Cost Effective High-Throughput Identification of Mycobacteriophages for Large-Scale Comparative Genomic Studies

Lauren A. Esposito[1], Swati Gupta[1], Lauren Mordukhaev[1], Brian E. Ford[1,2], and John J. Dennehy[1,2]

Biology Department, Queens College of the City University of New York, Flushing, NY[1]

Biology Department, The Graduate Center of the City University of New York, New York, NY

Abstract

Over 464 mycobacteriophages have been fully sequenced and assigned to one of 44 clusters and subclusters. However, over 3,400 collected phages remain unidentified. A strategy is needed to identify unknown mycobacteriophage prior to genomic sequencing. We identified mycobacteriophages in a panel of 15 unknown samples through the analysis of : phage morphology using transmission electron microscopy (TEM); and gel electrophoresis banding patterns of polymerase chain reaction (PCR) amplified phage DNA. PCR amplification was performed using a panel of oligonucleotide primers designed by researchers at University of Pittsburgh, University of Alabama at Birmingham, and CUNY Queens College. While few oligonucleotide primers were specific to a single phage cluster, most bound to a restricted subset of all phage clusters. By analyzing the patterns produced by multiple PCR reactions, we were able to use a process of elimination to identify the cluster assignment of these mycobacteriophages. This approach allows for cluster identification using only a set of 15 oligonucleotide primers, prior to genome sequencing.

Introduction

Mycobacteriophages are viruses that infect members of the *Mycobacterium* genus, including *M.smegmatis* and *M.tuberculosis*. Although mycobacteriophages have limited host ranges, they are highly genetically diverse, perhaps due to extensive gene shuffling among the phages in this group. The known phages are organized into clusters based upon evident genomic similarities. The original goal was to design primers that were specific to each phage cluster, but through subsequent PCR reactions and gel electrophoresis analysis, the primers found not to be specific to their cluster. A subset of 15 primers were selected and their binding affinities were analyzed to determine a characteristic pattern for each of the mycobacteriophage clusters.

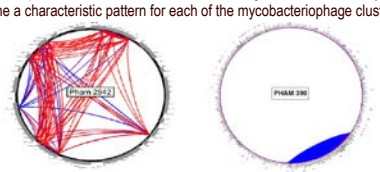

Fig. 1 - Examples of two representative Phams. A selected 80 genomes are shown on the circumference of each circle – arranged by cluster – with arcs indicating pairs of genomes containing a Pham member; thicker arcs indicate closer similarity. Red and blue arcs show BlastP and ClustalW comparisons respectively (from Pope et al. 2011).

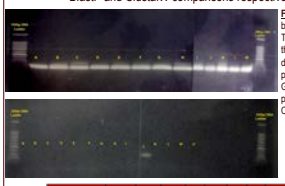

Fig. 2. Oligonucleotide primers were designed to be specific for each mycobacteriophage cluster. Through PCR and gel electrophoresis reactions, the specificity of each primer to its cluster was determined. (Top) Primers designed for G-Cluster phages were determined not to be specific to the G-Cluster. (Bottom) Primers designed for J-Cluster phages were determined to be specific to the J-Cluster.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A cluster	x						x		x	x	x				x
B Cluster	x						x		x	x	x				x
C Cluster	x	x			x		x		x	x	x				x
D Cluster	x		x	x	x	x	x	x	x	x	x		x		
E Cluster	x				x	x	x		x		x		x		x
F Cluster	x						x		x	x	x			x	x
G Cluster	x								x				x		
H Cluster	x			x			x		x		x		x		x
I Cluster	x						x	x	x	x	x				
J Cluster	x	x			x	x	x		x	x	x	x	x		x
K Cluster							x		x	x	x		x		x
L Cluster							x	x	x	x	x		x		x
M Cluster	x				x		x	x	x	x	x		x	x	
N Cluster	x														
O Cluster							x		x	x	x		x		x
P Cluster													x		

Table 1. The binding affinity of a subset of 15 oligonucleotide primers was used to construct a pattern for the identification of each cluster of mycobacteriophage. The horizontal axis represents the subset of 15 primers that were positive for this cluster. The vertical axis represents each of the phage clusters available at Queens College.

Identifying Unknowns by PCR Analysis

Table 2. The binding affinities of the subset of 15 oligonucleotide primers were analyzed to determine a pattern representative of each of the mycobacteriophage clusters. For this approach, the unknown phages were tested with each of the 15 primers. The patterns for clusters C through P are listed below.

