A 2024 Report on the Characteristics of Phages Assigned to Subcluster B4

Authors: Sean P. McClory[1], Kristen Clermont[1], Mary A. Ayuk[2], JoAnn Whitefleet-Smith[3], Bridgette Kirkpatrick[4], Sarah Reardon[5]

Reviewing Author: JoAnn Whitefleet-Smith[3]

Author Affiliations: [1]La Salle University, Philadelphia, PA; [2]Howard University, Washington, DC; [3]Worcester Polytechnic Institute, Worcester, MA; [4]Collin College, McKinney, TX; [5]Culver-Stockton College, Canton, MA

Isolation and Morphological Characteristics

As of July 2024, 34 phages have been assigned to Actinobacteriophage subcluster B4. Subcluster B4 consists of lytic phages isolated from environmental samples collected in 13 different US states and Taiwan, using *Mycobacterium smegmatis* mc^{2}155 as the host. A sampling of transmission electron micrographs reveals these phages to have siphoviral morphologies, while photographs of plaque assays reveal small- to medium-sized clear plaques. Examples of phages displaying these morphologies include phage Lolalove and Waleliano.

Genometrics

The genomes of phages in this subcluster have an average length of 70,904 bp (range: 69,258 bp – 72,139 bp) with an average GC content of 68.9% (range: 68.3% – 69.1%). The genomes have circularly permuted ends.

Gene Content

a. A survey of all manually annotated genomes in this subcluster (i.e., excluding draft genomes) reveals the number of genes called per genome to range between 91 – 100 genes, which represent a total of 159 gene phamilies (or phams).

b. Of these 159 phams, 51 are fully conserved across the subcluster. These phams are assigned putative functions that are involved in structure and assembly (e.g., portal protein, terminase, capsid maturation protease, major capsid hexamer protein, major capsid pentamer protein, major tail protein, head-to-tail adaptor, tape measure, and minor tail proteins), replication and recombination (e.g., exonuclease, DNA helicase, DNA primase/polymerase/helicase, and DNA polymerase I), and lysis (e.g., lysin A and lysin B). Some other notable putative functions include adenylate kinase, RuvC-like Resolvase, and toxin/antitoxin system, hicA-like. Notably, one pham appears twice within each individual genome, exemplified by Waleliano_33 and Waleliano_34.

c. Another 108 phams have varying levels of conservation. Notable here are 20 phams that are each only called once in this subcluster. One of these phams is assigned the function G-I-Y Y-I-G endonuclease.

d. Only one phage in subcluster B4 identified a tRNA gene (tRNA-Ile). There are also no identifiable integrase or repressor functions, suggesting that phages of this subcluster do not form lysogens and are, thus, lytic phages.

e. The genomes are organized with genes involved in structure and assembly located within the first half of the genome, while those involved in replication are scattered across the remaining half of the genome. About half of the genes are transcribed in the forward direction, with the remaining half being transcribed in the reserve direction.

f. The lysin genes are located at the mid-point of these genomes, but the putative holin genes are distally located between the capsid and tail structural genes. While holin genes are typically located proximally to the lysin genes in most phages, the unusual organization found in subcluster B4 phages has been described in other actinobacteriophages (Pollenz et al., 2022). Two genes downstream of the holin gene have putative transmembrane domains and may play a role in forming the holin complex, but we do not yet have enough evidence to assign the function, holin, to these putative genes.

g. Tail-assembly chaperone genes have not been confidently identified in these phages. While a gene upstream of the tape measure protein may have similarities to known tail-assembly chaperones, several differences make us uncertain of making a function assignment. Until new evidence is uncovered, it is recommended not to assign the tail assembly chaperone function.

h. 33 of the 36 finished genome sequences have a presumptive thymine hyper modification system as the first 4 genes (21 genomes) or 3 genes (12 genomes). The terminase gene immediately follows these genes. Current function assignments are as follows for Prince: gene 1 ParB-like nuclease domain, gene 2 thymidine hydroxylase, gene 3 DNA glycosylase, gene 4 adenylate kinase. Genomes with 3 genes preceding terminase, e.g., Stinger, are missing the DNA glycosylase. These gene assignments are currently based on sequence similarity. Function assignment has not yet been confirmed with wet lab data.

References

1. Pollenz RS, Bland J, Pope WH (2022) Bioinformatic characterization of endolysins and holin-like membrane proteins in the lysis cassette of phages that infect *Gordonia rubripertincta*. PLOS ONE 17(11): e0276603. https://doi.org/10.1371/journal.pone.0276603

Source Data: https://observablehq.com/d/5e5bc78c9b3ae2ed accessed July 2024.

Acknowledgments

We thank Maddie Bendele, Isobel Cobb, and Steve Cresawn for developing the tools used to generate this report.