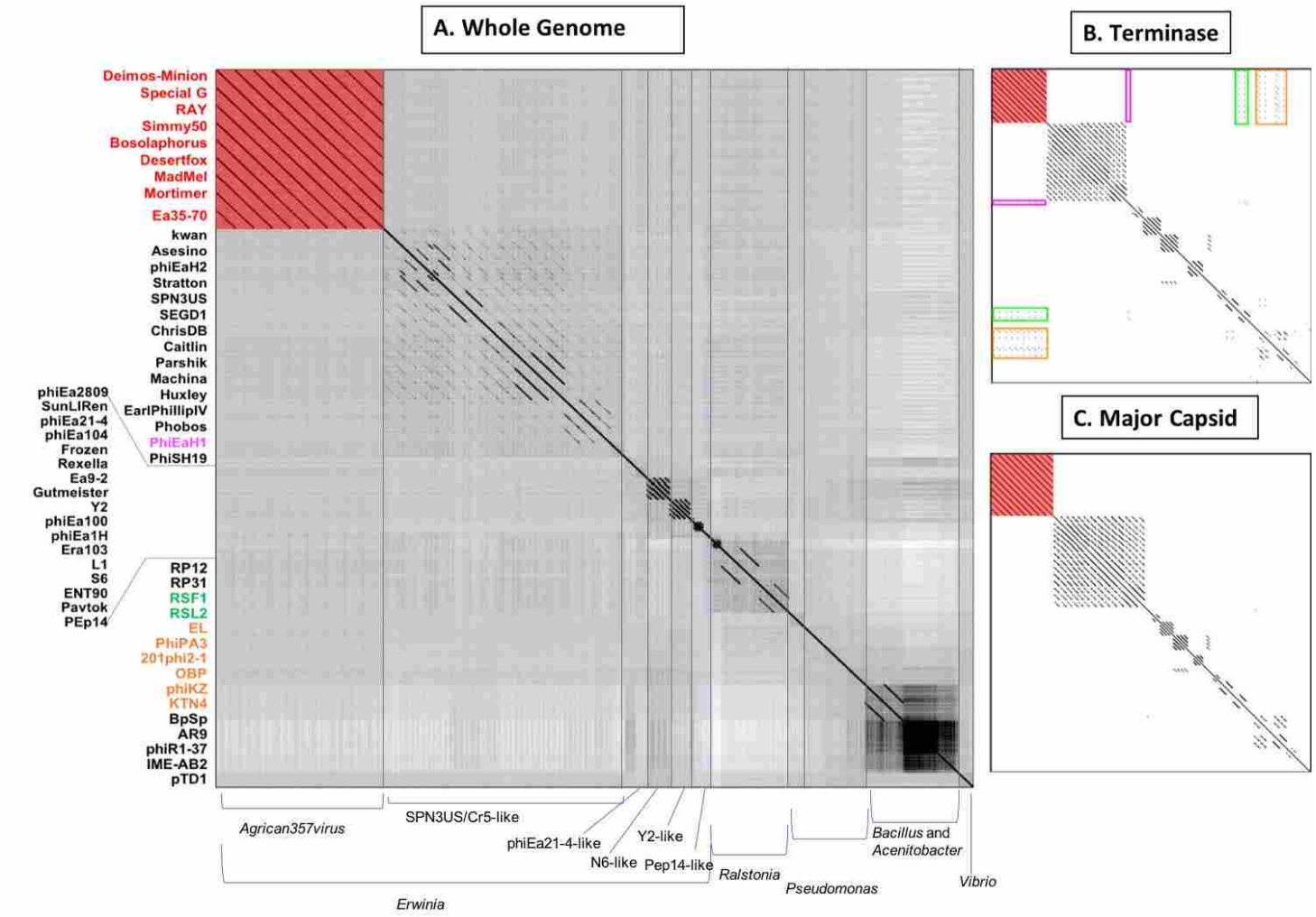


Figure 3.3 Whole-genome nucleotide (A) and protein terminase (B) or major capsid protein (C) dot plot analysis reveals a fairly isolated cluster of bacteriophages that includes Deimos-Minion, Special G, RAY, Simmy50, Bosolaphorus, Desertfox, MadMel and Mortimer and Ea35-70. Dot plots were constructed using Gepard.

can be seen in the terminase dot plot which was not apparent in the whole genome and major capsid protein dot plots. All of these bacteriophages are distantly related jumbo *Myoviridae*.

The two subunits of terminase protein; large and small, are an essential part of DNA packaging (88; 89). All eight of our *Agrican357virus* bacteriophages have a putative terminase gene with identical amino acid sequences: Deimos-Minion gp189, Special G gp185, RAY gp183, Simmy50 gp186, Desertfox gp184, Bosolaphorus gp185, MadMel gp185 and Mortimer gp188. This protein is also present in Ea35-70 gp181. This indicates that it is a highly conserved protein for this family. Considering the similarity between these bacteriophages, it can be inferred that all nine bacteriophages of *Agrican357virus* may have headful packaging (Figure 3.4). In support of this conclusion, blastp results demonstrated a match with *Pseudomonas* bacteriophage phiKZ with an e-value of 8.00E-156, a terminase large subunit from *Erwinia* bacteriophage PhiEaH2 with an e-value of 6.00E-122 and a terminase large subunit of *Pseudomonas* bacteriophage 201phi2-1 with an e-value of 5.00E-140. Bacteriophages phiKZ, phiEaH2 and 201phi2-1 are all known to have headful packaging (90). In addition to blastp, bacteriophage termini and packaging mode for six bacteriophages (excluding Deimos-Minion and Special G) was also determined using randomly fragmented next-generation sequencing (NGS) data with the help of software PhageTerm (91) https://galaxy.pasteur.fr. PhageTerm analysis indicated that RAY, MadMel, Desertfox, Bosolaphorus, Simmy50 and Mortimer have headful packaging without a pac site. Thus, the headful packaging strategy is supported by terminase homology and NGS sequencing data.

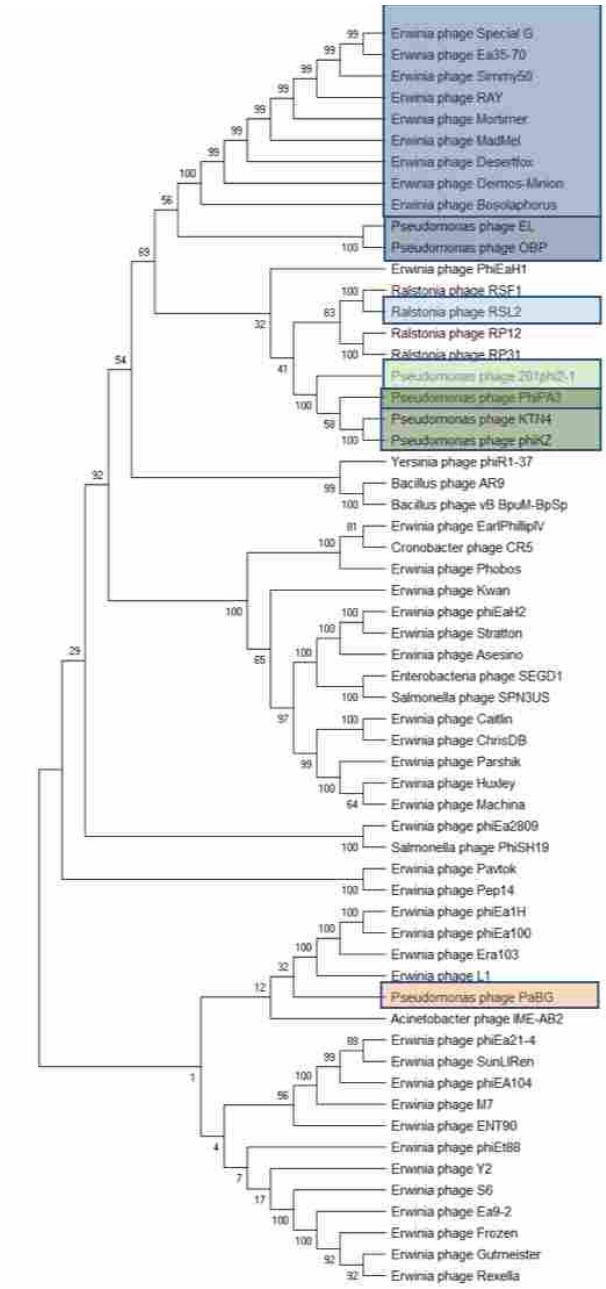


Figure 3.4 Phylogenetic analyses of phage terminase proteins supports the relationships depicted by dot plot analysis of the *Agrican357virus* bacteriophages. The evolutionary history was inferred using the Neighbor-Joining method in MEGAX. Unrooted tree was condensed to cutoff value of 50% where 2000 was set to be initial bootstrapping value.

3.4.5 Proteomic analysis of the *Agrican357virus* family

Due to great similarity between these bacteriophages we randomly chose RAY as a representative for the protein classification. Proteomic analysis of RAY reveals the novel nature

of these bacteriophages in that of 318 proteins, 202 proteins were considered to be novel with no BLAST hit (the e-value cutoff was <1.00E-04), 50 were hypothetical proteins with BLAST hits, and 67 were proteins with a putative function based on their BLAST hit (Supplementary figure 3.S1A). Thus, over half of the proteins had no BLAST hit outside of the *Agrican357virus* bacteriophages. These proteins represent a considerable proteomic "dark matter" (92), and underscore the vast biological richness harbored in bacteriophages. Of the 67 proteins with predicted function, a majority appear to be structural proteins (~41%), and DNA metabolism proteins (approximately 41%) (Supplementary figure 3.S1B)

The computer program Phamerator (70) was used to compare the entire genomes of the nine *Agrican357virus* bacteriophages that infect *E. amylovora*: Deimos-Minion, Special G, RAY, Simmy50, Bosolaphorus, Desertfox, MadMel, Mortimer and Ea35-70 (Figure 3.5). Despite their large size, these genomes display remarkable nucleotide sequence and proteomic conservation (>94% ANI, see supplementary table 3.S1). The genomes encode recognizable structural and enzymatic bacteriophage proteins vital to the bacteriophage life cycle, including terminase proteins, major capsid proteins, and tail fiber proteins as well as proteins involved in DNA transcription and translation, such as helicase proteins, DNA polymerase, and RNA polymerase. Though the genomes of these nine bacteriophages are virtually identical, a few genes are differentially present across these bacteriophage genomes. Most of these are hypothetical proteins, however, HNH endonucleases also differed consistently between the *Agrican357virus* bacteriophages. HNH endonucleases are proteins that splice DNA and assist in the movement of introns and other intron-like sequences (93).

Deimos-minion has two such HNH endonucleases, gp93 and gp234 that do not appear to be homologs based on protein similarity. Protein BLAST results of gp93 show that the HNH

endonuclease is also found in bacteriophages. Bosolaphorus, Desertfox, MadMel, RAY, Simmy50, Special G and Ea35-70, and is similar to those found in some *Pseudomonas* bacteriophages (phiKZ and KTN4) as well as both Gram-negative and Gram-positive strains of bacteria. However, only the HNH endonuclease domain (~amino acid 58-109 of bp93) is primarily conserved, the remaining 278 amino acid protein is not conserved in bacteria. On the other hand, homologs of HNH endonuclease gp234 are only found in Deimos-minion and MadMel, as well as several Gram-positive and Gram-negative bacteria. Genomes of Deimos-Minion, Desertfox and MadMel also displayed a reversed order of two proteins (gp93-gp94 in Deimos-Minion, gp88-gp89 in Desertfox and gp90-gp91 in MadMel) when compared to similar proteins in other bacteriophages of this family. The proteins involved are HNH endonuclease and ribonucleotide reductase. To search for repetitive sequences in the genome which may be involved in recombination, MEME (80) and FIMO (81) were used to locate motifs in the genomes of all eight of our *Agrican357virus* bacteriophages. Several common and unique motifs were discovered, however they had poor e-values with little or no significance and were not followed further.

Due to the large size of these bacteriophages, and their terminase similarity to bacteriophage phiKZ, these bacteriophages likely belong to the jumbo bacteriophages (21; 94) making it no surprise that the structural proteins are found in other bacteriophages. Along with hypothetical proteins, the proteins that are conserved with other phiKZ-like jumbo bacteriophages include: RNA polymerase beta subunit, nuclease RtcB-like, SbcC like, helicase, virion structural proteins, tail fiber, tail sheath, lysozyme domain, terminase, and major capsid protein.

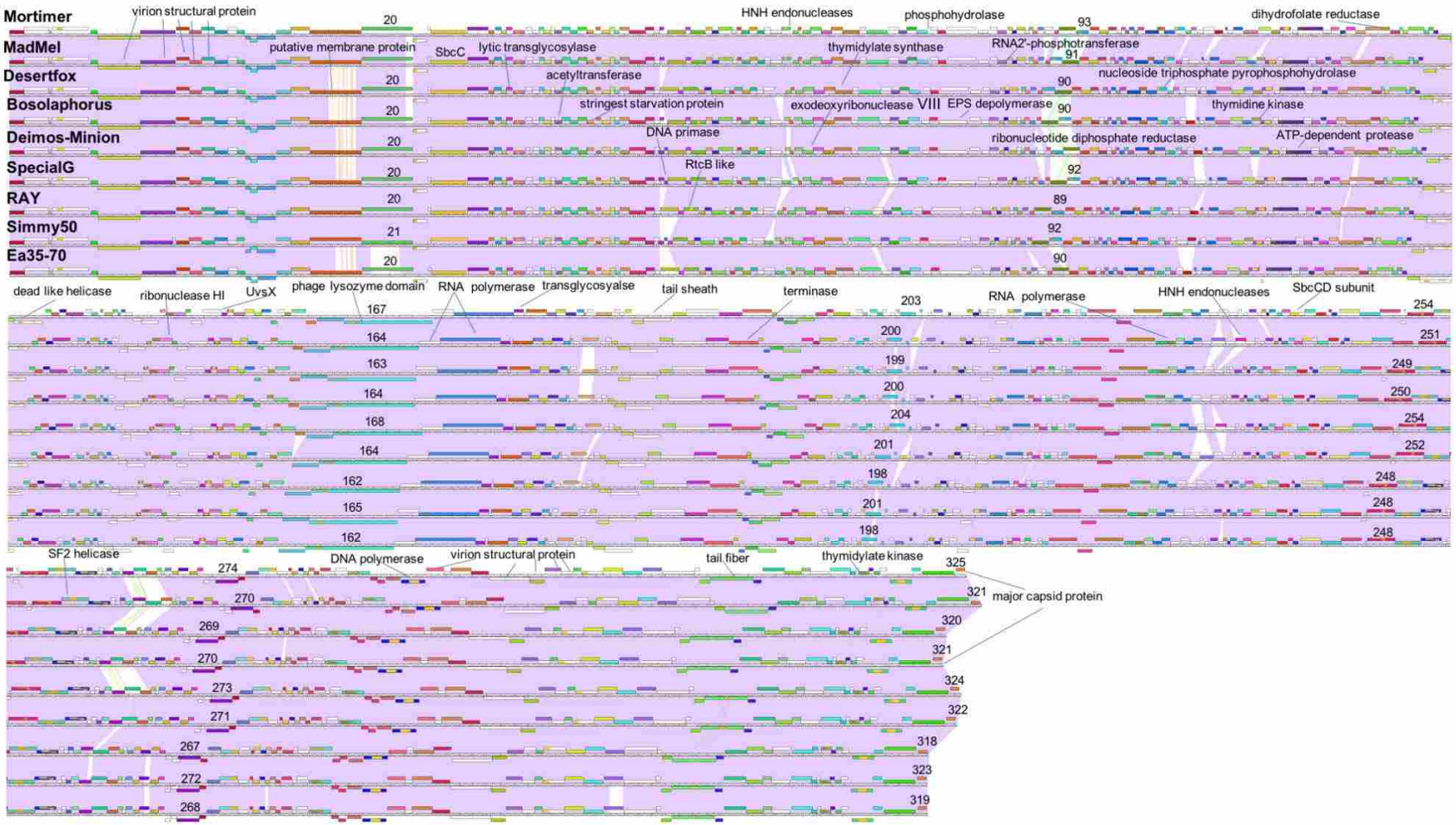


Figure 3.5 Whole genome Phamerator map of *E. amylovora* bacteriophages illustrates the high similarity of bacteriophages. Mortimer, MadMel, Desertfox, Bosolaphorus, Deimos-Minion, Special G, RAY, Simmy50, and Ea35-70. Bacteriophages were mapped using Phamerator and arranged based on highest protein similarity. Violet shading between genomes indicates genome nucleotide homology (with standard e-value cutoff of 1.00E-04) and the ruler indicates genome base pairs, while white spaces indicate areas without significant nucleotide similarity. Boxes above and below the genome ruler indicate ORFs going in the forward and reverse direction, respectively. They are labeled with predicted function, occasionally numbered, and colored to indicate protein homologs between the bacteriophages

A SplitsTree analysis showing the relationship of the related jumbo bacteriophages by protein conservation is displayed in Figure 3.6. This protein-based tree suggests seven groups of related jumbo *Myoviridae* bacteriophages, with the *Agrican357virus* group as the most distant group. It further confirms that proteins of *Agrican357virus* family are more similar to proteins from *Pseudomonas* bacteriophages EL and OBP and *Ralstonia* bacteriophage RSL2 than to other *Enterobacteriales* bacteriophages.

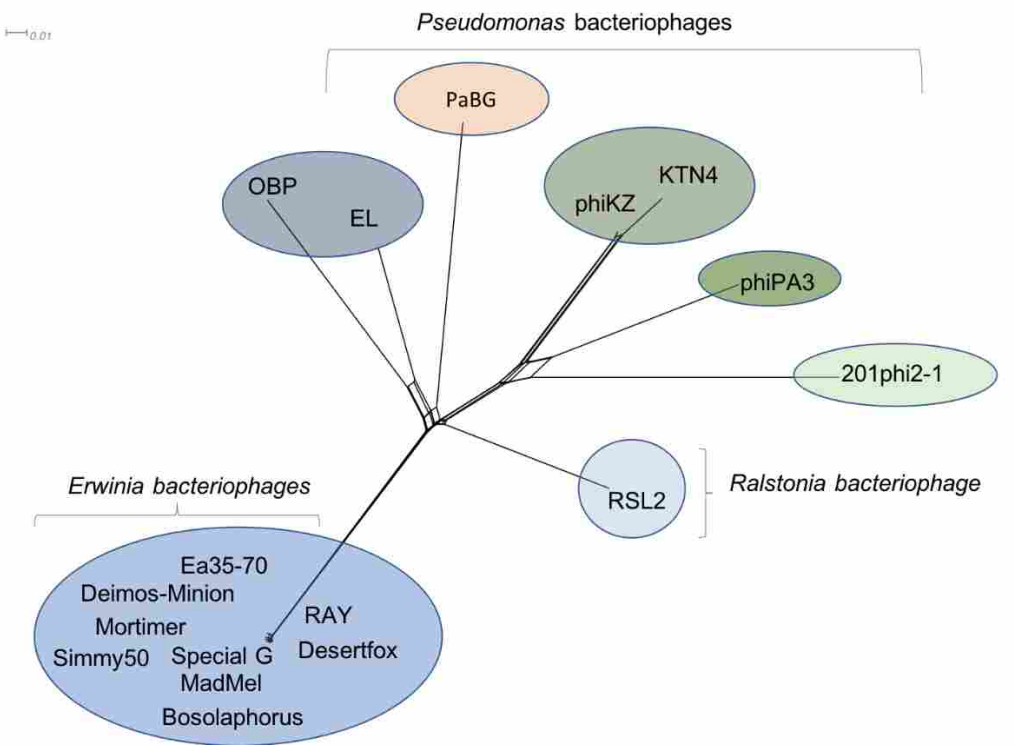


Figure 3.6 Protein-conservation analysis displayed by Splitstree of the *Agrican357virus* genus with related jumbo *Myoviridae* bacteriophages reveals *Agrican357virus* as a distant evolutionary group

3.4.6 Mass spectrometry validates 27 hypothetical proteins as proteins of unknown function

Further analysis of Deimos-Minion and RAY genomes via mass spectrometry (MS) detected several novel proteins, promoting the status of 27 proteins from hypothetical proteins to proteins of unknown function. In RAY and Deimos-Minion genomes collectively, MS analysis identified seventeen proteins with a putative function, eighteen novel hypothetical proteins specific to this bacteriophage family and nine hypothetical proteins (seven known bacteriophage

proteins and two other) with blastp hits to other bacteriophages (Table 3.3). The majority of proteins found through MS are novel hypothetical proteins found only in this family, followed by putative bacteriophage structural proteins, hypothetical bacteriophage proteins, proteins with putative functions and other hypothetical proteins (see Table 3.3). This analysis agrees with our predicted conservation of proteins depicted through Phamerator analyses.

3.4.7 Biofilm degradation (EPS) assays suggest specificity for *Pantoea*

Enzymatic proteins like extracellular polysaccharide (EPS) depolymerase and phage-related lysozyme are few of the annotated proteins with putative functions which were also predicted via mass spectrometry. EPS depolymerase (95) is an enzyme that degrades EPS and phage-related lysozyme is shown to lyse the bacterial cell wall (96). It has been shown that halo formation on the host could be a result of biofilm degradation assay (97, 98). The presence of halo zone after in infections of *Agrican357virus* family was first observed on *P. vagans* strain C9-1 (Figure 3.7A).

To investigate further the EPS- depolymerase gene was cloned into a plasmid pJG1700, amplified using *E. coli* DH5α, and spotted on *P.vagans* stain C-91 and *E. amylovora* strain ATCC 29780 (Figure 3.7B). Lysate from a similarly grown and prepared DH5α culture was used as a control. The clearing is indicative of EPS depolymerase activity on *P. vagans*. This activity was not seen on *E. amylovora* ATCC 29780.

Table 3.3 Mass Spectrometry reveals 27 hypothetical proteins as proteins of unknown function. Peptides detected by LC/MS/MS of a crude bacteriophage lysate of RAY and/or Deimos-Minion. Columns provide the gene product number corresponding to the peptide(s) detected, the putative function of the protein, the mass spectrometry retrieval number (which may reflect abundance), and the percent coverage for the protein. Gene products are grouped by putative function when available, and then by conservation. Deimos-Minion is abbreviated to DM

RAY	DM	Putative function	Retrieval #		% coverage	
			RAY	DM	RAY	DM
Putative Bacteriophage Structural Proteins						
	gp323	putative major capsid protein		4		62.65
gp178		putative virion structural protein	35		40.2	
gp154		putative virion structural protein	57		28.6	
gp179	gp185	putative tail sheath protein	105	45	22.5	43.45
gp18		putative virion structural protein	61		18.7	
	gp308	putative tail fiber protein		106		51.21
	gp9	putative virion structural protein		146		23.16
	gp19	putative virion structural protein		153		10.92
	gp188	putative virion structural protein		121		10.14
gp293		putative virion structural protein	104		31.89	
Putative Enzymatic Proteins						
gp76	gp79	putative EPS-depolymerase	58	89	23.6	20.16
gp162		putative phage-related lysozyme	94		29.2	
gp102	gp107	putative nucleotide triphosphatase	103	72	25.6	39.53
	gp127	putative dihydrofolate reductase		171		12.1
	gp23	putative SbcC-like protein		169		25.18
	gp228	putative DNA-directed RNA pol.		67		41.82
	gp94	putative ribonucleotide diphosphate reductase beta subunit		91		10.32
Novel hypothetical proteins found only in this bacteriophage family						
gp281	gp287	novel hypothetical protein	6	61	33.5	36.31
gp295	gp301	novel hypothetical protein	9	64	23.8	29.11
gp287		novel hypothetical protein	17		71.7	
gp185	gp191	novel hypothetical protein	18	68	66.0	68.49
gp188		novel hypothetical protein	33		35.5	
gp186		novel hypothetical protein	41		16.8	
gp196	gp202	novel hypothetical protein	44	137	34.9	21.7
gp55		novel hypothetical protein	46		37.3	
gp316		novel hypothetical protein	47		42.6	
gp110	gp114	novel hypothetical protein	49	116	21.6	33.99
gp298	gp304	novel hypothetical protein	50	70	42.8	27.72
gp173	gp179	novel hypothetical protein	55	55	58.1	61.49
gp166		novel hypothetical protein	78		40.9	
gp75		novel hypothetical protein	92		29.0	
gp99		novel hypothetical protein	62		28.0	
gp207	gp212	novel hypothetical protein	95	95	4.7	17.13
gp98	gp103	novel hypothetical protein	97	133	6.8	9.74
	gp140	novel hypothetical protein		84		18.49
Hypothetical bacteriophage proteins						
gp222	gp227	hypothetical phage protein	23	87	33.4	39.1
gp240	gp246	hypothetical phage protein	34	81	32.3	57.42
gp301	gp307	hypothetical phage protein	54	93	32.26	58.8
gp202		hypothetical phage protein	73		43.3	
gp292		hypothetical phage protein	79		37.9	
	gp251	hypothetical phage protein		88		25.88
	gp224	hypothetical phage protein		129		33.33
Other hypothetical proteins						
gp273		hypothetical protein	68		18.7	
gp41		hypothetical protein	56		34.2	

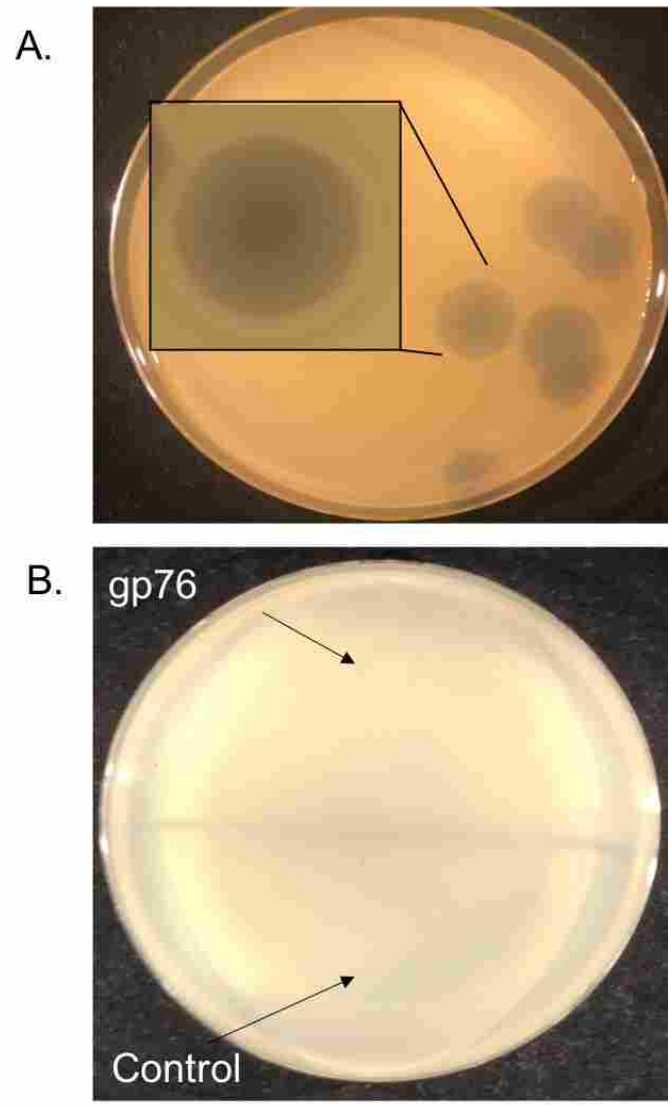


Figure 3.7 Halo formation on *P. vagans* by RAY (A) and bio-film degradation activity of gp76 on *P. vagans* (B).

3.4.8 Structural prediction supports the putative function of several proteins

To further understand *Agrican357virus* and verify their protein functions, we studied proteins involved in DNA metabolism (~ 45%), the largest group of functional proteins conserved in the *Agrican357virus*. Multiple mechanisms for DNA regulation and repair are evident with the presence of proteins that are hypothesized to aid DNA synthesis, repair, and recombination. These proteins may increase the stability and survival of these jumbo bacteriophages (supplementary table 3.S2, supplementary figure 3.S2). In order to proliferate in

host cells, bacteriophages need to be equipped with proteins that allow them to reproduce effectively. Although many bacteriophages harbor proteins for DNA damage repair and DNA reproduction inside a host bacteria cell, these large bacteriophages may require extremely viable progeny due to lower burst sizes (~4.6 functional virions compared to thousands reported for other bacteriophages).Two proteins with a conserved domain found in the nine *Agrican357virus* bacteriophages are SbcC and a SbcCD nuclease (see supplementary figure 3.S2A). The ability of SbcC and SbcCD to regulate and repair DNA has been shown to be essential for the stability and proliferation of some bacteriophages (99). During DNA replication, palindromic sequences will create hairpin-like structures that can inhibit the progression of DNA polymerase (100) SbcC and SbcCD proteins work together to cleave both double- stranded and single-stranded DNA and have been shown to recognize and specifically cleave hairpin structures.

This breaks down the replication fork, allowing the genome to be repaired through recombination, so replication can proceed (99, 100). The proteins SbcC and SbcCD nucleases preserve the viability of the genome by allowing replication without excising the palindromic sequences (100). There are many types of DNA damage that may occur within a genome, making recombination and repair of DNA important, such as mutations due to UV damage. UV damage creates kinks or abnormalities within a genome and prohibits proliferation. Exodeoxyribonuclease VIII breaks double stranded DNA, and degrades a genome on both 5' ends (101, 102). This allows the kinked and abnormal portions of a genome to be straightened and repaired through homologous recombination. Additionally, exodeoxyribonuclease VIII does not require ATP to perform DNA repair, enabling repair of the genome even in low-energy environments where the bacteriophage does not have access to ATP (102). We hypothesize that exodeoxyribonuclease VIII enables the bacteriophages to remain stable despite mutations from

UV damage. However, unique from our other predicted structure alignments, the protein from RAY does not match up well with other exodeoxyribonuclease VIII homologs (see supplementary figure 3.S2B). It is possible that since these proteins do not have the same protein folding and alignment, they may not have the same function but a related, adapted function.

In the *Agrican357virus* bacteriophages, there are several encoded proteins with conserved domains of the thymidylate kinase and thymidine kinase (see supplementary table 3.S2). Structural prediction and alignment confirm these proteins as likely thymidine kinases (see supplementary figure 3.S2C and 3.S2D), a necessary step due to the distant relationship (low e-values) of *Agrican357virus* bacteriophage proteins when compare to other biological entities. Thymidine kinase is an enzyme that catalyzes the phosphorylation of thymidine monophosphate (103). Thymidylate kinase then catalyzes the phosphorylation of thymidine diphosphate (104), which is an essential precursor for DNA (105). Therefore, these proteins are regulatory enzymes that make bacteriophage cell growth and survival possible by aiding proliferation through the synthesis of DNA (104, 105, 106). Other proteins shown in supplementary figure 3.S2.E and 3.S2.F are putative UvsX recombinase and a putative SF2 helicase with conserved helicase domain known as UvsW, which finishes the recombination (107, 108). UvsX and UvsW are proteins that have been known to work together to repair broken replication forks through homologous recombination (Maher and Morrical, 2013; Kadyrov and Drake, 2004). Homologous recombination is one of the most efficient ways to have error free DNA repair and is beneficial to bacteriophages to have this repair mechanism. These repair mechanisms would be important to the bacteriophages because it would not only help repair broken replication forks but it would also help repair damaged or broken DNA (109, 110). It has been shown that the absence of UvsX increases UV sensitivity (110).

3.5 Conclusion

Agrican357virus genus of bacteriophages are *Myoviridae* with dsDNA, large capsids, long contractile tails and high GC content. Their genomes are nearly identical (>94% ANI). All three dot plots (whole genome, major capsid protein, and terminase protein) show no close similarity between the *Agrican357virus* family and any of the other bacteriophages on NCBI (see fig. 3.3A, 3.3B, and 3.3C). We have also found that the *Agrican357virus* cluster is more closely related to bacteriophages infecting *Pseudomonas* and *Ralstoni*a, than those infecting *E. amylovora*. The contrast that we observe between this cluster of bacteriophages and the distantly related bacteriophage analyzed by dot plot contributes valuable information about evolutionary relationships between these other clusters (see figure 3.3), suggesting the distant relationship may emphasize the importance of ecological niche, since most other *Enterobacteriales* bacteriophages isolated infect animal pathogens rather than plant pathogens. It may also, however, simply indicate the abundance of unstudied bacteriophages. The *Agrican357virus* family of bacteriophages is a novel family, with very low similarity to any other viruses, providing approximately 250 novel proteins to add to the viral dark matter that have no homolog by blastp (92). To understand a bacteriophage, it is vital to understand the encoded proteome. A bacteriophage's proteins determine how effectively it can infect bacteria, and how stable and safe it would be to use in a phage cocktail (a mixture of bacteriophages used together for phage therapy). Of the proteins with predicted function, this family encodes primarily DNA metabolism and repair proteins. Since the bacteriophage host, *E. amylovora*, is found primarily on the blossoms of fruit trees of the *Rosaceae* family, these proteins may be particularly vital due to the onset of UV radiation including putative thymidine and thymidylate kinases which aid the production of the nucleotide thymine for DNA synthesis (104), putative SbcC and SbcCD

proteins which protect against DNA damage by cleaving harmful hairpin structures during replication (99), putative exodeoxyribonuclease VIII which makes double stranded DNA breaks to help repair DNA damage at low energy (101), and putative UvsX recombinase and putative SF2 helicase which aid in repair and recombination of DNA (109). The small burst size we report herein for these jumbo bacteriophages (~4.6 functional virions), may require a high level of fidelity to ensure success in the environment.

A paper published in 2003 on evolutionary pathways of *P. aeruginosa* bacteria demonstrated that phiKZ-like bacteriophages have a very broad host range (111). In 1995, Campbell *et al* (112) isolated bacteriophages from barley rhizosphere that infected *Pseudomonas spp.* other than *P. aeruginosa*. These bacteriophages displayed great morphological similarity to phiKZ-like bacteriophages despite low genomic similarity (89; 111, 29). Similarly, *Agrican357virus* bacteriophages display proteomic similarity to phiKZ-like bacteriophages, particularly with their structural proteins, with little genomic synteny. These results suggest the phiKZ-like bacteriophages are highly divergent, derived from a common ancestor and successful in a wide range of ecological niches. It is highly likely that *Agrican357virus* family evolved through both mutational divergence and modular evolution (acquisition of larger regions of DNA, or modules), which is a common phenomenon in bacteriophages (113), and yet there is extremely low variance in all isolates thus far (>94% ANI). Such high conservation in these large genomes may reflect selective forces on a majority of the genome, which is for the most part uncharacterized. The great challenge ahead is both the abundance of bacteriophages that are completely uncharacterized, and the abundance of novel proteins harbored in their genomes.

3.6 Acknowledgements

Genomics and Evolutionary Science (SEA-PHAGES) for support and training on bacteriophage analysis, Steven Cresawn training on Phamerator, Majid Ghassemian and BMPMSF at UCSD for carrying out mass spectrometry, George W Sundin for donating *E. amylovora* strains, Edward Wilcox from the BYU Sequencing Center, Michael Standing from the BYU Microscopy Lab, David Parady and Brandon Ekins from Life Science IT at BYU for hosting DNAmaster and PhamDB on BYU server, Tsz Ching Tam for media preparation, Daniel Arens for helping with cloning and Yomesh Sharma for helping with the formatting of manuscript. We would also like to thank the following undergraduates from the BYU Phage hunters' program for their contributions to isolation of the bacteriophages: Garrett Jensen, Jared Kruger, Madison Melville, Trevon Galbraith, Savannah Grossarth, Hannah Ferguson and Austin Simister.

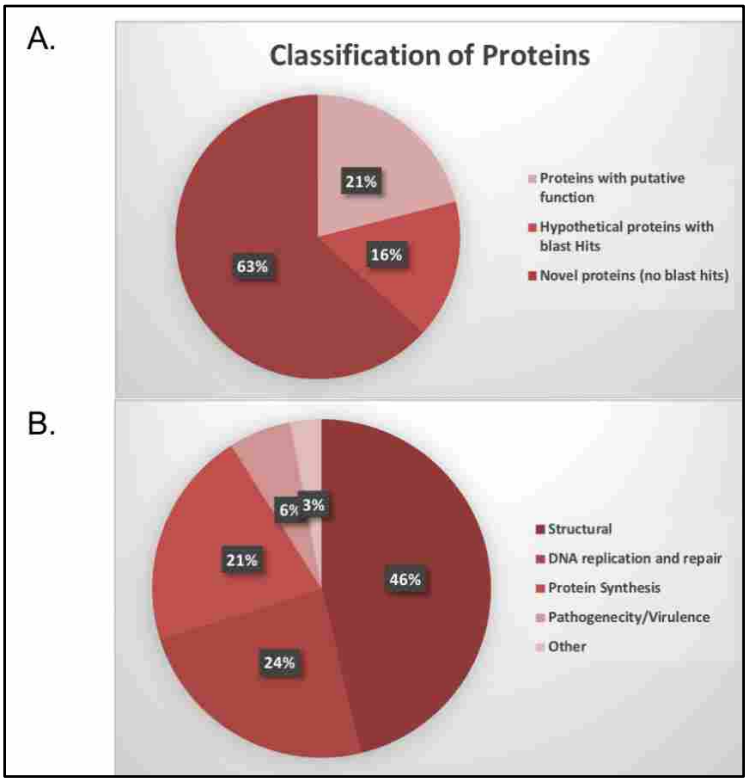
3.7 Supplementary data

Supplementary Table 3.S1 Average Nucleotide Identity of *Agrican357viruses*- Deimos-Minion, RAY, Special G, Simmy50, Desertfox, Mortimer, Bosolaphorus, MadMel and Ea35-70 suggests a single cluster of related bacteriophages with little or no similarity to other bacteriophages .
The genome sequences of all eight bacteriophages were compared against one phage from each potential cluster formed in whole genome dot plot analysis. Bacteriophages used in this analysis are *Erwinia* bacteriophages Deimos-Minion (KU886225) , RAY (KU886224), Simmy50 (KU886223), Special G (KU886222), Desertfox (MG655268), Bosolaphorus (MG6552687), MadMel (MG655269), Mortimer (MG655270), Ea35-70 (NC_023557), PhiEaH1 (NC_023610), Rexella (KX098390), Huxley (NC_031127), Yoloswag (KY448244), Joad (MF459647), SunLIRen (MH426725), phiEa21-4 (NC_011811) and Pavtok (MH426726) Salmonella phage SPN3US (NC_07402), Pseudomonas phage phiKZ (AF399011), 201phi2-1 (NC_010821) and Ralstonia phage RSL2 (AP014693).
Color coding based on decreasing order of similarity: dark grey (<100%) to light grey (<50%).

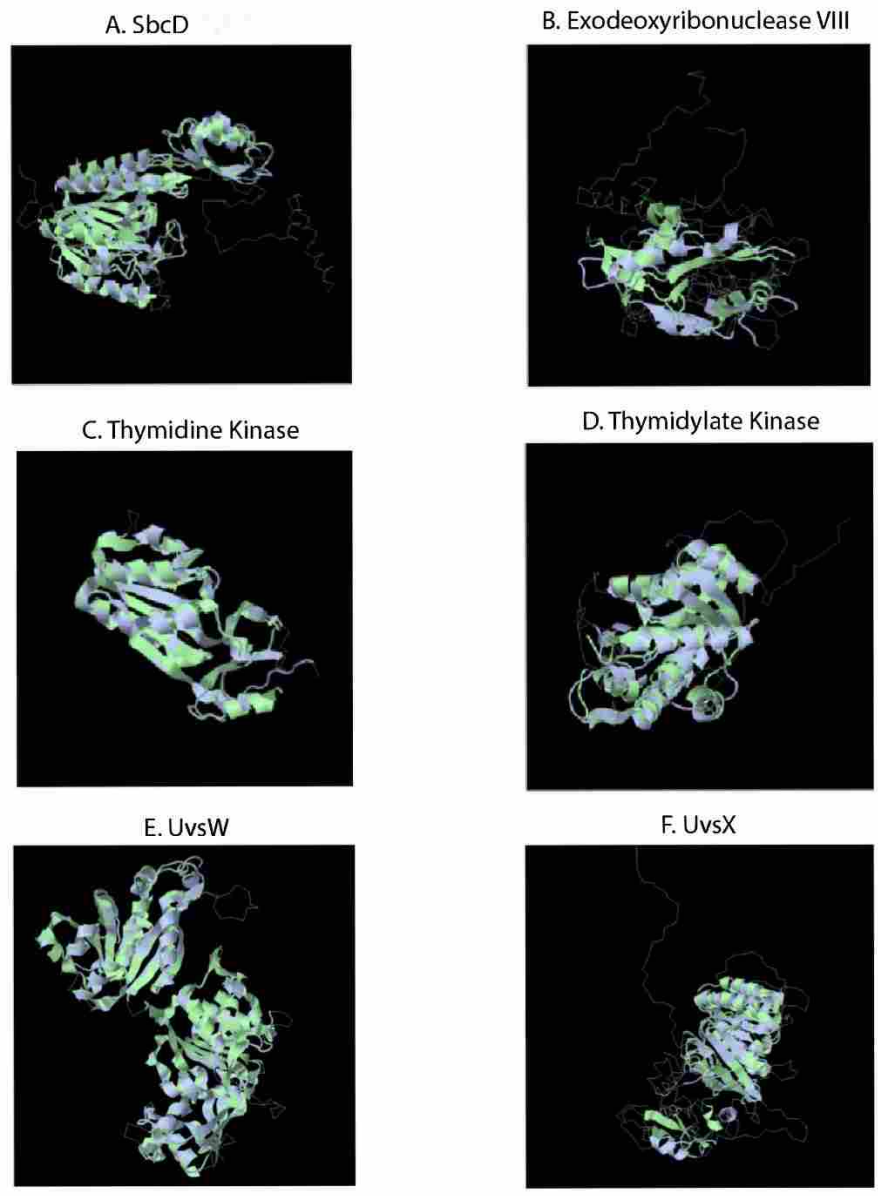
	Ea35-70	Special G	Simmy50	Deimos-Minion	RAY	Mortimer	Bosolaphorus	Desertfox	MadMel	Joad	SPN3US	Huxley	201phi2-1	Yoloswag	phiKZ	RSL2	phiEaH1	Rexella	SunLiRen	phiEa21-4	Pavtok
Ea35-70	100																				
Special G	95.3	100																			
Simmy50	95.5	96.7	100																		
Deimos-Minion	95.3	94.5	94.7	100																	
RAY	96	96.7	97.4	95.2	100																
Mortimer	95.6	96	96.6	95.3	96.9	100															
Bosolaphorus	94.8	95.7	96.2	95.9	96.8	97.1	100														
Desertfox	95.3	94.9	94.9	96.1	95.4	96.1	96.2	100													
MadMel	94.6	96.4	95.1	96.4	95.3	95.4	95.6	96.3	100												
Joad	33.4	33	33.1	33.1	33.3	32.9	33.1	33.1	32.8	100											
SPN3US	28	27.7	27.8	27.7	28	27.6	27.8	27.8	27.5	26.7	100										
Huxley	28.3	28	28	28	28.2	27.9	28	28.1	27.7	26.8	52.6	100									
201phi2-1	25.4	25.2	25.2	25.1	25.3	25.1	25.2	25.2	25	22.5	23.2	23.2	100								
Yoloswag	25.5	25.2	25.2	25.2	25.4	25.1	25.3	25.2	25	23.7	24.5	24.5	23.1	100							
phiKZ	24.4	24.2	24.2	24.2	24.4	24.2	24.2	24.2	24.1	22.5	23.2	23.1	24.8	24.4	100						
RSL2	24.4	24.1	24.1	24.1	24.3	24	24.2	24.1	23.9	23.8	24.7	25	21	24.5	21.5	100					
phiEaH1	20.7	20.5	20.5	20.5	20.7	20.5	20.6	20.6	20.4	19.6	20.2	20.3	19	20.4	35.1	19.3	100				
Rexella	15.8	15.6	15.6	15.6	15.7	15.5	15.6	15.6	15.5	12.8	10.3	10.3	7.7	8.6	7.9	9.9	7.8	100			
SunLiRen	9.8	9.7	9.8	9.7	9.8	9.7	9.8	9.8	9.7	10.2	10.2	10.2	15.4	9.4	8.9	10.4	8.5	7.7	100		
phiEa21-4	9.8	9.7	9.8	9.7	9.8	9.7	9.8	9.8	9.7	10.2	10.2	10.2	15.4	9.4	8.9	10.4	8.5	7.7	97.6	100	
Pavtok	7	6.9	6.9	6.9	7	6.9	7	7	6.9	7.1	7.3	7.4	5.7	13.6	6	8.1	6.6	6	6.1	6.1	100

Supplementary Table 3.S2 Putative gene products predicted to encode mechanisms of replication and DNA repair

Protein	Deimos-Minion	Simmy50	RAY	Special G	Mortimer	Desertfox	Bosolaphorus	MadMel
DNA repair proteins								
SbcC-like proteins	gp23	gp24	gp23	gp23	gp23	gp23	gp23	gp23
exodeoxyribonuclease VIII	gp67	gp67	gp64	gp67	gp66	gp64	gp65	gp66
SbcCD, D subunit	gp244	gp244	gp238	gp242	gp244	gp239	gp241	gp242
RADZ/SF2 Helicase	gp256	gp256	gp250	gp254	gp256	gp251	gp252	gp253
UvsX protein	gp155	gp153	gp150	gp152	gp154	gp150	gp151	gp151
Replication Proteins								
thymidine kinase	gp118	gp116	gp113	gp115	gp117	gp113	gp114	gp114
thymidylate kinase	gp317	gp316	gp311	gp315	gp318	gp313	gp314	gp314



Supplementary Figure 3.S1 Classification of proteins demonstrates the uniqueness of *Agrican357virus* bacteriophages due to A) abundance of novel proteins found in RAY and, B) by discovering majority of proteins with predicted function as structural genes. NCBI translated BLAST (blastx) was used to find the novel proteins in the genome of the bacteriophage RAY (BLAST hit of an e-value less than 1e-04 and no hit outside of the Agrican357virus bacteriophages).



Supplementary Figure 3.S2 Predicting putative protein structure of interesting proteins from *Agrican357virus* family via Raptor . Conserved domains for all bacteriophages in the family were found using the NCBI Conserved Domain Database (1-3) with the acceptable return threshold set at E-value < 3e-5. RaptorX (4-6) was used to predict tertiary structure and binding sites and to produce the possible images of Agrican357virus proteins. These predicted structures were used to show similarity between putative and known proteins as evidence that these proteins may indeed perform the given putative functions (4, 5, 7)The predicted fold of the proteins from the phage RAY is shown in blue and the known crystallography structures of these proteins are shown in green A) SbcCD protein of RAY with MRE11 B) Exodeoxyribonuclease VIII protein of RAY with exodeoxyribonuclease VIII of E. coli C) thymidine kinase of RAY with thymidine kinase from *Thermotoga maritime* D) Thymidylate kinase of RAY with thymidylate kinase of *Mycobacterium tuberculosis*. E) UvsW and F) UvsX of RAY with UvsW and UvsX from T4 help confirm putative functions.

3.7.1 References for supplementary material

1. Marchler-Bauer, A., J. B. Anderson, C. DeWeese-Scott, N. D. Fedorova, L. Y. Geer, S. He, D. I. Hurwitz*, et al.* "Cdd: A Curated Entrez Database of Conserved Domain Alignments." [In eng]. *Nucleic Acids Res* 31, no. 1 (Jan 1 2003): 383-7.

2. Marchler-Bauer, A., J. B. Anderson, P. F. Cherukuri, C. DeWeese-Scott, L. Y. Geer, M. Gwadz, S. He*, et al.* "Cdd: A Conserved Domain Database for Protein Classification." [In eng]. *Nucleic Acids Res* 33, no. Database issue (Jan 1 2005): D192-6.

3. Marchler-Bauer, A., M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R. C. Geer*, et al.* "Cdd: Ncbi's Conserved Domain Database." [In eng]. *Nucleic Acids Res* 43, no. Database issue (Jan 2015): D222-6.

4. Wang, S., J. Peng, and J. Xu. "Alignment of Distantly Related Protein Structures: Algorithm, Bound and Implications to Homology Modeling." [In eng]. *Bioinformatics* 27, no. 18 (Sep 15 2011): 2537-45.

5. Wang, S., J. Ma, J. Peng, and J. Xu. "Protein Structure Alignment Beyond Spatial Proximity." [In eng]. *Sci Rep* 3 (2013): 1448.

6. Peng, J., and J. Xu. "Raptorx: Exploiting Structure Information for Protein Alignment by Statistical Inference." [In eng]. *Proteins* 79 Suppl 10 (2011): 161-71.

7. Kallberg, M., H. Wang, S. Wang, J. Peng, Z. Wang, H. Lu, and J. Xu. "Template-Based Protein Structure Modeling Using the Raptorx Web Server." [In eng]. *Nat Protoc* 7, no. 8 (Jul 19 2012): 1511-2

REFERENCES

1. Whitman, W. B., D. C. Coleman, and W. J. Wiebe. "Prokaryotes: The Unseen Majority." [In eng]. *Proc Natl Acad Sci U S A* 95, no. 12 (Jun 9 1998): 6578-83.

2. Wommack, K. E., and R. R. Colwell. "Virioplankton: Viruses in Aquatic Ecosystems." [In eng]. *Microbiol Mol Biol Rev* 64, no. 1 (Mar 2000): 69-114.

3. Bergh, O., K. Y. Borsheim, G. Bratbak, and M. Heldal. "High Abundance of Viruses Found in Aquatic Environments." [In eng]. *Nature* 340, no. 6233 (Aug 10 1989): 467-8.

4. Hambly, E., and C. A. Suttle. "The Viriosphere, Diversity, and Genetic Exchange within Phage Communities." [In eng]. *Curr Opin Microbiol* 8, no. 4 (Aug 2005): 444-50.

5. Abedon, S. T., C. Thomas-Abedon, A. Thomas, and H. Mazure. "Bacteriophage Prehistory: Is or Is Not Hankin, 1896, a Phage Reference?" [In eng]. *Bacteriophage* 1, no. 3 (May 2011): 174-78.

6. Twort, F.W., L.R.C.P. LOND., M.R.C.S. "An Investigation on the Nature of Ultra-Microscopic Viruses." *The Lancet* 186, no. 4814 (1915): 1241-43.

7. Twort, F. W. "Further Investigations on the Nature of Ultra-Microscopic Viruses and Their Cultivation." [In eng]. *J Hyg (Lond)* 36, no. 2 (Jun 1936): 204-35.

8. D'Herelle, F. "On an Invisible Microbe Antagonistic toward Dysenteric Bacilli: Brief Note by Mr. F. D'herelle, Presented by Mr. Roux. 1917." [In eng]. *Res Microbiol* 158, no. 7 (Sep 2007): 553-4.

9. Keen, E. C. "A Century of Phage Research: Bacteriophages and the Shaping of Modern Biology." [In eng]. *Bioessays* 37, no. 1 (Jan 2015): 6-9.

10. Chen, J., and R. P. Novick. "Phage-Mediated Intergeneric Transfer of Toxin Genes." [In eng]. *Science* 323, no. 5910 (Jan 2 2009): 139-41.

11. Bollback, J. P., and J. P. Huelsenbeck. "Parallel Genetic Evolution within and between Bacteriophage Species of Varying Degrees of Divergence." [In eng]. *Genetics* 181, no. 1 (Jan 2009): 225-34.

12. Boyd, E. F. "Bacteriophage-Encoded Bacterial Virulence Factors and Phage-Pathogenicity Island Interactions." [In eng]. *Adv Virus Res* 82 (2012): 91-118.

13. Shivaswamy, V. C., S. B. Kalasuramath, C. K. Sadanand, A. K. Basavaraju, V. Ginnavaram, S. Bille, S. S. Ukken, and U. N. Pushparaj. "Ability of Bacteriophage in Resolving Wound Infection Caused by Multidrug-Resistant Acinetobacter Baumannii in Uncontrolled Diabetic Rats." [In eng]. *Microb Drug Resist* 21, no. 2 (Apr 2015): 171-7.

14. Sulakvelidze, A. "Phage Therapy: An Attractive Option for Dealing with Antibiotic-Resistant Bacterial Infections." [In eng]. *Drug Discov Today* 10, no. 12 (Jun 15 2005): 807-9.

15. Zwet TVD, Keil HL. *Fireblight: A Bacterial Disease of Rosaceous Plants.* 1979.

16. SV, Thomson. *Epidemiology of Fire Blight.* Fire Blight; the Disease and Its Causative Agent, Erwinia Amylovora. Edited by Vanneste JL. Wallingford, United Kingdom: CABI Publishing, 2000.

17. Esplin, I. N. D., J. A. Berg, R. Sharma, R. C. Allen, D. K. Arens, C. R. Ashcroft, S. R. Bairett*, et al.* "Genome Sequences of 19 Novel Erwinia Amylovora Bacteriophages." [In eng]. *Genome Announc* 5, no. 46 (Nov 16 2017).

18. Sharma R, Berg JA, Beatty NJ, Choi MC, Cowger AE, Cozzens BJR, Duncan SG, Fajardo CP, Ferguson HP, Galbraith T, Herring JA, Hoj TR, Durrant JL, Hyde JR, Jensen GL, Ke SY, Killpack S, Kruger JL, Lawrence EEK, Nwosu IO, Tam TC, Thompson DW, Tueller JA, Ward MEH, Webb CJ, Wood ME, Yeates EL, Baltrus DA, Breakwell DP, Hope S, Grose JH. "Genome Sequences of Nine Erwinia Amylovora Bacteriophages." *Microbiology Resource Announcements* 7 (October 11, 2018 2018).

19. Yagubi, A. I., A. J. Castle, A. M. Kropinski, T. W. Banks, and A. M. Svircev. "Complete Genome Sequence of Erwinia Amylovora Bacteriophage Vb_Eamm_Ea35-70." [In eng]. *Genome Announc* 2, no. 4 (Aug 21 2014).

20. Adams, M. J., E. J. Lefkowitz, A. M. Q. King, B. Harrach, R. L. Harrison, N. J. Knowles, A. M. Kropinski*, et al.* "Changes to Taxonomy and the International Code of Virus Classification and Nomenclature Ratified by the International Committee on Taxonomy of Viruses (2017)." [In eng]. *Arch Virol* 162, no. 8 (Aug 2017): 2505-38.

21. Yuan, Y., and M. Gao. "Jumbo Bacteriophages: An Overview." [In eng]. *Front Microbiol* 8 (2017): 403.

22. Effantin, G., R. Hamasaki, T. Kawasaki, M. Bacia, C. Moriscot, W. Weissenhorn, T. Yamada, and G. Schoehn. "Cryo-Electron Microscopy Three-Dimensional Structure of the Jumbo Phage Phirsl1 Infecting the Phytopathogen Ralstonia Solanacearum." [In eng]. *Structure* 21, no. 2 (Feb 5 2013): 298-305.

23. Yuan, Y., and M. Gao. "Proteomic Analysis of a Novel Bacillus Jumbo Phage Revealing Glycoside Hydrolase as Structural Component." [In eng]. *Front Microbiol* 7 (2016): 745.

24. Thomas, J. A., M. R. Rolando, C. A. Carroll, P. S. Shen, D. M. Belnap, S. T. Weintraub, P. Serwer, and S. C. Hardies. "Characterization of Pseudomonas Chlororaphis Myovirus 201varphi2-1 Via Genomic Sequencing, Mass Spectrometry, and Electron Microscopy." [In eng]. *Virology* 376, no. 2 (Jul 5 2008): 330-8.

25. Danis-Wlodarczyk, K., D. Vandenheuvel, H. B. Jang, Y. Briers, T. Olszak, M. Arabski, S. Wasik*, et al.* "A Proposed Integrated Approach for the Preclinical Evaluation of Phage Therapy in Pseudomonas Infections." [In eng]. *Sci Rep* 6 (Jun 15 2016): 28115.

26. Monson, R., I. Foulds, J. Foweraker, M. Welch, and G. P. Salmond. "The Pseudomonas Aeruginosa Generalized Transducing Phage Phipa3 Is a New Member of the Phikz-Like Group of 'Jumbo' Phages, and Infects Model Laboratory Strains and Clinical Isolates from Cystic Fibrosis Patients." [In eng]. *Microbiology* 157, no. Pt 3 (Mar 2011): 859-67.

27. Bhunchoth, A., R. Blanc-Mathieu, T. Mihara, Y. Nishimura, A. Askora, N. Phironrit, C. Leksomboon*, et al.* "Two Asian Jumbo Phages, Varphirsl2 and Varphirsf1, Infect Ralstonia Solanacearum and Show Common Features of Varphikz-Related Phages." [In eng]. *Virology* 494 (Jul 2016): 56-66.

28. Shaburova, O. V., K. Hertveldt, D. M. de la Crus, S. V. Krylov, E. A. Pleteneva, M. V. Burkaltseva, R. Lavigne, G. Volcaert, and V. N. Krylov. "[Comparison of New Giant Bacteriophages Obp and Lu11 of Soil Pseudomonads with Bacteriophages of Phikz-Supergroup of Pseudomonas Aeruginosa]." [In rus]. *Genetika* 42, no. 8 (Aug 2006): 1065-74.

29. Sokolova, O. S., O. V. Shaburova, E. V. Pechnikova, A. K. Shaytan, S. V. Krylov, N. A. Kiselev, and V. N. Krylov. "Genome Packaging in El and Lin68, Two Giant Phikz-Like Bacteriophages of P. Aeruginosa." [In eng]. *Virology* 468-470 (Nov 2014): 472-78.

30. Lecoutere, E., P. J. Ceyssens, K. A. Miroshnikov, V. V. Mesyanzhinov, V. N. Krylov, J. P. Noben, J. Robben*, et al.* "Identification and Comparative Analysis of the Structural Proteomes of Phikz and El, Two Giant Pseudomonas Aeruginosa Bacteriophages." [In eng]. *Proteomics* 9, no. 11 (Jun 2009): 3215-9.

31. Yamada, T., S. Satoh, H. Ishikawa, A. Fujiwara, T. Kawasaki, M. Fujie, and H. Ogata. "A Jumbo Phage Infecting the Phytopathogen Ralstonia Solanacearum Defines a New Lineage of the Myoviridae Family." [In eng]. *Virology* 398, no. 1 (Mar 1 2010): 135-47.

32. Kurochkina, L. P., P. I. Semenyuk, N. N. Sykilinda, and K. A. Miroshnikov. "The Unique Two-Component Tail Sheath of Giant Pseudomonas Phage Pabg." [In eng]. *Virology* 515 (Feb 2018): 46-51.

33. Lee, J. H., J. Bai, H. Shin, Y. Kim, B. Park, S. Heu, and S. Ryu. "A Novel Bacteriophage Targeting Cronobacter Sakazakii Is a Potential Biocontrol Agent in Foods." [In eng]. *Appl Environ Microbiol* 82, no. 1 (Jan 1 2016): 192-201.

34. Sullivan, M. B., M. L. Coleman, P. Weigele, F. Rohwer, and S. W. Chisholm. "Three Prochlorococcus Cyanophage Genomes: Signature Features and Ecological Interpretations." [In eng]. *PLoS Biol* 3, no. 5 (May 2005): e144.

35. Miller, E. S., J. F. Heidelberg, J. A. Eisen, W. C. Nelson, A. S. Durkin, A. Ciecko, T. V. Feldblyum*, et al.* "Complete Genome Sequence of the Broad-Host-Range Vibriophage

Kvp40: Comparative Genomics of a T4-Related Bacteriophage." [In eng]. *J Bacteriol* 185, no. 17 (Sep 2003): 5220-33.

36. Gibb, E. A., and D. R. Edgell. "Multiple Controls Regulate the Expression of Mobe, an Hnh Homing Endonuclease Gene Embedded within a Ribonucleotide Reductase Gene of Phage Aeh1." [In eng]. *J Bacteriol* 189, no. 13 (Jul 2007): 4648-61.

37. Kim, J. H., J. S. Son, Y. J. Choi, C. H. Choresca, Jr., S. P. Shin, J. E. Han, J. W. Jun, and S. C. Park. "Complete Genome Sequence and Characterization of a Broad-Host Range T4-Like Bacteriophage Phias5 Infecting Aeromonas Salmonicida Subsp. Salmonicida." [In eng]. *Vet Microbiol* 157, no. 1-2 (May 25 2012): 164-71.

38. Buttimer, C., H. Hendrix, H. Oliveira, A. Casey, H. Neve, O. McAuliffe, R. P. Ross*, et al.* "Things Are Getting Hairy: Enterobacteria Bacteriophage Vb_Pcam_Cbb." [In eng]. *Front Microbiol* 8 (2017): 44.

39. Gill, J. J., J. D. Berry, W. K. Russell, L. Lessor, D. A. Escobar-Garcia, D. Hernandez, A. Kane*, et al.* "The Caulobacter Crescentus Phage Phicbk: Genomics of a Canonical Phage." [In eng]. *BMC Genomics* 13 (Oct 10 2012): 542.

40. Arens, D. K., T. S. Brady, J. L. Carter, J. A. Pape, D. M. Robinson, K. A. Russell, L. A. Staley*, et al.* "Characterization of Two Related Erwinia Myoviruses That Are Distant Relatives of the Phikz-Like Jumbo Phages." [In eng]. *PLoS One* 13, no. 7 (2018): e0200202.

41. Simoliunas, E., L. Kaliniene, L. Truncaite, A. Zajanckauskaite, J. Staniulis, A. Kaupinis, M. Ger, M. Valius, and R. Meskys. "Klebsiella Phage Vb_Klem-Rak2 - a Giant Singleton Virus of the Family Myoviridae." [In eng]. *PLoS One* 8, no. 4 (2013): e60717.

42. Abbasifar, R., M. W. Griffiths, P. M. Sabour, H. W. Ackermann, K. Vandersteegen, R. Lavigne, J. P. Noben*, et al.* "Supersize Me: Cronobacter Sakazakii Phage Gap32." [In eng]. *Virology* 460-461 (Jul 2014): 138-46.

43. Thomas, J. A., S. C. Hardies, M. Rolando, S. J. Hayes, K. Lieman, C. A. Carroll, S. T. Weintraub, and P. Serwer. "Complete Genomic Sequence and Mass Spectrometric Analysis of Highly Diverse, Atypical Bacillus Thuringiensis Phage 0305phi8-36." [In eng]. *Virology* 368, no. 2 (Nov 25 2007): 405-21.

44. Lavysh, D., M. Sokolova, M. Slashcheva, K. U. Forstner, and K. Severinov. "Transcription Profiling of Bacillus Subtilis Cells Infected with Ar9, a Giant Phage Encoding Two Multisubunit Rna Polymerases." [In eng]. *MBio* 8, no. 1 (Feb 14 2017).

45. Zhao, Y., and M. Qi. "Comparative Genomics of Erwinia Amylovora and Related Erwinia Species-What Do We Learn?" [In eng]. *Genes (Basel)* 2, no. 3 (Sep 15 2011): 627-39

46. Abramoff, M.D., P.J. Magalhaes, and S.J. Ram. "Image Processing with Imagej." no. Biophotonics International (2004).

47. Delbruck, M. "The Burst Size Distribution in the Growth of Bacterial Viruses (Bacteriophages)." [In eng]. *J Bacteriol* 50, no. 2 (Aug 1945): 131-5.

48. Hockett, K. L., and D. A. Baltrus. "Use of the Soft-Agar Overlay Technique to Screen for Bacterially Produced Inhibitory Compounds." [In eng]. *J Vis Exp*, no. 119 (Jan 14 2017).

49. Ritchie, D.F , and Klos E. J. "Isolation of Erwinia Amylovora Bacteriophage from Aerial Parts of Apple Trees." *Phytopathology* 67 (1976): 101-04.

50. McGhee, G. C., and G. W. Sundin. "Erwinia Amylovora Crispr Elements Provide New Tools for Evaluating Strain Diversity and for Microbial Source Tracking." [In eng]. *PLoS One* 7, no. 7 (2012): e41706.

51. Sebaihia, M., A. M. Bocsanczy, B. S. Biehl, M. A. Quail, N. T. Perna, J. D. Glasner, G. A. DeClerck*, et al.* "Complete Genome Sequence of the Plant Pathogen Erwinia Amylovora Strain Atcc 49946." [In eng]. *J Bacteriol* 192, no. 7 (Apr 2010): 2020-1.

52. Pusey, P. L. "Crab Apple Blossoms as a Model for Research on Biological Control of Fire Blight." [In eng]. *Phytopathology* 87, no. 11 (Nov 1997): 1096-102.

53. Ishimaru, C.A, E. J. Klos, and R. R. Brubaker. "Multiple Antibiotic Production by Erwinia Herbicola." *Phytopathology* 78 (1987): 746-50.

54. Smits, T. H., F. Rezzonico, T. Kamber, A. Goesmann, C. A. Ishimaru, V. O. Stockwell, J. E. Frey, and B. Duffy. "Genome Sequence of the Biocontrol Agent Pantoea Vagans Strain C9-1." [In eng]. *J Bacteriol* 192, no. 24 (Dec 2010): 6486-7.

55. Datsenko, K. A., and B. L. Wanner. "One-Step Inactivation of Chromosomal Genes in Escherichia Coli K-12 Using Pcr Products." *Proc Natl Acad Sci U S A* 97, no. 12 (Jun 6 2000): 6640-5.

56. Gong, S., S. W. Bearden, V. A. Geoffroy, J. D. Fetherston, and R. D. Perry. "Characterization of the Yersinia Pestis Yfu Abc Inorganic Iron Transport System." [In eng]. *Infect Immun* 69, no. 5 (May 2001): 2829-37.

57. Fetherston, J. D., P. Schuetze, and R. D. Perry. "Loss of the Pigmentation Phenotype in Yersinia Pestis Is Due to the Spontaneous Deletion of 102 Kb of Chromosomal DNA Which Is Flanked by a Repetitive Element." [In eng]. *Mol Microbiol* 6, no. 18 (Sep 1992): 2693-704.

58. Ren, Y., Y. Ren, Z. Zhou, X. Guo, Y. Li, L. Feng, and L. Wang. "Complete Genome Sequence of Enterobacter Cloacae Subsp. Cloacae Type Strain Atcc 13047." [In eng]. *J Bacteriol* 192, no. 9 (May 2010): 2463-4.

59. Otman, J, M.E. Perugini, M.C.B. Tognim, and M.C. Vidotto. "Atypical Phenotypic Characteristics of Klebsiella Pneumoniae Isolates from an Outbreak in a Neonatal Intensive Care Unit in Brazil." *Brazilian journal of microbiology* 38 (2007): 273-77.

60. Nakamura, L. K., M. S. Roberts, and F. M. Cohan. "Relationship of Bacillus Subtilis Clades Associated with Strains 168 and W23: A Proposal for Bacillus Subtilis Subsp. Subtilis Subsp. Nov. And Bacillus Subtilis Subsp. Spizizenii Subsp. Nov." [In eng]. *Int J Syst Bacteriol* 49 Pt 3 (Jul 1999): 1211-5

61. Gicova, A., M. Orieskova, L. Oslanecova, H. Drahovska, and E. Kaclikova. "Identification and Characterization of Cronobacter Strains Isolated from Powdered Infant Foods." [In eng]. *Lett Appl Microbiol* 58, no. 3 (Mar 2014): 242-7.

62. Iversen, C., A. Lehner, N. Mullane, E. Bidlas, I. Cleenwerck, J. Marugg, S. Fanning, R. Stephan, and H. Joosten. "The Taxonomy of Enterobacter Sakazakii: Proposal of a New Genus Cronobacter Gen. Nov. And Descriptions of Cronobacter Sakazakii Comb. Nov. Cronobacter Sakazakii Subsp. Sakazakii, Comb. Nov., Cronobacter Sakazakii Subsp. Malonaticus Subsp. Nov., Cronobacter Turicensis Sp. Nov., Cronobacter Muytjensii Sp. Nov., Cronobacter Dublinensis Sp. Nov. And Cronobacter Genomospecies 1." [In eng]. *BMC Evol Biol* 7 (Apr 17 2007): 64.

63. Cartagena, E., O. A. Colom, A. Neske, J. C. Valdez, and A. Bardon. "Effects of Plant Lactones on the Production of Biofilm of Pseudomonas Aeruginosa." [In eng]. *Chem Pharm Bull (Tokyo)* 55, no. 1 (Jan 2007): 22-5.

64. Conway, H. F., W. C. Haynes, R. W. Jackson, J. M. Locke, T. G. Pridham, V. E. Sohns, and F. H. Stodola. "Pseudomonas Aureofaciens Kluyver and Phenazine Alpha-Carboxylic Acid, Its Characteristic Pigment." [In eng]. *J Bacteriol* 72, no. 3 (Sep 1956): 412-7.

65. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. "Basic Local Alignment Search Tool." [In eng]. *J Mol Biol* 215, no. 3 (Oct 5 1990): 403-10.

66. Gish, W., and D. J. States. "Identification of Protein Coding Regions by Database Similarity Search." [In eng]. *Nat Genet* 3, no. 3 (Mar 1993): 266-72.

67. Madden, T. L., R. L. Tatusov, and J. Zhang. "Applications of Network Blast Server." [In eng]. *Methods Enzymol* 266 (1996): 131-41.

68. Krumsiek, J., R. Arnold, and T. Rattei. "Gepard: A Rapid and Sensitive Tool for Creating Dotplots on Genome Scale." [In eng]. *Bioinformatics* 23, no. 8 (Apr 15 2007): 1026-8.

69. Lamine, J. G., R. J. DeJong, and S. M. Nelesen. "Phamdb: A Web-Based Application for Building Phamerator Databases." [In eng]. *Bioinformatics* 32, no. 13 (Jul 1 2016): 2026-8.

70. Cresawn, S. G., M. Bogel, N. Day, D. Jacobs-Sera, R. W. Hendrix, and G. F. Hatfull. "Phamerator: A Bioinformatic Tool for Comparative Bacteriophage Genomics." [In eng]. *BMC Bioinformatics* 12 (Oct 12 2011): 395.

71. Hauser, M., C. E. Mayer, and J. Soding. "Kclust: Fast and Sensitive Clustering of Large Protein Sequence Databases." [In eng]. *BMC Bioinformatics* 14 (Aug 15 2013): 248.

72. Huson, D. H., and D. Bryant. "Application of Phylogenetic Networks in Evolutionary Studies." [In eng]. *Mol Biol Evol* 23, no. 2 (Feb 2006): 254-67.

73. Katoh, K., K. Misawa, K. Kuma, and T. Miyata. "Mafft: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform." [In eng]. *Nucleic Acids Res* 30, no. 14 (Jul 15 2002): 3059-66.

74. Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton*, et al.* "Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data." [In eng]. *Bioinformatics* 28, no. 12 (Jun 15 2012): 1647-9.

75. Zuckerkandl E ; Pauling, L. *Evolutionary Divergence and Convergence in Proteins. Edited in Evolving Genes and Proteins.* Evolving Genes and Proteins. Edited by V. Bryson and H.J. Vogel. New York: Academic Press, 1965.

76. Felsenstein, J. "Confidence Limits on Phylogenies: An Approach Using the Bootstrap." [In eng]. *Evolution* 39, no. 4 (Jul 1985): 783-91.

77. Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura. "Mega X: Molecular Evolutionary Genetics Analysis across Computing Platforms." [In eng]. *Mol Biol Evol* 35, no. 6 (Jun 1 2018): 1547-49.

78. Guttman, M., G. N. Betts, H. Barnes, M. Ghassemian, P. van der Geer, and E. A. Komives. "Interactions of the Npxy Microdomains of the Low Density Lipoprotein Receptor-Related Protein 1." [In eng]. *Proteomics* 9, no. 22 (Nov 2009): 5016-28.

79. McCormack, A. L., D. M. Schieltz, B. Goode, S. Yang, G. Barnes, D. Drubin, and J. R. Yates, 3rd. "Direct Analysis and Identification of Proteins in Mixtures by Lc/Ms/Ms and Database Searching at the Low-Femtomole Level." [In eng]. *Anal Chem* 69, no. 4 (Feb15 1997): 767-76.

80. yuayuy, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li, and W. S. Noble. "Meme Suite: Tools for Motif Discovery and Searching." [In eng]. *Nucleic Acids Res* 37, no. Web Server issue (Jul 2009): W202-8.

81. Grant, C. E., T. L. Bailey, and W. S. Noble. "Fimo: Scanning for Occurrences of a Given Motif." [In eng]. *Bioinformatics* 27, no. 7 (Apr 1 2011): 1017-8.

82. Berg, J. A., B. D. Merrill, J. T. Crockett, K. P. Esplin, M. R. Evans, K. E. Heaton, J. A. Hilton*, et al.* "Characterization of Five Novel Brevibacillus Bacteriophages and Genomic Comparison of Brevibacillus Phages." [In eng]. *PLoS One* 11, no. 6 (2016): e0156838.

83. Lowe, T. M., and P. P. Chan. "Trnascan-Se on-Line: Integrating Search and Context for Analysis of Transfer Rna Genes." [In eng]. *Nucleic Acids Res* 44, no. W1 (Jul 8 2016): W54-7.

84. Deletoile, A., D. Decre, S. Courant, V. Passet, J. Audo, P. Grimont, G. Arlet, and S. Brisse. "Phylogeny and Identification of Pantoea Species and Typing of Pantoea Agglomerans Strains by Multilocus Gene Sequencing." [In eng]. *J Clin Microbiol* 47, no. 2 (Feb 2009): 300-10.

85. Palmer, M., P. de Maayer, M. Poulsen, E. T. Steenkamp, E. van Zyl, T. A. Coutinho, and S. N. Venter. "Draft Genome Sequences of Pantoea Agglomerans and Pantoea Vagans Isolates Associated with Termites." [In eng]. *Stand Genomic Sci* 11 (2016): 23.

86. Lassmann, T., and E. L. Sonnhammer. "Kalign--an Accurate and Fast Multiple Sequence Alignment Algorithm." [In eng]. *BMC Bioinformatics* 6 (Dec 12 2005): 298.

87. Smith, K. C., E. Castro-Nallar, J. N. Fisher, D. P. Breakwell, J. H. Grose, and S. H. Burnett. "Phage Cluster Relationships Identified through Single Gene Analysis." [In eng]. *BMC Genomics* 14 (Jun 19 2013): 410.

88. Kuebler, D., and V. B. Rao. "Functional Analysis of the DNA-Packaging/Terminase Protein Gp17 from Bacteriophage T4." [In eng]. *J Mol Biol* 281, no. 5 (Sep 4 1998): 803-14.

89. Mesyanzhinov, V. V., J. Robben, B. Grymonprez, V. A. Kostyuchenko, M. V. Bourkaltseva, N. N. Sykilinda, V. N. Krylov, and G. Volckaert. "The Genome of Bacteriophage Phikz of Pseudomonas Aeruginosa." [In eng]. *J Mol Biol* 317, no. 1 (Mar 15 2002): 1-19.

90. Merrill, B. D., A. T. Ward, J. H. Grose, and S. Hope. "Software-Based Analysis of Bacteriophage Genomes, Physical Ends, and Packaging Strategies." [In eng]. *BMC Genomics* 17 (Aug 26 2016): 679.

91. Garneau, Julian R, Florence Depardieu, Louis-Charles Fortier, David Bikard, and Marc Monot. "Phageterm: A Fast and User-Friendly Software to Determine Bacteriophage Termini and Packaging Mode Using Randomly Fragmented Ngs Data." *bioRxiv* (2017).

92. Hatfull, G. F. "Dark Matter of the Biosphere: The Amazing World of Bacteriophage Diversity." [In eng]. *J Virol* 89, no. 16 (Aug 2015): 8107-10.

93. Chevalier, B. S., and B. L. Stoddard. "Homing Endonucleases: Structural and Functional Insight into the Catalysts of Intron/Intein Mobility." [In eng]. *Nucleic Acids Res* 29, no. 18 (Sep 15 2001): 3757-74.

94. Hertveldt, K., R. Lavigne, E. Pleteneva, N. Sernova, L. Kurochkina, R. Korchevskii, J. Robben*, et al.* "Genome Comparison of Pseudomonas Aeruginosa Large Phages." [In eng]. *J Mol Biol* 354, no. 3 (Dec 2 2005): 536-45.

95. Kim, W. S., and K. Geider. "Characterization of a Viral Eps-Depolymerase, a Potential Tool for Control of Fire Blight." [In eng]. *Phytopathology* 90, no. 11 (Nov 2000): 1263-8.

96. Emrich, J., and G. Streisinger. "The Role of Phage Lysozyme in the Life Cycle of Phage T4." [In eng]. *Virology* 36, no. 3 (Nov 1968): 387-91.

97. Cornelissen, A., P. J. Ceyssens, V. N. Krylov, J. P. Noben, G. Volckaert, and R. Lavigne. "Identification of Eps-Degrading Activity within the Tail Spikes of the Novel Pseudomonas Putida Phage Af." [In eng]. *Virology* 434, no. 2 (Dec 20 2012): 251-6.

98. Majkowska-Skrobek, G., A. Latka, R. Berisio, B. Maciejewska, F. Squeglia, M. Romano, R. Lavigne, C. Struve, and Z. Drulis-Kawa. "Capsule-Targeting Depolymerase, Derived from Klebsiella Kp36 Phage, as a Tool for the Development of Anti-Virulent Strategy." [In eng]. *Viruses* 8, no. 12 (Dec 1 2016).

99. Connelly, J. C., L. A. Kirkham, and D. R. Leach. "The Sbccd Nuclease of Escherichia Coli Is a Structural Maintenance of Chromosomes (Smc) Family Protein That Cleaves Hairpin DNA." [In eng]. *Proc Natl Acad Sci U S A* 95, no. 14 (Jul 7 1998): 7969-74.

100. Leach, D. R., E. A. Okely, and D. J. Pinder. "Repair by Recombination of DNA Containing a Palindromic Sequence." [In eng]. Mol Microbiol 26, no. 3 (Nov 1997): 597-606.

101. Joseph, J. W., and R. Kolodner. "Exonuclease Viii of Escherichia Coli. I. Purification and Physical Properties." [In eng]. *J Biol Chem* 258, no. 17 (Sep 10 1983): 10411-7

102. Kolodner, R., S. D. Hall, and C. Luisi-DeLuca. "Homologous Pairing Proteins Encoded by the Escherichia Coli Rece and Rect Genes." [In eng]. *Mol Microbiol* 11, no. 1 (Jan 1994): 23-30.

103. Kokoris, M. S., and M. E. Black. "Characterization of Herpes Simplex Virus Type 1 Thymidine Kinase Mutants Engineered for Improved Ganciclovir or Acyclovir Activity." [In eng]. *Protein Sci* 11, no. 9 (Sep 2002): 2267-72.

104. Doharey, P. K., S. K. Singh, P. Verma, A. Verma, S. Rathaur, and J. K. Saxena. "Insights into the Structure-Function Relationship of Brugia Malayi Thymidylate Kinase (Bmtmk)." [In eng]. *Int J Biol Macromol* 88 (Jul 2016): 565-71.

105. Chaudhary, S. K., J. Jeyakanthan, and K. Sekar. "Cloning, Expression, Purification, Crystallization and Preliminary X-Ray Crystallographic Study of Thymidylate Kinase (Ttha1607) from Thermus Thermophilus Hb8." [In eng]. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 69, no. Pt 2 (Feb 1 2013): 118-21.

106. Cui, Q., W. S. Shin, Y. Luo, J. Tian, H. Cui, and D. Yin. "Thymidylate Kinase: An Old Topic Brings New Perspectives." [In eng]. *Curr Med Chem* 20, no. 10 (2013): 1286-305.

107. Gajewski, S., M. R. Webb, V. Galkin, E. H. Egelman, K. N. Kreuzer, and S. W. White. "Crystal Structure of the Phage T4 Recombinase Uvsx and Its Functional Interaction with the T4 Sf2 Helicase Uvsw." [In eng]. *J Mol Biol* 405, no. 1 (Jan 7 2011): 65-76.

108. Madej, T., J. F. Gibrat, and S. H. Bryant. "Threading a Database of Protein Cores." [In eng]. *Proteins* 23, no. 3 (Nov 1995): 356-69.

109. Maher, R. L., and S. W. Morrical. "Coordinated Binding of Single-Stranded and Double-Stranded DNA by Uvsx Recombinase." [In eng]. *PLoS One* 8, no. 6 (2013): e66654.

110. Kadyrov, F. A., and J. W. Drake. "Uvsx Recombinase and Dda Helicase Rescue Stalled Bacteriophage T4 DNA Replication Forks in Vitro." [In eng]. *J Biol Chem* 279, no. 34 (Aug 20 2004): 35735-40.

111. Krylov, V., E. Pleteneva, M. Bourkaltseva, O. Shaburova, G. Volckaert, N. Sykilinda, L. Kurochkina, and V. Mesyanzhinov. "Myoviridae Bacteriophages of Pseudomonas Aeruginosa: A Long and Complex Evolutionary Pathway." [In eng]. *Res Microbiol* 154, no. 4 (May 2003): 269-75.

112. Campbell, J.I.A, M Albrechtsen, and J Sorensen. "Large Pseudomonas Phages Isolated from Barley Rhizosphere." *FEMS Microbiol Ecology* 18 (1995): 63-74.

113. Botstein, D. "A Theory of Modular Evolution for Bacteriophages." [In eng]. *Ann N Y Acad Sci* 354 (1980): 484-90.

CHAPTER 4: Classification and Proteomic Analysis of *Enterobacteriales* Bacteriophages

4.1 Abstract

The *Enterobacteriaceae* family of bacteria contain many well characterized pathogens including *E. coli, Salmonella, Klebsiella* and *Shigella*. Bacteriophages that infect this family were many of the first identified and remain some of the best characterized such as T4, T7, Lambda and P22. Due to the sheer number of bacteriophages and their ability to transfer genetic material, bacteriophages play a central role in the evolution of bacteria, including pathogenic strains. Thus, analysis of bacteriophage genomes and proteomes can provide insight into specific host/phage interactions as well as the evolution of pathogenic strains. Herein we analyze and compare 1041 bacteriophage genomes from phages known to infect the *Enterobacteriaceae*. These bacteriophages fall into 92 clusters including 24 singletons of related phages. The proteomes of 597 lytic phages were examined further, revealing the highly unstudied nature of bacteriophages, with 84% of the proteins having unknown function.

4.2 Introduction

Phages are viruses that infect bacteria. With an estimated 10^{31} bacteriophages in the biosphere, they are the most abundant and diverse biological entities on the planet (1). Despite the astounding number and ubiquitous nature of bacteriophages, they have yet to be well characterized due in part to the eclectic nature of their diverse genetic makeup (2, 3). Insights into the tremendous impact that bacteriophages have in shaping microbial evolution and ecology has, in recent years, piqued an interest in bacteriophage as a method for vicariously studying their bacterial hosts (4). The development of many practical applications of phage, including the development of phage therapies to combat antibiotic resistant bacteria, has likewise contributed to what has been referred to as the renaissance of phage research (5, 6).

The *Enterobacteriaceae* phages were many of the first identified and remain some of the best characterized. These phages infect such well characterized hosts as *Salmonella* and *E. coli,* and their contributions to the evolution of the pathogenic members of the *Enterobacteriaceae are* poorly understood. Large scale analysis of bacteriophage nucleotides and proteomes may reveal unique host/phage interactions which may contribute heavily to the evolution and speciation of various strains, including pathogenic strains. As new phages continue to be isolated and sequenced, the need for bioinformatic analysis and a succinct form of phage classification and comparison becomes increasingly urgent (7).

This study builds upon the methods set forth by Graham Hatfull and coworkers for analyzing mycobacteriophage relationships (8-10). In 2014, we used similar methods to analyze and compare 337 *Enterobacteriaceae* infecting phages by genomic analysis, which divided convincingly into 56 distinct clusters based on >50% syntenic similarity. Within the established clusters, phages could be further segregated into 132 subclusters based on higher degrees of genomic homology. Having been studied in both Gram positive (*Mycobacteria*) and Gram negative (*Enterobacteriaceae)* hosts, these findings suggested that genomic comparison was a viable method for phage identification despite the mosaic nature of bacteriophage genome composition (11).

Since 2014, hundreds of additional *Enterobacteriaceae* phages have been isolated and sequenced making over a thousand available in GenBank. Here we discuss the classification and comparison of 1041 *Enterobacteriaceae* phage by whole genome nucleotide dot plot analysis (see supplementary figure S.41 for the phages analyzed). The cluster and subcluster classifications established in 2014 by Grose and Casjens are conserved and expanded. In addition to the nucleotide-based classifications, phage comparison is also considered by observing

similarities in the phage proteomes between and among cluster and subcluster groupings. The phage proteasome is also analyzed on a broader level, by identifying and categorizing all proteins found in 597 of the lytic *Enterobacteriaceae* phages based on known function.

4.3 Materials and methods

4.3.1 Comparative genomic nucleotide analysis using Gepard dot plots

All phages known to infect the bacterial family *Enterobacteriaceae* that had complete genomic sequences recorded on the National Center of Biotechnology Information (NCBI) website as of March 25, 2019 were recorded and preliminarily sorted into previously established clusters based on major capsid protein similarities by Julianne Grose and Sherwood Casjens. The correlation between MCP type and cluster membership was established by Julianne Grose and Sherwood Casjens in their publication: *Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family Enterobacteriaceae* (10).

For simplicity, granted the vast number of *Enterobacteriaceae* infecting phages and in order to facilitate a more detailed analysis of specific clusters, 7 of the 49 lytic phage clusters were the focus of this study and none of the temperate clusters were selected for subsequent exploration. Those chosen clusters included: Lytic 1, Lytic 3, Lytic 4, Lytic 13, Lytic 14, Lytic 15, and Lytic 16. FASTA files were obtained for each phage from the NCBI website, and the program Gepard (12) was used to create homology dot plots, compared with a Gepard word size of 10. Upon confirmation of the initial MCP cluster classification, phages were subsequently rearranged within the clusters, according to the nucleotide homology, to reveal neatly organized subclusters of higher nucleotide homology within each cluster.

4.3.2 Comparative proteome analysis of clusters and subclusters.

Proteomic similarities among the distinct clusters were then determined. This was achieved by first obtaining GenMark files from the NCBI website for one representative phage from each of the seven clusters and performing the necessary annotation corrections using the program DNAmaster (http://cobamide2.bio.pitt.edu/). Those phages that required annotation corrections are recorded in the table found in supplementary table 4.S3. A database was then created using PhamDB (13) containing these phages and run through the program Phamerator (14) to generate a pham map of the genomes, which compares nucleotide and protein similarities in a representative dot plot. This bioinformatic tool presents the unique proteins of each phage as multicolored boxes along the genome to create a simple and aesthetically pleasing way to facilitate the visual identification of proteins conserved among the distinct clusters. A similar process was also performed to compare the subclusters of Lytic 1 phages, by selecting one representative phage from each subcluster for comparison. To determine protein conservation within a subcluster, all phages pertaining to subcluster B of the Lytic 1 cluster were likewise compared in the same manner.

4.3.3 Genomic and proteome comparison of Lytic 1 cluster.

The average nucleotide identity (ANI) was obtained using MAFFT (15) plugin in Geneious (16) for all 74 phages of the Lytic 1 cluster. Similar to the Gepard dot plot, ANI is a tool for nucleotide comparison, but in addition to visually representing homology through shading, it also provides a numerical value for percent homology between genomes. Previous publications had determined that clusters should be identified by >50% syntenic homology. To facilitate comparison to the Lytic 1 whole genome pham maps (showing nucleotide and protein

conservation) and a Gepard dot plot were constructed with one phage from each of the 9 subclusters. The nucleotide homology represented in the Gepard dot plot and the proteomic homology shown in the pham map were then compared.

4.3.4 Proteome analysis among all *Enterobacteriaceae* infecting phages

On a much broader scale of proteomic comparison, all proteins from *Enterobacteriaceae* phages that had full genomic sequences recorded on the NCBI website as of March 25, 2019 were characterized, totaling 1041 lytic phages. Those that required annotation corrections are noted in the table found in supplementary table 4.S3. The program Phamerator (14) was then used to identify the proteins of each individual phage. Protein were separated into 14 categories based on function and protein conservation was identified. Those categories include: Phage structural proteins, bacterial structural proteins, proteases, chaperones, terminases, DNA metabolism, DNA recombination and repair, DNA binding proteins, CRISPR, virulence factors, cell lysis, unidentified proteins and other functions. The portion of the phage proteins pertaining to each category was noted. Those proteins conserved among 25 or more phage were also recorded.

4.3.5 SplitsTree analysis of singletons with one phage from each lytic cluster

SplitsTree (17) protein analysis was produced from the exported pham table of conserved proteins converted to a Nexus file using Janus (http://cobamide2.bio.pitt.edu). 49 phages (14 singletons and 1 phage each from remaining lytic clusters), were used in this analysis. The list of phages used in this analysis is provided in supplementary table 4.S4.

4.4 Results

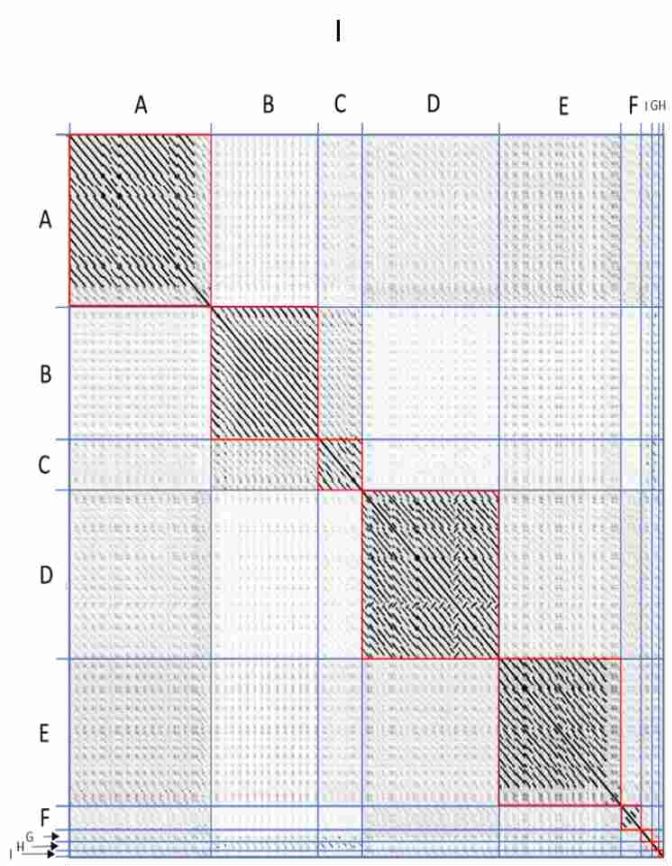
4.4.1 Comparative genomic nucleotide analysis using Gepard dot plots

All phages known to infect the bacterial family *Enterobacteriaceae* that had complete genomic sequences recorded on the NCBI website as of March 25, 2019 were sorted initially into preliminary clusters based on homology found in the major capsid protein of each phage. MCP sequence comparison suggested 49 lytic phage clusters and 39 temperate phage clusters pertaining to the bacterial family *Enterobacteriaceae*. These preliminary clusters were subsequently confirmed by Gepard dot plot comparison, with a Gepard word size of ten. The preliminary cluster allocations that were performed based on MCP similarity were found to be remarkably accurate, with only 5 (0.4%) phages misclassified by MCP. These rare examples appear to be recent exchanges of MCP with phage from other clusters. A complete list of clusters and subclusters is provided in supplementary table 4.S1.

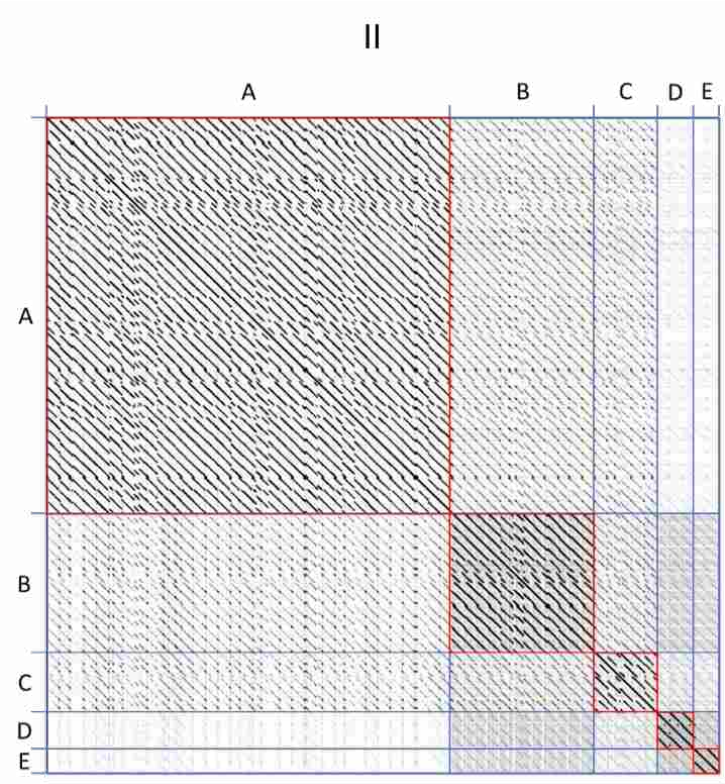
4.4.2 Comparative proteome analysis of clusters and sub clusters

As would be expected, a high degree of both proteomic and nucleotide similarity was displayed in the pham map comparing phages of Lytic 1 subcluster B (Figure 4.2). Shading indicates nucleotide sequence similarity determined by BLASTN, with purple shading indicated the highest level of genomic similarity and red shading indicating lower level of homology. Proteins are depicted as uniquely colored boxes. It can be visually noted that, while the alignment of the genomes may vary, they share a great deal of protein homology. In contrast to the strikingly similar genomes within the subcluster, the diversity between clusters of the *Enterobacteriaceae* infecting phages is quite different (Figure 4.3). Here, virtually no nucleotide homology is detected by BLASTN and a visual inspection of the proteins present reveals few

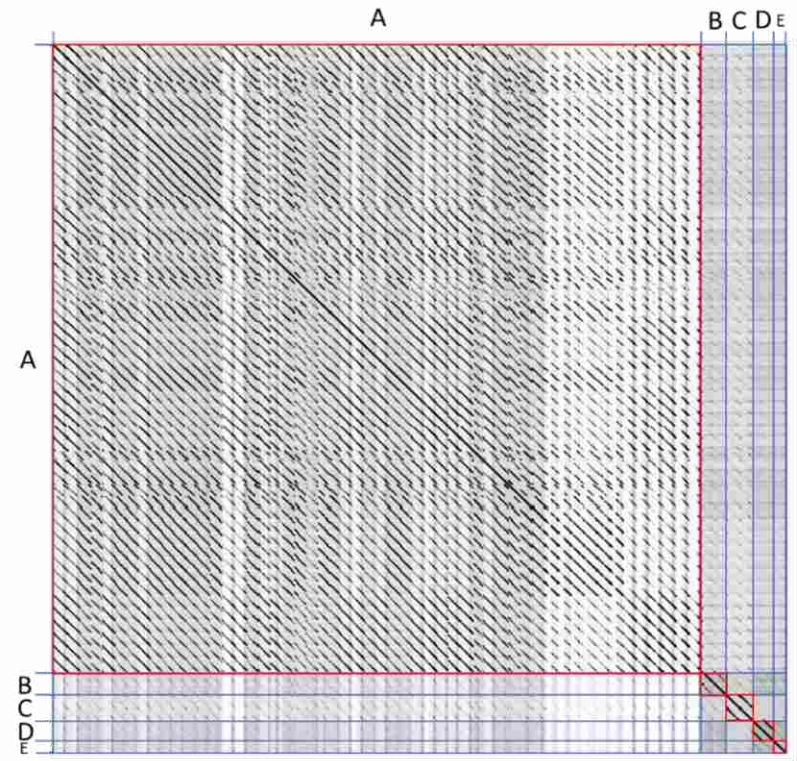
I



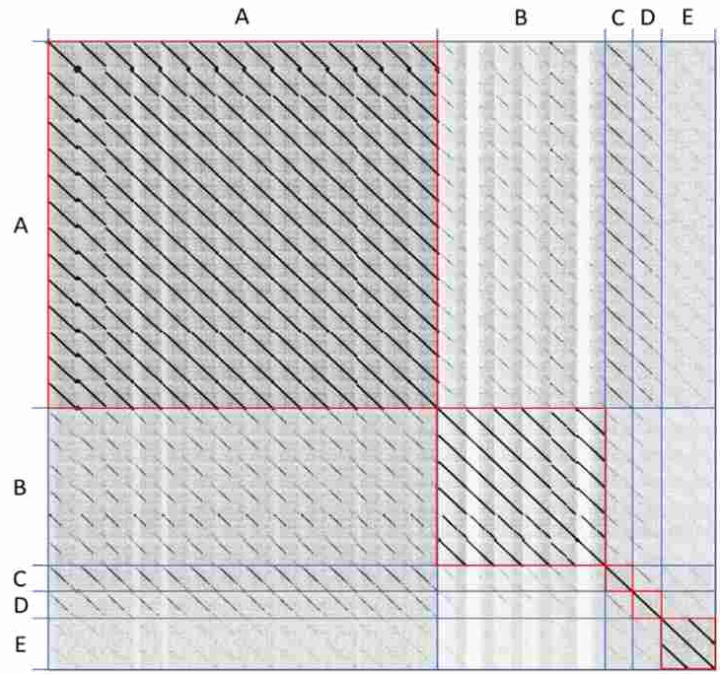
II



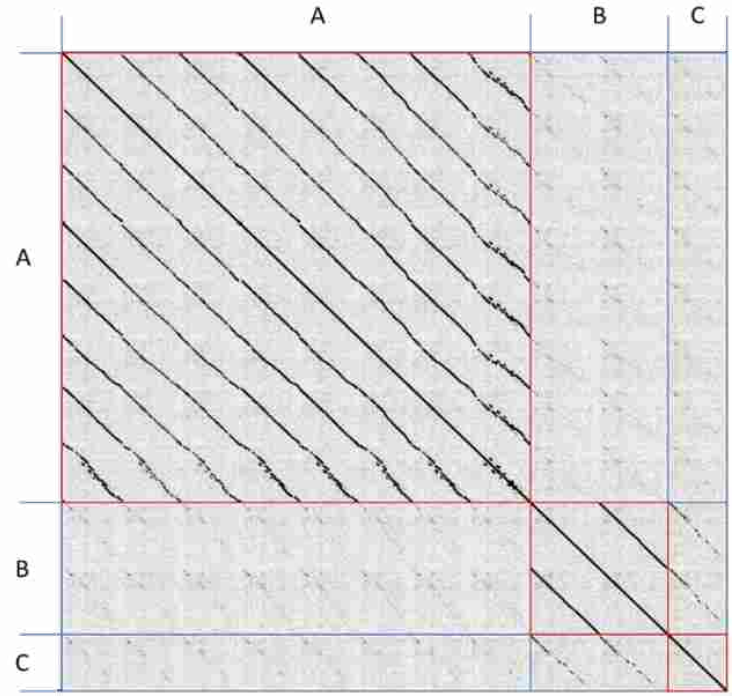
III



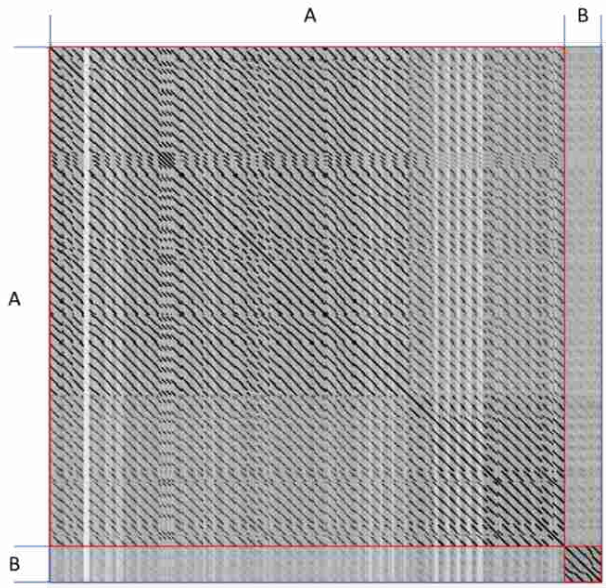
IV



V



VI



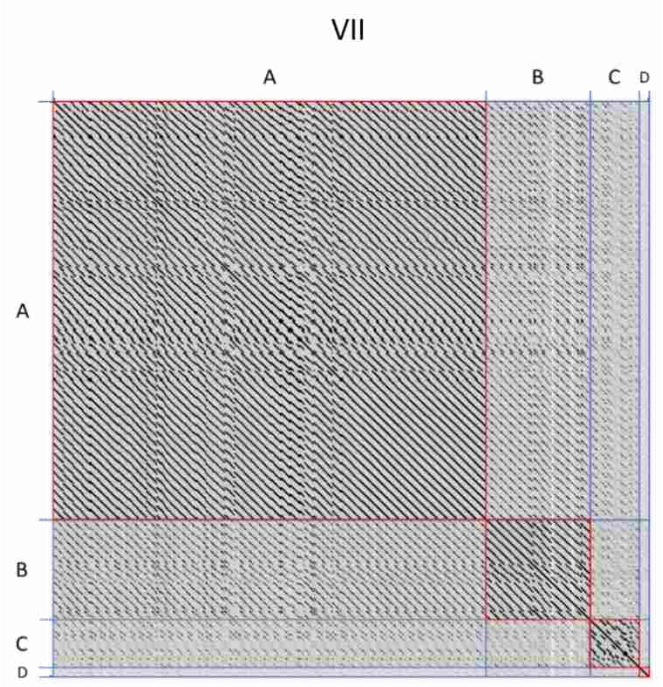


Figure 4.1 Panels I-VII display the Gepard dot plots of the seven clusters chosen to represent the 88 identified. Subcluster delineations, demarking closer relationships of phages, are seen in red. Blue lines help to visually distinguish homology between subcluster. (I) Lytic 1(T1-like) divided neatly into nine subclusters with three singletons. (II)Lytic3-(ViO1-like) divided neatly into five clusters. (III) Lytic 4 (T5-like) divided into five total subclusters, with subcluster E being a singleton and subcluster A containing the vast majority of phages. (IV) Lytic 13 (Chi-like) divided into five subclusters with two singletons, C and D.(V) Lytic 14 (Eco32-like) divided into three subclusters, with C being a singleton with homology to subcluster B. (VI) Lytic 15 (Felix-O1-like) divided into two subclusters, with A containing the majority of the phages. (VII) Lytic 16 (SETP3-like) divided into four subclusters, with D being a singleton.

if any common proteins among any two phages. Due to great variety in lengths, genomes have been broken into three section. Each section is labeled accordingly on the left side of the figure. As is clearly demonstrated in figure 4.3, there is a great deal of diversity that exists between phages not of the same cluster. Comparing figures 4.2 and 4.3 further emphasizes how remarkably similar phages are within a subcluster and solidifies the logic in categorizing them as such.

The most peculiar of the pham maps was that done for representative phages from each subcluster within the Lytic 1 cluster (Figure 4.4). These phages notably display less BLASTN identified nucleotide homology than phages within a subcluster. Most interesting, however, is that a visual inspection of the proteins reveals a protein homology similar to that shared within

the subcluster. This figured called for confirmation as to the lower than expected level of nucleotide homology experienced within this cluster.

Previous publications have established that, clusters should display genomic similarity over at least 50% of the genome. The lack of BLASTN homology displayed in the Lytic 1 pham map (figure 4.4) was supported by a Gepard dot plot comparison between the 7 representative Lytic 1 phages included in the pham map (Figure 4.5). An ANI was performed to calculate percent similarity between all 74 phages of the Lytic 1 cluster. The ANI revealed that, while there remained a high degree of nucleotide homology among subclusters, similarity between phages of distinct subclusters was on average found to be significantly lower than 50%. This figure can be found in supplementary table 4.S2. It would appear that the average nucleotide similarity within a cluster had been reduced with the addition of more phages. Proteomic similarity within a cluster, however, remained higher. Approximately 50% of all proteins were shared in ≥6 of the 9 representative phages, with 32% being conserved in all 9.

4.4.3 Proteome analysis among all *Enterobacteriaceae* infecting phages:

We now consider a much broader view of the bacteriophage proteome. The sequencing of many bacteriophage genomes has revealed remarkable diversity, including many novel proteins. It has been well noted that the phage genome is comprised of very few non-coding regions, but the function and essentiality of most phage proteins have yet to be identified [18]. In order to better quantify the wealth of diversity contained within phage proteomes, all proteins identified among 597 lytic *Enterobacteriaceae* infecting phages were categorized based on known function. Among this group, there was found to be a total of 11923 protein, 84 percent (1904 proteins) of which have functions yet to be discovered (Figure 4.6). Their categorization can be seen more clearly in Figure 4.7

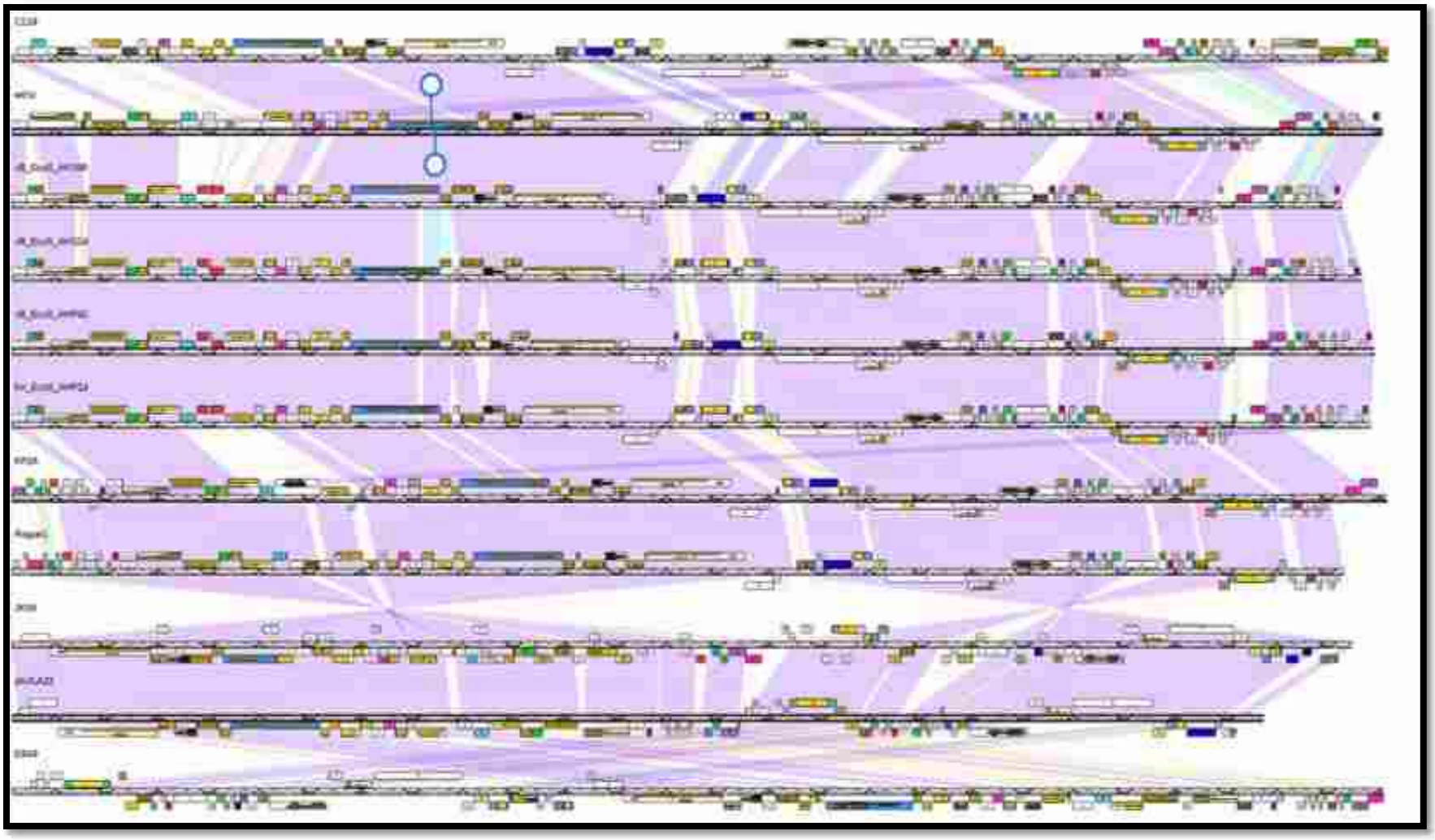


Figure 4.2 Lytic 1 subcluster B Pham Map. Comparison of all phages of Lytic 1 subcluster B. From top to bottom: C119, e4/1c, AKS96, AHS24, AHP42, AHP24, KP26, Rogue1, JK06, JLA23, EB49. Proteins are indicated by colored markers with homologous proteins being the same color, while conserved domains are indicated by yellow boxes within the shaded protein box. Protein homologs are defined as having either 35% similarity or an e-value less than e-7 according to clustal omega alignment. Shading indicates nucleotide sequence similarity determined by BLASTN, purple shading indicated the highest level of genomic similarity, red shading indicating the lowest

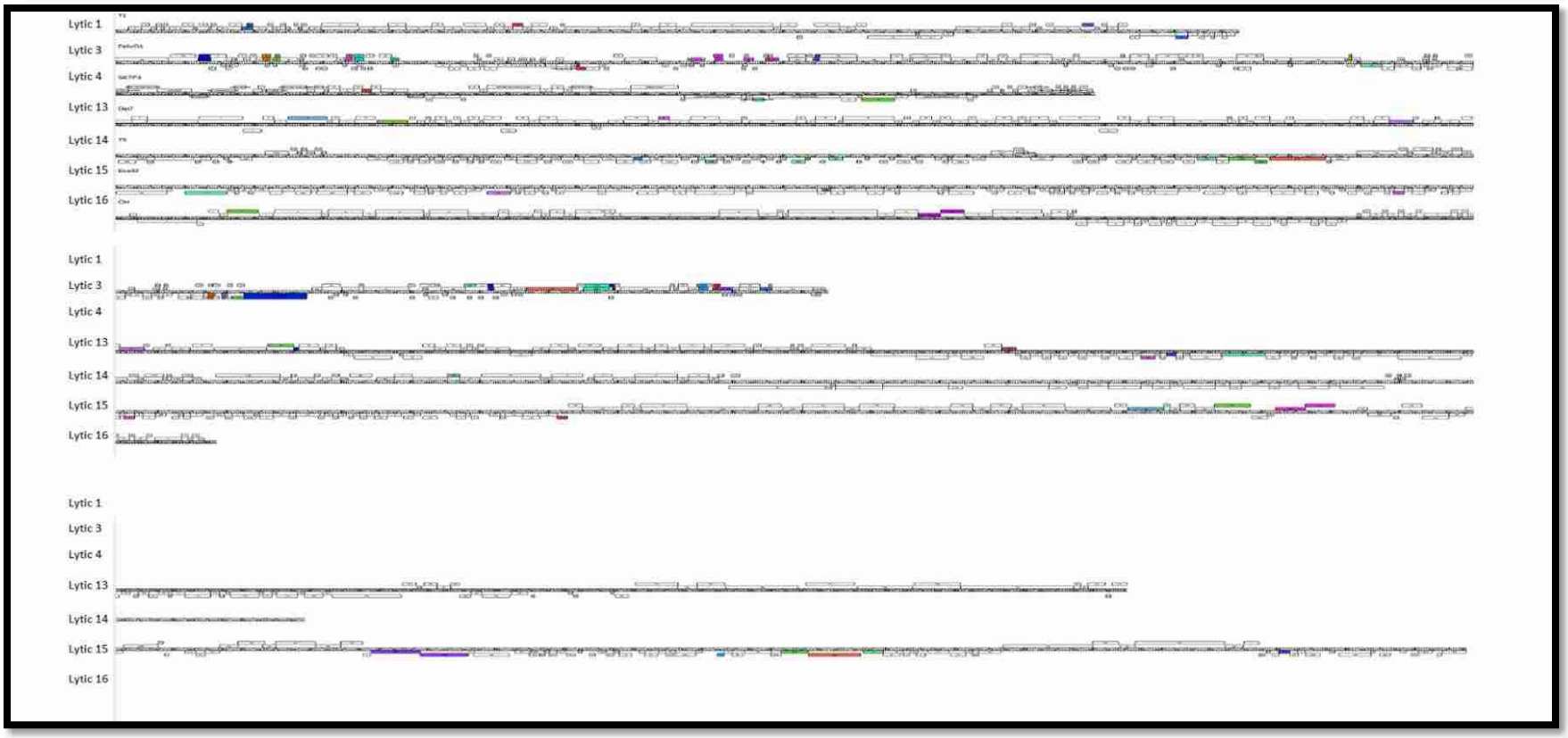


Figure 4.3 Pham Map of *Enterobacteriaceae* infecting 7 Lytic Clusters. Lytic 1, Lytic3, Lytic4, Lytic13, Lytic14, Lytic15 and Lytic16 are seven clusters displaying no nucleotide homology and very low level of protein homology only seen in conderved domains. Proteins are indicated by boxes with homologous proteins being the same color, while conserved domains are indicated by yellow boxes within the shaded protein box. Protein homologs are defined as having either 35% similarity or an e-value less than e-7 according to clustal omega alignment. Shading between genomes indicates nucleotide sequence similarity determined by BLASTN, purple shading indicated the highest level of genomic similarity, red shading indicating the lowest.

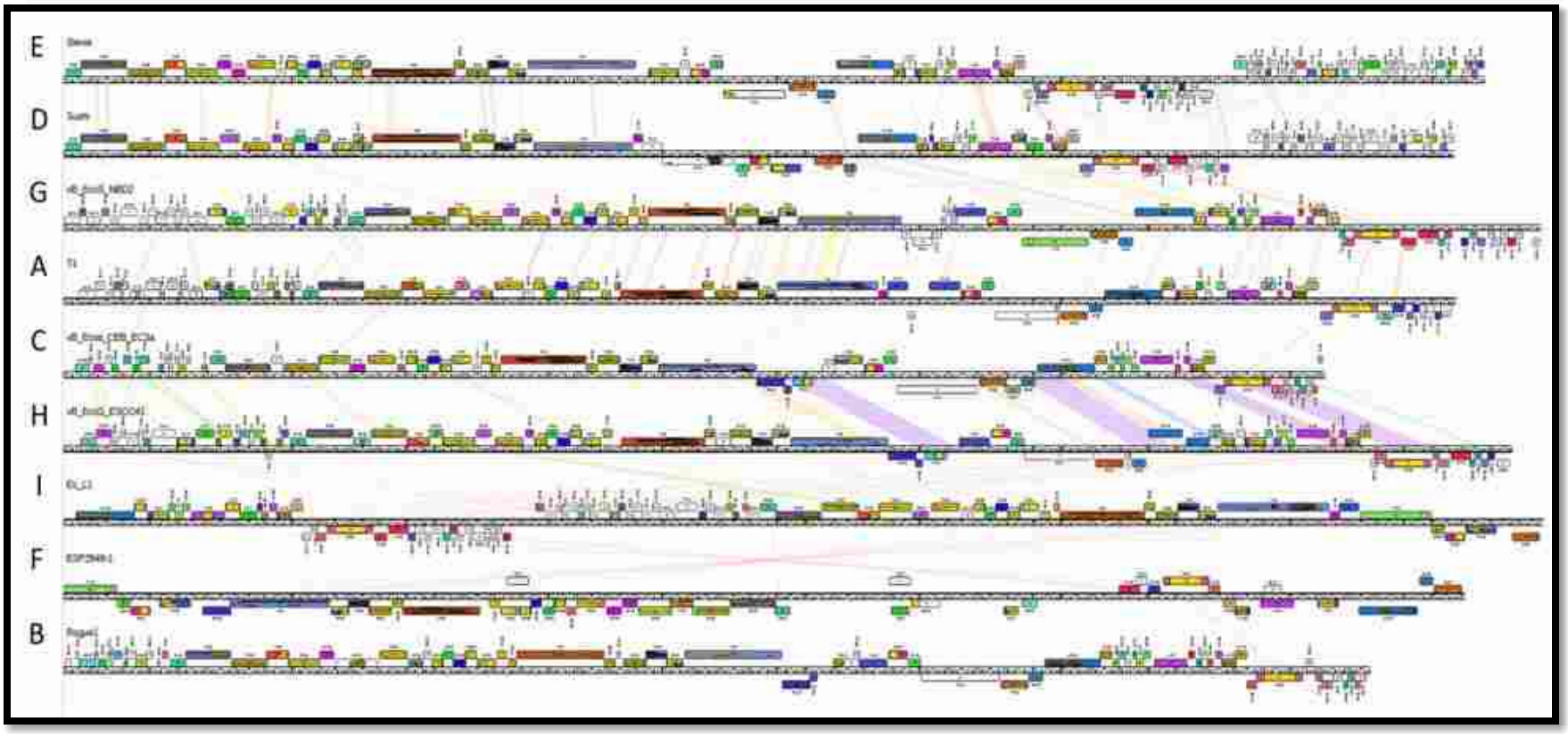


Figure 4.4 Lytic 1 Pham Map. Vertical letters on the left indicate which subcluster of cluster Lytic 1 each phage represents. Proteins are indicated by colored markers with homologous proteins being the same color, while conserved domains are indicated by yellow boxes within the shaded protein box. Protein homologs are defined as having either 35% similarity or an e-value less than e-7 according to clustal omega alignment. Shading indicates nucleotide sequence similarity determined by BLASTN, purple shading indicated the highest level of genomic similarity, red shading indicating the lowest

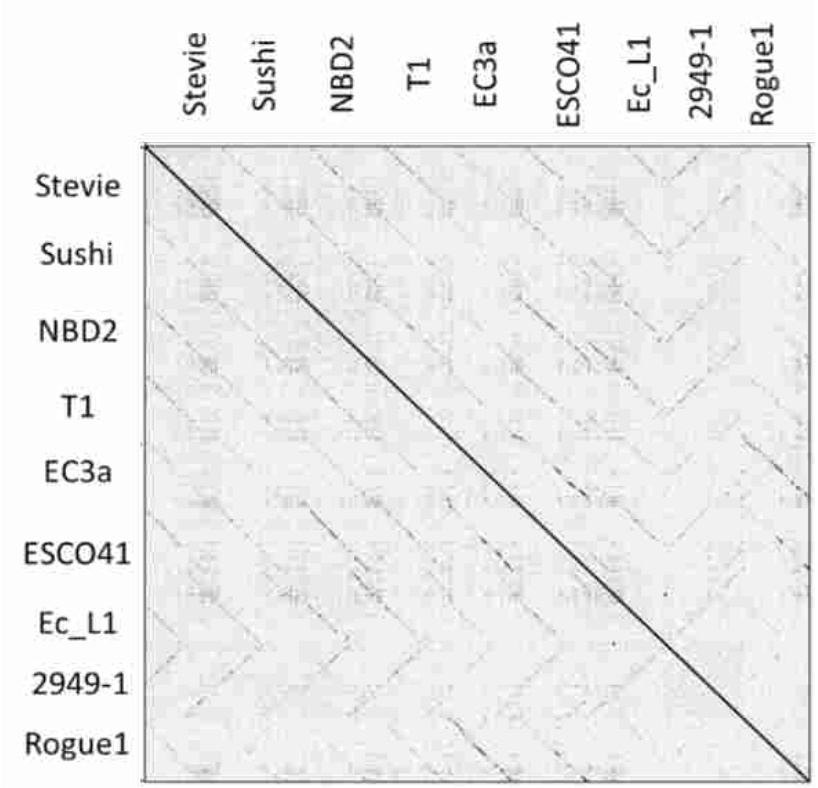


Figure 4.5 Gepard dot plot of phages representative of each subcluster within the Lytic 1 cluster

Of additional interest in this study was the number of proteins common among the *Enterobacteriaceae* infecting phages. Nearly 60 percent of the 11923 proteins were found to be unique among 1 or 2 phages. While it is possible that such proteins contribute to the fitness of these phages, it is more likely that these proteins are transitive and unessential. This is much less likely to be the case, however, with more commonly shared proteins. Found to be conserved among ≥25 phages were 614 of the 11923 proteins (Figure 4.8). Those common among ≥100 phages were 243, with the most common protein being found in 299 unique phages. Of the proteins conserved among ≥25 phages, more than half (325) have no known function. While their functions have yet to be revealed, the prevalence of these proteins serves as an indication as to their importance in phage proliferation. Discovery as to their function would contribute to our

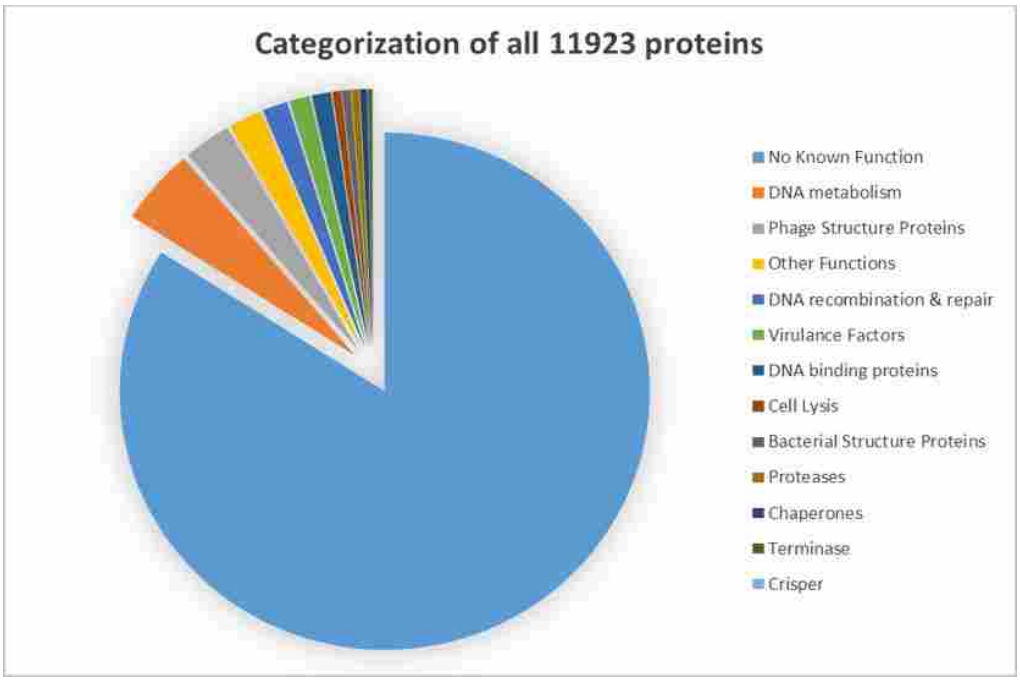


Figure 4.6 Categorization of 11923 unique proteins. Among the complete genomes of 597 *Enterobacteriaceae* infecting lytic phages, 11923 unique proteins were identified using the program Phamerator. Of those proteins, 84% (10019 proteins) have yet to be characterized.

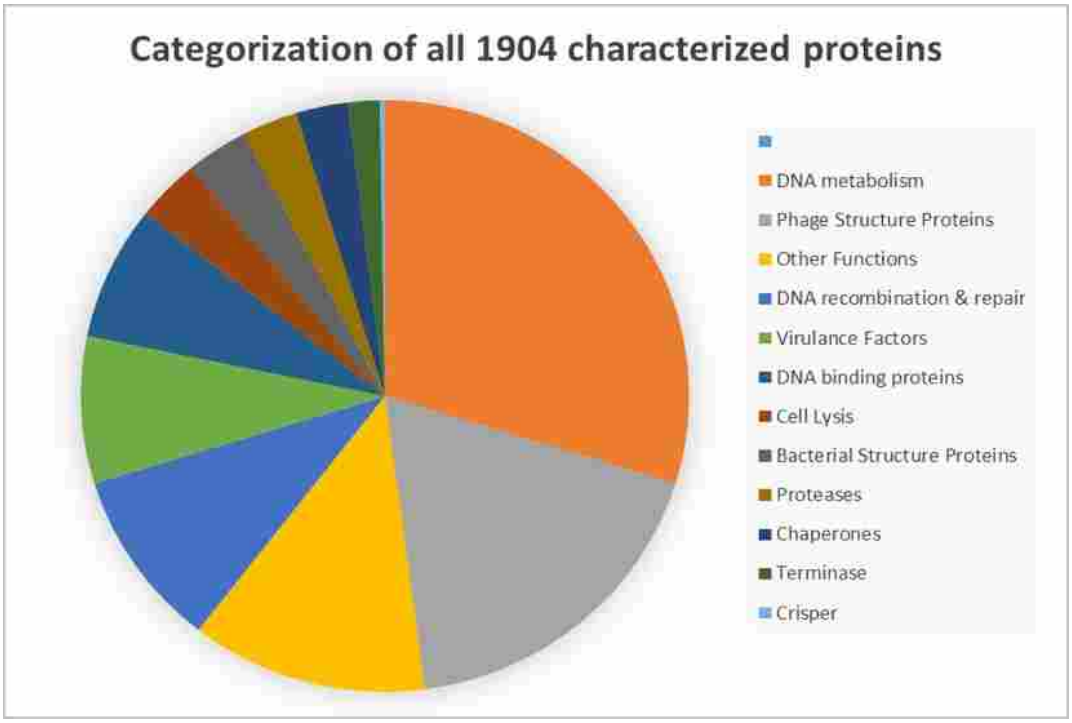


Figure 4.7 Distribution of all categorized proteins identified among 597 lytic *Enterobacteriaceae* infecting phages based on known function. Of the 1904 characterized proteins, the majority were found to be involved in DNA metabolism

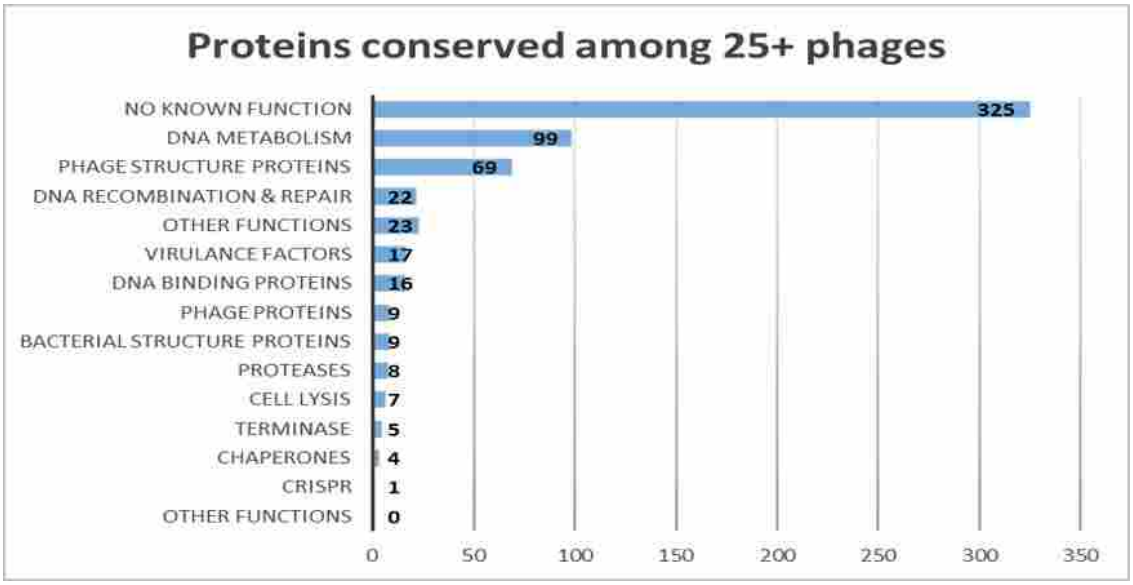


Figure 4.8 Identifying conserved proteins in 597 Lytic phages. Of the 11923 protein identified among 597 lytic *Enterobacteriaceae* infecting phages, 614 were found to be conserved among ≥25 phages. 325 of these have unknown function

understanding of the phage lifecycle and the ways in which phage interact with their bacterial hosts.

4.4.4 Cluster assignment confirmed by SplitsTree

5531 phams (protein families) from 49 phages were used to analyze the protein conservation between singleton and lytic clusters. A SplitsTree was inferred using the pham table created from Phamerator. The SplitsTree agrees with our current assignment of superclusters and clusters by putting phages of the same cluster closer to each other. Phages SETP3, Scapp, SO1, vB_PagS_MED16 and vB_Kp3 of supercluster SETP3 are seen much closer to each other than to other phages. Like wise phages from supercluster T7 (Peat1, KP34, vB_CskP_GAP227, SP6 and T7), rV5 (V5 and phi92) and N4 (N4 and vB_PatP_CB1) are seen branching of the same nodes. The most interesting thing is minimal number of conserved proteins. The most conserved of them was (number 2788) “bifunctional glutaredoxin”, found only in 8 phages out of 49 followed by (pham number 1729) “nicotinamide phosphoribosyl transferase”, found in 7 phages.

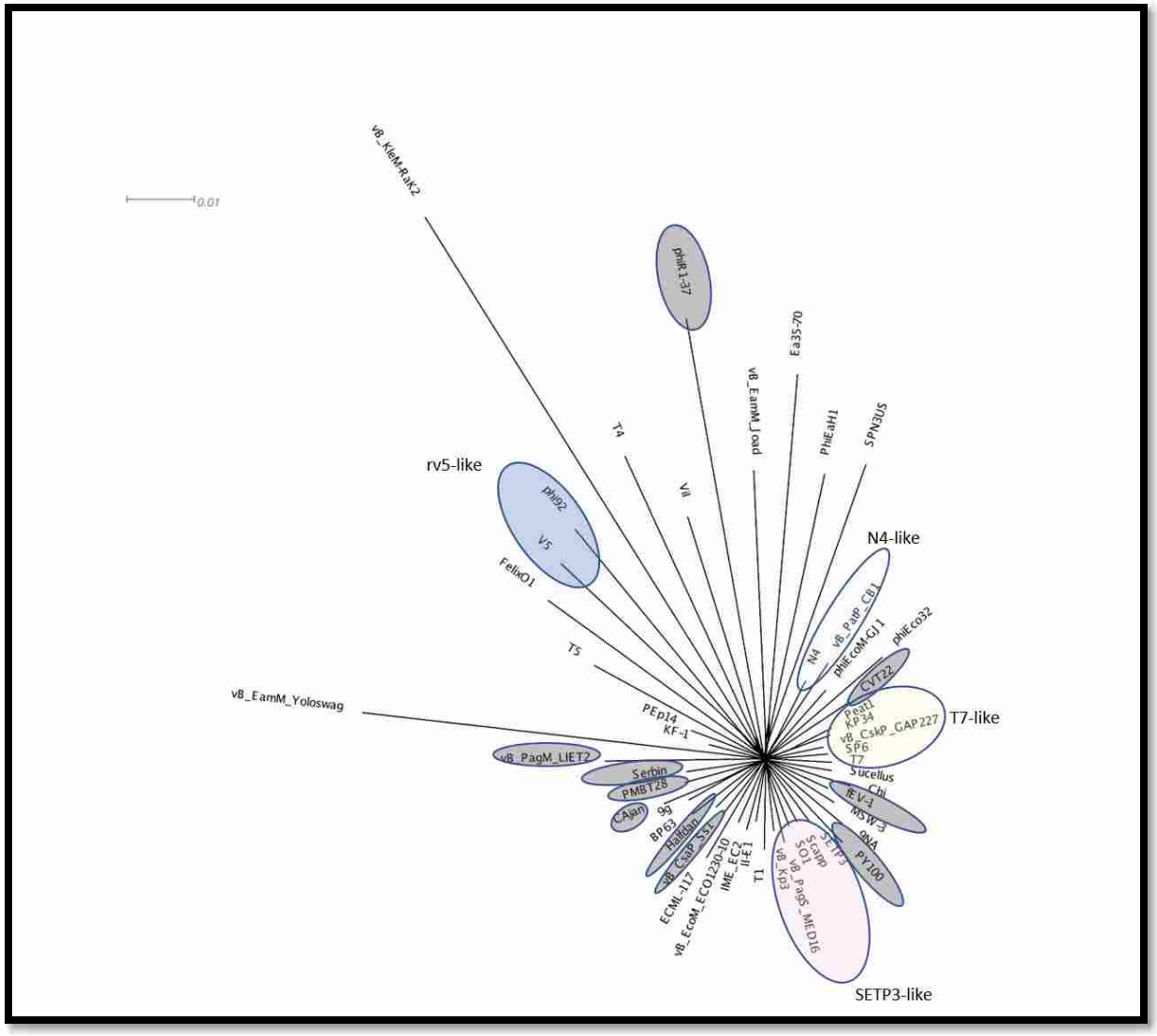


Figure 4.9 Protein conservation of 49 phages confirms the supercluster assignment by comparing their protein families. The singletons and supercluster of 2 or more phages are highlighted.

4.5 Conclusion

The initial cluster/subcluster classification pertaining to 1041 lytic phages that infect the bacterial family *Enterobacteriaceae* were initially performed using MCP comparison and then confirmed through Gepard dot plot analysis. Compared with the 337 phages analyzed in 2014 which fell into 53 clusters, these 1303 (1041 lytic and 262 temperate) form 88 new clusters. Thus a 3-fold increase in phages produces a 1.5-fold increase in clusters, suggesting we have in no way begun to tap the reservoir of phage diversity within the *Enterobacteriaceae family*. In

addition, the ability to expand those previously established clusters with newly sequenced phage supports such analyses as a viable method for categorizing bacteriophage. With the addition of more phage, it was noted that nucleotide similarity within a cluster fell below 50% genomic similarity between some phages, however nucleotide similarity remained 50% or greater with at least one other phage within a cluster. Nevertheless, proteomic similarity remained high (40% or greater), suggesting that protein similarity may be a significant factor to consider when making cluster/subcluster assignations. Several examples are provided with cases of phages having little nucleotide similarity, and overall proteomic conservation (>80%). With larger sample sizes, hallmark proteins of specific clusters may be identified and used to facilitate the categorization process.

The proteomic diversity displayed among *Enterobacteriaceae* infecting phage is impressive. While the bacteriophage genome is certainly eclectic, there are a high number of proteins (614 proteins or ~5%) commonly found to be conserved among a large number of phages (more than 25 phages). The frequency with which these proteins appear may be viewed as an indication of their importance in phage proliferation. Focusing research efforts on those most conserved proteins may prove revelatory in further explaining phage-host relationships.

4.6 Supplementary data

Supplementary Table 4.S1 Table containing all phages from the seven selected cluster. Corresponds with figures 4.1 panels I-VII.

CLUSTE	SubClust	Member phage	Host species	Accession number	Sequence publication	Genome size (bp)	Bacterial host family
Lytic1: T1-like	A	**T1**	*Escherichia coli*	AY216660	Virology 318:245	48836	*Enterobacteriaceae*
	A	**Shfl1**	*Shigella flexneri*	NC_015456	–	50661	*Enterobacteriaceae*
	A	**ADB-2**	*Escherichia coli*	JX912252	GenomeA 1:e00043-13	50552	*Enterobacteriaceae*
	A	**BIFF**	*Escherichia coli*	MH285980		49372	*Enterobacteriaceae*
	A	**SH2**	*Escherichia coli*	KY985004		49088	*Enterobacteriaceae*
	A	**IME18**	*Escherichia coli*	MH051911		50354	*Enterobacteriaceae*
	A	**IME167**	*Escherichia coli*	MH051912		49794	*Enterobacteriaceae*
	A	**ISF001**	*Shigella sonnei*	MG049919	J Food Sci Tech 55:550	50552	*Enterobacteriaceae*
	A	**ISF002**	*Shigella sonnei*	MF093736	JMedMico jan 2018 in press	50564	*Enterobacteriaceae*
	A	**JMPW1**	*Escherichia coli*	KU194206		49628	*Enterobacteriaceae*
	A	**JMPW2**	*Escherichia coli*	KU194205		50298	*Enterobacteriaceae*
	A	**ø2457T**	*Shigella flexneri*	MH917278		50219	*Enterobacteriaceae*
	A	**Sfin-1**	*Escherichia coli/Shigella*	MF468274		50403	*Enterobacteriaceae*
	A	**SH6**	*Shigella* sp.	KX828710	SciRep 7:40349	50552	*Enterobacteriaceae*
	A	**pSf-2**	*Shigella flexneri*	KP085586		50109	*Enterobacteriaceae*
	A	**SRT8**	*Escherichia coli*	MF996376		49579	*Enterobacteriaceae*
	A	**IME347**	*Escherichia coli*	MH051918	JBasicMicrobiol.2018Aug 26	50048	*Enterobacteriaceae*
	B	**Rogue1**	*Escherichia coli*	JQ182736	VirolJ 9:207	45805	*Enterobacteriaceae*
	B	**Sd1**	*Shigella dysenteriae*	MF158042	JVirol 92:e02117-17	48262	*Enterobacteriaceae*
	B	**Sf12**	*Shigella flexneri*	MF158039	JVirol 92:e02117-17	47647	*Enterobacteriaceae*
	B	**øKP26**	*Escherichia coli/S. enterica*	KC579452	ArchVirol 158:2395	47285	*Enterobacteriaceae*
	B	**JK06 (KP26?)**	*Escherichia coli*	DQ121662	–	46072	*Enterobacteriaceae*
	B	**øJLA23**	*Escherichia coli*	KC333879	GenAnn 1:e00219-12	43017	*Enterobacteriaceae*
	B	**øEB49**	*Escherichia coli*	JF770475	AEM 77:6630	47180	*Enterobacteriaceae*
	B	**øC119**	*Escherichia coli*	KT825490	PeerJ:e2423	47319	*Enterobacteriaceae*
	B	**AHS24**	*Escherichia coli*	KF771238	PLoSOne 9:100426	46440	*Enterobacteriaceae*
	B	**AHP42**	*Escherichia coli*	KF771237	PLoSOne 9:100426	46847	*Enterobacteriaceae*
	B	**AHP24**	*Escherichia coli*	KF771236	PLoSOne 9:100426	46719	*Enterobacteriaceae*
	B	**AKS96**	*Escherichia coli*	KF771239	PLoSOne 9:100426	45746	*Enterobacteriaceae*
	B	**C119**	*Escherichia coli*	KT825490		47319	*Enterobacteriaceae*
	B	**e4/1c**	*Escherichia coli*	KJ668713	–	47112	*Enterobacteriaceae*
	C	**Rtp**	*Escherichia coli*	AM156809	JBACT 188:1419	46219	*Enterobacteriaceae*
	C	**EC3a**	*Escherichia coli*	KY398841		44234	*Enterobacteriaceae*
	C	**IMM-001#**	*Escherichia coli*	MF630922		32486	*Enterobacteriaceae*
	C	**IME253**	*Escherichia coli*	KX130960		46717	*Enterobacteriaceae*
	C	**ACG-M12**	*Escherichia coli*	NC_019404	Viruses 4:471	46054	*Enterobacteriaceae*
	C	**DTL**	*Escherichia coli*	MG050172	J Ind Microbiol Biotechnol. 201	45814	*Enterobacteriaceae*
	D	**F20#**	*Enterobacter aerogenes*	JN67284	JGenVirol 93:2310	51543	*Enterobacteriaceae*
	D	**GML-KpCol1**	*Klebsiella pneumoniae*	MG552615		50249	*Enterobacteriaceae*
	D	**JY917**	*Klebsiella pneumoniae*	MG894052		37655	*Enterobacteriaceae*
	D	**KP36**	*Klebsiella pneumoniae*	NC_019781	VirolJ 10:100	49818	*Enterobacteriaceae*
	D	**KPN N141**	*Klebsiella pneumoniae*	MF415412		49090	*Enterobacteriaceae*
	D	**KpV522**	*Klebsiella pneumoniae*	KX237515		51099	*Enterobacteriaceae*
	D	**MezzoGao**	*Klebsiella pneumoniae*	MF612072		49807	*Enterobacteriaceae*
	D	**NJR15**	*Klebsiella pneumoniae*	MH633487		49468	*Enterobacteriaceae*
	D	**NJS1**	*Klebsiella pneumoniae*	MH445453		49292	*Enterobacteriaceae*
	D	**NJS2**	*Klebsiella pneumoniae*	MH633485		50132	*Enterobacteriaceae*
	D	**NJS3**	*Klebsiella pneumoniae*	MH633486		49387	*Enterobacteriaceae*
	D	**PKP126**	*Klebsiella pneumoniae*	KR269719	Park ArchVriol in press	50934	*Enterobacteriaceae*
	D	**1513**	*Klebsiella pneumoniae*	KP658157		49462	*Enterobacteriaceae*
	D	**Sushi**	*Klebsiella pneumoniae*	KR262148		49037	*Enterobacteriaceae*
	D	**TAH8**	*Klebsiella pneumoniae*	MH633484		49344	*Enterobacteriaceae*
	D	**KLPN1**	*Klebsiella pneumoniae*	KT001920	PeerJ 3:e1061	48754	*Enterobacteriaceae*
	D	**KOX1**	*Klebsiella pneumoniae*	KY780482		50526	*Enterobacteriaceae*
	E	**TLS**	*Escherichia coli*	AY308796	JMolBiol 308:579	49902	*Enterobacteriaceae*
	E	**FSL_SP-126 #**	*Salmonella enterica*	KC139521	BMCgenomics 14:481	51092	*Enterobacteriaceae*
	E	**YSP2**	*Salmonella enterica Pullorum*	MG241338		50316	*Enterobacteriaceae*
	E	**GJL01**	*Salmonella enterica Pullorum*	KY657202		50407	*Enterobacteriaceae*
	E	**LL5**	*Escherichia coli*	MH491968		49788	*Enterobacteriaceae*
	E	**PHB07**	*Salmonella enterica*	MH102284		51818	*Enterobacteriaceae*
	E	**phSE-2**	*Salmonella enterica*	KX015770		49167	*Enterobacteriaceae*
	E	**phSE-5**	*Salmonella enterica*	KX015771		49178	*Enterobacteriaceae*
	E	**Sazh**	*Citrobacter freundii*	MH729819		49665	*Enterobacteriaceae*
	E	**Stevie**	*Citrobacter freundii*	KM236241	GenomeA 3:e01434-14	49816	*Enterobacteriaceae*
	E	**36#**	*Salmonella enterica*	KR296690		41085	*Enterobacteriaceae*
	E	**CF1 DK-2017**	*Citrobacter freundii*	KY694971		50339	*Enterobacteriaceae*
	E	**CF-1**	*Citrobacter freundii*	KY694971		50339	*Enterobacteriaceae*
	E	**pSf-1**	*Shigella flexneri*	KC710998	ResMicro 164:979	51821	*Enterobacteriaceae*
	E	**swan01**	*Escherichia coli*	LT841304	GenomeA5:300501-17	50865	*Enterobacteriaceae*
	F	**ESP2949-1**	*Cronobacter sakazakii*	JF912400	ArchVirol 157:199	49116	*Enterobacteriaceae*
	F	**CS01**	*Cronobacter sakazakii*	MH845412		48195	*Enterobacteriaceae*
	G	**NBD2**	*Escherichia coli*	KX130668		51802	*Enterobacteriaceae*
	H	**ESCO41**	*Escherichia coli*	KY619305	ArchVirol2917 in press	50800	*Enterobacteriaceae*
	I	**Ec_L1**	*Enterobacter cloacae*	MG732930		51894	*Enterobacteriaceae*

Lytic3: Vi01-like	A	**38**	*Salmonella enterica*	KR296692	VirusGenes 52:117	156833	*Enterobacteriaceae*
	A	**BSP101**	*Salmonella enterica Typhimurium*	KY787213		157665	*Enterobacteriaceae*
	A	**CBA120**	*Escherichia coli*	JN593240	ViroIJ 8:430	157304	*Enterobacteriaceae*
	A	**Det7**	*Salmonella enterica* Typhimurium	KP797973	–	157498	*Enterobacteriaceae*
	A	**ECML-4**	*Escherichia coli*	JX128257	–	157308	*Enterobacteriaceae*
	A	**EP75**	*Escherichia coli*	MG748547		158143	*Enterobacteriaceae*
	A	**FEC14**	*Escherichia coli*	MG383452		158639	*Enterobacteriaceae*
	A	**FSL_SP-029 #**	*Salmonella enterica*	KC139566+other	BMCgenomics 14:481		*Enterobacteriaceae*
	A	**FSL_SP-063 #**	*Salmonella enterica*	KC139524+other	BMCgenomics 14:481		*Enterobacteriaceae*
	A	**GG32**	*Salmonella enterica*	KX245012	GenomeA 2016 Dec	157855	*Enterobacteriaceae*
	A	**Marshall**	*Salmonella enterica*	KF669653	GenomeA 1:e00867	156338	*Enterobacteriaceae*
	A	**Maynard**	*Salmonella enterica*	KF669654	GenomeA 1:e00866	154701	*Enterobacteriaceae*
	A	**Mooltan**	*Salmonella enterica Enteritidis*	MH688040		156882	*Enterobacteriaceae*
	A	**Mutine**	*Salmonella enterica* Typhimurium	MG428992		161502	*Enterobacteriaceae*
	A	**øSH19**	*Salmonella enterica*	JN126049	ViroIJ 8:498	157785	*Enterobacteriaceae*
	A	**PhaxI**	*Escherichia coli*	JN673056	Microbiology 159:1629	156628	*Enterobacteriaceae*
	A	**PM10**	*Salmonella enterica*	KX438380		158081	*Enterobacteriaceae*
	A	**PS5**	*Salmonella enterica* Typhimurium	MH940212		158400	*Enterobacteriaceae*
	A	**S8**	*Salmonella enterica Gallinarum*	KY630163		158432	*Enterobacteriaceae*
	A	**S115**	*Salmonella enterica Enteritidis*	MH370368		157946	*Enterobacteriaceae*
	A	**S117**	*Salmonella enterica* Typhimurium	MH370370		158110	*Enterobacteriaceae*
	A	**S118**	*Salmonella enterica Dublin*	MH370371		157013	*Enterobacteriaceae*
	A	**Sa157w**	*Escherichia coli*	MH939183		155887	*Enterobacteriaceae*
	A	**SeLz-1**	*Salmonella enterica*	MH709121		154811	*Enterobacteriaceae*
	A	**SeSz-3**	*Salmonella enterica*	MH709120		157630	*Enterobacteriaceae*
	A	**SenM-2**	*Salmonella* sp.	KX171211		158986	*Enterobacteriaceae*
	A	**SFP10**	*Salmonella enterica*	HQ259103	ApplEnvMicro 78:58	157950	*Enterobacteriaceae*
	A	**SJ2**	*Salmonella enterica*	KJ174317	FoodbornePahDis2016	152460	*Enterobacteriaceae*
	A	**SJ3**	*Salmonella enterica*	KJ174318	–	162910	*Enterobacteriaceae*
	A	**SKML-39**	*Salmonella enterica*	JX181829	–	159624	*Enterobacteriaceae*
	A	**SP1**	*Salmonella enterica*	MF001362		156585	*Enterobacteriaceae*
	A	**STP07**	*Salmonella entericaTyphimurium*	KY000003		160342	*Enterobacteriaceae*
	A	**STML-13-1#**	*Salmonella enterica*	JX181828	–	157235	*Enterobacteriaceae*
	A	**Vi01 (Vil)**	*Salmonella enterica*	FQ312032	JBact 192:5746	157061	*Enterobacteriaceae*
	B	**øD3**	*Dickeya* sp.	KM209228	Stand.Genomic Sci. fall2015	152308	*Pectobacteriaceae*
	B	**Coodle**	*Dickeya Solani*	MH807820	Viruses 10:621	152515	*Pectobacteriaceae*
	B	**JA15**	*Dickeya solani*	KY942056	Front Microbiol 8:1654	153757	*Pectobacteriaceae*
	B	**Kamild**	*Dickeya Solani*	MH807812	Viruses 10:621	152612	*Pectobacteriaceae*
	B	**øEM4**	*Enterobacter asburiae*	LC373201		160766	*Enterobacteriaceae*
	B	**øPD10.3 #**	*Dickeya solani et al.*	KM209270	PLoS One March 24, 2015	192291	*Pectobacteriaceae*
	B	**øPD23.1 #**	*Dickeya solani et al.*	KM209320	PLoS One March 24, 2015	188540	*Pectobacteriaceae*
	B	**RC_2014 (øD5)**	*Dickeya* sp.	KJ716335	ArchVirol in press 2014	155346	*Pectobacteriaceae*
	B	**XF4**	*Dickeya solani*	KY942057	Front Microbiol 8:1654	151519	*Pectobacteriaceae*
	B	**LIMEstone1**	*Dickeya solani*	HE600015	PLosOne 7:e33227	152472	*Pectobacteriaceae*
	B	**PP35**	*Dickeya solani*	MG266157		152048	*Pectobacteriaceae*
	B	**øSboM-AG3**	*Shigella boydii*	FJ373894	ViroIJ 8:242	158006	*Enterobacteriaceae*
	C	**0507-KN2-1**	*Klebsiella pneumoniae*	AB797215	–	159991	*Enterobacteriaceae*
	C	**KpS110**	*Klebsiella pneumoniae*	MG770379		156801	*Enterobacteriaceae*
	C	**Menlow**	*Klebsiella pneumoniae*	MG428990		157281	*Enterobacteriaceae*
	C	**May**	*Klebsiella pneumoniae*	MG428991		159631	*Enterobacteriaceae*
	C	**IME250**	*Serratia rubidaea*	KY073123		154938	*Yersiniaceae*
	D	**3M**	*Serratia marcescens*	MH929319		159398	*Yersiniaceae*
	D	**KSP90#**	*Serratia plymuthica*	AB452990	–		*Yersiniaceae*
	D	**øMAM1**	*Serratia plymuthica*	JX878496	JVirol 86:13872	157834	*Yersiniaceae*
	D	**2050H1**	*Serratia marcescens*	MF285619		159631	*Yersiniaceae*
	E	**Bue1**	*Erwinia amylovora*	MG973030		164037	*Erwiniaceae*
	E	**øEa2809**	*Erwinia amylovora*	KP037007	FEMS MicroLett 362:fnv031	162160	*Erwiniaceae*

Lytic4: T5-like	A	**100268_sal2**	*Salmonella enterica Enteritidis*	KU927497	GenomeA.00943-16	125114	*Enterobacteriaceae*
	A	**118970_sal2**	*Salmonella enterica Enteritidis*	KX017521		114180	*Enterobacteriaceae*
	A	**AKFV33**	*Escherichia coli*	NC_017969	PLoSONE e34585	108853	*Enterobacteriaceae*
	A	**APCEo03**	*Escherichia coli*	KR422353		103737	*Enterobacteriaceae*
	A	**BSP22A**	*Salmonella enterica Typhimurium*	KY787212		110741	*Enterobacteriaceae*
	A	**CEV-2 #**	*Escherichia coli*	HQ661859			*Enterobacteriaceae*
	A	**DT571/2**	*Escherichia coli*	KM979355	ArchVirol160:3133	108418	*Enterobacteriaceae*
	A	**DT57C**	*Escherichia coli*	KM979354	ArchVirol160:3133	108065	*Enterobacteriaceae*
	A	**EPS7**	*Salmonella enterica*	CP000917	FemsMicroLett 289:202	111382	*Enterobacteriaceae*
	A	**FFH1**	*Escherichia coli*	KJ190157	–	108483	*Enterobacteriaceae*
	A	**Gostya9**	*Escherichia coli*	MH203051		101665	*Enterobacteriaceae*
	A	**SPC35**	*Salmonella/Escherichia coli*	HQ406778	ApplEnvMicro 77:2042	118351	*Enterobacteriaceae*
	A	**SP3**	*Salmonella enterica Typhimurium*	MG387042		109306	*Enterobacteriaceae*
	A	**SP01**	*Salmonella enterica Enteritidis*	KY114934		117842	*Enterobacteriaceae*
	A	**SSP1**	*Shigella sonnei*	KY963424		113299	*Enterobacteriaceae*
	A	**STG2**	*Salmonella enterica Typhimurium*	MK005300		114275	*Enterobacteriaceae*
	A	**H8#**	*Salmonella enterica*	AC171169	JBact 189:5658	104373	*Enterobacteriaceae*
	A	**LVR16A**	*Salmonella enterica Kentucky*	MF681663		111601	*Enterobacteriaceae*
	A	**mar003J3**	*Escherichia coli*	LR027389		115471	*Enterobacteriaceae*
	A	**NR01**	*Salmonella enterica*	KR233164		111325	*Enterobacteriaceae*
	A	**øLLS**	*Escherichia coli*	KY677846	FrontMicro 8:in press	107263	*Enterobacteriaceae*
	A	**øR201**	*Yersinia enterocolitica*	HE956708	–	112795	*Yersiniaceae*
	A	**OSYSP**	*Escherichia coli O157:H7*	MF402939		110901	*Enterobacteriaceae*
	A	**PHB06#**	*Salmonella enterica Enteritidis*	MH102285		84406#	*Enterobacteriaceae*
	A	**S113**	*Salmonella enterica Typhimurium*	MH370366		112582	*Enterobacteriaceae*
	A	**S114**	*Salmonella enterica*	MH370367		110926	*Enterobacteriaceae*
	A	**S124**	*Salmonella enterica Derby*	MH370375		112564	*Enterobacteriaceae*
	A	**S126**	*Salmonella enterica Dublin*	MH370376		111999	*Enterobacteriaceae*
	A	**S130**	*Salmonella enterica Enteritidis*	MH370377		110091	*Enterobacteriaceae*
	A	**S131**	*Salmonella enterica Enteritidis*	MH370378		110091	*Enterobacteriaceae*
	A	**S132**	*Salmonella enterica*	MH370379		116832	*Enterobacteriaceae*
	A	**S133**	*Salmonella enterica*	MH370380		110926	*Enterobacteriaceae*
	A	**S147**	*Salmonella enterica Typhimurium*	MH370386		111447	*Enterobacteriaceae*
	A	**SH9**	*Salmonella enterica Hadar*	MF001363		111607	*Enterobacteriaceae*
	A	**Stitch**	*Salmonella enterica*	KM236244	GenomeA 3:e01435-14	123475	*Enterobacteriaceae*
	A	**Stp1 #**	*Salmonella enterica Typhimurium*	KY775453			*Enterobacteriaceae*
	A	**Sw2**	*Salmonella enterica Kentucky*	MH631454		114274	*Enterobacteriaceae*
	A	**Shivani**	*Salmonella enterica*	KP143763	GenomeA 3:e01443-14	120098	*Enterobacteriaceae*
	A	**SHSML-45**	*Shigella sonnei*	KX130863		108050	*Enterobacteriaceae*
	A	**slur09**	*Escherichia coli*	LN887948		111751	*Enterobacteriaceae*
	A	**T5**	*Escherichia coli*	AY543070	Virology 332:45	121752	*Enterobacteriaceae*
	A	**chee24**	*cow milk cheese*	MF431730	FrontMicrobiol2018In press	120622	*unknown host*
	A	**pork27**	*raw pork meat*	MF431731	FrontMicrobiol2018In press	120618	*unknown host*
	A	**pork29**	*raw pork meat*	MF431732	FrontMicrobiol2018In press	120622	*unknown host*
	A	**saus47N**	pork sausage	MF431733	FrontMicrobiol2018In press	120622	*unknown host*
	A	**saus111K**	pork sausage	MF431734	FrontMicrobiol2018In press	120620	*unknown host*
	A	**poul124**	poultry meat	MF431735	FrontMicrobiol2018In press	120629	*unknown host*
	A	**chee130_1**	cheese	MF431736	FrontMicrobiol2018In press	121986	*unknown host*
	A	**saus132**	pork sausage	MF431737	FrontMicrobiol2018In press	121986	*unknown host*
	A	**poul149**	poultry meat	MF431738	FrontMicrobiol2018In press	121986	*unknown host*
	A	**chee158**	*?*	MF431739	FrontMicrobiol2018In press	121986	*unknown host*
	A	**cott162**	?	MF431740	FrontMicrobiol2018In press	121986	*unknown host*
	A	**saus176N**	pork sausage	MF431741	FrontMicrobiol2018In press	121986	*unknown host*
	B	**My1**	*Pectobacterium carotovorum*	JX195166	JVirol 86:11410	122024	*Pectobacteriaceae*
	B	**DU_PP_V**	*Pectobacterium sp.*	MF979564		106185	*Pectobacteriaceae*
	C	**IME260**	*Klebsiella pneumoniae*	KX845404		123490	*Enterobacteriaceae*
	C	**Sugarland**	*Klebsiella pneumoniae*	MG459987		111103	*Enterobacteriaceae*
	D	**Stubb**	*Proteus mirabilis*	MH830339		104410	
	D	**PM135**	*Proteus mirabilis*	MG030347		104329	*Morganellaceae*
	E	**PreS_PR1**	*Providencia sp.*	KY363465		118537	*Morganellaceae*

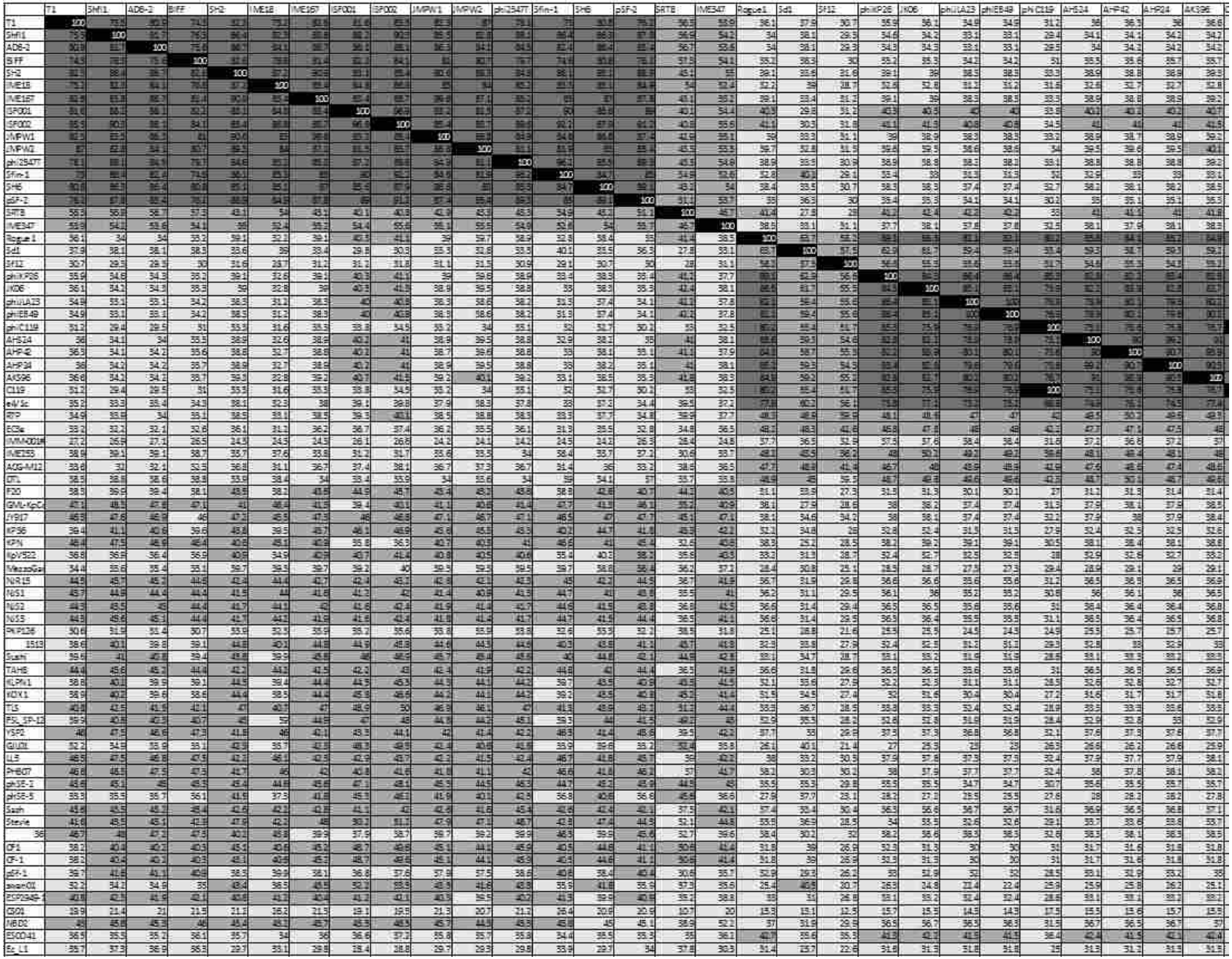
Lytic13: Chi-like	A	**37**	*Salmonella enterica*	KR296691		60216	*Enterobacteriaceae*
	A	**"Chi-DT104"**	*Salmonella enterica*	CVKM01000024		60058	*Enterobacteriaceae*
	A	**118970_sal1**	*Salmonella enterica*	KU927500		59518	*Enterobacteriaceae*
	A	**FSL_SP-030**	*Salmonella enterica*	KC139519	BMCgenomic 14:481	59746	*Enterobacteriaceae*
	A	**FSL_SP-039**	*Salmonella enterica*	KC139514	BMCgenomic 14:481	59815	*Enterobacteriaceae*
	A	**FSL_SP-088**	*Salmonella enterica*	KC139512	BMCgenomic 14:481	59454	*Enterobacteriaceae*
	A	**FSL_SP-124**	*Salmonella enterica*	KC139515	BMCgenomic 14:481	59245	*Enterobacteriaceae*
	A	**iEPS5**	*Salmonella enterica*	KC677662	ApplEnvMicro 79:4829	59214	*Enterobacteriaceae*
	A	**Siskin**	*Salmonella enterica*	MH631453		58476	*Enterobacteriaceae*
	A	**SPN19**	*Salmonella enterica*	JN871591	–	59203	*Enterobacteriaceae*
	A	**Utah**	*Escherichia coli*	KY014601	GenomeA5:e01494-16	59024	*Enterobacteriaceae*
	A	**35#**	*Salmonella enterica*	KR296689		55391	*Enterobacteriaceae*
	A	**BSPM4**	*Salmonella enterica*	KY620117		59097	*Enterobacteriaceae*
	A	**Chi (X)**	*Salmonella enterica*	JX094499; KM458633	ArchVirol 158:2179; GenomeA 3:e01229-14	59578	*Enterobacteriaceae*
	B	**KPN N137**	*Klebsiella pneumoniae*	MF415410		59100	*Enterobacteriaceae*
	B	**KPN N54**	*Klebsiella pneumoniae*	MF415413		59100	*Enterobacteriaceae*
	B	**KPN N98**	*Klebsiella pneumoniae*	MG835858			*Enterobacteriaceae*
	B	**KPN U2874**	*Klebsiella pneumoniae*	MF415411		59087	*Enterobacteriaceae*
	B	**Seifer**	*Klebsiella pneumoniae*	MH817999		58197	*Enterobacteriaceae*
	B	**YMC15/11/N53_KPN_BP**	*Klebsiella pneumoniae*	MF476924		59100	*Enterobacteriaceae*
	C	**Enc34**	*Enterobacter cancerogenus*	JQ340774	JVirol 86:11403	60364	*Enterobacteriaceae*
	D	**RedJac**	*Providencia stewartii*	NC_018832	PLoS One 8:e61762	58104	*Morganellaceae*
	E	**PM87**	*Proteus mirabilis*	MG030346		59128	*Morganellaceae*
	E	**pPM_01**	*Proteus mirabilis*	KP063118	Intervirology 59:243	58546	*Morganellaceae*

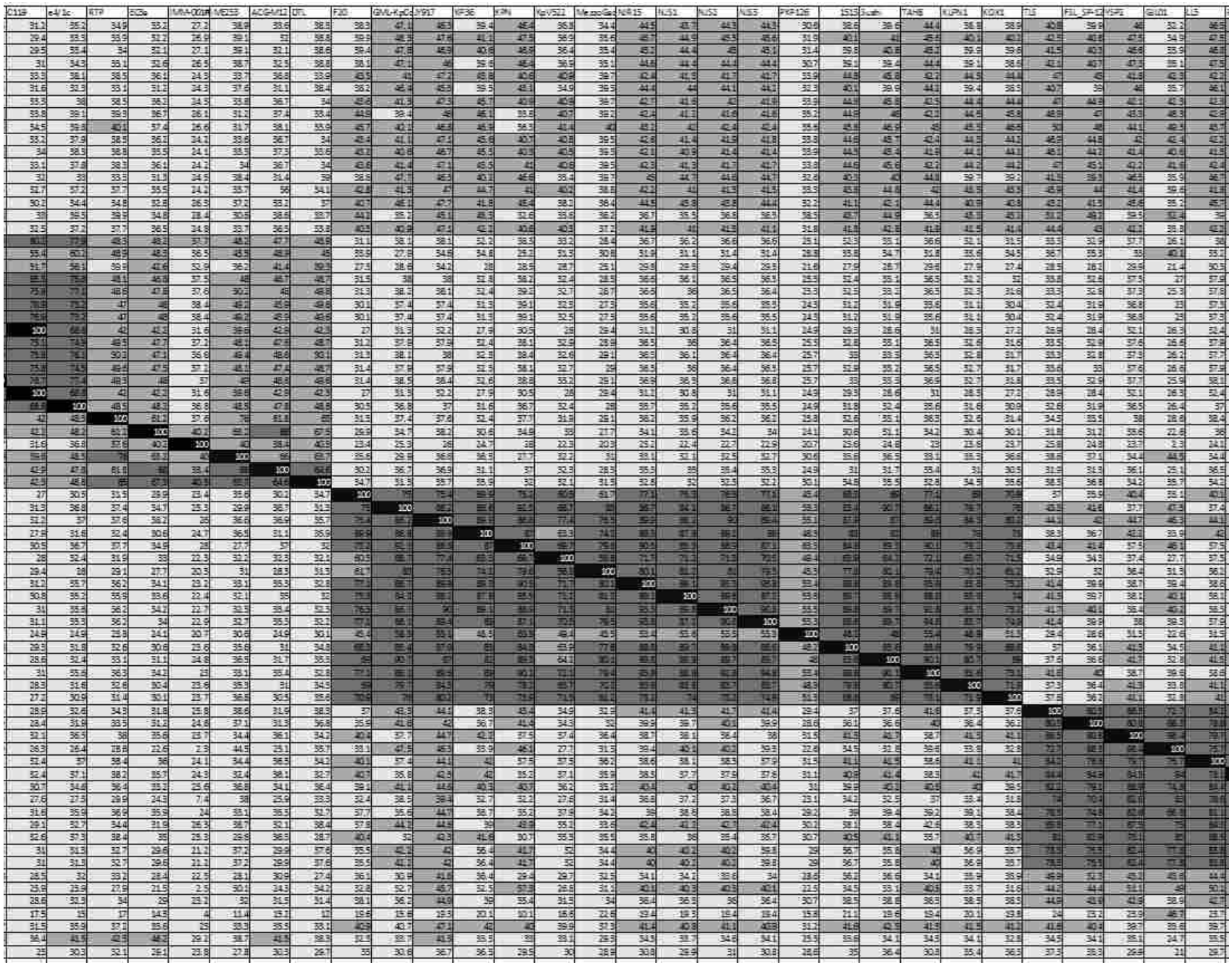
Lytic14: øEco32-like	A	**172-1**	*Escherichia coli*	KP308307		77266	*Enterobacteriaceae*
	A	**ECBP2 (KBNP135)**	*Escherichia coli*	JX415536	JVirol 86:12439	77315	*Enterobacteriaceae*
	A	**KBNP1711**	*Escherichia coli*	KF981730	–	76184	*Enterobacteriaceae*
	A	**NJ01**	*Escherichia coli*	JX867715	JVirol 86:13874	77448	*Enterobacteriaceae*
	A	**øEco32**	*Escherichia coli*	EU330206	JMolBiol 377:774	77554	*Enterobacteriaceae*
	A	**LAMP**	*Escherichia coli*	MG673519		68521	*Enterobacteriaceae*
	A	**SU10**	*Escherichia coli*	KM044272	PLoS One 9:e116294	77327	*Enterobacteriaceae*
	A	**myPSH1131**	*Escherichia coli*	MG983840	PLosOne 13:e0206278	76163	*Enterobacteriaceae*
	B	**7-11**	*Salmonella enterica*	HM997019	ArchVirol156:149	89916	*Enterobacteriaceae*
	B	**SE131**	*Salmonella enterica (Enteritidis)*	MG873442		89910	*Enterobacteriaceae*
	C	**GAP52**	*Cronobacter sakazakii*	JN882286	–	76631	*Enterobacteriaceae*

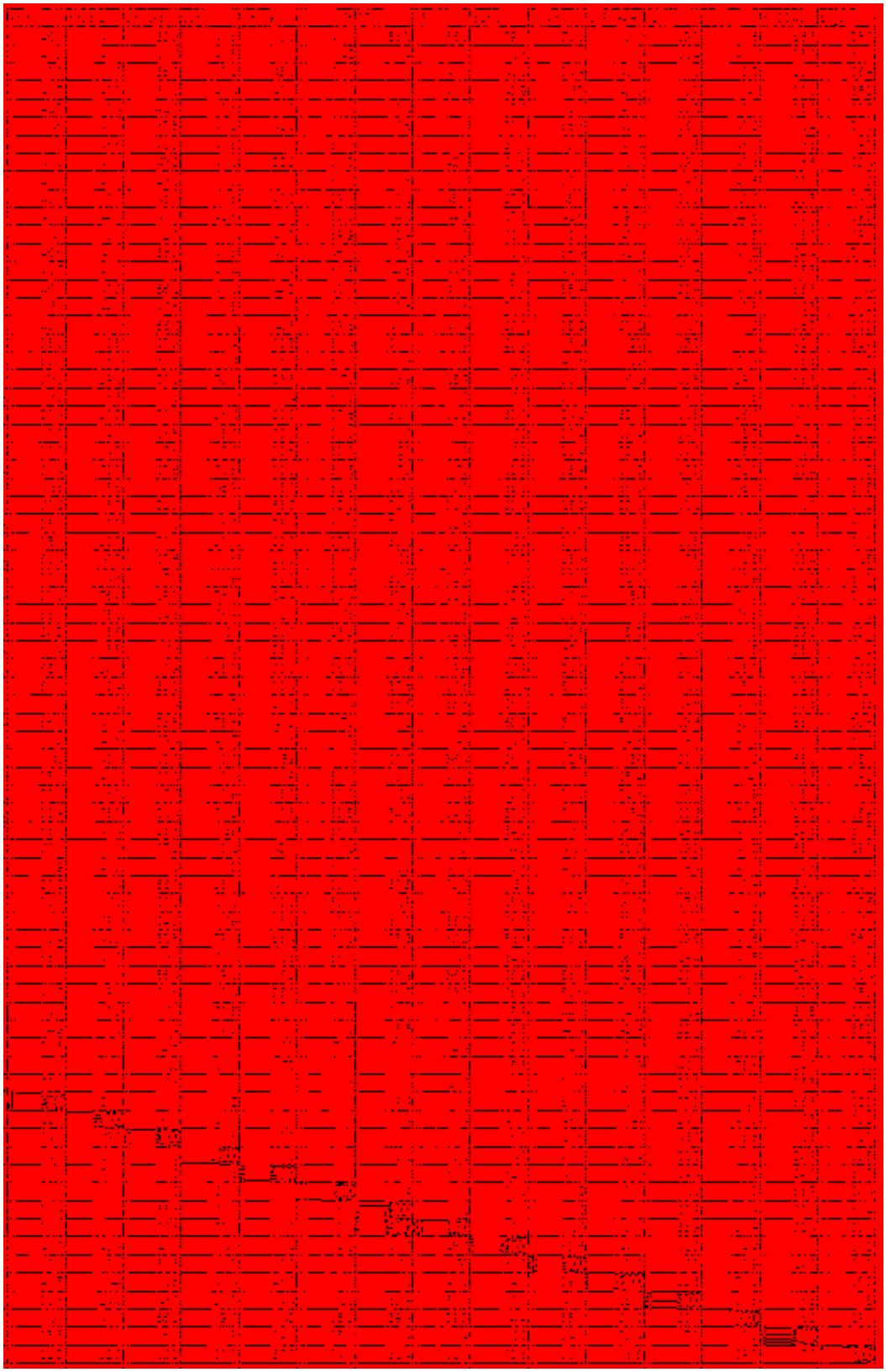
Lytic15: Felix-O1-like	A	**Alf5**	*Escherichia coli*	KX377933	GenomeA 5:e00315-17	87662	*Enterobacteriaceae*
	A	**AYO145A**	*Escherichia coli*	KR014248		87372	*Enterobacteriaceae*
	A	**BPS15Q2**	*Salmonella enterica Heidelberg*	KX405003	CurrBiol Dec 2016	89817	*Enterobacteriaceae*
	A	**BPS15S6**	*Salmonella enterica Heidelberg*	MG646670		87609	*Enterobacteriaceae*
	A	**BPS17L1**	*Salmonella enterica Shubra*	MG646672		84916	*Enterobacteriaceae*
	A	**BPS17W1**	*Salmonella enterica Shubra*	MG646669		87609	*Enterobacteriaceae*
	A	**BPS17S6**	*Salmonella enterica Shubra*	MG646671		87628	*Enterobacteriaceae*
	A	**Felix-O1**	*Salmonella enterica*	AF320576	Viruses 2:710	86155	*Enterobacteriaceae*
	A	**UAB_Phi87**	*Salmonella enterica*	JN225449	FrontMicro 7:545	87603	*Enterobacteriaceae*
	A	**FO1a**	*Salmonella enterica*	JF461087	–	88331	*Enterobacteriaceae*
	A	**FSL-SP-010#**	*Salmonella enterica*	KC139527+other	BMCgenomics 14:481	–	*Enterobacteriaceae*
	A	**FSL-SP-012#**	*Salmonella enterica*	KC139543+other	BMCgenomics 14:481	–	*Enterobacteriaceae*
	A	**FSL-SP-107#**	*Salmonella enterica*	KC139640+other	BMCgenomics 14:481	–	*Enterobacteriaceae*
	A	**EC6**	*Escherichia coli*	JX560968	GenomeA 1: e00085-12	86231	*Enterobacteriaceae*
	A	**HY02**	*Escherichia coli*	KM092515		86252	*Enterobacteriaceae*
	A	**IME338**	*"Enterobacteria"*	MH051914		85675	*Enterobacteriaceae*
	A	**wV8**	*Escherichia coli*	EU877232	VirolJ 6:41	88487	*Enterobacteriaceae*
	A	**XTG1**	*Escherichia coli*	KT184316		89635	*Enterobacteriaceae*
	A	**JH2**	*Escherichia coli*	KF055347	–	87721	*Enterobacteriaceae*
	A	**KhF1**	*Escherichia coli*	KT184313		88356	*Enterobacteriaceae*
	A	**KhF2**	*Escherichia coli*	KT184314		88309	*Enterobacteriaceae*
	A	**KhF3**	*Escherichia coli*	KT184315		88016	*Enterobacteriaceae*
	A	**Vpa-E1**	*Escherichia coli*	KM657822	–	88403	*Enterobacteriaceae*
	A	**VSe11**	*Salmonella enterica (Enteritidis)*	MG251391	GenomeA 6:e00398-18	86360	*Enterobacteriaceae*
	A	**VSe102**	*Salmonella enterica (Enteritidis)*	MG251392	GenomeA 6:e00398-18	86365	*Enterobacteriaceae*
	A	**Mushroom**	*Salmonella sp.*	KP143762	GenomeA 3:e00154	87709	*Enterobacteriaceae*
	A	**SBA-1781 #**	*Salmonella enterica*	JX181814			*Enterobacteriaceae*
	A	**Si3**	*Salmonella enterica Infantis*	KY626162		84419	*Enterobacteriaceae*
	A	**SP116**	*Salmonella enterica Typhimurium*	KP010413		87510	*Enterobacteriaceae*
	A	**SPT-1#**	*Salmonella sp.*	JX181822			*Enterobacteriaceae*
	A	**ST11**	*Salmonella enterica Pullorum*	MF370225		82101	*Enterobacteriaceae*
	A	**TP1**	*Escherichia coli*	KP869100	BMCGenom 16:271	88531	*Enterobacteriaceae*
	A	**TP8**	*Escherichia coli*	KP869106	BMCGenom 16:271	88998	*Enterobacteriaceae*
	A	**TP11**	*Escherichia coli*	KP869109	BMCGenom 16:271	88771	*Enterobacteriaceae*
	A	**TP12**	*Escherichia coli*	KP869110	BMCGenom 16:271	88632	*Enterobacteriaceae*
	A	**TP15**	*Escherichia coli*	KP869113	BMCGenom 16:271	92632	*Enterobacteriaceae*
	A	**SUSP1**	*Escherichia coli*	KT454805		90743	*Enterobacteriaceae*
	A	**SUSP2**	*Escherichia coli*	KT454806		88698	*Enterobacteriaceae*
	A	**Mijalis**	*Citrobacter freundii*	KY654690	GenomeA :5:e00228-17	87998	*Enterobacteriaceae*
	A	**Mordin**	*Citrobacter freundii*	KT363872	GenomeA 3:e01203-15	89596	*Enterobacteriaceae*
	A	**Moogle**	*Citrobacter freundii*	KM236239	GenomeA :3:e01426-14	87999	*Enterobacteriaceae*
	A	**Maleficent**	*Citrobacter freundii*	MH920362		89570	*Enterobacteriaceae*
	A	**Michonne**	*Citrobacter freundii*	KT001916	GenomeA 3:e01134-15	90000	*Enterobacteriaceae*
	A	**Sf13**	*Shigella*	MF158040	JVirol 92:e02117-17	87570	*Enterobacteriaceae*
	A	**Sf14**	*Shigella*	MF327003	JVirol 92:e02117-17	87575	*Enterobacteriaceae*
	A	**Sf15**	*Shigella*	MF158041	JVirol 92:e02117-17	88474	*Enterobacteriaceae*
	A	**Sf16**	*Shigella*	MF158043	JVirol 92:e02117-17	88580	*Enterobacteriaceae*
	A	**Sf17**	*Shigella*	MF327004	JVirol 92:e02117-17	90092	*Enterobacteriaceae*
	A	**Sf18**	*Shigella*	MF158044	JVirol 92:e02117-17	90270	*Enterobacteriaceae*
	A	**Sf19**	*Shigella*	MF327005	JVirol 92:e02117-17	90375	*Enterobacteriaceae*
	B	**M7**	*Erwinia amylovora*	HQ728263	ApplEnvMicro 77:5945	84694	*Erwiniaceae*
	B	**øEa104**	*Erwinia amylovora*	FQ482083	JBact 193:795	84565	*Erwiniaceae*
	B	**øEa116#**	*Erwinia amylovora*	FQ857195			*Erwiniaceae*
	B	**øEa21-4**	*Erwinia amylovora +others*	EU710883	ApplEnvMicro 75:2139	84567	*Erwiniaceae*

Lytic16: SETP3-like	A	**vB_sens_AG11**	*Salmonella enterica*	JX297445	–	41546	*Enterobacteriaceae*
	A	**BPS11Q3**	*Salmonella enterica*	KX405002	CurrBiol Dec 2016	43788	*Enterobacteriaceae*
	A	**BPS11T2**	*Salmonella enterica (Enteritidis)*	MG646668		43797	*Enterobacteriaceae*
	A	**SG2#**	*(Gallinarium)*	MF001356		33010#	*Enterobacteriaceae*
	A	**vB_SenS-Ent1**	*Salmonella enterica*	HE775250	JGenVirol 93:2046	42391	*Enterobacteriaceae*
	A	**vB_SenS-Ent2**	*Salmonella enterica*	HG934469		42093	*Enterobacteriaceae*
	A	**vB_SenS-Ent3**	*Salmonella enterica*	HG934470	–	42764	*Enterobacteriaceae*
	A	**SE2**	*Salmonella enterica*	JQ007353	JVirol 86:7712	43221	*Enterobacteriaceae*
	A	**SS3e (KS7)**	*Salmonella enterica*	AY730274	–	40794	*Enterobacteriaceae*
	A	**ST1**	*Salmonella enterica*	MF001366		42285	*Enterobacteriaceae*
	A	**ST3**	*Salmonella enterica*	MF001364		42266	*Enterobacteriaceae*
	A	**ST4**	*Salmonella enterica*	JX233783			*Enterobacteriaceae*
	A	**SE40#**	*Salmonella enterica (Enteritidis)*	KY626163			*Enterobacteriaceae*
	A	**SETP13**	*Salmonella enterica*	KF562864	–	42665	*Enterobacteriaceae*
	A	**SETP3**	*Salmonella enterica*	EF177456	JMedMicro 58:86	42572	*Enterobacteriaceae*
	A	**SETP7**	*Salmonella enterica*	KF562865	–	42789	*Enterobacteriaceae*
	A	**wksl3**	*Salmonella enterica*	JX202565	ApplEnvMicro 79:1958	42633	*Enterobacteriaceae*
	A	**FSL_SP-101**	*Salmonella enterica*	KC139511	BMCgenomics 14:481	41873	*Enterobacteriaceae*
	A	**Jersey**	*Salmonella enterica*	KF148055	–	43447	*Enterobacteriaceae*
	A	**STP03**	*(Typhimurium)*	KY176369		43428	*Enterobacteriaceae*
	A	**VSe103**	*Salmonella enterica (Enteritidis)*	MH424443		42262	*Enterobacteriaceae*
	A	**VSt10**	*(Typhimurium)*	MH424445		41581	*Enterobacteriaceae*
	A	**fSE1C**	*Salmonella enterica (Enteritidis)*	KT962832	StdGenomSci12:1	41720	*Enterobacteriaceae*
	A	**fSE4S**	*Salmonella enterica (Enteritidis)*	KT881477	StdGenomSci12:1	41768	*Enterobacteriaceae*
	A	**f18SE**	*Salmonella enterica (Pullorum)*	KR270151	GenomeA 3:00600-215	41868	*Enterobacteriaceae*
	A	**f2SE**	*Salmonella enterica (Enteritidis)*	KU951146		41865	*Enterobacteriaceae*
	A	**f3SE**	*Salmonella enterica (Enteritidis)*	KU951147		41867	*Enterobacteriaceae*
	A	**L13**	*Salmonella enterica*	KC832325			*Enterobacteriaceae*
	A	**LSPA1**	*Salmonella enterica* Paratyphi A	KM272358	GenomeA 3:01011-14	41880	*Enterobacteriaceae*
	A	**LPSE1**	*Salmonella enterica (Enteritidis)*	KY379853		41854	*Enterobacteriaceae*
	A	**LSHG-59"**	*Salmonella enterica (Enteritidis)*	LSHG01000059	bacterial genome project	41864	*Enterobacteriaceae*
	A	**MA12**	*(Enteritidis)****	KX245013	GenomeA e00810-16	41224	*Enterobacteriaceae*
	A	**phi135**	*Salmonella enterica (Enteritidis)*	MH992509		43142	*Enterobacteriaceae*
	A	**PVP_SE2**	*Salmonella enterica (Enteritidis)*	MF431252		42425	*Enterobacteriaceae*
	A	**S100**	*Salmonella enterica Typhimurium*	MH370358		43468	*Enterobacteriaceae*
	A	**S101**	*Salmonella enterica Typhimurium*	MH370359		42621	*Enterobacteriaceae*
	A	**S102**	*Salmonella enterica (Enteritidis)*	MH370360		42439	*Enterobacteriaceae*
	A	**S103**	*Salmonella enterica (Enteritidis)*	MH370361		42441	*Enterobacteriaceae*
	A	**S104**	*Salmonella enterica (Enteritidis)*	MH370362		43118	*Enterobacteriaceae*
	A	**S106**	*Salmonella enterica (Enteritidis)*	MH370363		42976	*Enterobacteriaceae*
	A	**S111**	*Salmonella enterica (Enteritidis)*	MH370365		43421	*Enterobacteriaceae*
	A	**S119**	*Salmonella enterica (Enteritidis)*	MH370372		43876	*Enterobacteriaceae*
	A	**S120**	*Salmonella enterica Typhimurium*	MH370373		43467	*Enterobacteriaceae*
	A	**S123**	*Salmonella enterica (Enteritidis)*	MH370374		43467	*Enterobacteriaceae*
	A	**S134**	*Salmonella enterica (Enteritidis)*	MH370381		43118	*Enterobacteriaceae*
	A	**S138**	*Salmonella enterica (Enteritidis)*	MH370384		43119	*Enterobacteriaceae*
	A	**S142**	*Salmonella enterica (Enteritidis)*	MH370385		43119	*Enterobacteriaceae*
	B	**K1-dep(4) / (K1G)**	*Escherichia coli*	GU196277	Virology 398:79	43587	*Enterobacteriaceae*
	B	**K1-dep(1) / (K1H)**	*Escherichia coli*	GU196278	Virology 398:79	41632	*Enterobacteriaceae*
	B	**K1-ind(1)**	*Escherichia coli*	GU196279	Virology 398:79	42292	*Enterobacteriaceae*
	B	**K1-ind(2)**	*Escherichia coli*	GU196280	Virology 398:79	42765	*Enterobacteriaceae*
	B	**K1-ind(3)**	*Escherichia coli*	GU196281ß	Virology 398:79	43461	*Enterobacteriaceae*
	B	**L AB-2017**	*Escherichia coli*	KY295896		41039	*Enterobacteriaceae*
	B	**P AB-2017**	*Escherichia coli*	KY295898		41184	*Enterobacteriaceae*
	B	**EcoS_MY**	*Escherichia coli*	MG099933		44829	*Enterobacteriaceae*
	B	**ST2**	*Escherichia coli*	MF153391		44517	*Enterobacteriaceae*
	B	**Golestan**	*Escherichia coli*	MG099933		44829	*Enterobacteriaceae*
	B	**G AB-2017**	*Escherichia coli*	KY295895		41519	*Enterobacteriaceae*
	C	**St161**	*Salmonella typhimurium*	MF158036	JVirol 92:e02117-17 - 10-2017	29178#	*Enterobacteriaceae*
	C	**St162**	*Salmonella typhimurium*	MF158037	JVirol 92:e02117-17	43701	*Enterobacteriaceae*
	C	**VSiP**	*Salmonella enterica (Infantis)*	MH424444		43110	*Enterobacteriaceae*
	C	**FSL_SP-031#**	*Salmonella enterica*	KC139518	BMCgenomics 14:481		*Enterobacteriaceae*
	C	**FSL_SP-038#**	*Salmonella enterica*	KC139652-66	BMCgenomics 14:481		*Enterobacteriaceae*
	C	**FSL_SP-049#**	*Salmonella enterica*	KC139557-59	BMCgenomics 14:481		*Enterobacteriaceae*
	C	**øEap-2**	*Enterobacter aerogenes*	KT287080	SciRep 6:28338	40491	*Enterobacteriaceae*
	D	**Eta (η)**	*Serraitia marcescens*	KC460990	VirolJ 11:6	42724	*Yersiniaceae*

Supplementary Table 4.S2 ANI chart comparing all 74 phages of Lytic 1 (T1-like) cluster







Supplementary Table 4.S3 List of phages used in genomic and proteomic analysis that needed annotation corrections . (Updated September 15, 2017)

CLUSTER	***PHAGE NAME***	***HOST***	***Accession number***	***Corrections made?***
Lytic1	*ADB-2*	*Escherichia coli*	*JX912252*	
	EcoS_SH2	*Escherichia coli*	*KY985004*	
	JMPW1	*Escherichia coli*	*KU194206*	
	JMPW2	*Escherichia coli*	*KU194205*	
	pSf-2	*Shigella flexneri*	*KP085586*	
	SH6	*Shigella sp.*	*KX828710*	
	Shfl1	*Shigella flexneri*	*NC_015456*	
	SsoS-ISF002	*Shigella sonnei*	*MF093736*	
	T1	*Escherichia coli*	*AY216660*	
	AHP24	*Escherichia coli*	*KF771236*	
	AHP42	*Escherichia coli*	*KF771237*	
	AHS24	*Escherichia coli*	*KF771238*	
	AKS96	*Escherichia coli*	*KF771239*	
	C119	*Escherichia coli*	*KT825490*	
	e4/1c	*Escherichia coli*	*KJ668713*	
	JK06	*Escherichia coli*	*DQ121662*	*Yes*
	øC119	*Escherichia coli*	*KT825490*	
	øEB49	*Escherichia coli*	*JF770475*	
	øJLA23	*Escherichia coli*	*KC333879*	

	øKP26	*Escherichia coli/S. enterica*	*KC579452*	*Yes*
	Rogue1	*Escherichia coli*	*JQ182736*	*Yes*
	ACG-M12	*Escherichia coli*	*NC_019404*	
	1513	*Klebsiella pneumoniae*	*KP658157*	
	KLPN1	*Klebsiella pneumoniae*	*KT001920*	
	KOX1	*Klebsiella pneumoniae*	*KY780482*	
	KPN N141	*Klebsiella pneumoniae*	*MF415412*	
	KpV522	*Klebsiella pneumoniae*	*KX237515*	
	MezzoGao	*Klebsiella pneumoniae*	*MF612072*	
	PKP126	*Klebsiella pneumoniae*	*KR269719*	
	Sushi	*Klebsiella pneumoniae*	*KR262148*	
	CF-1	*Citrobacter freundii*	*KY694971*	
	GJL01	*Salmonella enterica Pullorum*	*KY657202*	
	phSE-2	*Salmonella enterica*	*KX015770*	
	phSE-5	*Salmonella enterica*	*KX015771*	
	Stevie	*Citrobacter freundii*	*KM236241*	
	TLS	*Escherichia coli*	*AY308796*	
	ESP2949-1	*Cronobacter sakazakii*	*JF912400*	*Yes*
	pSf-1	*Shigella flexneri*	*KC710998*	
	swan01	*Escherichia coli*	*LT841304*	
	36#	*Salmonella enterica*	*KR296690*	
	FSL_SP-126 #	*Salmonella enterica*	*KC139521*	

	NBD2	*Escherichia coli*	*KX130668*	
	ESCO_41	*Escherichia coli*	*KY619305*	
	EcoS_CEB_EC3a	*Escherichia coli*	*KY398841*	
	EcoS-IME253	*Escherichia coli*	*KX130960*	
Lytic2	*7 (TP7)*	*Escherichia coli O157*	*KP869105*	
	ACG-C40	*Escherichia coli*	*JN986846*	*Yes*
	AR1	*Escherichia coli*	*AP011113*	
	CF2	*Escherichia coli*	*KY608967*	*Yes*
	e11/2 (EcoM_112)	*Escherichia coli*	*NC_024125*	
	ECML-134	*Escherichia coli*	*JX128259*	
	HY01	*Escherichia coli*	*KF925357*	
	HY03	*Escherichia coli*	*KR269718*	
	IME09	*Escherichia coli*	*JN202312*	
	PE37	*Escherichia coli*	*KU925172*	
	PEC04	*Escherichia coli*	*KR233165*	*Yes*
	RB10	*Escherichia coli*	*KM606999*	
	RB14	*Escherichia coli*	*NC_012638*	
	RB27	*Escherichia coli*	*KM607000*	*Yes*
	RB3	*Escherichia coli*	*KM606994*	
	RB32	*Escherichia coli*	*NC_008515*	
	RB33	*Escherichia coli*	*KM607001*	
	RB5	*Escherichia coli*	*KM606995*	

	RB51	*Escherichia coli*	*NC_012635*	
	RB55	*Escherichia coli*	*KM607002*	
	RB59	*Escherichia coli*	*KM607003*	
	RB6	*Escherichia coli*	*KM606996*	
	RB68	*Escherichia coli*	*KM607003*	
	RB7	*Escherichia coli*	*KM606997*	
	RB9	*Escherichia coli*	*KM606998*	
	slur02	*Escherichia coli*	*LN881726*	*Yes*
	slur03	*Escherichia coli*	*LN881728*	*Yes*
	slur04	*Escherichia coli*	*LN881729*	*Yes*
	slur07	*Escherichia coli*	*LN881732*	*Yes*
	slur08	*Escherichia coli*	*LN881733*	*Yes*
	slur11	*Escherichia coli*	*LN881734*	*Yes*
	slur13	*Escherichia coli*	*LN881737*	*Yes*
	slur14	*Escherichia coli*	*LN881736*	*Yes*
	T4	*Escherichia coli*	*AY318471 /AF158101*	*Yes*
	UFV-AREG1	*Escherichia coli*	*KX009778*	*Yes*
	UFV13	*Escherichia coli*	*KU867876*	
	wV7	*Escherichia coli*	*HM997020*	
	YUEEL01	*Escherichia coli*	*KY290975*	
	pSs-1	*Shigella sonnei*	*KM501444*	
	SH7	*Shigella sp.*	*KX828711*	

	SHBML-50-1	*Shigella sonnei*	*KX130864*	
	Shfl2	*Shigella flexneri*	*NC_015457*	
	SHFML-11	*Shigella sonnei*	*KX130861*	
	SHFML-26	*Shigella sonnei*	*KX130862*	
	øD1	*Yersinia pestis*	*HE956711*	*Yes*
	PST	*Yersinia pestis*	*KF208315*	
	13 (TP13)	*Escherichia coli O157*	*KP869111*	
	3 (TP3)	*Escherichia coli O157*	*KP869101*	
	6 (TP6)	*Escherichia coli O157*	*KP869104*	
	APCEc01	*Escherichia coli*	*KR422352*	
	HX01	*Escherichia coli*	*JX536493*	
	JS09	*Escherichia coli*	*NC_024124*	
	øC120	*Escherichia coli*	*KY703222*	
	øE142#	*Escherichia coli (& salmonella?)*	*KU255730*	
	PhAPEC2	*Escherichia coli*	*KF562341*	
	RB69	*Escherichia coli*	*NC_004928*	
	ST0	*Escherichia coli*	*MF044457*	
	SHBML-52-1	*Shigella sonnei*	*KX130865*	
	Shf125875	*Shigella flexneri*	*KM407600.*	
	Bp7	*Escherichia coli*	*HQ829472*	
	IME08	*Escherichia coli*	*NC_014260*	*Yes*
	JS10	*Escherichia coli*	*EU863409*	

	JS98	*Escherichia coli*	*EF469154*	
	MX01	*Escherichia coli*	*KU878969*	
	QL01	*Escherichia coli*	*KT176190*	*Yes*
	VR5	*Escherichia coli*	*KP007359*	
	WG01	*Escherichia coli*	*KU878968*	
	VR20	*Escherichia coli*	*KP007360*	
	VR25	*Escherichia coli*	*KP007361*	
	VR26	*Escherichia coli*	*KP007362*	
	VR7	*Escherichia coli*	*HM563683*	
	SP18	*Shigella sonnei*	*GQ981382*	
	CGG4-1	*Salmonella enterica Newport*	*NC_031065*	
	S16 (SenMS16)	*Salmonella enterica*	*HQ331142*	*Yes*
	STML_198	*Salmonella enterica*	*JX181825*	
	STP4-a	*Salmonella enterica*	*KJ000058*	
	fHe-Yen9-01	*Yersinia enterocolitica*	*KY593455*	
	øR1-RT	*Yersinia enterocolitica*	*HE956709*	
	TG1	*Yersinia enterocolitica*	*KP202158*	
	Pet-CM3-4	*Cronobacter malonaticus*	*LT614807*	
	PG7	*Enterobacter cloacae*	*KJ101592*	
	CC31	*Escherichia coli*	*GU323318*	
	ECD7	*Escherichia coli*	*KY683735*	
	GEC-3S	*Escherichia coli*	*HE978309*	

	JSE	*Escherichia coli*	*EU863408*	
	ø1	*Escherichia coli*	*NC_009821*	
	RB49	*Escherichia coli*	*NC_005066*	
	CfP1	*Citrobacter freundii*	*KX245890*	
	IME-CF2	*Citrobacter freundii*	*KR869820*	
	Margaery	*Citrobacter freundii*	*KT381880*	
	Miller	*Citrobacter freundii*	*KM236237*	
	GAP161	*Cronobacter sakazakii*	*JN882287*	*Yes*
	leB	*Cronobacter*	*KX443552*	
	leE	*Cronobacter*	*KX431559*	
	leN	*Cronobacter*	*KX431560*	
	Lw1	*Escherichia coli*	*NC_021344*	
	RB16	*Escherichia coli*	*HM134276*	
	RB43	*Escherichia coli*	*NC_007023*	
	PS2	*Serratia marcescens*	*KJ025957*	
	JD18	*Klebsiella pneumoniae*	*KT239446*	
	KpV477	*Klebsiella pneumoniae*	*KX258185*	
	PKO111	*Klebiella oxytoca*	*KR269720*	
	KPV15	*Klebsiella pneumoniae*	*KY000080*	
	MP1	*Morganella sp.*	*KX078569*	*Yes*
	Merlin	*Citrobacter freundii*	*KT001915*	*Yes*
	Moon	*Citrobacter freundii*	*KM236240*	
	øEap-3	*Enterobacter aerogenes*	*KT321315*	*Yes*

	phT4A#	*Escherichia coli*	*KX130727*	*Yes*
	KP15	*Klebsiella pneumoniae*	*GU295964*	*Yes*
	KP27	*Klebsiella pneumoniae*	*HQ918180*	*Yes*
	Matisse	*Klebsiella pneumoniae*	*KT001918*	
	Miro	*Klebsiella pneumoniae*	*KT001919*	
	PMBT1	*Klebsiella pneumoniae*	*LT607758*	
	PmiM_Pm5461	*Proteus mirabilis*	*NC_028762*	
	S13	*Cronobacter sakazakii*	*KC954775*	
	PEi20	*Edwardsiella ictaluri*	*AP014714*	
	PEi26	*Edwardsiella ictaluri*	*AP014715*	
	PM2	*Pectobacterium corotovforum*	*KF835987*	
	CBH8	*Serratia sp. ATCC 39006*	*MF036691*	
	CHI14	*Serratia sp. ATCC 39006*	*MF036690*	
	X20	*Serratia sp. ATCC 39006*	*MF036692*	
Lytic3	*38*	*Salmonella enterica*	*KR296692*	
	CBA120	*Escherichia coli*	*JN593240*	
	Det7	*Salmonella enterica*	*KP797973*	
	ECML-4	*Escherichia coli*	*JX128257*	*Yes*
	FSL_SP-029 #	*Salmonella enterica*	*KC139566+other*	*Yes*
	FSL_SP-063 #	*Salmonella enterica*	*KC139524+other*	

	GG32	*Salmonella enterica*	*KX245012*	*Yes*
	Marshall	*Salmonella enterica*	*KF669653*	
	Maynard	*Salmonella enterica*	*KF669654*	
	øSH19	*Salmonella enterica*	*JN126049*	
	PhaxI	*Escherichia coli*	*JN673056*	
	PM10	*Salmonella enterica*	*KX438380*	
	S8	*Salmonella enterica Gallinarum*	*KY630163*	
	SenM-2	*Salmonella sp.*	*KX171211*	
	SFP10	*Salmonella enterica*	*HQ259103*	
	SJ2	*Salmonella enterica*	*KJ174317*	
	SJ3	*Salmonella enterica*	*KJ174318*	
	SKML-39	*Salmonella enterica*	*JX181829*	
	SPT07	*Salmonella entericaTyphimurium*	*KY000003*	
	STML-13-1#	*Salmonella enterica*	*JX181828*	
	Vi01 (ViI)	*Salmonella enterica*	*FQ312032*	
	JA15	*Dickeya solani*	*KY942056*	
	LIMEstone1	*Dickeya solani*	*HE600015*	*Yes*
	øD3	*Dickeya sp.*	*KM209228*	
	øPD10.3 #	*Dickeya solani et al.*	*KM209270*	*Yes*
	øPD23.1 #	*Dickeya solani et al.*	*KM209320*	*Yes*
	øSboM-AG3	*Shigella boydii*	*FJ373894*	*Yes*

	RC_2014 (øD5)	*Dickeya sp.*	*KJ716335*	
	XF4	*Dickeya solani*	*KY942057*	
	0507-KN2-1	*Klebsiella pneumoniae*	*AB797215*	
	KSP90#	*Serratia plymuthica*	*AB452990*	*Yes*
	øMAM1	*Serratia plymuthica*	*JX878496*	
	Sru_IME250	*Serratia rubidaea*	*KY073123*	
	øEa2809	*Erwinia amylovora*	*KP037007*	
Lytic4	*100268_sal2*	*Salmonella enterica Enteritidis*	*KU927497*	
	118970_sal2	*Salmonella enterica Enteritidis*	*KX017521*	
	AKFV33	*Escherichia coli*	*NC_017969*	*Yes*
	APCEo03	*Escherichia coli*	*KR422353*	
	DT571/2	*Escherichia coli*	*KM979355*	
	DT57C	*Escherichia coli*	*KM979354*	
	EPS7	*Salmonella enterica*	*CP000917*	
	FFH1	*Escherichia coli*	*KJ190157*	
	H8#	*Salmonella enterica*	*AC171169*	*Yes*
	NR01	*Salmonella enterica*	*KR233164*	
	øLLS	*Escherichia coli*	*KY677846*	
	øR201	*Yersinia enterocolitica*	*HE956708*	
	OSYSP	*Escherichia coli O157:H7*	*MF402939*	
	Shivani	*Salmonella enterica*	*KP143763*	

	SHSML-45	*Shigella sonnei*	*KX130863*	
	slur09 #	*Escherichia coli*	*LN887948*	
	SPC35	*Salmonella/Escherichia coli*	*HQ406778*	
	Stitch	*Salmonella enterica*	*KM236244*	*Yes*
	Stp1 #	*Salmonella enterica Typhimurium*	*KY775453*	*Yes*
	T5	*Escherichia coli*	*AY543070*	
	My1	*Pectobacterium carotovorum*	*JX195166*	
	IME260	*Klebsiella pneumoniae*	*KX845404*	
	PreS_PR1	*Providencia sp.*	*KY363465*	
Lytic5	*64795_ec1*	*Escherichia coli*	*KU927499*	
	CICC 80001	*Escherichia coli*	*KM242061*	
	øA1122	*Yersinia pestis*	*AY247822*	
	R	*Yersinia pestis*	*JX000007*	
	T7	*Escherichia coli*	*V01146*	*Yes*
	Vi VI (VI06)	*Salmonella enterica*	*FR667955*	*Yes*
	Y	*Yersinia pestis*	*JQ957925*	
	YpP-R	*Yersinia pestis*	*JQ965701*	
	YpP-Y	*Yersinia pestis*	*JQ965700*	
	YpsP-G	*Yersinia pestis*	*JQ965703*	
	AP5	*Yersinia enterocolitica*	*KM253764*	

	E-2	*Enterobacter cloacae*	*KP791805*	
	E-3 #	*Enterobacter cloacae*	*KP791806*	
	E-4	*Enterobacter cloacae*	*KP791807*	
	ECA2	*Escherichia coli*	*KX130726*	
	øCFP-1	*Citrobacter freundii*	*KP313531*	
	øSG-JL2	*Salmonella enterica*	*EU547803*	
	øYe-F10	*Yersinia enterocolitica*	*KT008108*	
	øYeO3-12	*Yersinia enterocolitica*	*AJ251805*	*Yes*
	SH1	*Citrobacter freundii*	*KU687347*	
	SH2	*Citrobacter freundii*	*KU687348*	
	SM9-3Y	*Serratia marcescens*	*KX778611*	
	T3	*Escherichia coli*	*AJ318471*	
	285p	*Escherichia coli*	*GQ468526*	
	BA14	*Escherichia coli*	*NC_011040*	
	Berlin	*Yersinia pestis*	*NC_008694*	
	BP12A	*Salmonella enterica Hadar*	*KM366096*	
	FE44	*Erwinia (sp?)*	*NC_022744*	
	Kvp1	*Kluyvera cryocrescens*	*FJ194439*	
	P483	*Escherichia coli*	*KP090453*	
	P694	*Escherichia coli*	*KP090454*	
	PP74	*Pectobacterium wasabiae*	*KY084243*	*Yes*
	Yep-ø	*Yersinia pestis*	*HQ333270*	
	Yepe2	*Yersinia pestis*	*NC_011038*	

	YpP-G	*Yersinia pestis*	*JQ965702*	*Yes*
	BIS33	*Klebsiella pneumoniae*	*KY652725*	
	IL33	*Klebsiella pneumoniae*	*KY652724*	
	IME205	*Klebsiella(sp?)*	*KU183006*	
	K11	*Klebsiella sp. 390*	*EU734173*	
	K30	*Escherichia coli*	*HM480846*	*Yes*
	K5	*Klebsiella pneumoniae*	*KR149291*	
	K5-2	*Klebsiella pneumoniae*	*KY389315*	
	K5-4	*Klebsiella pneumoniae*	*KY389316*	
	Kp1	*Klebsiella pneumoniae*	*KT367885*	
	KP32	*Klebsiella pneumoniae*	*GQ413937*	
	KpV289	*Klebsiella pneumoniae*	*LN866626*	
	KpV763	*Klebsiella pneumoniae*	*KX591654*	
	KpV766	*Klebsiella pneumoniae*	*KX712071*	
	KpV767	*Klebsiella pneumoniae*	*KX712070*	
	PRA33	*Klebsiella pneumoniae*	*KY652723*	
	L1	*Erwinia amylovora*	*HQ728265*	
	MmP1	*Morganella morganii*	*EU652770*	
	MP2	*Morganella sp.*	*KX078568*	
	Dev2	*Cronobacter turicensis*	*HG813241*	
	EcoDS1	*Escherichia coli*	*NC_011042*	
	F AB-2017	*Escherichia coli*	*KY295894*	
	GA2A	*Escherichia coli*	*KT990215*	*Yes*

	JSS1	*Escherichia coli*	*KX689784*	
	K1F	*Escherichia coli*	*AM084414*	
	LM33_P1	*Escherichia coli*	*LT594300*	
	PE3-1	*Escherichia coli*	*KJ748011*	
	SH3	*Citrobacter freundii*	*KU687349*	
	SH4	*Citrobacter freundii*	*KU687350*	
	SH5	*Citrobacter freundii*	*KU687351*	
	ST31	*Escherichia coli*	*KY962008*	
	ZG49	*Escherichia coli*	*KX669227*	
	AP10	*Yersinia enterocolitica*	*KT852574*	
	øEAP-1	*Enterobacter aerogenes*	*KT321314*	
	PP47	*Pectobacterium carotovorum*	*KY250035*	
	PP81	*Pectobacterium carotovorum*	*KY124276*	
	PPWS4	*Pectobacterium carotovorum*	*LC216347*	
Lytic6	*AAPEc6*	*Escherichia coli*	*KX279892*	
	ACG-C91	*Escherichia coli*	*NC_019403*	
	B AB-2017	*Escherichia coli*	*KY295891*	
	BP12B	*Salmonella entericaHadar*	*KM366097*	
	C AB-2017	*Escherichia coli*	*KY295892*	
	D AB-2017	*Escherichia coli*	*KY295893*	

	K AB-2017	*Escherichia coli*	*KY295897*	
	K1-5 (K1-dep(3))	*Escherichia coli*	*AY370674*	
	K1E (K1dep(2))	*Escherichia coli*	*AM084415*	
	R AB-2017	*Escherichia coli*	*KY295899*	
	SP6	*Salmonella enterica*	*AY288927; AY370673*	
	UAB_Phi78	*Salmonella enterica*	*GU595417*	
	PP1	*Pectobacterium carotovorum*	*JQ837901*	
	ERA103	*Erwinia amylovora*	*EF160123*	
	øEa100	*Erwinia amylovora*	*FQ482086*	
	øEa1h	*Erwinia amylovora*	*FQ482084*	
	PM5460	*Proteus mirabilis*	*KP890822*	
	PM85	*Proteus mirabilis*	*KM819695*	
	PM93	*Proteus mirabilis*	*KM819696*	
	phD2B	*Lelliottia (was Enterobacter?)*	*KM370384*	
	ECBP5	*Escherichia coli*	*KJ749827*	
Lytic7	*AltoGau*	*Klebsiella pneumoniae*	*MF612071*	
	F19	*Klebsiella pneumoniae*	*KF765493*	
	KP-Rio/2015	*Klebsiella pneumoniae*	*KX856662*	*Yes*
	Kp2	*Klebsiella pneumoniae*	*KT367886*	
	KP34	*Klebsiella pneumoniae*	*NC_013649*	*Yes*

	KpV41	*Klebsiella pneumoniae*	*KT964103*	
	KpV475	*Klebsiella pneumoniae*	*KX211991*	
	KpV48	*Klebsiella pneumoniae*	*KX237514*	
	KpV71	*Klebsiella pneumoniae*	*KU666550*	
	KpV74	*Klebsiella pneumoniae*	*KY385423*	
	KPV811	*Klebsiella pneumoniae*	*KY000081*	
	NTUH-K2044-K1-1	*Klebsiella {sp?}*	*AB716666*	
	øBO1E	*Klebsiella pneumoniae*	*KM576124*	
	SU503	*Klebsiella pneumoniae*	*KP708985*	
	SU552A	*Klebsiella pneumoniae*	*KP708986*	
	LIMElight	*Pantoea agglomerans*	*FR687252*	
	øKDA1	*Enterobacter cloacae*	*JQ267518*	
	PM16	*Proteus mirabilis*	*KF319020*	
	PM75	*Proteus mirabilis*	*KM819694*	
Lytic8	*LIMEzero*	*Pantoea agglomerans*	*FR751545*	*Yes*
	J8-65	*Escherichia coli*	*NC_025445*	
Lytic9	*øKT*	*Escherichia coli*	*JN882298*	
Lytic10	*Dev-CD-23823*	*Coronobacter sakazakii*	*LN878149*	
	GAP227	*Coronobacter sakazakii*	*NC_020078*	*Yes*
	ISAO8	*Yersinia enterocolitica*	*KT184661*	
	øR8-01	*Yersinia entericolytica*	*HE956707*	
	fHe-Yen3-01	*Yersinia enterocolitica*	*KY318515*	

	ø80-18	*Yersinia entericolaytica*	*HE956710*	
	PP2	*Pectobacterium carotovorum*	*KX756572*	
Lytic39g	*Peat1*	*Pectobacterium atrosepticum*	*KR604693*	
	PP90	*Pectobacterium atrosepticum*	*KX278419*	
	øM1	*Pectobacterium atrosepticum*	*JX290549*	
	PP16	*Pectobacterium carotovorum*	*KX278418*	
	PPWS1	*Pectobacterium carotovorum*	*LC063634*	
	BF25/12	*Dickeya sp. B16*	*KT240186*	
Lytic11	*Bp4*	*Escherichia coli*	*KJ135004*	
	EC1-UPM	*Escherichia coli*	*KC206276.2*	
	ECBP1 (KNBP21?)	*Escherichia coli*	*JX415535*	
	G7C	*Escherichia coli*	*NC_015933*	*Yes*
	IME11	*Escherichia coli*	*NC_019423*	
	N4	*Escherichia coli*	*EF056009*	
	PhAPEC5	*Escherichia coli*	*KF192075*	
	PhAPEC7	*Escherichia coli*	*KF562340*	

	pSb-1	*Shigella boydii*	*KF620435*	
	FSL_SP-058	*Salmonella enterica*	*KC139517*	
	FSL_SP-076	*Salmonella enterica*	*KC139520*	
	Pollock	*Escherichia coli*	*KM236242*	*Yes*
	EcP1	*Enterobacter cloacae*	*NC_019485*	
	Ea9-2	*Erwinia amylovora*	*KF806588*	
	Frozen	*Erwinia amylovora*	*KX098389*	
	Gutmeister	*Erwinia amylovora*	*KX098390*	
	S6	*Erwinia amylovora*	*HQ728266*	
Lytic12	*9NA*	*Salmonella enterica*	*KJ802832*	
	Sasha	*Salmonella enterica*	*KX987158*	
	Sergiei	*Salmonella enterica*	*KY002061*	
Lytic13	*37*	*Salmonella enterica*	*KR296691*	
	"Chi-DT104"	*Salmonella enterica*	*CVKM01000024*	
	118970_sal1	*Salmonella enterica*	*KU927500*	
	35#	*Salmonella enterica*	*KR296689*	
	BP12C	*Salmonella enterica*	*AIT13784*	*Yes*
	Chi (X)	*Salmonella enterica*	*KM458633*	
	FSL_SP-030	*Salmonella enterica*	*KC139519*	
	FSL_SP-039	*Salmonella enterica*	*KC139514*	
	FSL_SP-088	*Salmonella enterica*	*KC139512*	
	FSL_SP-124	*Salmonella enterica*	*KC139515*	

	iEPS5	*Salmonella enterica*	*KC677662*	*Yes*
	SPN19	*Salmonella enterica*	*JN871591*	*Yes*
	pPM_01	*Proteus mirabilis*	*KP063118*	*Yes*
	KPN N137	*Klebsiella pneumoniae*	*MF415410*	
	KPN N54	*Klebsiella pneumoniae*	*MF415413*	
	KPN U2874	*Klebsiella pneumoniae*	*MF415411*	
	YMC15/11/N53_ KPN_BP	*Klebsiella pneumoniae*	*MF476924*	
Lytic14	*172-1*	*Escherichia coli*	*KP308307*	
	ECBP2 (KBNP135)	*Escherichia coli*	*JX415536*	*Yes*
	Eco32	*Escherichia coli*	*EU330206*	*Yes*
	KBNP1711	*Escherichia coli*	*KF981730*	*Yes*
	NJ01	*Escherichia coli*	*JX867715*	*Yes*
	SU10	*Escherichia coli*	*KM044272*	*Yes*
	7-11	*Salmonella enterica*	*HM997019*	*Yes*
	GAP52	*Cronobacter sakazakii*	*JN882286*	
Lytic15	*Alf5*	*Escherichia coli*	*KX377933*	*Yes*
	AYO145A	*Escherichia coli*	*KR014248*	
	BPS15Q2	*Salmonella enterica Heidelberg*	*KX405003*	
	EC6	*Escherichia coli*	*JX560968*	
	Felix-O1	*Salmonella enterica*	*AF320576*	

	FO1a	*Salmonella enterica*	*JF461087*	
	HY02	*Escherichia coli*	*KM092515*	
	JH2	*Escherichia coli*	*KF055347*	
	Mushroom	*Salmonella sp.*	*KP143762*	
	SBA-1781 #	*Salmonella enterica*	*JX181814*	
	Si3	*Salmonella enterica Infantis*	*KY626162*	
	SP116	*Salmonella enterica Typhimurium*	*KP010413*	
	SPT-1#	*Salmonella sp.*	*JX181822*	
	ST11	*Salmonella enterica Pullorum*	*MF370225*	
	TP1	*Escherichia coli*	*KP869100*	
	TP11	*Escherichia coli*	*KP869109*	
	TP12	*Escherichia coli*	*KP869110*	
	TP15	*Escherichia coli*	*KP869113*	
	TP8	*Escherichia coli*	*KP869106*	
	UAB_Phi87	*Salmonella enterica*	*JN225449*	
	Vpa-E1	*Escherichia coli*	*KM657822*	
	wV8	*Escherichia coli*	*EU877232*	*Yes*
	MM7	*Erwinia amylovora*	*HQ728263*	
	øEa104	*Erwinia amylovora*	*FQ482083*	
	øEa116#	*Erwinia amylovora*	*FQ857195*	

	øEa21-4	*Erwinia amylovora +others*	*EU710883*	*Yes*
	Michonne	*Citrobacter freundii*	*KT001916*	
	Mijalis	*Citrobacter freundii*	*KY654690*	
	Moogle	*Citrobacter freundii*	*KM236239*	
	Mordin	*Citrobacter freundii*	*KT363872*	
	øSUSP1	*Escherichia coli*	*KT454805*	
	øSUSP2	*Escherichia coli*	*KT454806*	
Lytic16	*AG11*	*Salmonella enterica*	*JX297445*	*Yes*
	BPS11Q3	*Salmonella enterica*	*KX405002*	
	Ent1	*Salmonella enterica*	*HE775250*	
	Ent2	*Salmonella enterica*	*HG934469*	
	Ent3	*Salmonella enterica*	*HG934470*	
	f18SE	*Salmonella enterica (Pullorum)*	*KR270151*	
	f2SE	*Salmonella enterica (Enteritidis)*	*KU951146*	
	f3SE	*Salmonella enterica (Enteritidis)*	*KU951147*	
	fSE1C	*Salmonella enterica (Enteritidis)*	*KT962832*	
	fSE4S	*Salmonella enterica (Enteritidis)*	*KT881477*	

	Jersey	*Salmonella enterica*	*KF148055*	*Yes*
	L13#	*Salmonella enterica*	*KC832325*	
	LSPA1	*Salmonella entericaParatyphi A*	*KM272358*	
	LPSE1	*Salmonella enterica (Enteritidis)*	*KY379853*	
	LSHG-59"	*Salmonella enterica (Enteritidis)*	*LSHG01000059*	*Yes*
	MA12	*Salmonella enterica (Enteritidis)****	*KX245013*	
	PVP_SE2	*Salmonella enterica (Enteritidis)*	*MF431252*	
	SE2	*Salmonella enterica*	*JQ007353*	
	SE40#	*Salmonella enterica (Enteritidis)*	*KY626163*	
	SETP13	*Salmonella enterica*	*KF562864*	
	SETP3	*Salmonella enterica*	*EF177456*	
	SETP7	*Salmonella enterica*	*KF562865*	
	SSe3 (KS7)	*Salmonella enterica*	*AY730274*	
	ST4#	*Salmonella enterica*	*JX233783*	
	STP03	*Salmonella enterica (Typhimurium)*	*KY176369*	
	wksl3	*Salmonella enterica*	*JX202565*	

	FSL_SP-031#	*Salmonella enterica*	*KC139518*	
	FSL_SP-038#	*Salmonella enterica*	***KC139652**-66*	*Yes*
	FSL_SP-049#	*Salmonella enterica*	*KC139557-59*	
	G AB-2017	*Escherichia coli*	*KY295895*	
	K1-dep(1) / (K1H)	*Escherichia coli*	*GU196278*	
	K1-dep(4) / (K1G)	*Escherichia coli*	*GU196277*	
	K1-ind(3)	*Escherichia coli*	*GU196281ß*	
	L AB-2017	*Escherichia coli*	*KY295896*	
	P AB-2017	*Escherichia coli*	*KY295898*	
	ST2	*Escherichia coli*	*MF153391*	
	Eta (h)	*Serraitia marcescens*	*KC460990*	*Yes*
	øEap-2	*Enterobacter aerogenes*	*KT287080*	
Lytic17	*EK99P-1*	*Escherichia coli*	*KM233151*	
	Envy	*Escherichia coli*	*KX534335*	
	EP23	*Shigella sonnei*	*JN984867*	
	Gluttony	*Escherichia coli*	*KX534336*	
	HK578	*Escherichia coli*	*JQ086375*	*Yes*
	JL1	*Escherichia coli*	*JX865427*	
	Lust	*Escherichia coli*	*KX534338*	
	Pride	*Escherichia coli*	*KX534341*	
	Sloth	*Escherichia coli*	*KX534339*	

	slur05	*Escherichia coli*	*LN881730*	*Yes*
	slur06	*Escherichia coli*	*LN881731*	
	SO-1	*Sodalis glossinidius*	*GQ502199*	*Yes*
	XSSL-2009a(EEP)	*Escherichia coli*	*FJ750948*	
	YD-2008.s	*Escherichia coli*	*KM896878*	
	eiAu-183	*Edwardsiella ictaluri*	*KF772234*	
Lytic40	*Kp3*	*Klebsiella pneumoniae*	*KT367887*	
Lytic18	*ECO1230-10*	*Escherichia coli*	*GU903191*	
	EcoM_ECOO78	*Escherichia coli*	*KY705409*	
	ep3	*Escherichia coli*	*KM360178*	
	AyrA	*Enterobacter sp. CT7*	*KX231828*	
Lytic19	*Gj1 (øEcoM-Gj1)*	*Escherichia coli*	*EF460875*	*Yes*
	PM1	*Pectobacterium carotovorum*	*KF534715*	
	PP101	*Pectobacterium carotovorum*	*KY087898*	
	Y2	*Erwinia amylovora*	*HQ728264*	
Lytic20	*PY100*	*Yersinia enterocolitica*	*AM076770*	*Yes*
Lytic21	*ECGD1*	*Escherichia coli & Salmonella*	*KU522583*	
	ø92	*Escherichia coli*	*FR775895*	
	ESCO13	*Escherichia coli*	*KX552041*	*Yes*

	ESCO5	*Escherichia coli*	*KX664695*	*Yes*
	phAPEC8	*Escherichia coli*	*JX561091*	*Yes*
Lytic22	*2_JES-2013*	*Escherichia coli*	*NC_022323*	*Yes*
	APCEc02	*Escherichia coli*	*KR698074*	
	FFH2	*Escherichia coli*	*KJ190158*	
	FV3	*Escherichia coli*	*JQ031132*	
	Murica	*Escherichia coli*	*KT001917*	
	rV5	*Escherichia coli*	*NC_011041*	*Yes*
	slur12	*Escherichia coli*	*LN881735*	*Yes*
	slur16	*Escherichia coli*	*LN881727*	*Yes*
	TP14	*Escherichia coli*	*KP869112*	
	TP5	*Escherichia coli*	*KP869103*	
	V18	*Escherichia coli*	*KY683736*	
	4MG	*Escherichia coli*	*KF550303*	
	GAP31	*Cronobacter sakazakii*	*JN882284*	*Yes*
	PVP-SE1	*Salmonella enterica*	*GU070616*	*Yes*
	SSE-121	*Salmonella enterica*	*JX181824*	
	CR3	*Cronobacter sakazakii*	*JQ691612*	
	CR8	*Cronobacter sakazakii*	*KC954774*	
	CR9	*Cronobacter sakazakii*	*JQ691611*	
	øTE	*Pectobacterium atrosepticum*	*NC_020201*	
	PBES 02	*Cronobacter sakazakii*	*KT353109*	

	BIS47	*Klebsiella pneumoniae*	*KY652726*	
	KB57	*Klebsiella pneumoniae*	*KT934943*	
	19 #	*Salmonella enterica*	*KR296684*	
	41 #	*Salmonella enterica*	*KR296695*	
	Av-05	*Escherichia coli*	*KM190144*	
Lytic23	*NAFV-136*	*Escherichia coli*	*NAFV010001 36*	*Yes*
	SEGD1	*Salmonella enterica*	*KU726251*	
	SPN3US	*Salmonella enterica*	*JN641803*	
	Asesino	*Erwinia amylovora*	*KX397364*	
	øEaH2	*Erwinia amylovora*	*JX316028*	
	Stratton	*Erwinia amylovora*	*KX397373*	
	CR5	*Cronobacter sakazakii*	*NC_021531*	
	EarlPhillipIV	*Erwinia amylovora*	*KX397367*	
	Phobos	*Erwinia amylovora*	*KX397372*	
	Kwan	*Erwinia amylovora*	*KX397369*	
	Huxley	*Erwinia amylovora*	*KX397368*	
	Machina	*Erwinia amylovora*	*KX397370*	
	Parshik	*Erwinia amylovora*	*KX397371*	
	Caitlin	*Erwinia amylovora*	*KX397365*	
	ChrisDB	*Erwinia amylovora*	*KX397366*	
Lytic24	*K64-1*	*Klebsiella pneumoniae*	*AB897757*	
	RaK2	*Klebsiella sp. KV-3*	*JQ513383*	

	BF	*Seratia marcescens*	*KY630187*	
	CBB	*Pectobacterium(Erwinia+ Cronobacter)*	*KU574722*	
	GAP32	*Cronobacter sakazakii*	*JN882285*	*Yes*
	121Q	*Escherichia coli*	*KM507819*	
	PBECO4	*Escherichia coli*	*KC295538*	
	slurp01	*Escherichia coli*	*LT603033*	
Lytic25	*øR1-37*	*Yersinia entericolytica*	*AJ972879*	*Yes*
Lytic26	*E1*	*Salmonella enterica*	*AM491472*	
	64795_sal3	*Salmonella enterica (Typhimurium)*	*KX017520*	
	LPST10	*Salmonella enterica (Typhimurium)*	*KY860935*	
	IME207	*Klebsiella pneumoniae*	*KX523699*	
Lytic27	*ECML-117*	*Escherichia coli*	*JX128258*	
	øFenriz	*E. coli and Pseudomonas!*	*KT254133*	
	øHabibi	*E. coli and Pseudomonas!*	*KT254132*	
	øMoody	*E. coli and Pseudomonas!*	*KT254131*	
	øVader	*E. coli and Pseudomonas!*	*KT254130*	
Lytic28	*K1-F*	*Edwardsiella tarda*	*AB757800*	
	IW-1	*Edwardsiella tarda*	*AB757801*	
Lytic29	*MSW-3*	*Edwardsiella tarda*	*AB767244*	
	PEi2	*Edwardsiella ictaluri*	*NC_021342*	

	JD001	*Klebsiella pneumoniae*	*JX866719*	
	Kpn112 #	*Klebsiella pneumoniae*	*KJ021043*	*Yes*
	KpV52	*Klebsiella pneumoniae*	*KX237516*	
Lytic30	*Deimos-Minion*	*Erwinia amylovora*	*KU886225*	
	Ea35-70	*Erwinia amylovora*	*KF806589*	*Yes*
	RAY	*Erwinia amylovora*	*KU886224*	
	Simmy50	*Erwinia amylovora*	*KU886223*	
	Special G	*Erwinia amylovora*	*KU886222*	
Lytic31	*øEaH1*	*Erwinia amylovora*	*KF623294*	
Lytic32	*9g*	*Escherichia coli*	*NC_024146.1*	*Yes*
	JenK1	*Escherichia coli*	*KP719134*	
	JenP1	*Escherichia coli*	*KP719132*	
	JenP2	*Escherichia coli*	*KP719133*	
	SE1	*Salmonella enteritidis*	*KY926791*	
	CAjan	*Escherichia coli*	*KP064094*	
	Greed	*Bladder microbiota assembly*	*KX534337*	
	Seurat	*Escherichia coli*	*KM236243*	
	slur01	*Escherichia coli*	*LN881725*	
Lytic33	*IME-EC2*	*Escherichia coli*	*KF591601*	
Lytic34	*Ss1*	*Cronobacter sakazakii*	*KM058087*	
Lytic35	*CTV22*	*Citrobacter sp.*	*KP774835*	

Lytic36	*BP63*	*Salmonella enterica Infaantis*	*KM366099*	
	UPF_BP2	*Salmonella enterica Bredney*	*KX826077*	
	øEC1	*Escherichia coli*	*KY608966*	*Yes*
Lytic37	*pEP-14*	*Erwinia pyrifoliae*	*JN585957*	
	SopranoGao	*Klebsiella pneumoniae*	*MF612073*	
Lytic38	*Yoloswag*	*Erwinia amylovora*	*KY448244*	

Supplementary Table 4.S4 Phages used in creating SplitsTree

Supercluster	Cluster	Phage name	Accession number	changes made to annotation
T1	Lytic1	T1	AY216660	
T4	Lytic2	T4	AY318471	YES
Vi01	Lytic3	Vi01	FQ312032	
T5	Lytic4	T5	AY543070	
	Lytic5	T7	V01146	YES
	Lytic6	SP6	AY288927	
	Lytic7	KP34	NC_013649	YES
T7	Lytic10	GAP227	NC_020078	YES
N4	Lytic11	N4	EF056009	
9NA	Lytic12	9NA	KJ802832	
Chi	Lytic13	Chi	JX094499	
phiECO32	Lytic14	øECO32	EU330206	YES
Felix01	Lytic15	Felix01	AF320576	
	Lytic16	SETP3	EF177456	
SETP3	Lytic17	SO-1	GQ502199	YES
ECO1230	Lytic18	ECO1230	GU903191	
GJ1	Lytic19	GJ1	EF460875	YES
singleton	Lytic20	PY100	AM076770	YES
rv5	Lytic21	ø92	FR775895	
rv5	Lytic22	rv5	NC_011041	YES
SPN3US	Lytic23	SPN3US	JN641803	
Rak2	Lytic24	Rak2	JQ513383	
singleton	Lytic25	øR1-37	AJ972879	YES
E1	Lytic26	E1	AM491472	YES
ECML-117	Lytic27	ECML-117	JX128258	
KF-1	Lytic28	KF-1	AB757800	
MSW-3	Lytic29	MSW-3	AB767244	
Ea35-70	Lytic30	Ea35-70	KF806589	YES
PhiEaH1	Lytic31	øEaH1	KF623294	
9g	Lytic32	9g	KJ419279	
IME_EC2	Lytic33	IME_EC2	KF591601	
singleton	Lytic34	Ss1	KM058087	
singleton	Lytic35	CVT22	KP774835	
BP63	Lytic36	BP63	KM366099	
Pep14	Lytic37	Pep14	JN585957	
Yoloswag	Lytic38	Yoloswag	KY448244	
T7	Lytic39	Peat1	KR604693	
SETP3	Lytic40	Kp3	KT367887	
Joad	Lytic41	Joad	MF459647	
singleton	Lytic42	PMBT28	MG641885	
Jello	Lytic43	Sucellus	MH059634	
singleton	Lytic44	fEV-1	LT992259	
N4	Lytic45	CB1	KY514264	
singleton	Lytic47	Halfdan	MH362766	
	Lytic48	Med16	MK095605	
SETP3	Lytic49	Scapp	MH553517	
singleton	Lytic50	LIET2	MK388689	
singleton	Lytic52	Serbin	MK608336	
singleton	Lytic53	CAjan	KP064094	

REFERENCES

1. Keen E. C. (2015). A century of phage research: bacteriophages and the shaping of modern biology. *BioEssays: news and reviews in molecular, cellular and developmental biology*, *37*(1), 6-9.

2. Clokie, M. R., Millard, A. D., Letarov, A. V., & Heaphy, S. (2011). Phages in nature. *Bacteriophage*, *1*(1), 31-45.

3. S.R. Casjens, P.A. Thuman-Commike. Evolution of mosaically related tailed bacteriophage genomes seen through the lens of phage P22 virion assembly. Virology, 411 (2011), pp. 393–415.

4. Canchaya et al., 2003C. Canchaya, G. Fournous, S. Chibani-Chennoufi, M.L. Dillmann, H. Brüssow (2003). Phage as agents of lateral gene transfer. *El Sevier: current opinion in microbiology,* 6(4), 417-424.

5. Potera C. (2013). Phage renaissance: new hope against antibiotic resistance. *Environmental health perspectives*, *121*(2), a48-53.

6. Kuchment, A. (2013). The forgotten cure: The past and future of phage therapy. 10.1007/978-1-4614-0251-0.

7. Nelson D. (2004). Phage taxonomy: we agree to disagree. *Journal of bacteriology*, *186*(21), 7029-31.

8. Hatfull G. F. (2014). Molecular Genetics of Mycobacteriophages. *Microbiology spectrum*, *2*(2), 1-36.

9. Hatfull, G.F., Jacobs-Sera, D., Lawrence, J.G., Pope, W.H., Russell, D.A., Ko, C.C., Weber, R.J., Patel, M.C., Germane, K.L., Edgar, R.H., Hoyte, N.N., Bowman, C.A., Tantoco, A.T., Paladin, E.C., Myers, M.S., Smith, A.L., Grace, M.S., Pham, T.T., O'Brien, M.B., Vogelsberger, A.M., Hryckowian, A.J., Wynalek, J.L., Donis-Keller, H., Bogel, M. W., Peebles, C.L., Cresawn, S.G., Hendrix, R.W. (2010). Comparative genomic analysis of 60 Mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. *J. Mol. Biol*ogy, 397: 119–143.

10. Grose, J. H., Casjens, S. R. (2014). Understanding the Enormous Diversity of Bacteriophages: The Tailed Phages That Infect the Bacterial Family Enterobacteriaceae. *Virology*, 421-43.

11. Casjens S.R., Thuman-Commike, P.A. (2011). Evolution of mosaically related tailed bacteriophage genomes seen through the lens of phage P22 virion assembly. *Virology.*

12. Krumsiek, J., R. Arnold, and T. Rattei. "Gepard: A Rapid and Sensitive Tool for Creating Dotplots on Genome Scale." [In eng]. *Bioinformatics* 23, no. 8 (Apr 15 2007): 1026-8.

13. Lamine, J. G., R. J. DeJong, and S. M. Nelesen. "Phamdb: A Web-Based Application for Building Phamerator Databases." [In eng]. *Bioinformatics* 32, no. 13 (Jul 1 2016): 2026-8.

14. Cresawn, S. G., M. Bogel, N. Day, D. Jacobs-Sera, R. W. Hendrix, and G. F. Hatfull. "Phamerator: A Bioinformatic Tool for Comparative Bacteriophage Genomics." [In eng]. *BMC Bioinformatics* 12 (Oct 12 2011): 395.

15. Katoh, K., K. Misawa, K. Kuma, and T. Miyata. "Mafft: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform." [In eng]. *Nucleic Acids Res* 30, no. 14 (Jul 15 2002): 3059-66.

16. Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton*, et al.* "Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data." [In eng]. *Bioinformatics* 28, no. 12 (Jun 15 2012): 1647-9.

17. Huson, D. H., and D. Bryant. "Application of Phylogenetic Networks in Evolutionary Studies." [In eng]. *Mol Biol Evol* 23, no. 2 (Feb 2006): 254-67.

18. Hatfull, G. F., & Hendrix, R. W. (2011). Bacteriophages and their genomes. *Current opinion in virology*, *1*(4), 298-303

CHAPTER 5: Conclusion and Future Directions

The bacteriophages are forces of nature that drive evolution of bacterial strains. With the ability of horizontal gene transfer (1) and specialized transduction as in shiga toxin (2) they are able to share genes with bacteria which in turn can make bacteria either resistant or pathogenic. For e.g. *Vibrio cholerae* and *Cornebacterium diptheriae* would not be pathogenic if it were not for their prophages, CTXφ (3) and Beta (4). Understanding the role of lytic and temperate bacteriophages in a family is the most direct way of studying their emergence and evolution. This can be achieved in two ways: a) by characterizing the genomes and proteomes of bacteriophages, and b) by tapping into the enormous diversity of bacteriophages by studying their relationship on a broader level.

We started this process with characterization of 8 bacteriophages of a new genus *Agrican357virus* (5) and compared it with another bacteriophage from the same family Ea35-70 (6), found in Ontario, Canada. We found that the genomes of these bacteriophages are highly related to each other with >97% genomic and proteomic similarity. Out of 319-324 genes they harbor, 80% of them have no known function. They have broad host range and incredibly small burst size of 4.5-4.9 phages per bacterium, as compared to other phages of same genome size. It contains survival proteins like SbcC, SbcD, exodeoxyribonuclease VIII, UvsX, UvsW etc., that may aid in DNA repair and metabolism. We also found virulence factors like EPS depolymerase which is a biofilm degradation protein. In our analyses, it was observed that bacteriophages of *Agrican357virus* has EPS-degradation activity against *P. vagans*. This activity was not seen in *E. amylovora*.

The most curious aspect of this study would be determining the function of hypothetical proteins. In our mass spectrometry analysis, we were able to detect 32 proteins for bacteriophage

RAY and 27 for Deimos-Minion (including some hypothetical proteins). Although 202 hypothetical proteins with no known blast hits or 50 with blast hits considerably contribute to the viral dark matter (7) which needs to be further analyzed.

What is most interesting about this family is they share more proteomic similarity to *Pseudomonas* and *Ralstonia* phages than their *Erwinia* counterparts. This unique feature brings back the question of what drives evolution of phages, is it their hosts or ecological niche? In a recent study done on healthy blossoms it was found that *Pseudomonas* is the most prevalent bacteria after *Erwinia* and *Pantoea* found on tree blossoms (personal communication) which may indicate that it is may indeed be the ecological niche that is driving evolution, at least in case of *Agrican357virus.*

The idea of studying their evolutionary behavior laid the foundation of enhancing our current understanding of bacteriophage diversity. Our next study is built upon a previous study done by Grose and Casjens in 2014 based (8) on methods set forth by Graham Hatfull (9-11). In this study they grouped 337 tailed bacteriophages isolated on 18 genera of bacteria from *Enterobacteriaceae* into 56 diverse clusters (32 lytic and 24 temperate). We further expanded this study to 1303 tailed bacteriophages (49 lytic and 39 temperate clusters) from the order *Enterobacteriales,* submitted to GenBank as of March 25, 2019. With addition of new phages, we observed that phages with <50% genomic similarity may fall in the same cluster as long a they have >50% genomic similarity with at least one phage within the same cluster. The subclusters on the other hand, now have a more substantiated definition of having >80% proteomic similarity (12) with rest of the phages in the same subcluster. The outlined definition of superclusters remains largely unchanged.

Out of the 11923 proteins we studied, nearly 60 % of them were unique between 1 or 2 phages. We also found that 614 them were conserved between ≥25 phages and 243 were conserved between ≥100. Of the 614 more than half have no known function. Discovering their functions would be the next step to add more to the understanding of these bacteriophages.

REFERENCES

1. Boyd, E. F., K. E. Moyer, L. Shi, and M. K. Waldor. "Infectious Ctxphi and the Vibrio Pathogenicity Island Prophage in Vibrio Mimicus: Evidence for Recent Horizontal Transfer between V. Mimicus and V. Cholerae." [In eng]. *Infect Immun* 68, no. 3 (Mar 2000): 1507-13.

2. Melton-Celsa, A. R. "Shiga Toxin (Stx) Classification, Structure, and Function." [In eng]. *Microbiol Spectr* 2, no. 4 (Aug 2014): Ehec-0024-2013.

3. Faruque, S. M., and J. J. Mekalanos. "Phage-Bacterial Interactions in the Evolution of Toxigenic Vibrio Cholerae." [In eng]. *Virulence* 3, no. 7 (Nov 15 2012): 556-65.

4. Gill, D. M., T. Uchida, and R. A. Singer. "Expression of Diphtheria Toxin Genes Carried by Integrated and Nonintegrated Phage Beta." [In eng]. *Virology* 50, no. 3 (Dec 1972): 664-8.

5. Adams, M. J., E. J. Lefkowitz, A. M. Q. King, B. Harrach, R. L. Harrison, N. J. Knowles, A. M. Kropinski*, et al.* "Changes to Taxonomy and the International Code of Virus Classification and Nomenclature Ratified by the International Committee on Taxonomy of Viruses (2017)." [In eng]. *Arch Virol* 162, no. 8 (Aug 2017): 2505-38.

6. Yagubi, A. I., A. J. Castle, A. M. Kropinski, T. W. Banks, and A. M. Svircev. "Complete Genome Sequence of Erwinia Amylovora Bacteriophage Vb_Eamm_Ea35-70." [In eng]. *Genome Announc* 2, no. 4 (Aug 21 2014).

7. Hatfull, G. F. "Dark Matter of the Biosphere: The Amazing World of Bacteriophage Diversity." [In eng]. *J Virol* 89, no. 16 (Aug 2015): 8107-10.

8. Grose, J. H., Casjens, S. R. (2014). Understanding the Enormous Diversity of Bacteriophages: The Tailed Phages That Infect the Bacterial Family Enterobacteriaceae. *Virology*, 421-43.

9. Hatfull G. F. (2014). Molecular Genetics of Mycobacteriophages. *Microbiology spectrum*, *2*(2), 1-36.

10. Hatfull, G.F., Jacobs-Sera, D., Lawrence, J.G., Pope, W.H., Russell, D.A., Ko, C.C., Weber, R.J., Patel, M.C., Germane, K.L., Edgar, R.H., Hoyte, N.N., Bowman, C.A., Tantoco, A.T., Paladin, E.C., Myers, M.S., Smith, A.L., Grace, M.S., Pham, T.T., O'Brien, M.B., Vogelsberger, A.M., Hryckowian, A.J., Wynalek, J.L., Donis-Keller, H., Bogel, M. W., Peebles, C.L., Cresawn, S.G., Hendrix, R.W. (2010). Comparative genomic analysis of 60 Mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. *J. Mol. Biol*ogy, 397: 119–143.

11. Casjens S.R., Thuman-Commike, P.A. (2011). Evolution of mosaically related tailed bacteriophage genomes seen through the lens of phage P22 virion assembly. *Virology.*

12. Pope, W. H., T. N. Mavrich, R. A. Garlena, C. A. Guerrero-Bustamante, D. Jacobs-Sera, M. T. Montgomery, D. A. Russell, M. H. Warner, and G. F. Hatfull. "Bacteriophages of Gordonia Spp. Display a Spectrum of Diversity and Genetic Relationships." [In eng]. *MBio* 8, no. 4 (Aug 15 2017).

APPENDIX I: Genome Sequences of 19 Novel *Erwinia amylovora* Bacteriophages

The following Appendix is taken from an article submitted to Genome Announcements Journal. All content and figures have been formatted for this dissertation, but it is otherwise unchanged

I.1 Abstract

Erwinia amylovora is the causal agent of fire blight, a devastating disease affecting some plants of the *Rosaceae* family. We isolated bacteriophages from samples collected from infected apple and pear trees along the Wasatch Front in Utah. We announce 19 high-quality complete genome sequences of *E. amylovora* bacteriophages.

I.2 Discussion

Erwinia amylovora is a Gram-negative facultative anaerobic rod-shaped bacterium and the causative agent of fire blight (1), a disease that affects some members of the plant family Rosaceae and causes the infected areas of the plant to appear burnt (2, 3). E. amylovora is a member of the Enterobacteriaceae family, which includes many well-characterized pathogenic bacteria such as *Salmonella* enterica and *Escherichia coli*. Thus, understanding the evolution of this plant pathogen and the bacteriophages that infect it may provide insight into the evolution of the Enterobacteriaceae family, including other pathogenic strains. Herein, we announce the genome sequences of 19 novel *E. amylovora* bacteriophages, vB_EamP_Frozen, vB_EamP_Gutmeister, vB_EamP_Rexella, vB_EamM_Deimos-Minion, vB_EamM_RAY, vB_EamM_Simmy50, vB_EamM_Special G, vB_EamM_Caitlin,vB_EamM_ChrisDB, vB_EamM_EarlPhillipIV, vB_EamM_Huxley, vB_EamM_Kwan, vB_EamM_Machina, vB_EamM_Parshik, vB_EamM_Phobos, vB_EamM_Stratton, vB_EamM_Joad, vB_EamM_RisingSun, and vB_EamM_Yoloswag. Samples were collected from apple and

pear trees bearing symptoms of fire blight infection that were found along the Wasatch Front of Utah. Phages were amplified via enrichment culture of these samples, and resulting phages were then plaque purified by a minimum of three passages. All phages reported in this announcement infect the *Erwinia amylovora* ATCC 29780 strain. Genomic DNA was extracted using the Phage DNA isolation kit (Norgen Biotek Corporation) and sequenced using 454 pyrosequencing (454 Life Sciences, Roche Diagnostics) or Illumina HiSeq 2500 sequencing (Illumina, 250-bp reads). Contigs were assembled using Newbler version 2.9 (Roche Diagnostics, Branford, CT) and Consed (4) for 454 pyrosequencing reads or Geneious version R8 (5) for Illumina reads. Assembled genomes were annotated using DNA Master (6) and other programs as described previously (7, 8). The 19 phages fell into five distinct clusters according to genomic analysis. The first group included the jumbo myoviruses vB_EamM_Deimos-Minion, vB_EamM_RAY, vB_EamM_Simmy50, and vB_EamM_Special G, which share a minimum of 97.2% average nucleotide identity to one another. The second group included two jumbo myoviruses, vB_EamM_RisingSun and vB_EamM_Joad, which differ by only two putative gene products. The third group included diverse jumbo myoviruses vB_EamM_Caitlin, vB_EamM_ChrisDB, vB_EamM_EarlPhillipIV, vB_EamM_Huxley, vB_EamM_Kwan, vB_EamM_Machina, vB_EamM_Parshik, vB_EamM_Phobos, and vB_EamM_Stratton, which share a minimum of 50.5% average nucleotide identity. An additional jumbo myovirus, vB_EamM_Yoloswag, did not have any close phage relatives. Podovirus phages vB_EamP_Frozen, vB_EamP_Gutmeister, and vB_EamP_Rexella share at least 97.2% average nucleotide identity. The four jumbo myovirus groups package DNA by headful packaging based on homology of their putative terminase genes to the phiKZ terminase (9). Three of these genomically permuted myovirus groups were assigned their base pair (bp) 1 by alignment to previously published genomes by use

of BLASTN (10) and Gepard (11) (Ea35-70 for the Deimos-Minion group [12], EL [13, 14] for the RisingSun group, and SPN3US [15] for the Caitlin group). vB_EamM_Yoloswag shared very little DNA homology with any other phage; therefore, its bp 1 was assigned to position its putative terminase at the beginning of the genome. The podovirus group genomes were assigned bp 1 by their relation to N4, in terms of both terminase similarity and whole-genome alignment, suggesting they have small terminal repeats.

I.3 Accession number(s)

GenBank accession numbers for the 19 *Erwinia* bacteriophages are listed in Table II.1.

Table I.1 Properties of 19 novel *Erwinia amylovora* bacteriophage genomes . ORFs, open reading frames. NA, no tRNAS were identified

Phage Name	GenBank accession no.	Sequencing Type	Minimum-maximum fold coverage (avg read depth)	Genome Length (bp)	No. of ORFs	No. of tRNAs	G+C content (%)
vB_EamP_Gutmeister	KX098391	Illumina	423–2,415 (662)	71,173	84	8	46.9
vB_EamP_Frozen	KX098389	454	79–1,779 (862)	75,147	92	8	46.9
vB_EamP_Rexella	KX098390	454	69–1,780 (885)	75,448	92	7	46.9
vB_EamM_Deimos-Minion	KU886225	454	61–1,780 (873)	273,501	326	NA	49.9
vB_EamM_RAY	KU886224	Illumina	335–910 (677)	271,182	319	1	49.9
vB_EamM_Special G	KU886222	454	19–1,779 (874)	273,224	324	NA	49.8
vB_EamM_Simmy50	KU886223	Illumina	150–831 (282)	271,088	322	1	49.9
vB_EamM_Caitlin	KX397365	Illumina	84–249 (174)	241,147	271	7	52.2
vB_EamM_ChrisDB	KX397366	454	66–1,780 (874)	244,840	277	11	49.4
vB_EamM_EarlPhilliplV	KX397367	Illumina	75–243 (164)	223,935	241	NA	50.6
vB_EamM_Huxley	KX397368	454	75–1,779 (880)	240,761	271	9	51.1
vB_EamM_Kwan	KX397369	Illumina	192–554 (362)	246,390	285	8	52.1
vB_EamM_Machina	KX397370	454	65–1,780 (879)	241,654	272	9	51
vB_EamM_Parshik	KX397371	454	64–1,779 (880)	241,050	271	10	51
vB_EamM_Phobos	KX397372	454	59–1,779 (873)	229,501	247	NA	49.1
vB_EamM_Stratton	KX397373	454	64–1,779 (874)	243,953	276	12	51.3
vB_EamM_Yoloswag	KY448244	Illumina	5–265 (99.5)	259,700	334	NA	46.91
vB_EamM_RisingSun	MF459646	Illumina	50–293 (138.6)	235,108	243	NA	48.32
vB_EamM_Joad	MF459647	Illumina	232–1,065 (522.2)	235,374	245	NA	48.29

I.4 Acknowledgements

We extend a special thanks to the Howard Hughes Medical Institute Science Education Alliance–Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) for

support and training on phage analysis. We appreciate the help of Ed Wilcox (BYU DNA Sequencing Center), Michael Standing (BYU Microscopy Lab), Alisa L. Buchanan, Brett Buhler, and Brianna R. Keele.

I.5 Funding statement

This work was graciously funded by the Department of Microbiology and Molecular Biology and the College of Life Sciences at Brigham Young University, as well as by a private donor.

I.6 Conflict of interest

J.H.G. is in the process of submitting a patent for using *Erwinia* phages for the treatment of fire blight. J.H.G., S.H., and D.P.B. have a license agreement with a company for distribution of *Erwinia* phages.

REFERENCES

1. Khan MA, Zhao Y, Korban SS. 2012. Molecular mechanisms of pathogenesis and resistance to the bacterial pathogen Erwinia amylovora, causal agent of fire blight disease in Rosaceae. Plant Mol Biol Rep 30:247–260. https://doi.org/10.1007/s11105-011-0334-1

2. Schroth MN, Thomson SV, Hildebrand DC, Moller WJ. 1974. Epidemiology and control of fire blight. Annu Rev Phytopathol 12:389 – 412. https://doi.org/10.1146/annurev.py.12.090174.002133.

3. Thomson SV. 2000. Epidemiology of fire blight 2, p 9 –36. In Fire Blight: the disease and its causative agent Erwinia amylovora. CABI Publishing, Wallingford, United Kingdom.

4. Gordon D, Green P. 2013. Consed: a graphical editor for next-generation sequencing. Bioinformatics 29:2936 –2937. https://doi.org/10.1093/ bioinformatics/btt515.

5. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/ bts199.

6. Lawrence J. 2012. DNA Master, v5.0.2. http://cobamide2.bio.pitt.edu/ computer.htm.

7. Grose JH, Jensen JD, Merrill BD, Fisher JN, Burnett SH, Breakwell DP. 2014. Genome sequences of three novel Bacillus cereus bacteriophages. Genome Announc 2(1):e01118-13. https://doi.org/10.1128/ genomeA.01118-13.

8. Sheflo MA, Gardner AV, Merrill BD, Fisher JN, Lunt BL, Breakwell DP, Grose JH, Burnett SH. 2013. Complete genome sequences of five Paenibacillus larvae bacteriophages. Genome Announc 1(6):e00668-13. https:// doi.org/10.1128/genomeA.00668-13.

9. Mesyanzhinov VV, Robben J, Grymonprez B, Kostyuchenko VA, Bourkaltseva MV, Sykilinda NN, Krylov VN, Volckaert G. 2002. The genome of bacteriophage phiKZ of Pseudomonas aeruginosa. J Mol Biol 317:1–19. https://doi.org/10.1006/jmbi.2001.5396.

10. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403– 410. https://doi.org/10.1016/ S0022-2836(05)80360-2.

11. Krumsiek J, Arnold R, Rattei T. 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics 23:1026 –1028. https://doi.org/10.1093/bioinformatics/btm039.

12. Yagubi AI, Castle AJ, Kropinski AM, Banks TW, Svircev AM. 2014. Complete genome sequence of Erwinia amylovora bacteriophage vB_EamM_ Ea35-70. Genome Announc 2(4):e00413-14. https://doi.org/10.1128/ genomeA.00413-14.

13. Hertveldt K, Lavigne R, Pleteneva E, Sernova N, Kurochkina L, Korchevskii R, Robben J, Mesyanzhinov V, Krylov VN, 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.

14. Sokolova OS, Shaburova OV, Pechnikova EV, Shaytan AK, Krylov SV, Kiselev NA, Krylov VN. 2014. Genome packaging in EL and Lin68, two giant phiKZ-like bacteriophages of P. aeruginosa. Virology 468 – 470: 472– 478. https://doi.org/10.1016/j.virol.2014.09.002.

15. Lee JH, 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.

APPENDIX II: Genomic Comparison of 60 Completely Sequenced Bacteriophages that Infect *Erwinia* and/or *Pantoea* Bacteria

The following Appendix is taken from an article submitted to Virology Journal. All content and figures have been formatted for this dissertation, but it is otherwise unchanged.

II.1 Abstract

Erwinia and *Pantoea* are closely related bacterial plant pathogens in the Gram negative *Enterobacteriales* order. Sixty tailed bacteriophages capable of infecting these pathogens have been completely sequenced by investigators around the world and are in the current databases, 30 of which were sequenced by our lab. These 60 were compared to 991 other *Enterobacteriales* bacteriophage genomes and found to be, on average, just over twice the overall average length. These *Erwinia* and *Pantoea* phages comprise 20 clusters based on nucleotide and protein sequences. Five clusters contain only phages that infect the *Erwinia* and *Pantoea* genera, the other 15 clusters are closely related to bacteriophages that infect other *Enterobacteriales*; however, within these clusters the *Erwinia* and *Pantoea* phages tend to be distinct, suggesting ecological niche may play a diversification role. The failure of many of their encoded proteins to have predicted functions highlights the need for further study of these phages.

II.2 Introduction

Bacteriophages are viruses that infect bacteria. Their virions are comprised of a protein shell containing genetic material that can be dsDNA, ssDNA, dsRNA or ssRNA. Their genomes can contain as few as 3.3 kb or as many as 500 kb (*2, 3*). They are the most abundant and diverse biological entities, with an estimate of about 10^{32} tailed bacteriophages on Earth (*4*). Since bacteriophages are parasites of bacteria, they have played an important role in the evolution of bacteria. Bacteriophages can have two alternate lifestyles when infecting a

bacterium, lytic and temperate. Lytic phages simply replicate to form progeny virions which are released to infect other host cells. Temperate phages can also propagate lytically but may instead enter a semi-dormant "prophage" state in which the phage DNA either replicates as a plasmid or integrates into the host chromosome and replicates passively as part of that replicon. Prophages can be stable indefinitely, but environmental triggers can cause their "induction" into the lytic growth cycle.

Erwinia and *Pantoea* are very closely related Gram negative bacteria in the E*rwiniacea*e family of the *Enterobacteriales* order (*5*) that are often plant pathogens, causing necrosis in the tissues of the infected plant. These pathogens are a large burden on the agricultural community of the United States and are currently listed as possible bio-terrorism agents. For example, *Pantoea agglomerans* is the causative agent of potato blight and has also been documented as an opportunistic human pathogen (*6*), and *Erwinia amylovora* is the causative agent of fruit tree fire blight, which is responsible for an average of 100 million US dollars damage annually to apple orchards in the United States (*7*). Fire blight infections are currently treated with antibiotics; however, up to 70% of these bacteria found in nature are resistant to the currently used antibiotics (*8*). Due to their ability to kill their bacterial hosts, phages are projected to provide an alternative anti-bacterial therapy for these plant diseases.

A number of bacteriophages that infect *Erwinia* or *Pantoea* (*Erwiniacea*e phages) have been isolated by a variety of investigators from several continents, and 60 of their complete genome sequences are available at the National Center for Biological Information (NCBI) GenBank database (*9*). Thirty of these phages were isolated and characterized in our laboratory (*10, 11*). Host range studies of phages that infect *Erwiniaceae* suggest relatedness between the two host genera in that several phages isolated on *Erwinia* can infect both *Erwinia* and *Pantoea*

strains including phages Joad and RisingSun (*12*), Y3 (*13*), ØEa2809 (*14*) and CBB (*15*). Herein, we compare these 60 bacteriophages and place them in 20 different clusters based on their genomic and proteomic traits. An analysis of each of these clusters is provided, along with comparisons to known phages. The purpose of our analysis is four fold: 1) to gain insight into the relationship and interaction between different bacteriophage types and their host bacterial species, 2) to further understand the relationships among members of the *Enterobacteriales* order by comparing their bacteriophages, 3) to contribute to our understanding of overall bacteriophage diversity, and 4) to provide information that will aid in the treatment of the above plant diseases by development of improved phage therapy cocktail design and safety.

II.3 Materials and methods

II.3.1 Isolation, sequencing and assembly of phages

Thirty *Erwinia* phages were isolated by our laboratory at Brigham Young and 28 of these have been previously described (*10, 11*). The genomes of the two previously undescribed phages Rebecca (accession No. MK514281) and Derbicus (MK514282) were sequenced from libraries made with the Illumina TruSeq DNA Nano kit and Illumina HiSeq 2500 sequencing (250-bp paired end). Genomes were assembled with Geneious (*16*) version 8.1 using *de novo* assembly with medium-low sensitivity as described previously in Sharma *et al (11).* Coverage depths were 527-1615 (1015.6 average) for Rebecca and 183-1758 (572.3 average) for Derbicus, both phages circularized their genomes upon assembly.

II.3.2 Genomic analysis and comparison

Gepard (17) was used to generate dot plots that compare nucleotide sequences of multiple genomes. Default settings (word size 10) were used to generate dot plots, however lower and upper color limit were increased in order to allow better image viewing.

Geneious (*16*) was used to align the sequences in an identity matrix using MAFFT plugin and setting parameters to auto-algorithm, a scoring matrix of 200PAM/k=2, a gap open penalty of 1.53 and an offset value of 0.123. Phamdb (*18*), an online version of Phamerator (*19*), a bioinformatic tool designed to compare bacteriophage genomes was used to visualize both nucleotide and protein similarity using kClust (*20*). The default settings of PhamDB were used in this comparison. The cluster file generated by Phamerator was aligned using Janus (available on the DNA-master website https://phagesdb.org/DNAMaster/) and then used to generate a phylogenetic tree of the proteins using the SPLITStree program (*21*). BLASTp (*22, 23*) was used from the NCBI website except for when accessed through Phamerator.

II.4 Results and discussion

II.4.1 Genomic and proteomic analyses separate the 60 *Erwiniaceae* bacteriophages into 20 clusters

A summary of the 60 *Erwiniaceae* phage genomes available in GenBank as of January 1, 2019 is provided in Table 1. These 60 phages were isolated in 10 countries, and 30 were isolated and characterized by our laboratory (*10, 11*). Three phages, LIMElight, LIMEzero and Vid5, were isolated on *Pantoea* hosts, and 56 were isolated on *Erwinia* hosts (*24*). One phage, CBB, was isolated on *Pectobacterium* but forms plaques on a strain of *Erwinia* (*15*). Among those with a reported isolation location, many were found in infected trees or in the soil around them. Of the 30 phages we isolated, the genomes of only two, Joad and RisingSun, have been fully discussed in the literature (*12*), 26 have been reported in only genome announcements, and two (Rebecca and Derbicus) are first reported here.

Table II.1 Sixty *Erwinia* and *Pantoea* bacteriophages. Phages are organized by clusters (see text for definition of “clusters”) which are indicated by different colored cells. The clusters are listed in order of descending genome size, and this group color scheme is carried throughout this report. The first column is the cluster as defined by Grose and Casjens (*25*) when applicable. Pre-existing clusters are named according to the founding *Enterobacteriales* phage, and bold phage names in the second column indicate

the first *Erwiniaceae* member of that cluster. N/A (not available) indicates phage genomes in GenBank that are otherwise not published. ǂ LS-2018a has a reported genome length of 59,759 bp, but this appears to be an untrimmed partial concatemer of the true sequence; the genome length given in the table putative properly trimmed sequence.

Phage Name	Isolation Location	Isolation Source	Gene Bank Accession	Genome Length	Number of ORF's	Reference
CBB	Little Island, Ireland	Waste water sludge	KU574722	378,379	605	15
RAY	UT, USA	Leaves and Stem	KU886224	271,182	319	10
Deimos-minion	UT, USA	Branches and Blossom	KU886225	273,501	326	10
Special G	UT, USA	Branches and Blossom	KU886222	273,224	324	10
Simmy50	UT, USA	Bark	KU886223	271,088	322	10
Ea35-70	Canada	Soil	KF806589	271,084	314	76
Desertfox	UT, USA	Soil	MG655268	272,485	320	N/A
Bosolaphorus	UT, USA	Soil	MG655267	272,228	321	N/A
Rebecca	UT, USA	Tree	MK514281	273,731	320	N/A
MadMel	UT, USA	Soil	MG655269	275,000	321	N/A
Mortimer	UT, USA	Unknown	MG655270	273,914	325	N/A
Yoloswag	UT, USA	Unknown	KY448244	259,700	334	10
Y3	Sursee, Switzerland	Soil, Apple tree	KY984068	261,365	333	13
Alexandra	UT, USA	Unknown	MH248138	266,532	349	N/A
Asesino	UT, USA	Branches and Blossom	KX397364	246,291	289	N/A
phiEaH2	Hungary	Unknown	JX316028	243,050	263	41
Stratton	UT, USA	Unknown	KX397373	243,953	276	10
Huxley	UT, USA	Branches and Blossom	KX397368	240,761	271	10
Machina	UT, USA	Unknown	KX397370	241,654	272	10
Parshik	UT, USA	Unknown	KX397371	241,050	271	10
ChrisDB	UT, USA	Unknown	KX397366	244,840	277	10
Caitlin	UT, USA	Branches and Blossom	KX397365	241,147	271	10
Phobos	UT, USA	Unknown	KX397372	229,501	247	10
EarlPhilipIV	UT, USA	Apple tree	KX397367	223,935	241	10
Derbicus	UT, USA	Pear tree	MK514282	223,950	240	N/A
Wellington	UT, USA	Unknown	MH426724	244,950	295	11
Kwan	UT, USA	Unknown	KX397369	246,390	285	10
phiEaH1	NCAIM, Hungary	Aerial tissue	KF623294	218,339	244	41
Joad	UT, USA	Pear tree	MF459647	235,374	245	10
RisingSun	UT, USA	Apple tree	MF459646	235,108	243	10
Cronus	Denmark	Organic waste	MH059636	175,774	295	N/A
øEa2809	Belarus	Leaves of apple tree	KP037007	162,160	145	14
Bue1	Switzerland	Soil from apple orchard	MG973030	164,037	178	N/A
øEa21-4	Canada	Unknown	EU710883	84,576	117	51
øEa104	Germany	Unknown	FQ482083	84,565	118	69
M7	Switzerland	Unknown	HQ728263	84,694	117	1
SunLiRen	USA	Unknown	MH426725	84,559	142	N/A
S6	Switzerland	Unknown	HQ728266	74,669	115	1
Frozen	UT, USA	Branches and Blossom	KX098389	75,147	92	10
Rexella	UT, USA	Branches and Blossom	KX098390	75,448	92	10
Gutmeister	UT, USA	Apple tree	KX098391	71,173	84	10
Ea9-2	Canada	Soil	KF806588	75,568	89	N/A
øEaP-8	South Korea	Unknown	MH160392	75,929	78	56
Vid5	Lithuania	Thicket shadbush	MG948468	61,437	99	72
PEp14	Korea	Unknown	JN585957	60,714	64	N/A
Pavtok	UT, USA	Unknown	MH426726	61,401	62	N/A
Faunus	Denmark	Organic waste	MH191398	54,065	78	N/A
Y2	Switzerland	Unknown	NC019504	56,621	92	55
øEt88	USA	Unknown	FQ482085	47,279	68	69
Era103	USA	Unknown	EF160123	45,445	53	33
øEa100	USA	Unknown	FQ482086	45,554	51	69
S2	Switzerland	Soil	MG736918	45,495	49	N/A
øEa1H	USA	Unknown	FQ482084	45,522	50	69
LIMElight	Merelbeke, Belgium	Soil from potato	FR687252	44,546	55	24
LIMEzero	Merelbeke, Belgium	Soil from potato	FR751545	43,032	57	24
FE44	Ukraine	T2 phage contamination	KF700371	39,860	47	N/A
L1	Switzerland	Unknown	HQ728265	39,282	51	1
LS-2018a	MD, USA	Unknown	CP013974	31,798	N/A	N/A
ENT90	South Korea	Unknown	HQ110084	29,564	60	N/A
EtG	USA	Cucumber	MF276773	30,413	45	N/A

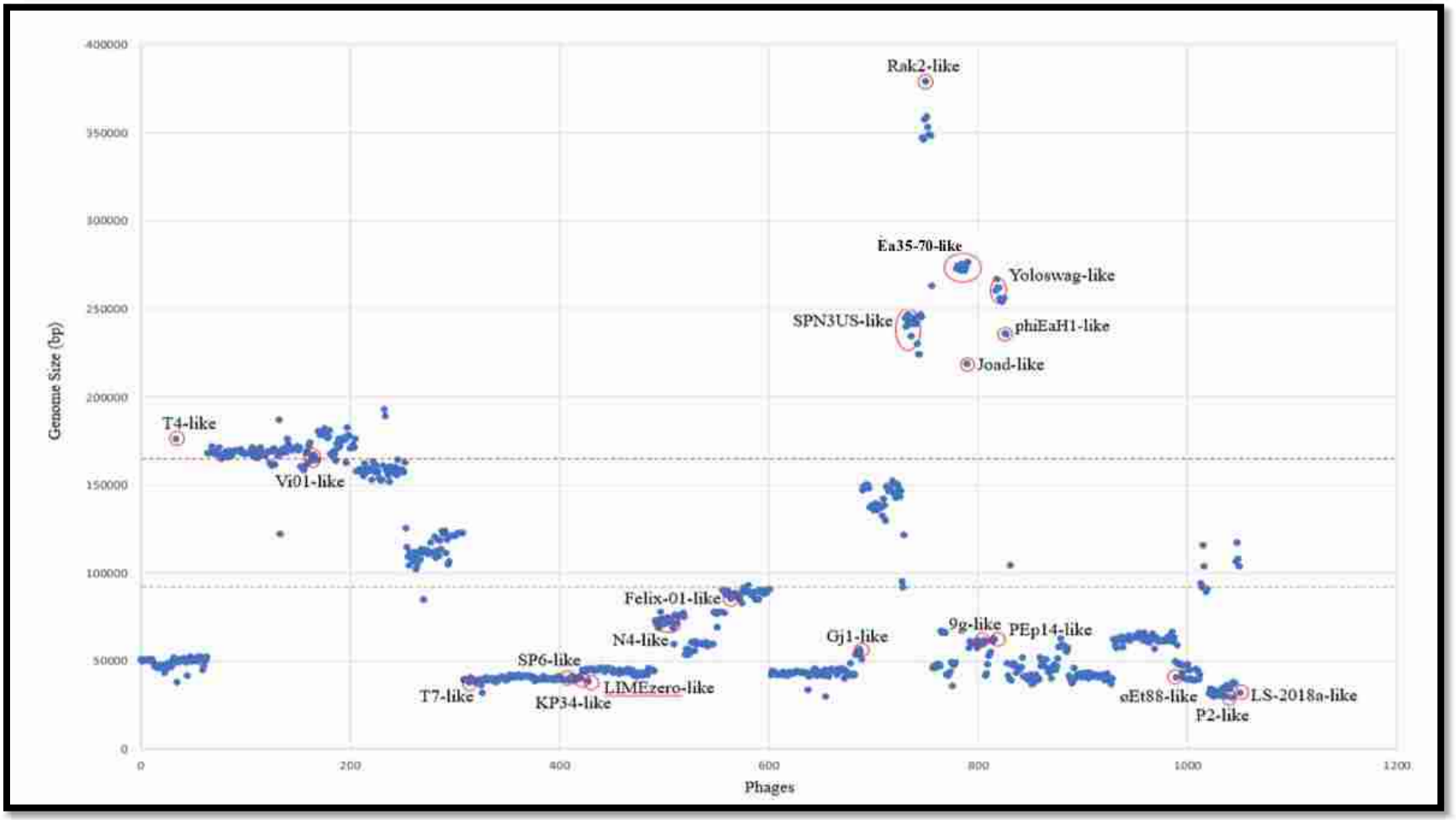


Figure II.1 Comparison of *Enterobacteriales* bacteriophage average genome size with the average *Erwiniaceae* phage genome reveals large *Erwiniaceae* phage genomes. Phage genome size is plotted on the y-axis for each of 1134 *Enterobacteriales* phages on the x-axis. The green dashed line represents the average genome length of all *Enterobacteriales* phages, and the red dashed line represents the average of all *Erwiniaceae* phage genome lengths. The red circles mark *Erwiniaceae* clusters.

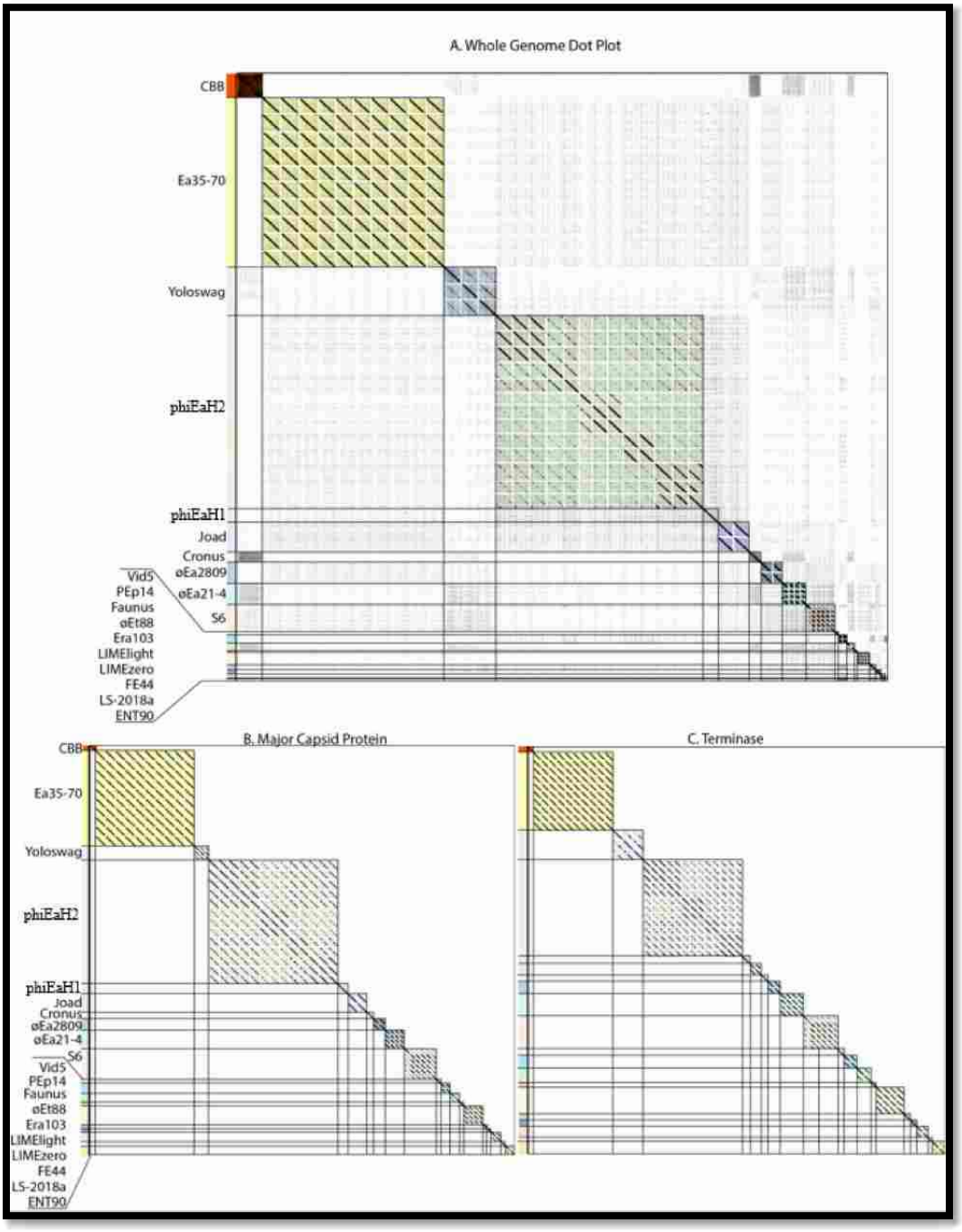


Figure II.2 Dot plots that compare the *Erwiniaceae* tailed phages reveal 20 clusters of related phages. (A) Whole genome nucleotide sequence dot plot. Sequences were reoriented to make parallel genome alignments within each cluster; the founding phage of each cluster (bold in table 1) labels each whole cluster. (B) Major capsid protein (MCP) amino acid sequence dot plot. (C) Large terminase amino acid sequence dot plot. Horizontal and vertical black lines separate clusters, and white lines within the colored cluster boxes mark the ends of each phage genome. Dot plots were constructed using Gepard (11). Note that the phage LS-2018a sequence was not annotated, but putative MCP and terminase were identified using tBLASTn (22).The genomes of the *Erwiniaceae* phages range from 378,379 bp (phage CBB)

to 29,564 bp (phage ENT90). The average genome length of the 991 other *Enterobacteriales* tailed phage genomes currently in the NCBI database is 81,187 bp, but the *Erwiniaceae* phages have an average genome length of 162,734 bp. Thus, *Erwiniaceae* bacteriophages comprise about five percent of the sequenced *Enterobacteriales* tailed phages, and the average genome size is almost double the overall average. Figure II.1 plots the length of all the *Enterobacteriales* tailed phage genomes and indicates the locations of the *Erwiniaceae* phages. The *Erwiniaceae* phage genome lengths are within the previously known extremes, but it is not known if their large average size is the result of isolation methods used, properties of the hosts or the skew in isolation sources toward trees and the soil around them.

The 60 *Erwiniaceae* phage whole genome nucleotide sequences were compared with Genome Pair Rapid Dotter (Gepard) (*17*) (figure II.2A). By the criterion of diagonal line strength, these phages fall into 20 clusters that have similarity over 50% of the phage genome as previously described (*26, 27*). The clusters in figure 2A are indicated by the founding *Erwiniaceae* phage in the group (the first sequence released in GenBank) unless the phage belongs to a previously-described *Enterobacteriales* cluster, in which case the previously published name for that cluster is used (*26*). An Average Nucleotide Identity (*10*) matrix was also constructed using Geneious (*16*), and if phage clusters are defined so that each phage has ≥50% ANI with at least one other phage in the group and ≤24% ANI with phages from other clusters (supplementary Table II.S1), the ANI grouping matches the dot plot-defined clusters perfectly. Our clusters correspond in general to genera or subfamilies that have been defined by the International Committee on Virus Taxonomy (ICTV), but a number of our clusters have not yet been formalized by that group.

In addition to genome nucleotide sequence analysis, whole proteome and single protein analyses support these 20 clusters. Whole proteome analysis was performed using Phamerator (*19*) to group the phage-encoded proteins into related "Phamilies", and SPLITSTree (*28*) was used to infer relationships based on the Phamily content among the 59 annotated bacteriophages (figure II.3; phage LS-2018a is not included because it has not been annotated). The SPLITSTree analysis perfectly parallels the cluster assignments generated by whole genome dot plot and ANI analysis above. It also points out the previously observed distant relationship between LIMEzero and LIMElight, which have previously been assigned to separate clusters within the T7 supercluster. Superclusters are groups of related phage clusters that share genome size and synteny (genes that have similar functions and have similar orders) that is not observed at the nucleotide level (*26*). In addition to whole proteome analysis, single protein dot plot analysis was performed using the major capsid (MCP) (figure II.2B) and large terminase (figure II.2C) protein sequences, which have been previously used to place phages into related clusters (*26, 29*). Both of these plots agree with the clustering by the above methods and show similarities within each of the *Erwiniaceae* clusters and differences among them. The fact that all the above analyses give identical phage groupings demonstrates the robustness of such cluster determinations and indicates that the extent of past horizontal exchange of genetic information among these phages was not sufficient to disrupt their overall grouping. Thus, all these methods can be useful tools for determining phage relationships, but the fact that all but dot plots do not point out mosaic relationships should not be forgotten, and in situations where horizontally exchanged, mosaically related sequences occur at higher frequency an ANI comparison may be less informative. A summary of the 20 *Erwiniaceae* phage clusters is provided in Table II.2, which shows that they range from eight singleton clusters to two clusters that contain nine or more phages.

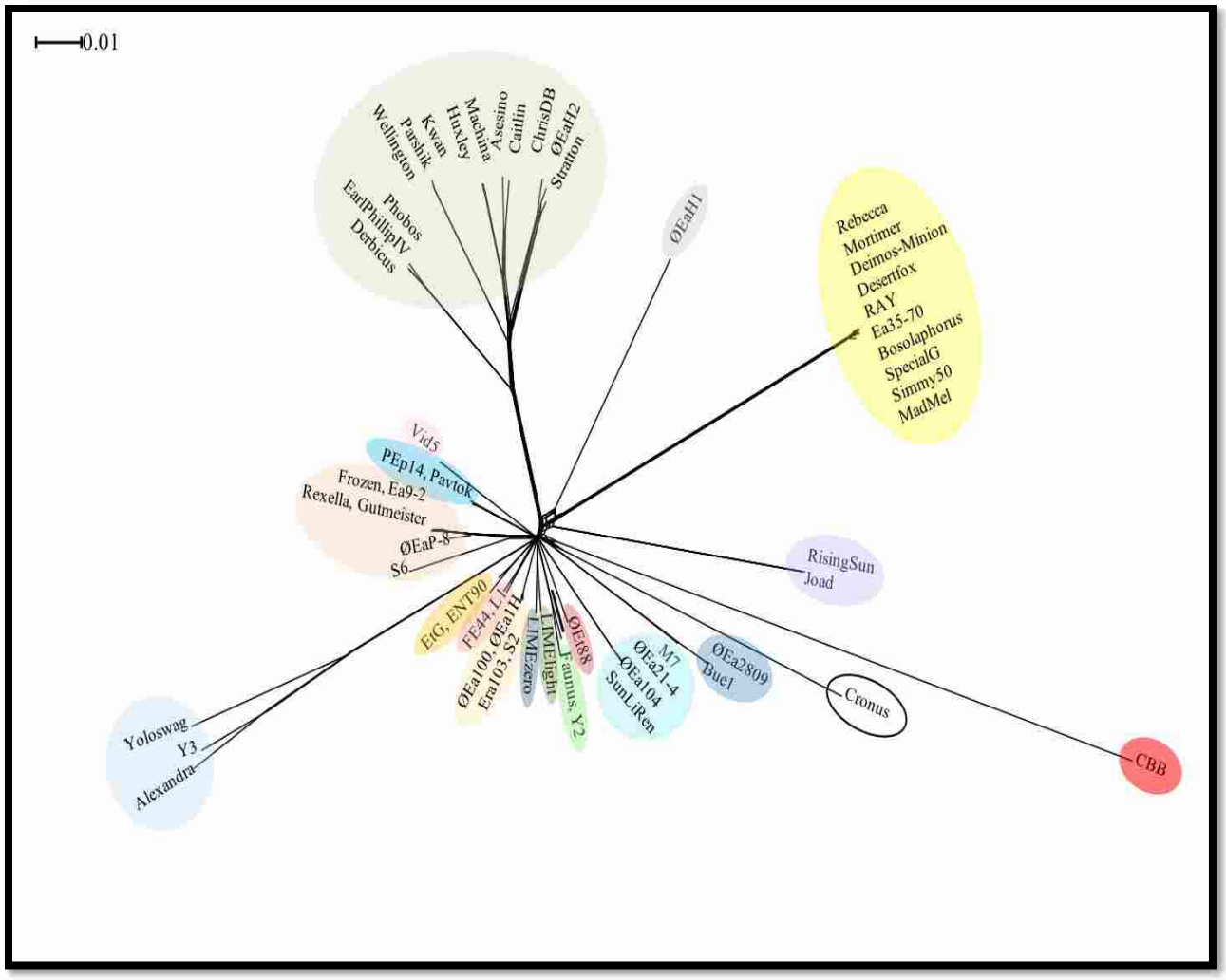


Figure II.3 A proteome phylogeny of 59 of *Erwiniaceae* tailed phages reveals 19 clusters of phages. Phamerator (30) was used to group phage proteins into phams of related proteins. SPLITStree software (21) was used to generate the tree from each pham's absence or presence in each phage genome. The phage LS-2018a genome has not been annotated and was therefore not used in this analysis.

Phages from all three families of *Caudovirales* (*Podoviradae, Myoviradae* and *Siphoviridae*) have been isolated that infect *Erwiniaceae* bacteria. The number of annotated genes ranges from 47 (phage FE44) to 605 (CBB). The genome length is quite constant within each cluster, varying by at most 9%. As seen with other bacteriophages, *Erwiniaceae* phage genes are tightly packed with an average gene density of 1.2 ORFs (open reading frames)/kb. In Figure II.4 we plot the number of ORFs against the genome size of the founding *Erwiniaceae* phage of each group. Most lie close to the trend line, and we note that since this analysis is dependent on the annotation practices of different research groups, phages furthest from the line may not be as different as their locations suggest. A genomic map comparison of the founding phage members of each cluster is provided in supplementary Figure II.S1. It clearly shows the densely packed genomes of all 19 clusters that have annotated members.

The average G+C content of the *Erwiniaceae* tailed phage genomes is 48.5% and individual phages range from 38.4% to 55.4%. *Erwinia amylovora* is the most common host species for these phages, and its G+C content is 53.6%. *Pantoea agglomerans,* the most common *Pantoea* host is 55.1% G+C. With a few exceptions, the G+C content of bacteriophage genomes is closely related to their target host (*31*), making this drastic difference interesting. We note that phage Cronos belongs to the T4-like cluster (see below) in which other members are known to have substantially lower G+C contents than their hosts (*32*). Although the purpose for alternate G+C content is unknown, it has been suggested by some authors that lower G+C phages differ from their host in order to introduce their own set of tRNA's which favor the viral genome and the associated preferred codons (*32*).

Table II.2 A summary of the 20 clusters of *Erwiniaceae* phages. The columns contain the group's name (given by founding phage from that group), the number of phages within a group, the average genome length within the group (with standard deviation), the average number of ORF's (with standard deviation), the ORF's/Genome Length (calculated from the average and supplied in ORF's/Kb), the average GC content of each cluster (with standard error), the reported morphology, and closest non-*Erwiniaceae* phage relative of the cluster (well-known phages are selectively provided when available as a relative, otherwise less well-known phages are given). Note that phage LS-2018a has not been annotated, however we determined morphology bioinformatically. None – has no close relatives (i.e., defines a novel cluster).

Erwinaceae Group Name	Number of *Erwiniaceae* phages included	Average genome length	Number of ORF's	Number of ORF's/Genome length*1000	GC content	Morphology	Close Outside Relative of Cluster
CBB	1	378,379	605	1.6 ± 0.2	36.0	*Myoviridae*	RaK2
Ea35-70	10	272,744	321	1.2 ± 0.2	49.7	*Myoviridae*	None
Yoloswag	3	262,532	339	1.3 ± 0.2	48.1	*Myoviridae*	JA11
phiEaH2	13	239,344	269	1.1 ± 0.2	50.9	*Myoviridae*	SPN3US
phiEaH1	1	218,339	244	1.1 ± 0.2	52.3	*Myoviridae*	2050HW
Joad	2	235,241	244	1.0 ± 0.2	48.3	*Myoviridae*	None
Cronus	1	175,774	295	1.7 ± 0.2	38.4	*Myoviridae*	T4
øEa2809	2	163,099	162	1.0 ± 0.2	50.3	*Myoviridae*	Vi01
øEa21-4	4	84,599	124	1.5 ± 0.2	41.8	*Myoviridae*	Felix-01
S6	6	74,656	92	1.3 ± 0.2	47.8	*Podoviridae*	N4
Vid5	1	61,437	99	1.6 ± 0.2	48.8	*Siphoviridae*	9g
PEp14	2	61,058	63	1.0 ± 0.2	50.0	*Podoviridae*	SopranoGao
Faunus	2	55,343	85	1.5 ± 0.2	43.9	*Myoviridae*	EcoM-GJ1
øEt88	1	47,279	68	1.4 ± 0.2	47.3	*Myoviridae*	T1
Era103	4	45,504	51	1.1 ± 0.2	49.8	*Podoviridae*	SP6
LIMElight	1	44,546	55	1.3 ± 0.2	54.0	*Podoviridae*	KP34
LIMEzero	1	43,032	57	1.3 ± 0.2	55.4	*Podoviridae*	J8-65
FE44	2	39,571	49	1.2 ± 0.2	50.3	*Podoviridae*	T7
LS-2018a	1	31,798	-	-	51.0	*Siphoviridae*	None
ENT90	2	29,989	53	1.8 ± 0.2	55.0	*Myoviridae*	P2
Overall average:	**3**	**162,734**	**198**	**1.2**	**48.5**		

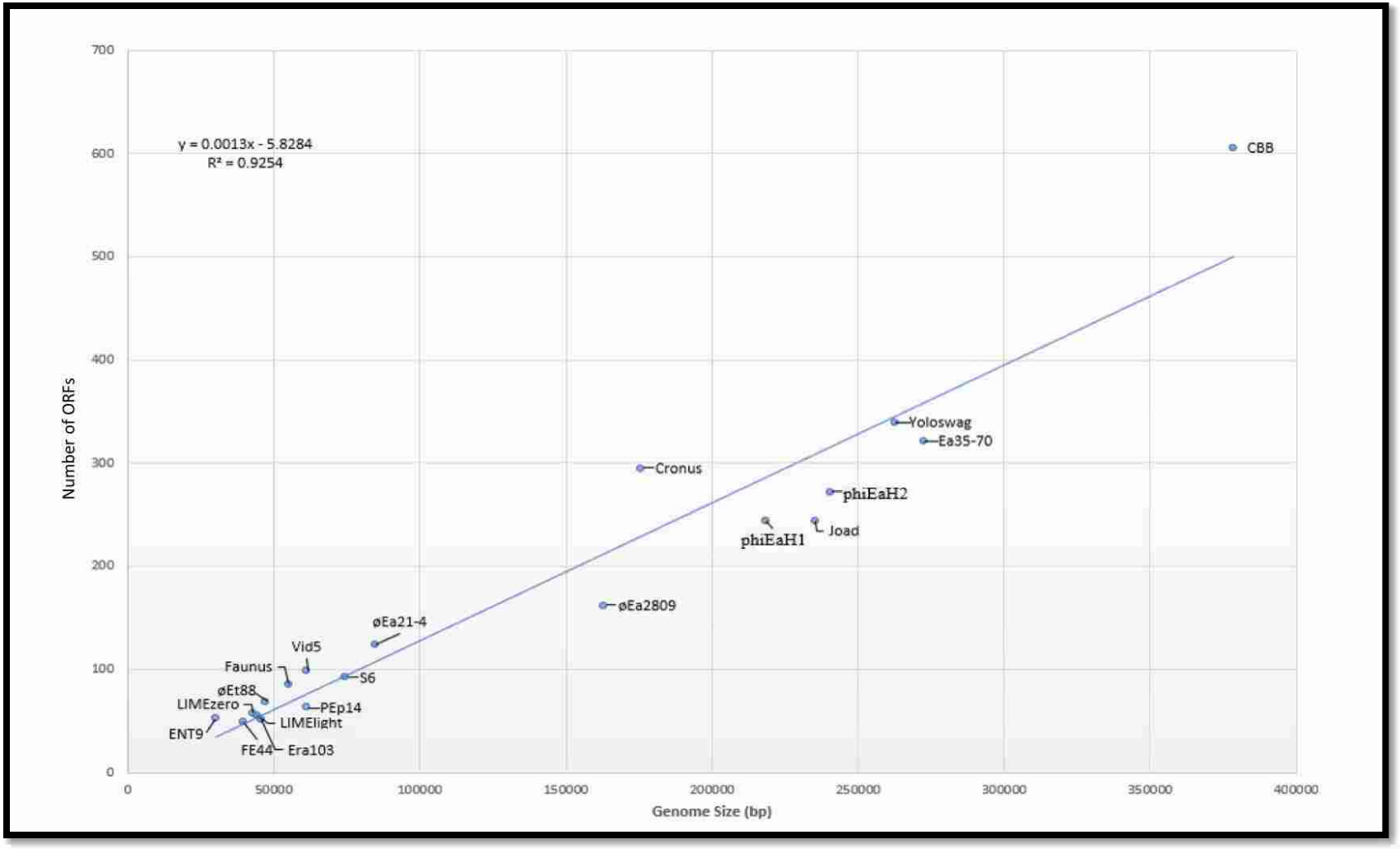


Figure II.4 Open reading frame density in *Erwiniaceae* bacteriophage clusters. Each cluster is labeled by the founding phage but represents the whole cluster's average. Equation and R2 value are displayed on the chart. The line represents a linear regression model of the average number of ORFs per phage compared to the average genome size of 19 *Erwinaceae* clusters.

II.4.2 Protein function among the *Erwiniaceae* phages

We selected one representative bacteriophage from each of the 19 annotated *Erwiniaceae* phage clusters and examined their predicted protein functions. Table II.3 shows that of the 2667 genes annotated in these 19 phage genomes only 793 (30%) have a predicted function. Since BLASTp detected homology is commonly used to identify putative function, this means that 70% of the annotated genes have no database match or match a protein whose function is unknown. Phage Era103 (*33*) had the highest percent (63%) of genes called with a putative protein function, which may in part be due to its smaller genome. In most of the *Erwiniaceae* jumbo phages only 20-30% of the encoded proteins have predicted functions. Among those with putative functions, DNA replication and recombination genes are most abundant (35-52% of total proteins with function). Phage structural proteins were also commonly annotated, with the major capsid and large terminase proteins being identified in all 19 clusters.

II.4.3 Lifestyles of *Erwiniaceae* tailed phages

We attempted to determine whether the *Erwiniaceae* phages in this study are lytic or temperate by bio-informatic means. Most appear not to carry genes such as integrase that might be indicative of the temperate lifestyle, but since prophages may not be integrated this lack does not prove a lytic lifestyle. In our 2016 study (25) we showed that the number of bacterial genome sequences that are available in the extant database is high enough that virtually all known prophage types are represented in those sequences. Therefore, if a newly identified phage is temperate it should have close relatives in extant bacterial genome sequences, especially in genomes of its host species or close relatives. Indeed, nearly all previously examined temperate. *Enterobacteriales* phage clusters encode MCP relatives with ≥97% amino acid sequence identity to proteins encoded by prophages in *Enterobacteriales* bacterial host genomes, while no

Table II.3 Putative gene functions reported from representative phages of each of the 19 annotated *Erwiniaceae* phage clusters. One representative phage from each of the 19 clusters was selected to analyze the protein function annotation. Protein function was sorted into four sections shown in different colors: structural proteins are in blue, DNA replication and recombination are in orange, cell lysis genes are in yellow, and host related genes are in green. Numbers refer to the number of proteins annotated for that function. We are aware of some possible overlap among protein function categories, this is due to the use of original annotations. LS-2018a is not represented in this table since it had no annotation.

	CBB	RAY	Yoloswag	Huxley	phiEaH1	Joad	Cronus	øEa2809	øEa21-4	S6	Vid5	PEp14	Faunus	øEt88	Era103	LIMElight	LIMEzero	FE44	ENT90
Head protein	3	1	1	1	1	5	9	3	1	2	3	2	1	3	2	2	2	4	4
Tail Fiber	5	1	11	3	3	4	6	11	5		5	2	1	2	3	3	5	4	7
Baseplate	3	1					8	3	2				2	2					3
Putative virion structural protein	60	26		38	29	26	1		3		1		2		1	2	2		
Neck/whisker	1						2	2			1								
Procapsid							2	3	1										
Terminase	2	1	1	2	1	1	2	2	1	1	2	2	1	2	2	1	1	1	2
DNA Polymerase	3	1	4	1		2	3	3	2	2	2		1		1	2	1	1	
RNA Polymerase	1	7	1	7	5	7	3			1			1		1	1	1	2	
Helicase	3	3	2	2	3	3	2	4	1	1	1		1	1	1	1	1	1	
Nuclease	9	5	6	3	3	5	9	8	2		3		2	2	5	2	2	4	
Hydrolase	4	2	2	2	3	3		2		3	2	1	2		1			2	
Recombination/ Repair	4	1	2		1	2	3	3						3					
Thymidine kinase/synthase	2	3	2	2	2	4	2	2	1	1			1						
Nucleotide reductase	5	2		1		1	4	2	3										
Topoisomerase	1		2				1	2											
Ligase	3		2			1	2	1	1		1		1		1	1		1	
Primase	2		1				1	1	1		1					1	1	1	
DNA-binding protein	1		5	1	1		3	4		1	1	2	1	1				2	
Lysin			2	1		2		1	1	1	1	3	2	1	2	1	1	2	2
Lysozyme	3	1	2				2								1				1
Holin							1		1		1	2	1	1	1	1			
Lysis inhibitor/regulator							4		1		1								
Lytic transglycosylase		2	2		2	2													
Integrase												1		1					1
Transcriptional/ Translational repressor protein	2		3	1		1	1	1										2	3
Nucleoid disruptor protein							1												
Secretion systems			4																
EPS		1	1	1	1			1			1	1			1	1			

purely lytic phages had such closely related homologs (25). We therefore searched for MCP genes similar to those of the 20 clusters described here in bacterial host genomes (Table II.4). Of the 20 clusters, only three have homologues with >80% identity in bacterial genomes. *Erwinia* phage ENT90's MCP was 100% identical to a gene (locus tag C2E16_18005) in a similar prophage in a *Pantoea* sp. PSNIH2, suggesting it is most likely temperate in nature which is consistent with its similarity to temperate *E. coli* phage P2. *Erwinia* phage øET88 MCP has a 97% identical homologue (locus_tag SAMN05216522_1056) as a putative prophage in the genome of Rosenbergiella nectarea strain 8N4; this species is a close relative of *Pantoea* and *Erwinia* (34). In addition, Müller et al. (35) reported that øET88 was isolated after mitomycin C treatment of an *Erwinia tasmaniensis* strain, a treatment that often results in prophage induction. Finally, we have previously argued from genomic analysis that øET88 should be considered a member of the phage lambda supercluster (26), and all other members of this large group are temperate. The putative phage LS-2018a MCP has 94% identical homologues encoded by the genomes of several *Yersinia* pestis isolates (e.g., strain I-2638). At least one of these genes is present on a circular 34 kb plasmid (Acc. No. KT020860) that is largely homologous to the LS-2018a genome. Thus, we suggest that LS-2018a is very likely a temperate phage with a circular plasmid prophage. In addition, phage PEp-14 encodes a protein with some similarity to phage integrases, suggesting that it could be temperate in spite of the fact that its closest MCP matches in the reported bacterial genome sequences are ≤88% identical and are in very distantly related bacteria; however, it is possible that by chance no host genomes with PEp-14-like prophages have been sequenced.

We also note that the *Burkholderia* phages BcepIL02 and Bcep22 have substantial genome synteny with PEp-14 and also carry an apparent integrase gene.

Table II.4 The closest tBLASTn match to the MCP of 20 *Erwiniaceae* bacteriophage clusters. The MCP of the founding phage for each group (see Table 2) was used in a tBLASTn search for closest relatives in bacterial genomes that were greater than 1 megabase. *Erwiniaceae* bacteriophage clusters that are not represented in this table had no significant tBLASTn hits. *Closest *Enterobacteriales* bacteria

Phage	Best tBLASTn Bacterial Match	Accession Number	Identity
øEa21-4	*Polyangium brachysporum* strain DSM 7029	CP011371	32%
S6	*Alteromonas sp.* RKMC-009	CP031010	56%
Vid5	*Nitrosomonas ureae* strain Nm10	CP013341	47%
	**Enterococcus faecalis* strain TY1	CP031027	35%
PEp14	*Martelella* sp AD-3	CP014275	75%
Faunus	*Rhizobiales* strain PAMC 29148	CP036515	29%
	**Enterobacter cloacae* strain 20710	CP030076	28%
øEt88	*Rosenbergiella nectarea* strain 8N4	CP009706	97%
Era103	*Pandoraea faecigallinarum* strain DSM 23572	CP011807	30%
LIMElight	*Cronobacter sakazakii* strain ATCC 29544	CP011047	41%
LIMEzero	*Enterobacter kobei* strain DSM 13645	CP017181	52%
LS-2018a	*Yersinia pestis* strain I-2638	CP013974	94%
ENT90	*Pantoea sp.* PSNIH2	CP009866	100%

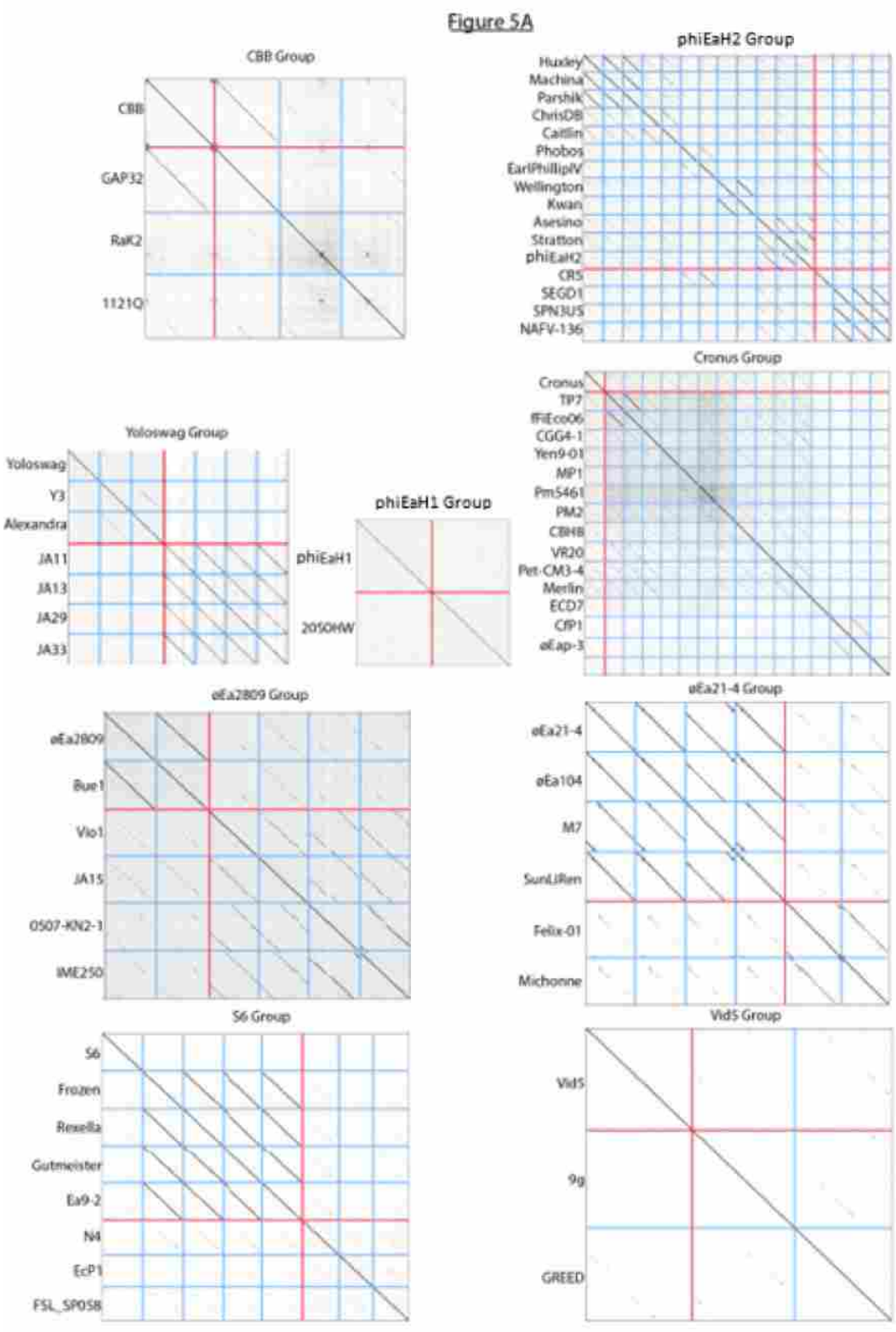
They have been reported not to form stable lysogens but may be able to form a transient benign association with the host (36); on the other hand, in a single counter-example we find a protein that is 93% identical to Bcep22 MCP encoded by a gene (locus tag WS71_20305) in an integrated prophage that is quite similar to Bcep22 in the genome of *Burkholderia* sp. DU8 (Acc. No. CP0013389). Thus, definitive determination whether phage PEp-14 is lytic or temperate awaits further study, but we conclude ENT90, øET88 and LS-2018a are almost certainly temperate, and the other 16 clusters discussed here most likely contain lytic phages. Bacterial matches included in Table II.4 are all clearly inserted in the bacterial chromosome or in a known plasmid (it is possible that finding a fragment of a phage genome in a bacterial draft genome can be a result of a lytic phage infection at the time of sequencing).

II.4.4 The 20 clusters of *Erwiniaceae* tailed phages

Since we have shown that in the *Enterobacteriales* clusters MCP sequence clustering nearly always reflects whole genome clustering (*26*), BLASTp searches with MCPs from each of these clusters were first used to identify the most closely related non-*Erwiniaceae* phages. These results and subsequent whole genome nucleotide comparisons showed that 17 of the *Erwiniaceae* clusters can be placed in previously defined *Enterobacteriales* phage clusters (summarized in Table II.2). Figures II.5A and B show nucleotide sequence dot plots that compare phages from each of the 17 non-singleton clusters with their most closely related *Enterobacteriales* phages. Subcluster designations, indicating closer relationships, are provided in Table II.1 (see Grose and Casjens for *Enterobacteriales* cluster/sub-cluster assignments).

Three *Erwiniaceae* phage clusters typified by phages Yoloswag, Joad and LS-2018a represent novel *Enterobacteriales* tailed phage clusters that have not been previously described.

Figure 5A


CBB Group
CBB
GAP32
RalK2
1121Q
phiEaH2 Group
Huxley
Machina
Parshik
ChrisDB
Caitlin
Phobos
EarlPhillipIV
Wellington
Kwan
Asesino
Stratton
phiEaH2
CR5
SEGD1
SPN3US
NAFV-136
Cronus Group
Cronus
TP7
fFiEco06
CGG4-1
Yen9-01
MP1
Pm5461
PM2
CBH8
VR20
Pet-CM3-4
Merlin
ECD7
CfP1
øEap-3
Yoloswag Group
Yoloswag
Y3
Alexandra
JA11
JA13
JA29
JA33
phiEaH1 Group
phiEaH1
20SOHW
øEa2809 Group
øEa2809
Bue1
Vio1
JA15
0507-KN2-1
IME250
øEa21-4 Group
øEa21-4
øEa104
M7
SunLiRen
Felix-01
Michonne
S6 Group
S6
Frozen
Rexella
Gutmeister
Ea9-2
N4
EcP1
FSL_SP058
Vid5 Group
Vid5
9g
GREED

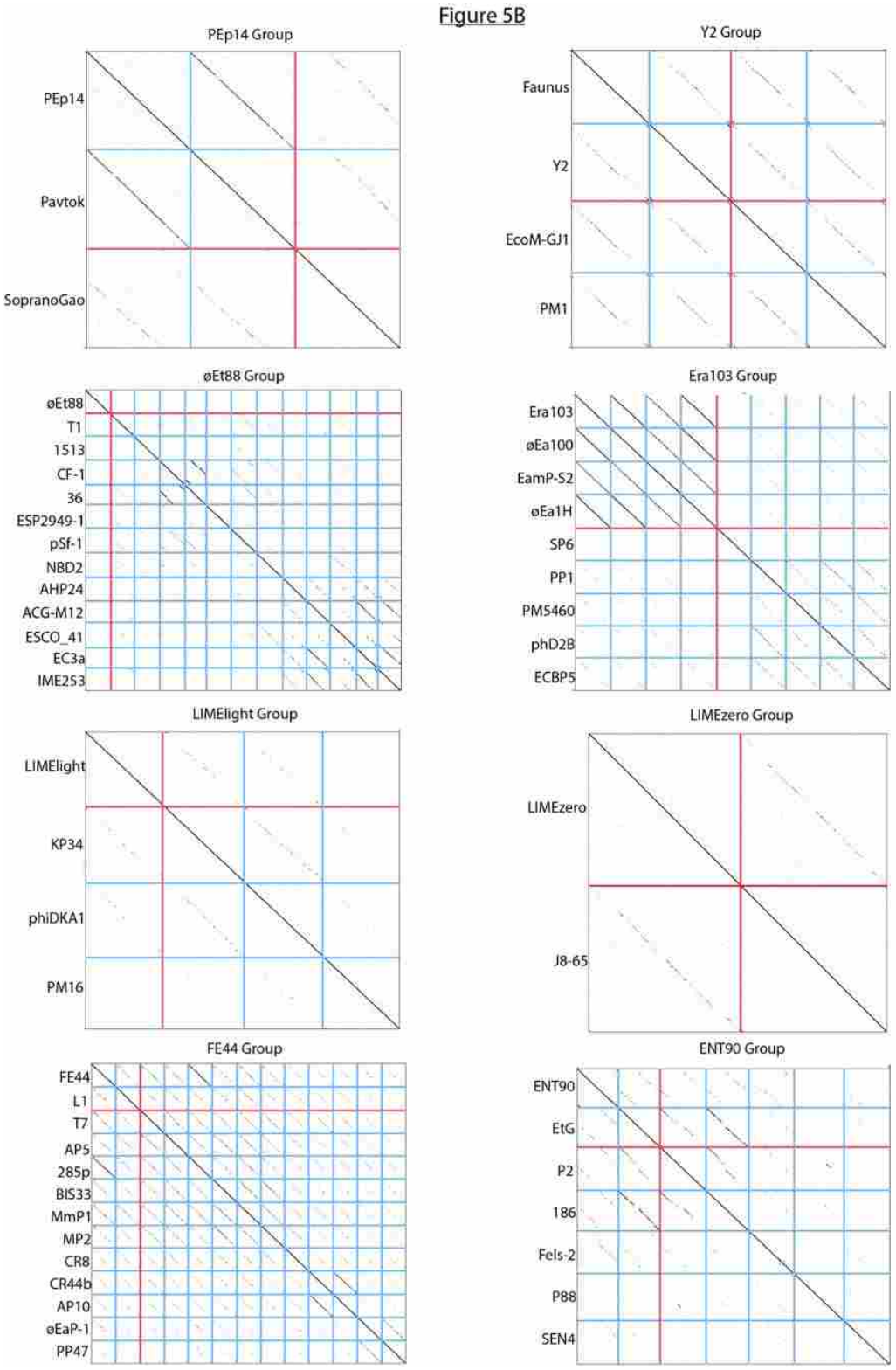


Figure II.5 Dot plots of 17 *Erwiniaceae* phage clusters with their relatives. Red lines separate *Erwiniaceae* bacteriophage clusters and homologous Enterobacteriales phage genomes. Blue lines indicate the ends of each genome. Parts A and B depict nine and eight phage clusters, respectively. Due to the large number of phages in some of the phage clusters only representative phages are shown.

The following paragraphs examine the molecular lifestyles of the 20 phage clusters with members that infect the *Erwiniaceae*:

II.4.4.1 Jumbo phages with genomes larger than 200 kb

1) CBB was originally isolated on *Pectobacterium* but forms plaques on an *Erwinia* strain *(15).* It is the largest *Erwiniaceae* phage reported to date (*15*). CBB fits in the RaK2-like *Enterobacteriales* phage cluster and is most similar to *Cronobacter* phage GAP32. The nine known phages in the RaK2-like cluster form three subclusters and a representative from each cluster is shown in Figure II.5A. The RaK2-like phages are jumbo *Myoviridae* phages, and many or all are "hairy" with unusual whisker-like structural proteins along the contractile tail. The Phamerator map (supplementary Figure II.S1 indicates that the terminal regions of the CBB genome (as it is currently oriented in GenBank) share some similarity to other *Erwiniaceae* phages, specifically with the Cronus and ØEa2809 clusters. These related regions encode proteins annotated as hypothetical proteins and structural proteins. (For more information see reference (*37*) for phage GAP32 characterization)

2) The phiEaH2-like *Erwiniaceae* group fits into the previously defined *Enterobacteriales* SPN3US-like phage cluster (*26*). This cluster consists of jumbo myoviruses with genomes in the 229-247 kb range (note that an error in Table 1 of reference (*26*) places the SPN3US-like and Rak2-like clusters inside the rV5 supercluster, but this is incorrect). The SPN3US-like cluster also includes phages that infect *Salmonella, Escherichia* and *Cronobacter* hosts, and the dot plot in Figure 5A shows that the 16 phages currently in this cluster separate into 9 subclusters, of which only subcluster A includes phage from multiple host genera (*Escherichia* and *Salmonella* including phages *SPN3US, SEGD1, NAFV-136*). The 13 *Erwinia*

phages form 7 different subclusters that contain no other phages, highlighting the strong correlation between phage subclusters and host genus.

One of the noteworthy features of phage SPN3US is that it encodes a five subunit RNA polymerase that is packaged into the virion and injected into the host cell with the phage DNA and the *Erwinia* members carry similar genes. This group also shares a number of gene homologies with *Pseudomonas aeruginosa* phage øKZ (91 genes), including 61 virion structural genes (*38, 39*). (For more information see reference (*40*) for phage SPN3US characterization)

3) The jumbo *Erwinia Siphoviridae* phage phiEaH1 has a 218 kb genome and is the prototypical member of the *Enterobacteriales* phiEaH1-like phage cluster (*41, 42*). The only other phage in this cluster is *Serratia* phage 2050HW. These two phages are moderately distant relatives sharing syntenic proteomes (with their MCP's sharing 56% identity and 71% similarity) and are only very distantly related at the nucleotide level (see Figure II.5A). (For more information see reference (*43*) for phage 2050HW characterization)

II.4.4.2 *Myoviridae* with genomes between 50 and 180 kb

4) *Erwinia* phage Cronus forms a singleton subcluster in the *Enterobacteriales* T4-like *Myoviridae* cluster. Its genome size of 175 kb is typical of phages in this cluster, and like many other phages in this cluster its DNA has a substantially lower G+C content than its host. This cluster currently contains 169 completely sequenced genomes of phages that infect 14 host genera from six of the families within the *Enterobacteriales* order (supplementary Table II.S2). The dot plot in supplementary Figure II.S2 Part B shows that there are 21 subclusters (A through U) in this cluster, one of which is defined by phage Cronus. Six of the subclusters are singletons, but of the 15 subclusters with more than one member, 11 contain members that all infect the same host genus (assuming that *Escherichia* and *Shigella* are actually one genus (*44*); and all but

one has members that infect a single host family. Thus, subcluster membership is far from random, with many genus-specific or family-specific subclusters at this level of analysis. We also note that diversity within this cluster is still quite incompletely understood (*30*), since (i) the almost 30% singleton subclusters implies the existence of numerous undiscovered subclusters, (ii) individual genera are often infected by multiple phage subclusters, and (iii) phages of a number of *Enterobacteriales* families and genera remain unexplored.

5) The *Erwinia* phages øEa2809 (*14*) and Bue1 (accession No. MG973030) share similarity to the *Enterobacteriales* Vi01-like cluster of *Myoviridae*, which is currently comprised of 51 *Enterobacteriales* phages, including *E. coli* phage CBA120 and *Salmonella enterica* phage Det7, that typically have genomes in the 150-165 kb range. These phages have virion structural genes that are moderately distant relatives of those of phage T4, but their virion heads are isomorphic rather than elongated, and their homologous genes are not syntenic with the T4-like phages. They encode a thymidylate synthase that suggests they may incorporate hydroxymethyldeoxyuracil into their DNA (*45*), and they encode multiple tailspikes that allow them to adsorb to several different hosts (*46-48*). This cluster was previously separated into at least six subclusters, one of which is comprised of only the two *Erwinia* phages øEa2809 (*14*) and Bue1, phages that were isolated in Belarus and Switzerland, respectively (Figure 5A). (For more information see reference (*45*) for phage CBA120 characterization)

6) The øEa21-4-like *Erwinia* phage group lies within the previously defined Felix-O1-like *Enterobacteriales* phage cluster (*26, 50, 51*). This cluster of contractile tailed phages have genome sizes that range from 82 to 91 kb and carry a number tRNA genes which are highly conserved across the øEa21-4 group. The Felix-O1-like cluster currently contains 46 completely sequenced *Enterobacteriales* phages that fall into three subclusters (*26*), and the four *Erwinia*

phages in this cluster form one of these subclusters (Figure 5A). The known phages in this cluster infect six different *Enterobacteriales* host genera, and there are fairly close relatives that infect *P. aeruginosa* in the *Pseudomonadales* order of Gamma-Proteobacteria (*50*). (For more information see reference (*52*) for phage TP1 characterization)

7) The phage Y2-like *Erwinia* group has similarity to the previously defined *Enterobacteriales* lytic Myoviridae phage øEcoM-Gj1-like cluster (26), currently containing four subclusters. This cluster is currently comprised of two *Escherichia* phages, øEcoM-Gj1 (53) and ST32, two *Pectobacterium* phages, PM1 (54) and PP101, and two *Erwinia* phages, Faunus and Y2 (55). The last two are sufficiently different that they each form a distinct singleton subcluster (Figure 5B). These phages have genomes in the 52-57 kb range and encode a single subunit RNA polymerase like phage T7 (53). (No phages in this cluster have been extensively characterized)

II.4.4.3 Lytic *Podoviridae* phage

8) The *Erwinia* S6-like group fits into the previously defined *Enterobacteriales* N4-like cluster of *Podoviriadae* phages. The 26 currently known completely sequenced members of this cluster fall into six subclusters, three of which, typified by phages Ea9-2 , S6 (*1*) and ØEaP-8 (*56*), are made up by the seven *Erwinia* phages and no others; the last two are singleton subclusters (Figure II.5A). The larger group of N4-like phages appears to be a very successful group of phages whose members infect other Gamma-Proteobacteria orders as well as Beta-Proteobacteria hosts (*e.g.,* N4-like phage JWDelta infects the Beta-Proteobacteria *Achromobacter xylosoxidans (57)).* A unique feature of this group is its large (about 3500 amino acid) single subunit RNA polymerase that is present in the virion and is injected with the DNA into the host cell (*58*).

9) Phages PEp-14 and Pavtok define an *Erwinia* group that expands the previously defined *Enterobacteriales* PEp-14-like *Podoviridae* singleton cluster (*26*). *Klebsiella* phage SopranoGao is also a recently sequenced member of this cluster, but the two *Erwinia* phages form a unique subcluster (Figure 5A). As discussed above it remains unclear whether these phages are temperate or lytic. A striking feature of these phages, that have genomes about 61 kb long, is that they encode an exceptionally large putative protein that is 4915, 5007 and 4369 amino acids long in the PEp-14 (Acc. No. YP005098431), Pavtok (AXF51455) and SopranoGao (ASV45029) homologues, respectively. These single genes occupy about a quarter of their genomes, and their products are the longest bacteriophage encoded proteins that we are aware of. Other classes of large phage proteins are the virion RNA polymerases of the N4-like phages (above) (*59*) and a possible tail fiber of øKO2 at 3433 AA (*60*). BLASTp searches with the large Pavtok protein (locus_tag PAVTOK_25) have shown that it shares patches of convincing similarity to large proteins in the following phages that infect diverse hosts: ≥50% identity to *Vibrio* phage VvAW1 (3640 AA; Gamma-Proteobacteria host), *Pseudomonas* phage Skulduggery (3695 AA; Gamma-Proteobacteria host), *Agrobacterium* phage atu_ph08 (4877 AA; Alpha-Proteobacteria), and *Sinorhizobium* phage PBC5 (2849 AA; Alpha-Proteobacteria host), as well as 35% identity to proteins from several Beta-Proteobacteria phages including *Burkholderia* phages Bcep22 (4602 AA). The function of these large proteins has not been studied directly, but two sequence matches are informative. First, amino acids 70-170 of all three of the PEp-14-like cluster phages' large protein contain a lysozyme motif and are 33% identical to a section of phage T7 gene *16* protein. There are a small number of molecules of *16* protein in the T7 virion, and they are released into the host with the DNA (*61, 62*). Many tailed phage that infect Gram negative bacteria are thought to inject proteins with lysozyme activity that cleave the

peptidoglycan so that DNA can pass through it to reach the cytoplasm during injection (*63, 64*), and the T7 gene *16* protein has been shown to have such an activity (*65*). Second, a region between amino acids 2700 and 3200 of the PEp-14-like large proteins have weak but convincing similarity to parts of *E. coli* phage P1 DarB protein (2255 AA; accession No. YP_006479), which has also been shown to be injected with the DNA (*66*) and is involved in defense against host restriction endonucleases (*67*). We conclude that it is very likely that these large PEp-14-like cluster proteins are present in the virions and are injected into the host with the DNA. Gill *et al.* (*36*) have made a similar argument with the homologous large gp75 protein of *Burkholderia* phage Bcep22 (Acc. No. NP944303), which has been shown to be a virion protein. Why are these PEp-14-like phage proteins and their homologues so large? We speculate that when a phage "finds" a new protein function that is advantageous to inject from the virion, it may be evolutionarily simplest to fuse it to an existing protein that is injected. Thus, such proteins may accumulate new polypeptide sections and become large multidomain proteins over time. This would also explain the patchy nature of the relationships between such proteins in different phages. We note that the distantly related phages mentioned above all have similarity to the leftmost approximately 37 kb of the PEp-14-like phages (in the Pavtok GenBank orientation), a region that contains the putative virion assembly genes (supplementary Material figure S3 shows a comparison of phage Pavtok with *Burkholderia* phage DC1/Bcep22). (No phage in this cluster has been extensively characterized)

(10-13) The four *Erwiniaceae* phage clusters discussed in this section fall into the previously defined *Enterobacteriales* T7-, SP6-, KP34- and LIMEZERO-like clusters (*26*), which in turn all reside within the T7 supercluster (classified by the International Committee on Virus Taxonomy as the *Autographivirinae* subfamily of the *Podoviridae*). They all have apparent

lytic life cycles similar to phage T7 which infects *E. coli* (*68*) and is one of the best characterized and most prolific tailed bacteriophages. It has many known relatives that infect a wide variety of bacterial hosts, even outside of the *Enterobacteriales*. Hallmarks of these phages include a phage encoded single subunit RNA polymerase.

Erwinia phage ERA103 fits into the *Enterobacteriales* SP6-like cluster, where it, along with *Erwinia* phages øEa100 (*69*), øEa1H and S2, form the *Erwinia* specific subcluster D. *Pantoea* phage LIMElight belongs to the KP34-like cluster where it forms the singleton subcluster B. *Pantoea* phage LIMEzero is the prototype phage for the LIMEzero-like cluster, which also contains *Escherichia* phage J8-65. These two phages define different subclusters. Finally, *Erwinia* phage FE44 shares its highest overall nucleotide sequence identity of 91-94% to *Escherichia* phages 285P, BA14 and S523 (*70, 71*) and is a member of the T7-like cluster. FE44, along with phages that infect the *Escherichia, Yersinia, Salmonella, Kluyvera* and *Pectobacterium* genera, form subcluster C of this *Enterobacteriales* cluster (Figure 5B).

II.4.4.4 Lytic *Siphoviridae* phage

14) *Pantoea* phage Vid5 is a member of the *Enterobacteriales* 9g-like cluster of lytic phages (*26, 72*). This *Siphoviridae* cluster's founding member phage 9g has deoxy-archaeosine (modified guanosine) nucleotides in its DNA that make it resistant to many restriction endonucleases (*73*). Vid5 has a similar but not identical set of genes predicted to be involved in this or a similar DNA modification, and its DNA is similarly resistant to such nucleases (*72*). The 15 phages with available complete genomes in this cluster fall into three subclusters, two of which have been called the *Nonagvirus* and *Seuratvirus* genera (*74*), and the third is Vid5 which forms a singleton subcluster. The dot plot in Figure 5A compares representatives of these three subclusters; subclusters A and B are all *Escherichia* phages except for one *Salmonella* phage

(phage SE1; accession No. KY926791) in subcluster A. (For more information see reference (*73*) for phage 9g characterization)

II.4.4.5 Temperate *Myoviridae* phage

15) *Erwinia* phage EtG is quite closely related to *Escherichia* phage 186 and *Salmonella* phage PsP3, and although ENT90 is more distantly related, both of these *Erwinia* phages are clearly members of the P2-like *Enterobacteriales* temperate phage cluster (*26*) (Figure 5B) (*75*). Phages in this cluster are widely distributed with phages that infect many different types of *Enterobacteriales* (*25*), and EtG belongs to subcluster B that also contains phages that infect *Escherichia* and *Salmonella*, while ENT90 defines singleton subcluster D.

II.4.4.6 Clusters that currently contain only *Erwinia/Pantoea* phages

Although most of the currently known *Erwiniaceae* tailed phages fall into to one of the over 70 previously defined *Enterobacteriales* phage clusters (*25, 26*), five of 20 *Erwiniaceae* phage-containing clusters contain only phages that infect the *Erwinia* and/or *Pantoea* genera (the Ea35-70-, Yoloswag-, Joad-, LS-2018a- and øET88-like phages). Three of these five clusters (Yoloswag-, Joad-, and LS02018a-like) form novel *Enterobacteriales* clusters that have not been previously described.

16) The *Erwiniaceae* phages within the previously defined Ea35-70-like *Enterobacteriales* phage cluster (*26*) form the most highly conserved of all of the *Erwiniaceae* clusters we analyzed, with less than 3% ANI variance among the phages within this cluster. It is comprised of jumbo *Myoviridae* phages typified by phage Ea35-70 that was isolated from soil beneath a fire blight-infected pear tree in Ontario, Canada (*76*). No similar phages are known that infect other host species. More than 60% of their 271-275 kb genomes are made up of novel genes without significant BLASTp matches in the current database, and like other jumbo phages

their small fraction of genes with predicted functions encode mainly virion structural proteins and DNA metabolism proteins. (No phage in this cluster has been extensively characterized)

17) Phage Yoloswag represents an *Erwinia* jumbo phage group that includes two closely related phages, Alexandra and Y3. Five *Dickeya* phages have recently been described who's putative MCPs are about 74% identical to that of Yoloswag, and Figure 5A shows a dot plot analysis of these eight phages. Long weak diagonal similarity lines confirm that they all have similar genome organization and belong in this previously undefined *Enterobacteriales* phage cluster which we call the Yoloswag-like cluster. It also shows that substantial diversity is present within the cluster, and we define three subclusters, A, B and C. Subclusters A and B are quite different from C, and the three phages within B are more diverse (weaker diagonal similarity line) than those within C. Interestingly, the clusters do not correlate perfectly with host genus, since subcluster B contains phages with *Erwinia* and *Dickeya* hosts, and these two genera have recently been placed in the two different but rather closely related families, *Erwiniaceae* and *Pectobacteriaceae*. This suggest that one of these phages, perhaps AD1, has switched hosts in the relatively recent distant past.

The proteins that are expressed by the conserved core-genome of the Yoloswag-like *Erwinia* phages are mostly virion structural proteins and DNA replication and repair proteins. Recent publications describing the phage Y3 genome sequence (*13*) and the four *Dickeya* relative genomes (*77*) have presented various aspects of this group of phages, so we will not discuss them in detail here but will only briefly mention some of this cluster's salient features. Our Phamerator (*19*) analysis shows that there are 176 protein Phamilies conserved among the three *Erwinia* members of this cluster, with only 46 of these proteins having a predicted function, including four secretion system proteins (products of the phage Yoloswag genes *88, 107, 152*

and *154*) that are not present in any of the other *Erwinia* phages. A conserved Cas4-like protein is also encoded by all three *Erwinia* phages. Cas4 has no well-defined function but is known to be present in CRISPR/Cas gene clusters. It could be carried by the phage to modify CRISPR systems in its hosts. The virions of this cluster have large isometric heads about 130 nm in diameter and a contractile tail about 190 nm long. An interesting reported feature of at least Y3 and the *Dickeya* members of this cluster is the presence of unusual curly hair-like fibrils of unknown function extending from the sheath along the length of the tail similar to those seen in the RaK2-like phages (above). (No phage in this cluster has been extensively characterized)

18) The jumbo *Myoviridae* phages Joad and RisingSun represent a new *Enterobacteriales* tailed phage cluster with 235 kb genomes (see Figure 2 of ref. (*12*)). These two *Erwinia* phages share 96.6% whole genome ANI, with Joad encoding two genes not present in RisingSun (a putative HNH endonuclease and a hypothetical protein). The RisingSun genome encodes 243 predicted proteins, ~43% of which have no significant BLASTP database match (e-value of $\geq 10^{-7}$); another 24% of its genes have no known function but do have BLASTP matches to hypothetical proteins (*12*). This novel cluster shares some homology with *Pseudomonas* phages EL and OBP as well as *Vibrio* phages P4B and pTD1, with 112 genes that had corresponding BLASTp hits with these *Pseudomonas* phages, indicating these phages are clearly related.

19) *Erwinia* phage LS-2018a also represents (as a singleton) a new *Enterobacteriales* tailed phage cluster. Its sequence in GenBank contains very large terminal redundancy (if a small amount of sequence imprecision is allowed), and we believe it very likely has a circular genome that is 31,789 bp long. Its sequence in GenBank is unannotated, but we find a 97% identical homologue of its putative MCP (bp 29319-30479 of accession No. CP013974) and a similar

terminase encoded by several isolates of *Yersinia pestis*. In *Y. pestis* biovar Medievalis strain I-2638 the 33,778 bp long circular plasmid pTP33 (MCP encoded between bp 11182 and 12342 of accession No. KT020860 (*78*)) encodes such an MCP homologue (the other matches are on *Yersinia* sequence contigs that are the same size or smaller but are not annotated as plasmids). The dot plot in supplementary Material figure S2 shows that LS-2018a and pTP33 share considerable syntenic similarity and that (with some, not unexpected, mosaicism) they have nearly identical genome organizations. We conclude that pTP33 is very likely a circular plasmid prophage and that LS-2018a may have a similar prophage (although we note that both carry a possible integrase gene). The *Yersinia* genus is a member of the newly defined *Yersiniaceae* family in the *Enterobacteriales* (*79*), and LS-2018a and pTP33 represent two singleton subclusters, each of which infects a different host family.

20) *Erwinia* phage øET88 is the singleton representative of its *Enterobacteriales* cluster (*26*). Although its MCP is up to 49% identical to some phages in the T1-like lytic phage cluster, it is likely a temperate member of the phage lambda supercluster (*26*) (see above).

Eighteen of the 20 *Erwiniaceae* clusters contain more than one authentic phage member (øET88 and LS-2018a comprise singleton clusters). Within each of these clusters, most of the *Erwiniaceae* phages are more closely related to one another than to phages that infect other host genera and so form distinct subclusters. Nonetheless, eight of the 33 *Erwiniaceae* phage-containing subclusters also contain phages that infect other genera in addition to *Erwinia* and *Pantoea* (Figure 5), indicating a few possible examples of relatively recent host switching.

II.5 Conclusions

The purpose of this analysis was to gain insight into the relationships among the 60 *Erwiniaceae* bacteriophages that have completely sequenced genomes and to further

understanding of their host interactions. We found that on average *Erwiniaceae* phages have much larger genomes than the average *Enterobacteriales* phage, which may be due to their isolation source (trees and the soil surrounding them) or may be driven by the bacterial host. We note that only dsDNA tailed phages infective of *Erwinia* and *Pantoea* have been isolated to date. When the nucleotide and protein sequences of these 60 phages are compared, they naturally separate into 20 clusters, or 3 phages/cluster on average. This ratio highlights the diversity present in these phages in spite of the fact that they share highly related hosts. In comparison, 472 *E. coli* phages currently in GenBank fall into 50 clusters with an average of 9.4 *E. coli* phages/cluster. The lower phages/cluster ratio for *Erwiniaceae* phages (3 phages/cluster) may not be due to the decreased number of total phages isolated because *Paenibacillus* phages have a comparable number of isolates, but a phages/cluster ratio more similar to *E. coli* (9.6 phages/cluster). Comparison of the 60 *Erwiniaceae* phage genomes with all the other *Enterobacteriales* phage genomes, showed that 17 of the *Erwiniaceae* clusters belong to previously defined *Enterobacteriales* phage clusters that include phages with hosts outside this family, and three form clusters whose known members infect only *Erwiniaceae*. The *Erwiniaceae* phages in the 17 *Enterobacteriales* clusters tend form their own subcluster within their clusters. This latter distinction is perhaps due to the plant-based ecological niche of *Erwinia* and *Pantoea*. A majority of the proteins encoded by the *Erwiniaceae* phages (~70% or 1874 proteins) have unknown function, highlighting the need for further characterization of these phages. Each of the 19 analyzed *Erwiniaceae* bacteriophage clusters encodes unique proteins, including tail fibers, lysins, holins, and CRISPER proteins, which likely contribute to the phage host range and will be important considerations in the development and improvement of phage therapy cocktail design and safety.

II.6 Declarations

II.6.1 Competing interests

We declare that there were no competing interests with any author or institution responsible for this manuscript.

II.6.2 Funding

This research was funded by the department of Microbiology and Molecular Biology and the College of Life Sciences at Brigham Young University.

II.6.3 Authors contributions

DWT, SRC and JHG all performed analysis of the phage genomes and wrote the manuscript. RS annotated phages Rebecca and Derbicus and performed analysis of the Ea35-70 phage group.

II.7 Acknowledgements

The authors extend a special thanks to the Science Education Alliance – Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) for support and training on phage analysis, Steven Cresawn training on Phamerator, Edward Wilcox from the BYU Sequencing Center, and David Parady and Brandon Ekins from Life Science IT at BYU for providing DNAmaster and PhamDB on BYU servers. We would also like to thank the following undergraduates from the BYU Phage hunters program for their contributions to isolation of the phages Derbicus and Rebecca: Nolan Beatty, Jordan Berg, Minsey Choi, Steven Duncan, Rebecca Eardley, Christopher Fajardo, Hannah Ferguson, Jared Kruger, Ashley Payne and C.J. Webb, as well as Professors Donald Breakwell and Sandra Hope from the BYU Phage Hunters program.

II.8 Supplementary tables and figures

Supplementary Table II.S1 Average Nucleotide Identity (5) matrix using Geneious (6) reveals 20 clusters from 48 representative *Erwinia/Pantoea* bacteriophages . Due to the ANI program restraints and high similarity of phages within some clusters, 13 phages were removed from the ANI. The 13 phages that were removed were more than 90% similar to others within their specific cluster and therefore will not affect the overall ANI table. The individual phages that were removed from the table: SpecialG, Simmy50, Desertfox, Bosolaphorus, Rebecca, MadMel, Mortimer, Assesino, ChrisDB, Derbicus, Wellington, Kwan, ØEaP-8.

	Ray	Deimos-minion	Ea35-70	Yoloswag	Y3	Alexandra	Huxley	Machina	Parshik	Caitlin	Phobos	EarlPhillipIV	Stratton	phiEaH2	phiEaH1	Joad	RisingSun	Cronus	ØEa2809	Bue1
Ray		95.207	96.029	19.718	20.177	20.185	21.273	21.261	21.244	21.199	20.607	20.683	21.425	21.058	18.854	21.44	21.517	15.941	14.827	14.852
Deimos-minion	95.207		95.337	19.569	20.036	20.069	21.12	21.11	21.092	21.029	20.444	20.516	21.259	20.87	18.723	21.287	21.363	15.788	14.713	14.74
Ea35-70	96.029	95.337		19.778	20.234	20.258	21.385	21.369	21.356	21.306	20.709	20.771	21.522	21.13	18.852	21.47	21.547	15.93	14.823	14.85
Yoloswag	19.718	19.569	19.778		47.478	43.39	20.055	20.093	20.032	19.823	19.171	19.101	20.034	20.22	18.804	20.507	20.6	16.359	14.96	14.963
Y3	20.177	20.036	20.234	47.478		55.715	20.524	20.531	20.501	20.32	19.807	19.796	20.472	20.489	19.57	21.045	21.136	16.757	15.381	15.31
Alexandra	20.185	20.069	20.258	43.39	55.715		20.394	20.412	20.363	20.256	19.748	19.771	20.377	20.28	19.417	20.874	20.975	16.344	15.179	15.184
Huxley	21.273	21.12	21.385	20.055	20.524	20.394		97.727	98.351	60.345	38.536	38.237	51.88	46.188	20.001	23.261	23.403	17.451	16.345	16.379
Machina	21.261	21.11	21.369	20.093	20.531	20.412	97.727		97.026	60.704	38.587	38.313	52.051	46.292	19.999	23.315	23.457	17.485	16.342	16.376
Parshik	21.244	21.092	21.356	20.032	20.501	20.363	98.351	97.026		60.197	38.48	38.196	51.859	46.202	19.989	23.248	23.396	17.438	16.311	16.342
Caitlin	21.199	21.029	21.306	19.823	20.32	20.256	60.345	60.704	60.197		38.837	38.481	51.724	46.7	20.037	23.21	23.306	17.355	16.304	16.36
Phobos	20.607	20.444	20.709	19.171	19.807	19.748	38.536	38.587	38.48	38.837		66.579	38.679	36.739	20.462	23.365	23.438	18.125	16.677	16.592
EarlPhillipIV	20.683	20.516	20.771	19.101	19.796	19.771	38.237	38.313	38.196	38.481	66.579		38.34	36.211	20.552	23.486	23.506	18.119	16.805	16.728
Stratton	21.425	21.259	21.522	20.034	20.472	20.377	51.88	52.051	51.859	51.724	38.679	38.34		95.729	20.017	23.465	23.597	17.509	16.235	16.28
phiEaH2	21.058	20.87	21.13	20.22	20.489	20.28	46.188	46.292	46.202	46.7	36.739	36.211	95.729		20.579	23.713	23.892	18.212	17.342	17.415
phiEaH1	18.854	18.723	18.852	18.804	19.57	19.417	20.001	19.999	19.989	20.037	20.462	20.552	20.017	20.579		21.146	21.247	23.478	22.149	22.002
Joad	21.44	21.287	21.47	20.507	21.045	20.874	23.261	23.315	23.248	23.21	23.365	23.486	23.465	23.713	21.146		95.783	18.488	16.912	16.817
RisingSun	21.517	21.363	21.547	20.6	21.136	20.975	23.403	23.457	23.396	23.306	23.438	23.506	23.597	23.892	21.247	95.783		18.528	16.956	16.899
Cronus	15.941	15.788	15.93	16.359	16.757	16.344	17.451	17.485	17.438	17.355	18.125	18.119	17.509	18.212	23.478	18.488	18.528		26.903	26.559
øEa2809	14.827	14.713	14.823	14.96	15.381	15.179	16.345	16.342	16.311	16.304	16.677	16.805	16.235	17.342	22.149	16.912	16.956	26.903		91.338
Bue1	14.852	14.74	14.85	14.963	15.31	15.184	16.379	16.376	16.342	16.36	16.592	16.728	16.28	17.415	22.002	16.817	16.899	26.559	91.338	
øEa21-4	8.18	8.117	8.222	18.61	14.414	12.683	9.236	9.224	9.206	9.132	9.093	9.085	9.047	9.054	8.945	9.257	9.307	9.253	8.608	8.49
øEa104	8.177	8.116	8.219	18.616	14.422	12.688	9.235	9.227	9.204	9.124	9.096	9.08	9.046	9.068	8.95	9.263	9.309	9.244	8.622	8.502
M7	8.241	8.162	8.269	18.685	14.751	12.916	9.519	9.509	9.507	9.312	9.176	9.246	9.271	10.082	9.037	9.327	9.367	9.538	8.916	8.846
SunLiRen	8.19	8.124	8.229	18.626	14.447	12.708	9.249	9.239	9.219	9.144	9.074	9.078	9.041	9.027	8.917	9.276	9.326	9.228	8.606	8.496
S6	7.097	7.03	7.079	7.054	7.352	7.227	8.027	8.015	8.016	7.994	8.174	8.363	7.946	8.276	19.079	8.072	8.082	13.015	13.447	13.239
Frozen	7.194	7.136	7.194	7.204	7.404	7.3	8.075	8.082	8.039	7.999	8.287	8.352	7.978	8.435	11.116	8.225	8.221	15.3	23.794	23.458
Rexella	7.212	7.155	7.213	7.237	7.436	7.338	8.091	8.096	8.055	8.019	8.318	8.351	7.995	8.5	11.161	8.267	8.263	15.327	23.832	23.498
Gutmeister	6.801	6.746	6.805	6.933	7.005	6.907	7.651	7.659	7.616	7.56	7.845	7.91	7.561	8.434	10.518	7.801	7.794	14.533	22.667	22.35
Ea9-2	7.207	7.15	7.209	7.259	7.47	7.346	8.105	8.11	8.071	8.048	8.32	8.385	8.011	8.445	11.179	8.259	8.258	15.362	23.965	23.64
Vid5	6.314	6.261	6.319	6.119	6.011	6.007	9.08	9.054	9.077	9.076	9.212	9.327	9.193	13.209	7.204	7.404	7.393	8.09	8.258	8.172
Pep14	5.782	5.733	5.794	5.718	5.822	5.888	6.597	6.608	6.601	6.658	6.741	6.919	6.603	7.111	9.356	6.646	6.654	16.983	13.329	13.197
Pavtok	5.872	5.826	5.874	5.792	5.894	5.981	6.642	6.655	6.639	6.719	6.804	6.946	6.676	7.188	9.405	6.736	6.732	17.204	13.468	13.355
LS-2018a	6.25	6.196	6.251	6.056	5.985	6.001	9.099	9.064	9.111	9.001	8.974	9.038	9.122	12.643	7.221	7.23	7.228	7.907	8.181	8.097
Faunus	5.435	5.364	5.433	5.214	5.269	5.124	6.29	6.28	6.281	6.309	6.458	6.545	6.291	6.793	5.805	13.334	13.358	6.403	6.097	6.047
Y2	5.642	5.581	5.641	5.423	5.494	5.443	7.843	7.816	7.827	7.838	10.928	11.2	7.728	7.665	6.309	6.678	6.69	7.042	6.703	6.655
ØEt88	4.857	4.828	4.874	4.745	4.643	4.674	7.786	7.754	7.783	7.722	6.779	6.714	7.887	11.563	5.592	5.725	5.726	6.253	6.306	6.254
Era103	4.116	4.084	4.122	6.494	7.579	10.212	4.727	4.719	4.724	4.676	4.805	4.821	4.646	4.799	4.833	4.751	4.768	4.917	4.88	4.877
ØEa100	4.129	4.096	4.132	6.535	7.627	10.26	4.743	4.735	4.739	4.708	4.839	4.845	4.667	4.741	4.848	4.753	4.766	4.928	4.896	4.883
EamP-s2	4.127	4.104	4.136	5.995	6.87	8.716	4.69	4.69	4.694	4.656	4.769	4.786	4.619	4.773	4.811	4.686	4.706	4.891	4.867	4.862
ØEa1H	4.127	4.094	4.131	6.53	7.621	10.255	4.739	4.731	4.736	4.705	4.836	4.842	4.665	4.742	4.843	4.751	4.763	4.927	4.898	4.885
LIMElight	6.49	6.444	6.497	4.33	4.321	4.288	5.31	5.307	5.309	5.338	5.36	5.482	5.238	6.011	5.11	5.221	5.237	5.604	5.887	5.871
LIMEzero	7.278	7.198	7.299	4.225	4.167	4.141	5.19	5.173	5.195	5.173	5.276	5.353	5.184	5.965	4.93	5.053	5.07	5.387	5.682	5.658
FE44	4.124	4.087	4.134	4.024	3.936	3.9	6.904	6.886	6.91	6.895	6.015	5.958	7.074	10.376	4.702	4.87	4.892	5.378	5.345	5.297
EamP-L1	4.173	4.13	4.179	3.986	3.914	3.853	7.808	7.783	7.813	7.779	6.381	6.36	8.041	11.784	4.734	4.885	4.891	5.19	5.491	5.454
ENT90	6.807	6.736	6.824	2.9	2.927	2.882	3.669	3.66	3.661	3.684	3.75	3.793	3.662	4.072	3.466	3.529	3.539	3.826	4.01	3.979
EtG	3.15	3.118	3.145	2.958	3.015	2.987	4.92	4.911	4.914	4.966	7.559	7.776	4.841	4.75	3.56	3.715	3.719	3.783	3.946	3.922

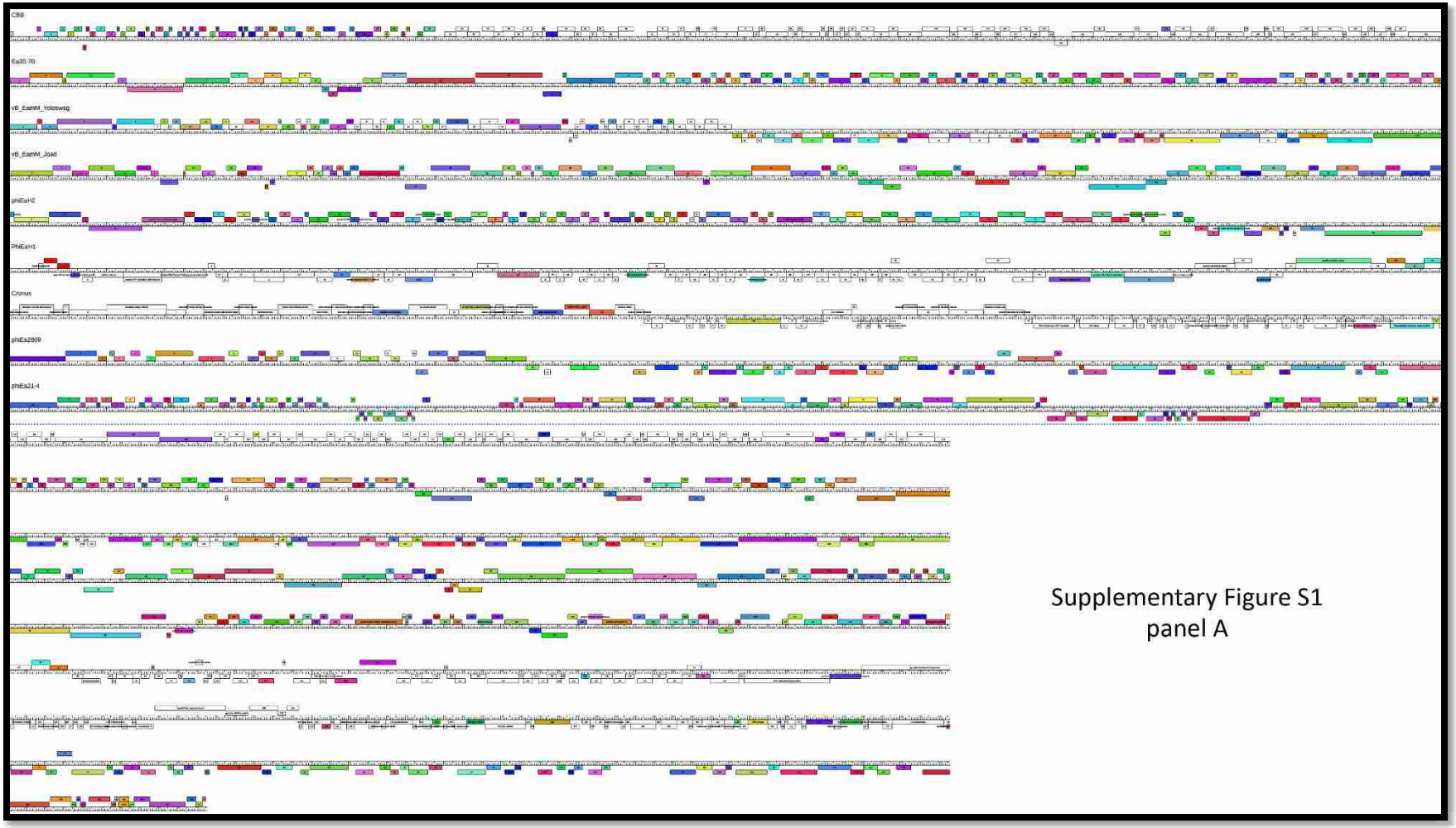
ØEa21-4	phiEa104	M7	SunLiRen	Eamp-S6	Frozen	Rexella	Gutmeister	Ea9-2	Vid5	Pep14	Pavtok	LS-2018a	Faunus	Y2	ØEt88	Era103	phiEa100	EamP-s2	ØEa1H	LIMElight	LIMEzero	FE44	EamP-L1	ENT90	EtG
8.18	8.177	8.241	8.19	7.097	7.194	7.212	6.801	7.207	6.314	5.782	5.872	6.25	5.435	5.642	4.857	4.116	4.129	4.127	4.127	6.49	7.278	4.124	4.173	6.807	3.15
8.117	8.116	8.162	8.124	7.03	7.136	7.155	6.746	7.15	6.261	5.733	5.826	6.196	5.364	5.581	4.828	4.084	4.096	4.104	4.094	6.444	7.198	4.087	4.13	6.736	3.118
8.222	8.219	8.269	8.229	7.079	7.194	7.213	6.805	7.209	6.319	5.794	5.874	6.251	5.433	5.641	4.874	4.122	4.132	4.136	4.131	6.497	7.299	4.134	4.179	6.824	3.145
18.61	18.616	18.685	18.626	7.054	7.204	7.237	6.933	7.259	6.119	5.718	5.792	6.056	5.214	5.423	4.745	6.494	6.535	5.995	6.53	4.33	4.225	4.024	3.986	2.9	2.958
14.414	14.422	14.751	14.447	7.352	7.404	7.436	7.005	7.47	6.011	5.822	5.894	5.985	5.269	5.494	4.643	7.579	7.627	6.87	7.621	4.321	4.167	3.936	3.914	2.927	3.015
12.683	12.688	12.916	12.708	7.227	7.3	7.338	6.907	7.346	6.007	5.888	5.981	6.001	5.124	5.443	4.674	10.212	10.26	8.716	10.255	4.288	4.141	3.9	3.853	2.882	2.987
9.236	9.235	9.519	9.249	8.027	8.075	8.091	7.651	8.105	9.08	6.597	6.642	9.099	6.29	7.843	7.786	4.727	4.743	4.69	4.739	5.31	5.19	6.904	7.808	3.669	4.92
9.224	9.227	9.509	9.239	8.015	8.082	8.096	7.659	8.11	9.054	6.608	6.655	9.064	6.28	7.816	7.754	4.719	4.735	4.69	4.731	5.307	5.173	6.886	7.783	3.66	4.911
9.206	9.204	9.507	9.219	8.016	8.039	8.055	7.616	8.071	9.077	6.601	6.639	9.111	6.281	7.827	7.783	4.724	4.739	4.694	4.736	5.309	5.195	6.91	7.813	3.661	4.914
9.132	9.124	9.312	9.144	7.994	7.999	8.019	7.56	8.048	9.076	6.658	6.719	9.001	6.309	7.838	7.722	4.676	4.708	4.656	4.705	5.338	5.173	6.895	7.779	3.684	4.966
9.093	9.096	9.176	9.074	8.174	8.287	8.318	7.845	8.32	9.212	6.741	6.804	8.974	6.458	10.928	6.779	4.805	4.839	4.769	4.836	5.36	5.276	6.015	6.381	3.75	7.559
9.085	9.08	9.246	9.078	8.363	8.352	8.351	7.91	8.385	9.327	6.919	6.946	9.038	6.545	11.2	6.714	4.821	4.845	4.786	4.842	5.482	5.353	5.958	6.36	3.793	7.776
9.047	9.046	9.271	9.041	7.946	7.978	7.995	7.561	8.011	9.193	6.603	6.676	9.122	6.291	7.728	7.887	4.646	4.667	4.619	4.665	5.238	5.184	7.074	8.041	3.662	4.841
9.054	9.068	10.082	9.027	8.276	8.435	8.5	8.434	8.445	13.209	7.111	7.188	12.643	6.793	7.665	11.563	4.799	4.741	4.773	4.742	6.011	5.965	10.376	11.784	4.072	4.75
8.945	8.95	9.037	8.917	19.079	11.116	11.161	10.518	11.179	7.204	9.356	9.405	7.221	5.805	6.309	5.592	4.833	4.848	4.811	4.843	5.11	4.93	4.702	4.734	3.466	3.56
9.257	9.263	9.327	9.276	8.072	8.225	8.267	7.801	8.259	7.404	6.646	6.736	7.23	13.334	6.678	5.725	4.751	4.753	4.686	4.751	5.221	5.053	4.87	4.885	3.529	3.715
9.307	9.309	9.367	9.326	8.082	8.221	8.263	7.794	8.258	7.393	6.654	6.732	7.228	13.358	6.69	5.726	4.768	4.766	4.706	4.763	5.237	5.07	4.892	4.891	3.539	3.719
9.253	9.244	9.538	9.228	13.015	15.3	15.327	14.533	15.362	8.09	16.983	17.204	7.907	6.403	7.042	6.253	4.917	4.928	4.891	4.927	5.604	5.387	5.378	5.19	3.826	3.783
8.608	8.622	8.916	8.606	13.447	23.794	23.832	22.667	23.965	8.258	13.329	13.468	8.181	6.097	6.703	6.306	4.88	4.896	4.867	4.898	5.887	5.682	5.345	5.491	4.01	3.946
8.49	8.502	8.846	8.496	13.239	23.458	23.498	22.35	23.64	8.172	13.197	13.355	8.097	6.047	6.655	6.254	4.877	4.883	4.862	4.885	5.871	5.658	5.297	5.454	3.979	3.922
	98.31	87.879	97.627	6.036	5.941	5.971	5.877	5.979	5.003	4.989	5.131	5.141	4.674	4.985	4	5.93	5.923	5.673	5.924	4.401	4.303	3.818	3.655	3.075	3.104
98.31		88.385	97.362	6.036	5.954	5.981	5.89	5.992	5.008	5.007	5.154	5.149	4.649	4.99	3.991	5.925	5.919	5.676	5.92	4.38	4.267	3.808	3.659	3.077	3.122
87.879	88.385		87.336	5.784	5.894	5.924	5.825	5.915	6.664	5.135	5.287	6.572	4.706	5.179	5.51	5.807	5.774	5.582	5.775	4.408	4.272	5.022	4.846	3.038	3.275
97.627	97.362	87.336		6.038	5.945	5.975	5.857	5.983	4.984	4.992	5.14	5.131	4.668	4.989	3.986	5.921	5.915	5.686	5.916	4.402	4.294	3.817	3.654	3.076	3.124
6.036	6.036	5.784	6.038		12.466	12.437	12	12.499	6.437	12.2	12.166	6.445	4.897	5.221	5.348	4.45	4.444	4.413	4.447	5.316	5.241	5.017	5.062	4.122	3.619
5.941	5.954	5.894	5.945	12.466		97.287	94.707	93.527	6.675	15.456	15.496	6.754	5.147	5.633	5.554	4.455	4.454	4.418	4.456	5.538	5.309	5.166	5.205	4.159	3.762
5.971	5.981	5.924	5.975	12.437	97.287		92.557	92.689	6.698	15.376	15.401	6.782	5.125	5.582	5.558	4.437	4.434	4.4	4.436	5.514	5.309	5.176	5.197	4.155	3.719
5.877	5.89	5.825	5.857	12	94.707	92.557		88.566	6.915	14.893	14.906	7.016	5.064	5.522	5.775	4.183	4.183	4.129	4.184	5.498	5.257	5.387	5.427	4.15	3.767
5.979	5.992	5.915	5.983	12.499	93.527	92.689	88.566		6.614	15.462	15.489	6.715	5.15	5.621	5.538	4.46	4.459	4.426	4.46	5.485	5.311	5.157	5.171	4.185	3.797
5.003	5.008	6.664	4.984	6.437	6.675	6.698	6.915	6.614		6.235	6.338	21.666	6.111	8.833	17.785	4.437	4.372	4.47	4.374	7.771	7.637	16.414	16.251	5.747	6.699
4.989	5.007	5.135	4.992	12.2	15.456	15.376	14.893	15.462	6.235		78.934	6.254	4.591	5.028	5.335	4.083	4.073	4.036	4.074	5.325	5.52	4.901	5.032	4.418	3.76
5.131	5.154	5.287	5.14	12.166	15.496	15.401	14.906	15.489	6.338	78.934		6.453	4.598	5.065	5.492	4.108	4.091	4.122	4.093	5.412	5.585	4.993	5.083	4.399	3.798
5.141	5.149	6.572	5.131	6.445	6.754	6.782	7.016	6.715	21.666	6.254	6.453		6.373	8.783	17.915	4.348	4.28	4.251	4.277	8.225	7.869	16.678	16.485	6.101	6.648
4.674	4.649	4.706	4.668	4.897	5.147	5.125	5.064	5.15	6.111	4.591	4.598	6.373		51	5.496	3.356	3.376	3.436	3.38	5.381	5.291	5.412	5.296	4.165	3.889
4.985	4.99	5.179	4.989	5.221	5.633	5.582	5.522	5.621	8.833	5.028	5.065	8.783	51		4.848	3.837	3.886	3.885	3.885	5.846	5.554	4.559	4.755	4.542	19.414
4	3.991	5.51	3.986	5.348	5.554	5.558	5.775	5.538	17.785	5.335	5.492	17.915	5.496	4.848		3.926	3.856	3.855	3.858	6.948	7.199	22.357	22.298	5.525	3.543
5.93	5.925	5.807	5.921	4.45	4.455	4.437	4.183	4.46	4.437	4.083	4.108	4.348	3.356	3.837	3.926		97.605	74.519	97.682	3.841	3.619	3.61	3.679	3.113	2.865
5.923	5.919	5.774	5.915	4.444	4.454	4.434	4.183	4.459	4.372	4.073	4.091	4.28	3.376	3.886	3.856	97.605		73.683	99.917	3.825	3.594	3.552	3.613	3.095	2.89
5.673	5.676	5.582	5.686	4.413	4.418	4.4	4.129	4.426	4.47	4.036	4.122	4.251	3.436	3.885	3.855	74.519	73.683		73.741	3.861	3.706	3.628	3.694	3.152	2.846
5.924	5.92	5.775	5.916	4.447	4.456	4.436	4.184	4.46	4.374	4.074	4.093	4.277	3.38	3.885	3.858	97.682	99.917	73.741		3.824	3.594	3.554	3.616	3.097	2.888
4.401	4.38	4.408	4.402	5.316	5.538	5.514	5.498	5.485	7.771	5.325	5.412	8.225	5.381	5.846	6.948	3.841	3.825	3.861	3.824		31.309	6.982	7.033	23.322	4.683
4.303	4.267	4.272	4.294	5.241	5.309	5.309	5.257	5.311	7.637	5.52	5.585	7.869	5.291	5.554	7.199	3.619	3.594	3.706	3.594	31.309		6.982	7.059	26.891	4.628
3.818	3.808	5.022	3.817	5.017	5.166	5.176	5.387	5.157	16.414	4.901	4.993	16.678	5.412	4.559	22.357	3.61	3.552	3.628	3.554	6.982	6.982		41.786	5.705	3.345
3.655	3.659	4.846	3.654	5.062	5.205	5.197	5.427	5.171	16.251	5.032	5.083	16.485	5.296	4.755	22.298	3.679	3.613	3.694	3.616	7.033	7.059	41.786		5.829	3.593
3.075	3.077	3.038	3.076	4.122	4.159	4.155	4.15	4.185	5.747	4.418	4.399	6.101	4.165	4.542	5.525	3.113	3.095	3.152	3.097	23.322	26.891	5.705	5.829		39.8
3.104	3.122	3.275	3.124	3.619	3.762	3.719	3.767	3.797	6.699	3.76	3.798	6.648	3.889	19.414	3.543	2.865	2.89	2.846	2.888	4.683	4.628	3.345	3.593	39.8	

Supplementary Table II.S2 Members of the T4-like cluster of phages that infect the *Enterobacteriales*

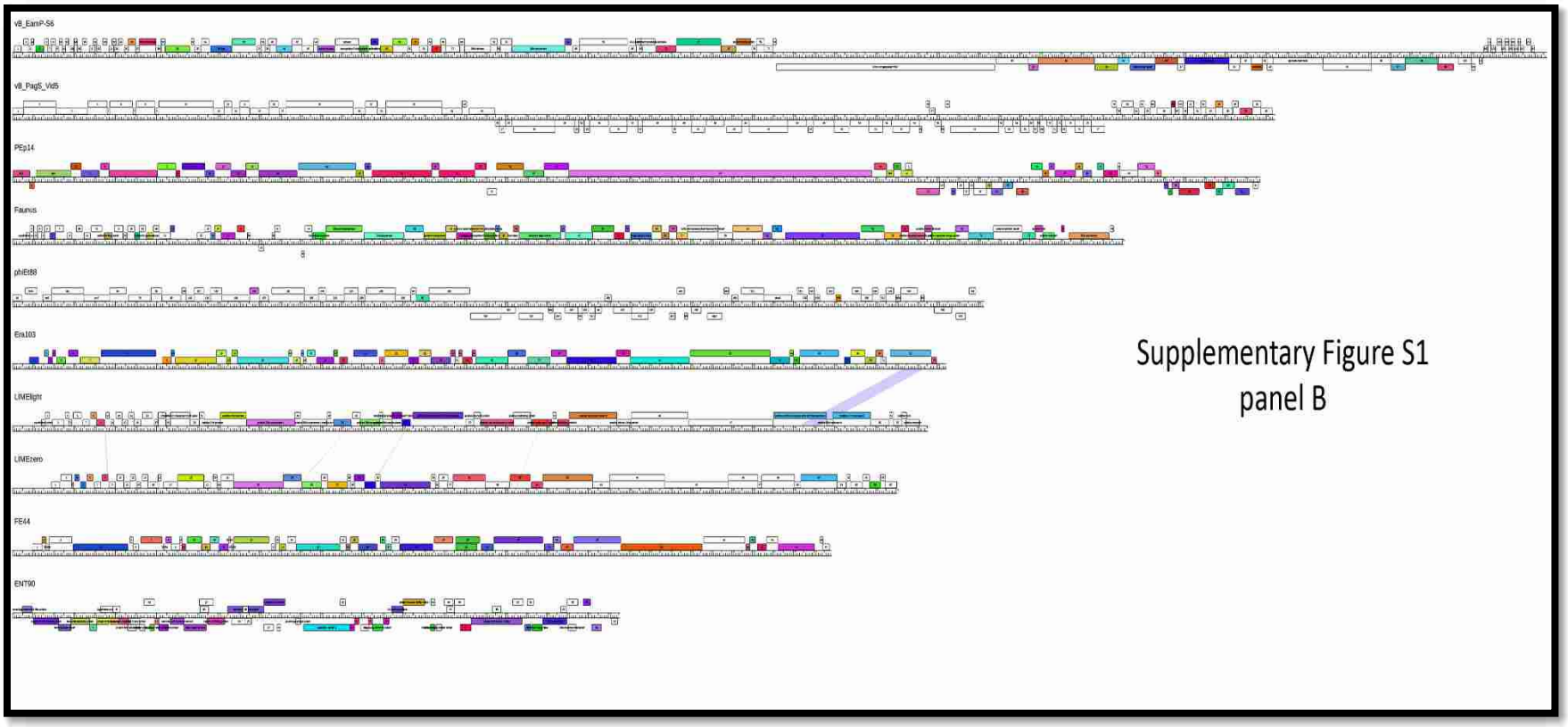
SubCluster	Member phages	Host species	Accession number	Sequence publication	Genome size (bp)	Host family
A	**ACG-C40**	*Escherichia coli*	JN986846	Viruses 4:471	167396	*Enterobacteriaceae*
A	**Apig8**	*Escherichia coli*	KT184308	–	168496	*Enterobacteriaceae*
A	**AR1**	*Escherichia coli*	AP011113	JVirol 85:6567	167435	*Enterobacteriaceae*
A	**CF2**	*Escherichia coli*	KY608967	SciRep7:46151	168188	*Enterobacteriaceae*
A	**CrRp10**	*Citrobacter rodentium*	MG775043	–	171500	*Enterobacteriaceae*
A	**e11/2 (EcoMI_112)**	*Escherichia coli*	NC_024125	–	168470	*Enterobacteriaceae*
A	**EC121**	*Escherichia coli*	MF001359	–	168805	*Enterobacteriaceae*
A	**ECML-134**	*Escherichia coli*	JX128259	–	166783	*Enterobacteriaceae*
A	**ECO4**	*Escherichia coli*	MF001360	–	168638	*Enterobacteriaceae*
A	**fHoEco02**	*Escherichia coli*	MG781191	–	167064	*Enterobacteriaceae*
A	**fHiEco06**	*Escherichia coli*	MG781190	–	167076	*Enterobacteriaceae*
A	**GiZh**	*Escherichia coli*	KT184311	–	166514	*Enterobacteriaceae*
A	**HY01**	*Escherichia coli*	KF925357	PLoS One 11:e0168985	166977	*Enterobacteriaceae*
A	**HY03**	*Escherichia coli*	KR269718	–	170770	*Enterobacteriaceae*
A	**IME09**	*Escherichia coli*	JN202312	–	166499	*Enterobacteriaceae*
A	**IME339**	*Escherichia coli*	MH051915	–	164366	*Enterobacteriaceae*
A	**IME340**	*Escherichia coli*	MH051916	–	165549	*Enterobacteriaceae*
A	**JB75**	*Escherichia coli*	MH355584	–	167208	*Enterobacteriaceae*
A	**Kha5h**	*Escherichia coli*	KT184312	–	167318	*Enterobacteriaceae*
A	**KIT103**	*Escherichia coli*	AP018932	–	166848	*Enterobacteriaceae*
A	**NGB2**	*Escherichia coli*	MH243439	–	166083	*Enterobacteriaceae*
A	**PD112**	*Escherichia coli*	MH837626	–	168084	*Enterobacteriaceae*
A	**PE37**	*Escherichia coli*	KU925172	–	166423	*Enterobacteriaceae*
A	**PEC04**	*Escherichia coli*	KR233165	–	167552	*Enterobacteriaceae*
A	**φD1**	*Yersinia pestis*	HE956711	–	167063	*Yersiniaceae*
A	**PP01**	*Escherichia coli*	LC348379	–	167812	*Enterobacteriaceae*
A	**pSs-1**	*Shigella sonnei*	KM501444	–	164999	*Enterobacteriaceae*
A	**PST**	*Yersinia pestis*	KF208315	–	167785	*Yersiniaceae*
A	**PYPS2T**	*Yersinia pseudotuberculosis*	MH809535	–	169604	*Yersiniaceae*
A	**RB10**	*Escherichia coli*	KM606999	GenomeA 3:e01122-14	168401	*Enterobacteriaceae*
A	**RB14**	*Escherichia coli*	NC_012638	–	165429	*Enterobacteriaceae*
A	**RB18**	*Escherichia coli*	MH553563	–	166677	*Enterobacteriaceae*
A	**RB27**	*Escherichia coli*	KM607000	GenomeA 3:e01122-14	165179	*Enterobacteriaceae*
A	**RB3**	*Escherichia coli*	KM606994	GenomeA 3:e01122-14	168402	*Enterobacteriaceae*
A	**RB32**	*Escherichia coli*	NC_008515	–	165890	*Enterobacteriaceae*
A	**RB33**	*Escherichia coli*	KM607001	GenomeA 3:e01122-14	166007	*Enterobacteriaceae*
A	**RB5**	*Escherichia coli*	KM606995	GenomeA 3:e01122-14	168394	*Enterobacteriaceae*
A	**RB51**	*Escherichia coli*	NC_012635	–	168394	*Enterobacteriaceae*
A	**RB55**	*Escherichia coli*	KM607002	GenomeA 3:e01122-14	168896	*Enterobacteriaceae*
A	**RB59**	*Escherichia coli*	KM607003	GenomeA 3:e01122-14	168966	*Enterobacteriaceae*
A	**RB6**	*Escherichia coli*	KM606996	GenomeA 3:e01122-14	168394	*Enterobacteriaceae*
A	**RB68**	*Escherichia coli*	KM607003	GenomeA 3:e01122-14	168401	*Enterobacteriaceae*
A	**RB7**	*Escherichia coli*	KM606997	GenomeA 3:e01122-14	168395	*Enterobacteriaceae*
A	**RB9**	*Escherichia coli*	KM606998	GenomeA 3:e01122-14	168395	*Enterobacteriaceae*
A	**Sf21**	*Shigella flexneri + E. coli*	MF327007	JVirol 92:e02117-17	166002	*Enterobacteriaceae*
A	**Sf22**	*Shigella sonnei*	MF158045	JVirol 92:e02117-17	166283	*Enterobacteriaceae*
A	**Sf23**	*Shigella boydii*	MF158046	JVirol 92:e02117-17	167678	*Enterobacteriaceae*
A	**Sf24**	*Shigella flexneri*	MF327008	JVirol 92:e02117-17	168112	*Enterobacteriaceae*
A	**Sf25**	*Shigella*	MF327009	JVirol 92:e02117-17	168573	*Enterobacteriaceae*
A	**SG1**	*Salmonella enterica*	MF001354	–	169805	*Enterobacteriaceae*
A	**SH7**	*Shigella sp.*	KX828711	–	164870	*Enterobacteriaceae*
A	**SHBML-50-1**	*Shigella sonnei*	KX130864	–	166634	*Enterobacteriaceae*
A	**Shf2**	*Shigella flexneri*	NC_015457	–	165919	*Enterobacteriaceae*
A	**SHFML-11**	*Shigella sonnei*	KX130861	–	170650	*Enterobacteriaceae*
A	**SHFML-26**	*Shigella sonnei*	KX130862	–	168993	*Enterobacteriaceae*
A	**slur02**	*Escherichia coli*	LN881726	–	167298	*Enterobacteriaceae*
A	**slur03**	*Escherichia coli*	LN881728	–	167467	*Enterobacteriaceae*
A	**slur04**	*Escherichia coli*	LN881729	–	167298	*Enterobacteriaceae*
A	**slur07**	*Escherichia coli*	LN881732	–	167124	*Enterobacteriaceae*
A	**slur08**	*Escherichia coli*	LN881733	–	167467	*Enterobacteriaceae*
A	**slur11**	*Escherichia coli*	LN881734	–	167298	*Enterobacteriaceae*
A	**slur13**	*Escherichia coli*	LN881737	–	167299	*Enterobacteriaceae*
A	**slur14**	*Escherichia coli*	LN881736	–	167467	*Enterobacteriaceae*

SubCluster	Member phages	Host species	Accession number	Sequence publication	Genome size (bp)	Host family
A	T2	*Escherichia coli*	MH751506	–	163832	*Enterobacteriaceae*
A	T4	*Escherichia coli*	AY318471 / AF158101	MMBR 67:87	168903	*Enterobacteriaceae*
A	T6	*Escherichia coli*	MH550421	–	168706	*Enterobacteriaceae*
A	TP7 (7)	*Escherichia coli O157*	KP869105	BMCgenomics 16:271	167747	*Enterobacteriaceae*
A	UFV-AREG1	*Escherichia coli*	KX009778	GenomeA 4:e00412-16	170787	*Enterobacteriaceae*
A	UFV13	*Escherichia coli*	KU867876	Sci Rep. 2018 in press	165772	*Enterobacteriaceae*
A	wV7	*Escherichia coli*	HM997020	JVirol 88:1026	166452	*Enterobacteriaceae*
A	YUEEL01	*Escherichia coli*	KY290975	–	169621	*Enterobacteriaceae*
B	APCEc01	*Escherichia coli*	KR422352	–	168771	*Enterobacteriaceae*
B	ATK47	*Escherichia coli*	KT184309	–	170020	*Enterobacteriaceae*
B	AYK48	*Escherichia coli*	KT184310	–	169729	*Enterobacteriaceae*
B	HP3 {øEC1}	*Escherichia coli*	KY608965	SciRep7:46151	170254	*Enterobacteriaceae*
B	HX01	*Escherichia coli*	JX536493	JVirol 86:13871	161158	*Enterobacteriaceae*
B	JS09	*Escherichia coli*	NC_024124	Intervirol 58:218	169148	*Enterobacteriaceae*
B	NBG1	*Escherichia coli*	MH243438	GenomeA e00586-18	168869	*Enterobacteriaceae*
B	OLB35	*Escherichia coli*	MH992122	–	169140	*Enterobacteriaceae*
B	p000y	*Escherichia coli*	MK047718	–	169872	*Enterobacteriaceae*
B	PhAPEC2	*Escherichia coli*	KF562341	VetMicro171:470	167318	*Enterobacteriaceae*
B	phi25-307	*Shigella sonnei*	MG589383	–	167544	*Enterobacteriaceae*
B	øC120	*Escherichia coli*	KY703222	–	186570	*Enterobacteriaceae*
B	RB69	*Escherichia coli*	NC_004928	VirolJ 3:30	167560	*Enterobacteriaceae*
B	SHBML-52-1	*Shigella sonnei*	KX130865	–	169621	*Enterobacteriaceae*
B	SF	*Escherichia coli*	MH359124	–	168695	*Enterobacteriaceae*
B	Shf125875	*Shigella flexneri*	KM407600		169062	*Enterobacteriaceae*
B	ST0	*Escherichia coli*	MF044457	StdGenSci 12:85	170496	*Enterobacteriaceae*
B	TP13 (13)	*Escherichia coli O157*	KP869111	BMCgenomics 16:271	162417	*Enterobacteriaceae*
B	TP3 (3)	*Escherichia coli O157*	KP869101	BMCgenomics 16:271	168733	*Enterobacteriaceae*
B	TP6 (6)	*Escherichia coli O157*	KP869104	BMCgenomics 16:271	160570	*Enterobacteriaceae*
C	Bp7	*Escherichia coli*	HQ829472	JVirol 86:13832	168066	*Enterobacteriaceae*
C	EcS1	*Escherichia coli*	LC371242	Arcu Virol 2018 in press	175437	*Enterobacteriaceae*
C	IME08	*Escherichia coli*	NC_014260	ArchVirol 156:1489	172253	*Enterobacteriaceae*
C	IME281	*Escherichia coli*	MH051913	–	170531	*Enterobacteriaceae*
C	IME341	*Escherichia coli*	MH051917	–	172379	*Enterobacteriaceae*
C	JS10	*Escherichia coli*	EU863409	Virology 388:21	171451	*Enterobacteriaceae*
C	JS98	*Escherichia coli*	EF469154	JBact 189:8206	170523	*Enterobacteriaceae*
C	MX01	*Escherichia coli*	KU878969		168929	*Enterobacteriaceae*
C	QL01	*Escherichia coli*	KT176190	JBasicMicrobiol 55:1	170527	*Enterobacteriaceae*
C	VR5	*Escherichia coli*	KP007359	AEM in press 2015	170473	*Enterobacteriaceae*
C	WG01	*Escherichia coli*	KU878968	–	169936	*Enterobacteriaceae*
D	SP18	*Shigella sonnei*	GQ981382	JMicrobiol 48:213	170605	*Enterobacteriaceae*
D	VR7	*Escherichia coli*	HM563683	ArchVirol 155:871	169285	*Enterobacteriaceae*
D	VR20	*Escherichia coli*	KP007360	AEM in press 2015	170336	*Enterobacteriaceae*
D	VR25	*Escherichia coli*	KP007361	AEM in press 2015	170822	*Enterobacteriaceae*
D	VR26	*Escherichia coli*	KP007362	AEM in press 2015	171541	*Enterobacteriaceae*
E	CGG4-1	*Salmonella enterica Newport*	NC_031065	–	159878	*Enterobacteriaceae*
E	Melville	*Salmonella enterica Newport*	MF957259	–	159323	*Enterobacteriaceae*
E	S16 (SenMS16)	*Salmonella enterica*	HQ331142	MolMicro 87:818	160221	*Enterobacteriaceae*
E	STML_198	*Salmonella enterica*	JX181825	–	158099	*Enterobacteriaceae*
E	STP4-a	*Salmonella enterica*	KJ000058	–	159914	*Enterobacteriaceae*
F	fHe-Yen9-01	*Yersinia enterocolitica*	KY593455	–	167773	*Yersiniaceae*
F	øR1-RT	*Yersinia enterocolitica*	HE956709	–	168809	*Yersiniaceae*
F	TG1	*Yersinia enterocolitica*	KP202158	–	162101	*Yersiniaceae*
G	Pet-CM3-4	*Cronobacter malonaticus*	LT614807	–	171975	*Enterobacteriaceae*
G	PG7	*Enterobacter cloacae*	KJ101592	–	173276	*Enterobacteriaceae*
G	CC31	*Escherichia coli*	GU323318	–	166540	*Enterobacteriaceae*
G	myPSH1140	*Enterobacter cloacae*	MG999954	–	172614	*Enterobacteriaceae*
H	ECD7	*Escherichia coli*	KY683735	–	164706	*Enterobacteriaceae*
H	GEC-3S	*Escherichia coli*	HE978309	–	163424	*Enterobacteriaceae*
H	JSE	*Escherichia coli*	EU863408	Virology 388:21	166418	*Enterobacteriaceae*
N	KFS-EC	*Escherichia coli*	MH560358	–	164715	*Enterobacteriaceae*
H	ø1	*Escherichia coli*	NC_009821	–	164270	*Enterobacteriaceae*
H	RB49	*Escherichia coli*	NC_005066	VirolJ 3:30	164018	*Enterobacteriaceae*

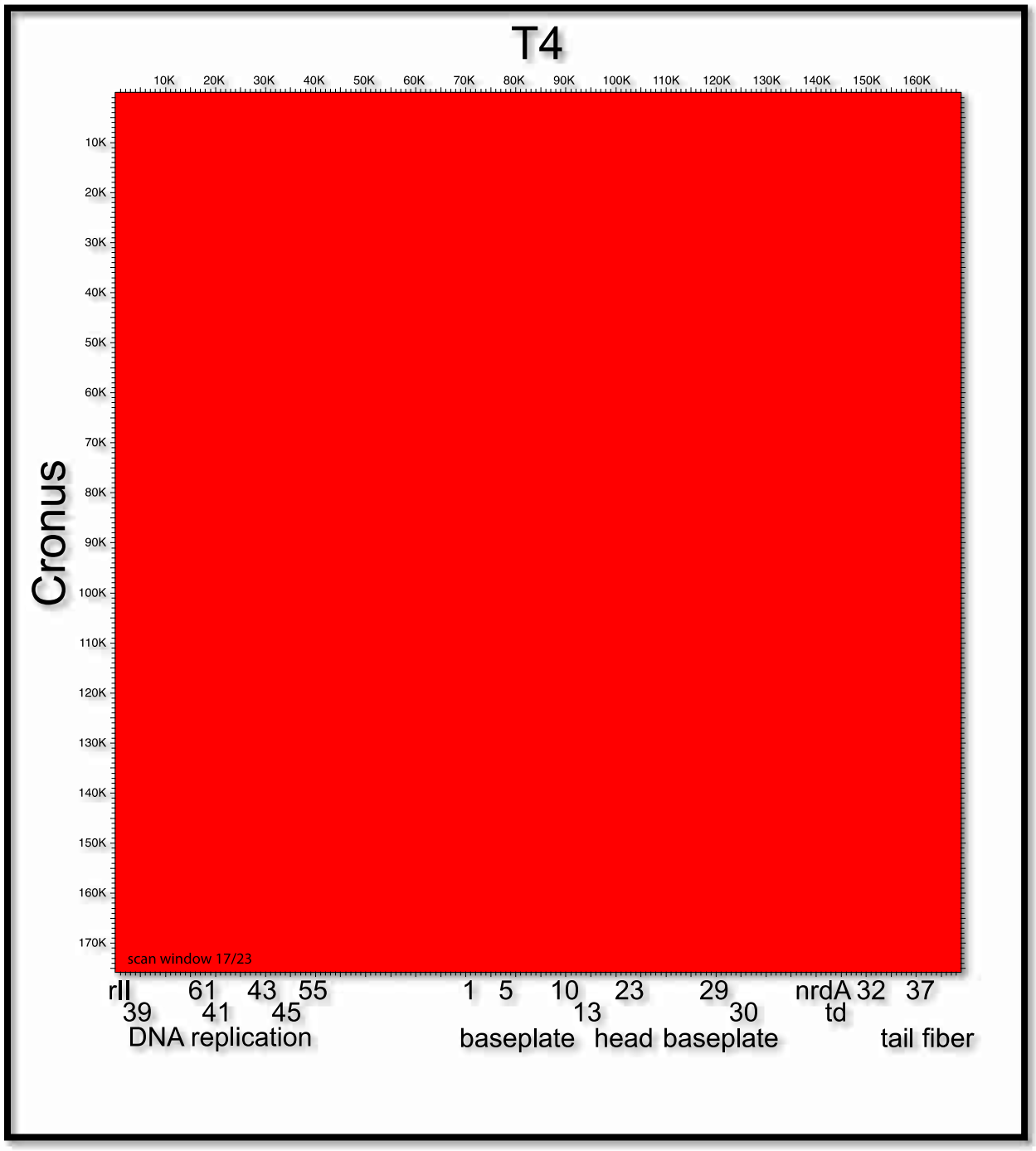
SubCluster	Member phages	Host species	Accession number	Sequence publication	Genome size (bp)	Host family
N	**Sf20**	*Shigella flexneri*	MF327006	JVirol 92:e02117-17	163982	*Enterobacteriaceae*
I	**CfP1**	*Citrobacter freundii*	KX245890	–	180219	*Enterobacteriaceae*
I	**IME-CF2**	*Citrobacter freundii*	KR869820	–	177688	*Enterobacteriaceae*
I	**Margaery**	*Citrobacter freundii*	KT381880	–	178182	*Enterobacteriaceae*
I	**Maroon**	*Citrobacter freundii*	MH823906	–	178830	*Enterobacteriaceae*
I	**Miller**	*Citrobacter freundii*	KM236237	GenomeA 3:.01425-14	178171	*Enterobacteriaceae*
I	**GAP161**	*Cronobacter sakazakii*	JN882287	JVirol 86:13806	178193	*Enterobacteriaceae*
I	**IeB**	*Cronobacter*	KX443552	Int. J. Food Microbiol. 253:1	181570	*Enterobacteriaceae*
I	**IeE**	*Cronobacter*	KX431559	Int. J. Food Microbiol. 253:1	177907	*Enterobacteriaceae*
I	**IeN**	*Cronobacter*	KX431560	Int. J. Food Microbiol. 253:1	179516	*Enterobacteriaceae*
I	**Lw1**	*Escherichia coli*	NC_021344	GenomeA 1:e00743-13	176227	*Enterobacteriaceae*
I	**RB16**	*Escherichia coli*	HM134276	–	176789	*Enterobacteriaceae*
I	**RB43**	*Escherichia coli*	NC_007023	ViroIJ 3:30	180500	*Enterobacteriaceae*
J	**PS2**	*Serratia marcescens*	KJ025957	ArchVirol 2018 in press	167266	*Yersiniaceae*
K	**JD18**	*Klebsiella pneumoniae*	KT239446	–	166313	*Enterobacteriaceae*
K	**KpV477**	*Klebsiella pneumoniae*	KX258185	GenomeA14:e00694-17	168272	*Enterobacteriaceae*
K	**Mineola**	*Klebsiella pneumoniae*	MH333064		166130	*Enterobacteriaceae*
K	**PKO111**	*Klebiella oxytoca*	KR269720	Park ArchVriol in press	168758	*Enterobacteriaceae*
K	**KP1**	*Klebsiella pneumoniae*	MG751100	–	167989	*Enterobacteriaceae*
K	**KPV15**	*Klebsiella pneumoniae*	KY000080	–	167034	*Enterobacteriaceae*
K	**KPV179**	*Klebsiella pneumoniae*	MH729874	–	162630	*Enterobacteriaceae*
L	**MP1**	*Morganella sp.*	KX078569	SciRep 7:46157	163201	*Morganellaceae*
M	**CF1 ERZ-2017**	*Citrobacter freundii*	MG250484	–	171911	*Enterobacteriaceae*
M	**Merlin**	*Citrobacter freundii*	KT001915	GenomeA 3:e01133	172733	*Enterobacteriaceae*
M	**Moon**	*Citrobacter freundii*	KM236240	–	170341	*Enterobacteriaceae*
N	**øEap-3**	*Enterobacter aerogenes*	KT321315	–	175814	*Enterobacteriaceae*
N	**phT4A**	*Escherichia coli*	KX130727	–	171598	*Enterobacteriaceae*
N	**KP15**	*Klebsiella pneumoniae*	GU295964	ViroIJ 10:100	174436	*Enterobacteriaceae*
N	**KP27**	*Klebsiella pneumoniae*	HQ918180	ViroIJ 10:100	174413	*Enterobacteriaceae*
N	**Matisse**	*Klebsiella pneumoniae*	KT001918	–	176081	*Enterobacteriaceae*
N	**Miro**	*Klebsiella pneumoniae*	KT001919	GenomeA 3:e01137	176055	*Enterobacteriaceae*
N	**phT4A**	*Escherichia coli*	KX130727	–	171598	*Enterobacteriaceae*
N	**PMBT1**	*Klebsiella pneumoniae*	LT607758	–	175206	*Enterobacteriaceae*
O	**phiP4-3**	*Proteus penneri*	MG696114	–	167849	*Morganellaceae*
O	**Pm5461**	*Proteus mirabilis*	NC_028762	–	161989	*Morganellaceae*
P	**S13**	*Cronobacter sakazakii*	KC954775	–	182145	*Enterobacteriaceae*
Q	**PEi20**	*Edwardsiella ictaluri*	AP014714	–	177643	*Hafniaceae*
Q	**PEi26**	*Edwardsiella ictaluri*	AP014715	–	177215	*Hafniaceae*
R	**PM2**	*Pectobacterium corotovforum*	KF835987	PlantPathol 31:83	170286	*Pectobacteriaceae*
S	**CBH8**	*Serratia sp. ATCC 39006*	MF036691	–	171175	*Yersiniaceae*
S	**CHI14**	*Serratia sp. ATCC 39006*	MF036690	–	171175	*Yersiniaceae*
S	**X20**	*Serratia sp. ATCC 39006*	MF036692	–	172450	*Yersiniaceae*
T	**F48**	*Klebsiella pneumoniae*	MG746602	Viruses 10:E482	170764	*Enterobacteriaceae*
U	**Cronus**	*Erwinia amylovora*	MH059636	–	175774	*Erwiniaceae*



Supplementary Figure S1
panel A

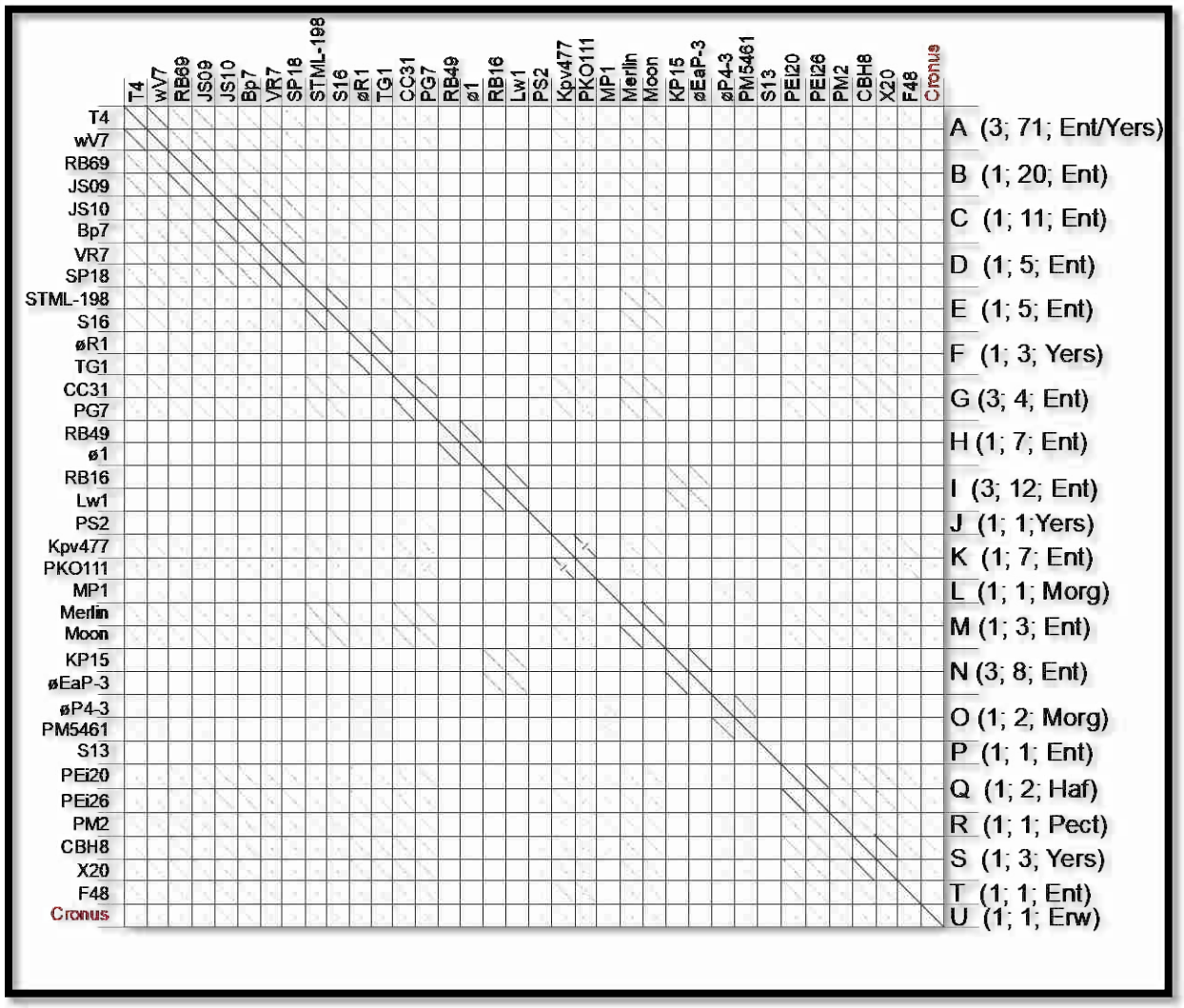


Supplementary Figure II.S1 Phamerator map of the genomes of 19 *Erwiniaceae* clusters using the founding phage of each cluster. Overall genomic and proteomic structure of each phage can be seen, with ittle to no protein and nucleotide homology is seen between clusters. The coloring of the boxes shows homologous proteins, and lines between strands show homology in the nucleotides. Boxes on the top of the ruler are expressed on the forward strand and boxes below the ruler are expressed on the reverse strand. Proteins are labeled with annotated function and proteins with no known function are given by gene number. Nine phages are shown in panel A, with a dotted line indicating where the genomes are continued on a new line due to the large nature of the genomes. The remaining ten phages are shown in panel B.

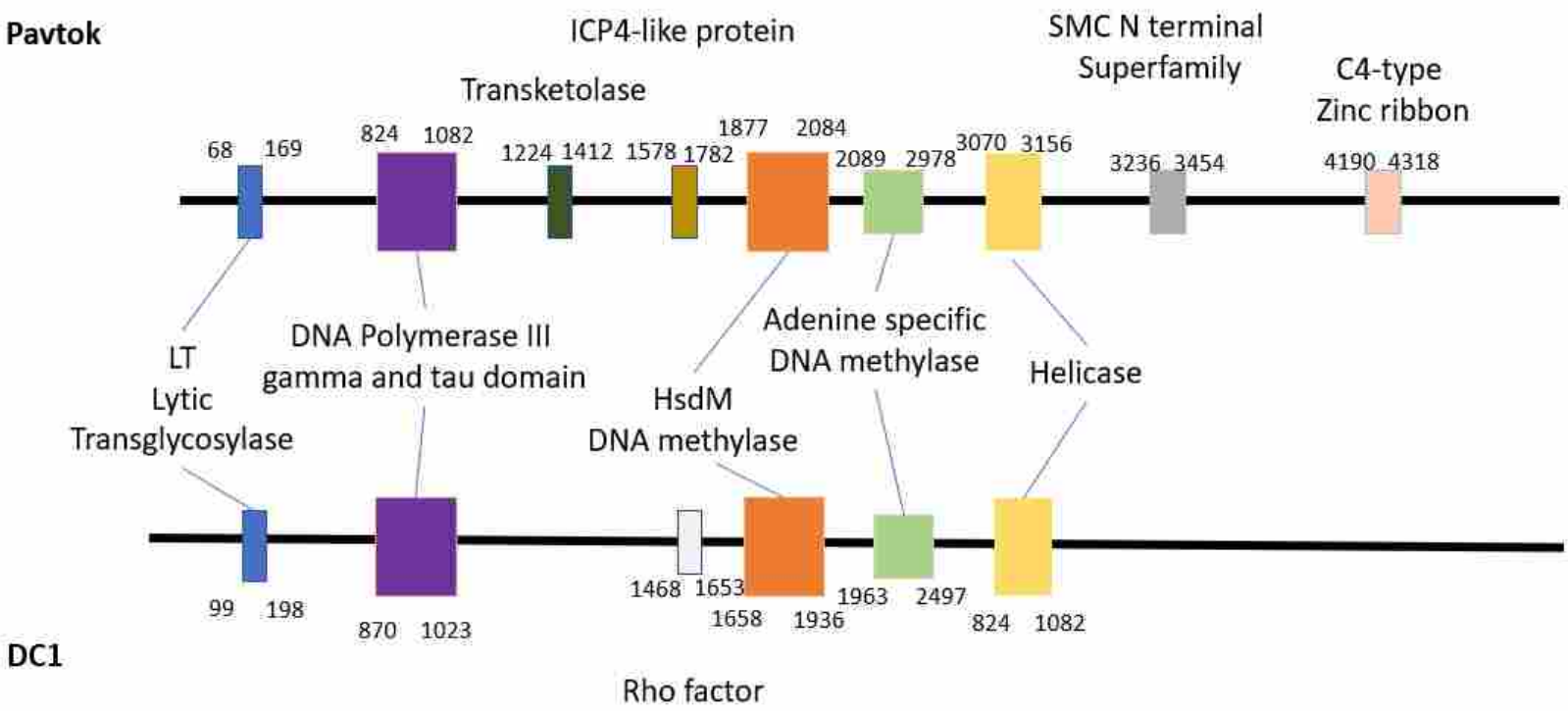


Supplementary Figure II.S2.A Cronus defines a unique subcluster within the *Enterobacteriales* T4-like phage cluster. A. A dot plot comparing phages T4 and Cronus created by DNA Strider (1). The Cronus and T4 genomes show a typical mosaic relationship; the Cronus regions between 35 and 75 kb and between 105 and 118 kb are particularly different from the parallel T4 regions. The most closely related regions encode head and tail virion assembly proteins. T4 genes and genome regions are indicated

below.



Supplementary Figure II.S2.B Cronus defines a unique subcluster within the *Enterobacteriales* T4-like phage cluster. The whole genome dot plot of selected members of all 20 subclusters was created by Gephard (*2*). Two phage genomes were chosen a random from each of the subclusters with more than one member. Phage names are shown at the left and top; on the right the letters are subcluster names and in parentheses the "number of host species known to be infected; number of phages in the subcluster; families infected" are shown (assuming that *Escherichia* and *Shigella* are actually one genus; (*3*)). The family names are abbreviated as follows: Ent, *Enterobacteriaceae*; Yers*, Yersiniaceae*; Morg, *Morganellaceae*; Haf, *Hafniaceae*; Pect, *Pectobacteriaceae*; Erw, *Erwinaceae* (according to ref (*4*))



Supplementary Figure II.S3 Comparison of Pavtok and DC1, two novel, large proteins and their predicted domains. Boxes represent the domains with the amino acid start and stop that correspond. Protein function is included, instances where both genes contained the same domain the function is placed in between the genes. Instances where only one gene had a domain the function is written on the outside of the genes. Distances are not exact and only show gene order not necessarily distance between domains.

II.8.1 References for supplementary material

1. S. E. Douglas, DNA Strider. An inexpensive sequence analysis package for the Macintosh. *Mol Biotechnol* 3, 37-45 (1995).

2. J. Krumsiek, R. Arnold, T. Rattei, Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 23, 1026-1028 (2007).

3. G. Zuo, Z. Xu, B. Hao, Shigella strains are not clones of Escherichia coli but sister species in the genus Escherichia. *Genomics Proteomics Bioinformatics* 11, 61-65 (2013).

4. M. Adeolu, S. Alnajar, S. Naushad, S. G. R, Genome-based phylogeny and taxonomy of the 'Enterobacteriales': proposal for Enterobacterales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morganellaceae fam. nov., and Budviciaceae fam. nov. *Int J Syst Evol Microbiol* 66, 5575-5599 (2016).

5. I. N. D. Esplin *et al.*, Genome Sequences of 19 Novel Erwinia amylovora Bacteriophages. *Genome Announc* 5, (2017).

6. M. Kearse *et al.*, Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647-1649 (2012).

REFERENCES

1. W. Fiers *et al.*, Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* 260, 500-507 (1976).

2. G. F. Hatfull, R. W. Hendrix, Bacteriophages and their genomes. *Curr Opin Virol* 1, 298-303 (2011).

3. C. A. Suttle, Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* 5, 801-812 (2007).

4. K. C.I., *Erwinia and Related Genera*. F. S. Dworkin M., Rosenberg E., Schleifer KH., Stackebrandt E., Ed., The Prokaryotes (Springer, New York, NY, 2006).

5. M. Sengupta, S. Banerjee, N. K. Das, P. Guchhait, S. Misra, Early Onset Neonatal Septicaemia Caused by Pantoea agglomerans. *J Clin Diagn Res* 10, DD01-02 (2016).

6. J. L. Norelli, A. L. Jones, H. S. Aldwinckle, Fire Blight Management in the Twenty-first Century: Using New Technologies that Enhance Host Resistance in Apple. *Plant Dis* 87, 756-765 (2003).

7. H. Forster, G. C. McGhee, G. W. Sundin, J. E. Adaskaveg, Characterization of Streptomycin Resistance in Isolates of Erwinia amylovora in California. *Phytopathology* 105, 1302-1310 (2015).

8. I. N. D. Esplin *et al.*, Genome Sequences of 19 Novel Erwinia amylovora Bacteriophages. *Genome Announc* 5, (2017).

9. R. Sharma *et al.*, Genome Sequences of Nine Erwinia amylovora Bacteriophages. *Microbiology Resource Announcements* 7, (2018).

10. NC_015249.

11. J. Krumsiek, R. Arnold, T. Rattei, Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 23, 1026-1028 (2007).

12. M. Kearse *et al.*, Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647-1649 (2012).

13. S. G. Cresawn *et al.*, Phamerator: a bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinformatics* 12, 395 (2011).

14. A. W. Dress, D. H. Huson, Constructing splits graphs. *IEEE/ACM Trans Comput Biol Bioinform* 1, 109-115 (2004).

15. M. Adeolu, S. Alnajar, S. Naushad, S. G. R, Genome-based phylogeny and taxonomy of the 'Enterobacteriales': proposal for Enterobacterales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae

fam. nov., Hafniaceae fam. nov., Morganellaceae fam. nov., and Budviciaceae fam. nov. *Int J Syst Evol Microbiol* 66, 5575-5599 (2016).

16. D. K. Arens *et al.*, Characterization of two related Erwinia myoviruses that are distant relatives of the ØKZ-like Jumbo phages. *PLoS One* 13, e0200202 (2018).

17. C. Buttimer *et al.*, Erwinia amylovora phage vB_EamM_Y3 represents another lineage of hairy Myoviridae. *Res Microbiol*, (2018).

18. A. L. Lagonenko, O. Sadovskaya, L. N. Valentovich, A. N. Evtushenkov, Characterization of a new ViI-like Erwinia amylovora bacteriophage ØEa2809. *FEMS Microbiol Lett* 362, (2015).

19. C. Buttimer *et al.*, Things Are Getting Hairy: Enterobacteria Bacteriophage vB_PcaM_CBB. *Front Microbiol* 8, 44 (2017).

20. S. R. Casjens, J. H. Grose, Contributions of P2- and P22-like prophages to understanding the enormous diversity and abundance of tailed bacteriophages. *Virology* 496, 255-276 (2016).

21. J. H. Grose, S. R. Casjens, Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family Enterobacteriaceae. *Virology* 468-470, 421-443 (2014).

22. G. F. Hatfull *et al.*, Comparative genomic analysis of 60 Mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. *J Mol Biol* 397, 119-143 (2010).

23. D. H. Huson, D. Bryant, Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23, 254-267 (2006).

24. K. C. Smith *et al.*, Phage cluster relationships identified through single gene analysis. *BMC Genomics* 14, 410 (2013).

25. I. Bahir, M. Fromer, Y. Prat, M. Linial, Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol Syst Biol* 5, 311 (2009).

26. K. Limor-Waisberg, A. Carmi, A. Scherz, Y. Pilpel, I. Furman, Specialization versus adaptation: two strategies employed by cyanophages to enhance their translation efficiencies. *Nucleic Acids Res* 39, 6016-6028 (2011).

27. M. Halpern, S. Fridman, N. Atamna-Ismaeel, I. Izhaki, Rosenbergiella nectarea gen. nov., sp. nov., in the family Enterobacteriaceae, isolated from floral nectar. *Int J Syst Evol Microbiol* 63, 4259-4265 (2013).

28. R. A. King *et al.*, Newly discovered antiterminator RNAs in bacteriophage. *J Bacteriol* 193, 5784-5792 (2011).

29. J. J. Gill *et al.*, Genomes and characterization of phages Bcep22 and BcepIL02, founders of a novel phage type in Burkholderia cenocepacia. *J Bacteriol* 193, 5300-5313 (2011).

30. B. Ali *et al.*, 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 (2017).

31. J. A. Thomas *et al.*, Identification of Essential Genes in the Salmonella Phage SPN3US Reveals Novel Insights into Giant Phage Head Structure and Assembly. *J Virol* 90, 10284-10298 (2016).

32. K. Meczker *et al.*, The genome of the Erwinia amylovora phage PhiEaH1 reveals greater diversity and broadens the applicability of phages for the treatment of fire blight. *FEMS Microbiol Lett* 350, 25-27 (2014).

33. J. H. Grose, S. R. Casjens, Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family Enterobacteriaceae. *Virology* 468-470, 421-443 (2014).

34. G. Zuo, Z. Xu, B. Hao, Shigella strains are not clones of Escherichia coli but sister species in the genus Escherichia. *Genomics Proteomics Bioinformatics* 11, 61-65 (2013).

35. S. Iida, M. B. Streiff, T. A. Bickle, W. Arber, Two DNA antirestriction systems of bacteriophage P1, darA, and darB: characterization of darA- phages. *Virology* 157, 156-166 (1987).

36. E. M. Kutter *et al.*, Characterization of a ViI-like phage specific to Escherichia coli O157:H7. *Virol J* 8, 430 (2011).

37. E. M. Adriaenssens *et al.*, A suggested new bacteriophage genus: "Viunalikevirus". *Arch Virol* 157, 2035-2046 (2012).

38. C. Chen *et al.*, Crystal structure of ORF210 from E. coli O157:H1 phage CBA120 (TSP1), a putative tailspike protein. *PLoS One* 9, e93156 (2014).

39. S. R. Casjens, D. Jacobs-Sera, G. F. Hatfull, R. W. Hendrix, Genome Sequence of Salmonella enterica Phage Det7. *Genome Announc* 3, (2015).

40. D. J. Magill *et al.*, Pf16 and ØPMW: Expanding the realm of Pseudomonas putida bacteriophages. *PLoS One* 12, e0184307 (2017).

41. S. M. Lehman, A. M. Kropinski, A. J. Castle, A. M. Svircev, Complete genome of the broad-host-range Erwinia amylovora phage ØEa21-4 and its relationship to Salmonella phage felix O1. *Appl Environ Microbiol* 75, 2139-2147 (2009).

42. J. Park, G. M. Lee, D. Kim, D. H. Park, C. S. Oh, Characterization of the Lytic Bacteriophage ØEaP-8 Effective against Both Erwinia amylovora and Erwinia pyrifoliae Causing Severe Diseases in Apple and Pear. *Plant Pathol J* 34, 445-450 (2018).

43. J. Wittmann *et al.*, First genome sequences of Achromobacter phages reveal new members of the N4 family. *Virol J* 11, 14 (2014).

44. S. C. Falco, K. V. Laan, L. B. Rothman-Denes, Virion-associated RNA polymerase required for bacteriophage N4 development. *Proc Natl Acad Sci U S A* 74, 520-523 (1977).

45. E. Simoliunas *et al.*, Pantoea Bacteriophage vB_PagS_Vid5: A Low-Temperature Siphovirus That Harbors a Cluster of Genes Involved in the Biosynthesis of Archaeosine. *Viruses* 10, (2018).

46. E. E. Kulikov *et al.*, Genomic sequencing and biological characteristics of a novel Escherichia coli bacteriophage 9g, a putative representative of a new Siphoviridae genus. *Viruses* 6, 5077-5092 (2014).

47. P. Sazinas *et al.*, Comparative Genomics of Bacteriophage of the Genus Seuratvirus. *Genome Biol Evol* 10, 72-76 (2018).

48. K. M. Kazmierczak, E. K. Davydova, A. A. Mustaev, L. B. Rothman-Denes, The phage N4 virion RNA polymerase catalytic domain is related to single-subunit RNA polymerases. *EMBO J* 21, 5815-5823 (2002).

49. S. R. Casjens *et al.*, The pKO2 linear plasmid prophage of Klebsiella oxytoca. *J Bacteriol* 186, 1818-1832 (2004).

50. P. Kemp, L. R. Garcia, I. J. Molineux, Changes in bacteriophage T7 virion structure at the initiation of infection. *Virology* 340, 307-317 (2005).

51. C. Y. Chang, P. Kemp, I. J. Molineux, Gp15 and gp16 cooperate in translocating bacteriophage T7 DNA into the infected cell. *Virology* 398, 176-186 (2010).

52. S. R. Casjens, I. J. Molineux, Short noncontractile tail machines: adsorption and DNA delivery by podoviruses. *Adv Exp Med Biol* 726, 143-179 (2012).

53. M. Moak, I. J. Molineux, Peptidoglycan hydrolytic activities associated with bacteriophage virions. *Mol Microbiol* 51, 1169-1183 (2004).

54. M. Moak, I. J. Molineux, Role of the Gp16 lytic transglycosylase motif in bacteriophage T7 virions at the initiation of infection. *Mol Microbiol* 37, 345-355 (2000).

55. D. Piya, L. Vara, W. K. Russell, R. Young, J. J. Gill, The multicomponent antirestriction system of phage P1 is linked to capsid morphogenesis. *Mol Microbiol* 105, 399-412 (2017).

56. N. Jamalludeen *et al.*, Complete Genomic Sequence of Bacteriophage φEcoM-GJ1, a Novel Phage That Has Myovirus Morphology and a Podovirus-Like RNA Polymerase. *Applied and Environmental Microbiology* 74, 516-525 (2008).

57. M. Kalischuk, J. Hachey, L. Kawchuk, Complete Genome Sequence of Phytopathogenic Pectobacterium atrosepticum Bacteriophage Peat1. *Genome Announc* 3, (2015).

58. Y. Born, L. Bosshard, B. Duffy, M. J. Loessner, L. Fieseler, Protection of Erwinia amylovora bacteriophage Y2 from UV-induced damage by natural compounds. *Bacteriophage* 5, e1074330 (2015).

59. M. Demerec, U. Fano, Bacteriophage-Resistant Mutants in Escherichia Coli. *Genetics* 30, 119-136 (1945).

60. I. Muller, M. Kube, R. Reinhardt, W. Jelkmann, K. Geider, Complete genome sequences of three Erwinia amylovora phages isolated in north america and a bacteriophage induced from an Erwinia tasmaniensis strain. *J Bacteriol* 193, 795-796 (2011).

61. B. Xu, X. Ma, H. Xiong, Y. Li, Complete genome sequence of 285P, a novel T7-like polyvalent E. coli bacteriophage. *Virus Genes* 48, 528-533 (2014).

62. J. Michalewicz, E. Hsu, J. J. Larson, A. W. Nicholson, Physical map and genetic early region of the T7-related coliphage, BA14. *Gene* 98, 89-93 (1991).

63. B. Kalionis, I. B. Dodd, J. B. Egan, Control of gene expression in the P2-related template coliphages. III. DNA sequence of the major control region of phage 186. *J Mol Biol* 191, 199-209 (1986).

64. A. I. Yagubi, A. J. Castle, A. M. Kropinski, T. W. Banks, A. M. Svircev, Complete Genome Sequence of Erwinia amylovora Bacteriophage vB_EamM_Ea35-70. *Genome Announcements* 2, (2014).

65. A. Day, J. Ahn, G. P. C. Salmond, Jumbo Bacteriophages Are Represented Within an Increasing Diversity of Environmental Viruses Infecting the Emerging Phytopathogen, Dickeya solani. *Front Microbiol* 9, 2169 (2018).

66. V. V. Kutyrev *et al.*, Phylogeny and Classification of Yersinia pestis Through the Lens of Strains From the Plague Foci of Commonwealth of Independent States. *Front Microbiol* 9, 1106 (2018).

67. C. Stamereilers *et al.*, Genomic Analysis of 48 Paenibacillus larvae Bacteriophages. *Viruses* 10, (2018).

APPENDIX III: Presentations

1. **Sharma R** and Grose JH. Understanding the relationship between bacteriophages of *Enterobacteriaceae* and *Pseudomonadaceae* family. ASM tri-branch meeting, April 2015, Durango CO.

2. **Sharma R**, Mooni HI, Putnam MJ and Grose JH. Genomic characterization and comparison of five different phamilies of bacteriophages infecting *Erwinia amylovora*. Phage conference, April 2016, Provo UT.

3. Colby BA, Stubbs OA, Bell KA, Rader KA, **Sharma R**, Duncan S and Grose JH. Analysis of interesting proteins in Deimos-Minion bacteriophage family. ASM tri-branch meeting, April 2017, Weber UT.

4. Cardinal J, Gille J, Kyle Ke, Salazar EG, **Sharma R** and Grose JH. Discovery of Likely Transcriptional Regulons and Hypothesized Protein Function in Phage RAY of the Deimos-Minion Family through Motif Analysis. ASM tri-branch meeting, April 2017, Weber UT.

5. Hughes JF, Loertscher E, **Sharma R**, Duncan S and Grose JH. Genome Comparison of Five *Erwinia amylovora* Bacteriophages. ASM tri-branch meeting, April 2017, Weber UT.

6. Judge L, Harley K, **Sharma R**, Duncan S, and Grose JH. Comparative Genomics of Four *Erwinia* Bacteriophages and N4, a Pathogenic Driving Force in E. coli. ASM tri-branch meeting, April 2017, Weber UT.

7. Ballard T, Withers J, **Sharma R**, Duncan S and Grose JH. Dots, Dots, Lines: A Dot Plot Comparison of the *Erwinia* Phage Frozen. ASM tri-branch meeting, April 2017, Weber UT.

8. Nieman T, Yeates E, Hovenden T, **Sharma R**, and Grose JH. Phinding Phamily for Phage Deimos-Minion: A Phylogenetics Study. ASM tri-branch meeting, April 2017, Weber UT.

9. **Sharma R** and Grose JH. Deimos-Minion: A Phage So Big it is Hard to See. ASM tri-branch meeting, April 2017, Weber UT.

10. Melhado E, Sarabia R, Loertscher E, **Sharma R**, Hope S, Breakwell DP and Grose JH. Bacteriophage diversity revelaed by nine novel *Enterobacteriaceae* phages isolated from sewage samples. April 2018, Durango, CO.

11. Yeates, EL, Nieman TB, **Sharma R**, and Grose JH. Comparison of three new bacteriophage phamilies infecting *Erwinia amylovora*. April 2018, Durango, CO\\

12. Carr E, Melhado E, Loertscher E, Thurgood TL, **Sharma R**, and Grose JH. Discovery of geographical gene variants in related *Pseudomonas aeruginosa* bacteriophages. ASM tri-branch meeting, April 2019, Provo UT.