Cluster Determination Through PCR Analysis	
C Cluster	**Test with primer 2**: if the DNA binds to it then it is a **C cluster phage**
D Cluster	**Test with primer 3**: if the DNA binds to it then it is a **D cluster phage**
E Cluster	**Test with primers 6, 3 & 12:** If DNA binds to primer 6 but not the primers 3 or 12, then it is an **E cluster phage**
F Cluster	**Test with primers 6 &14:** if the DNA binds to primer 14 but not to primer 6, then it then it is a **M cluster phage**
G Cluster	**Test with primers 9, 11, 7:** if DNA only binds to primer 9 but not primers 7 or 11, then it is a **G cluster phage**
H Cluster	**Test with primers 3 & 4**: If the DNA binds to primer 4 but not to 3, then it is an **H cluster Phage**
I Cluster	**Test with primers 7, 10, 11, & 13**: If the DNA binds to primers 7, 10, and 11 and not to primer 13 then it is an **I cluster phage**
J Cluster	**Test with primer 12**: if DNA binds to it then it is a **J cluster phage**
K Cluster	**Test with primers 7, 8, 10, 11, 12 & 15 primers**: if the DNA binds to all of these primers except for 8, and 12 then it is a **K cluster phage**
L Cluster	**Test with primers 8 & 15**: if the DNA binds to both primers 8 and 15 then it is a **L cluster phage**
M Cluster	**Test with primers 5 & 14**: if the DNA binds to both primers 5 and 14, then it is a **M cluster phage**
N Cluster	**Test with primers 1, 5, & 13**: If the DNA only binds to the primer 1 and not to primers 5 and 13, then it is a **N cluster Phage**
O Cluster	**Test with the primers 2, 3 and 15**: if the DNA binds to the primer 15 but not to primer 2 and 3, then it is an **O cluster phage**
P Cluster	**Test with primers 1, 5, & 13:** If the DNA binds primer 13 but not to primer 1 or 5, then it is a **P cluster phage**

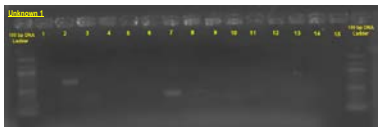

Fig.3. Gel electrophoresis of Unknown Mycobacteriophage 1. Unknown mycobacteriophage 1 was identified to be a C cluster phage, indicated by primer 2's ability to amplify the genomic DNA.

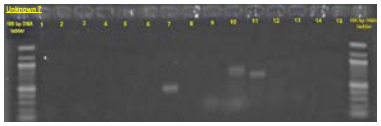

Fig.4. Gel electrophoresis of Unknown Mycobacteriophage 7. Unknown mycobacteriophage 7 was identified to be a I cluster phage, indicated by the ability of primers 7, 10, and 11 to amplify the genomic DNA, but not primer 13.

Unknown Mycobacteriophage Panel

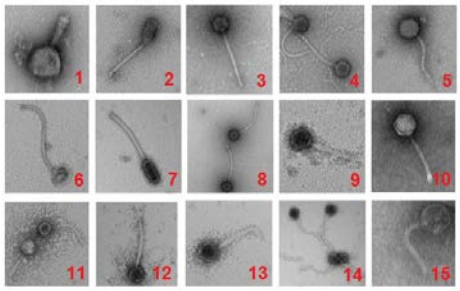

Fig.5. Transmission Electron Microscopy Images of Queens College's Unknown Phage Panel. The micrographs of each unknown phage lysate sent to Queens College, provided by the Science Education Alliance, aided analyzing of the morphologies of the unknown phages, and lead to the determination that Unknown 13 was a mixed culture, and that there is a discrepancy with the identity of Unknown 10.

Results

Table 3. Guesses made through PCR analysis of the oligonucleotide primers and actual results of the 15 unknown mycobacteriophages. Green indicating correct guesses, red indicating an accurate identification but for only one of the phages in a mixed culture.

Unknown	1	2	3	4	5	6	7	8	9	10*	11	12	13	14	15
Guess	C1	I1	G	F1	A1	F	I1	H	N	C	K	E	J	M	J
Actual	C1	L*	K3	F1	A1 & G	B1	I1	B3	A2	H	Q	E	J	R	A4

Unknown Mycobacteriophage 10 | **Mycobacteriophage Patience**

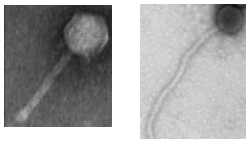

Fig.6. Initially, Unknown 10 (indicated in blue in Table 3) was identified to be a C cluster phage due to a misleading positively stained micrograph. After identities of the 15 unknown phages were released, TEM was performed for a second time where it was discovered that there is a discrepancy in the identity of Unknown 10. To the left is the micrograph of our Unknown 10 sent with the panel; and to the right is a micrograph obtained from phagesdb.org. Their morphologies are strikingly different.

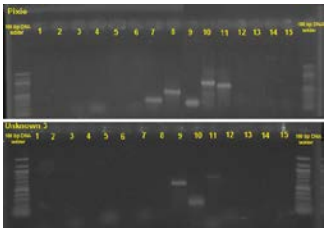

Fig. 7. Gel electrophoresis comparing amplified DNA of phage Pixie, from the Queens College collection, and Unknown 3, expected to be Pixie. Differences in binding patterns were also seen when comparing the following Queens College stocks to the Unknown samples from the Panel: Chah to Unknown 6; Babsiella to Unknown 7; Giles to Unknown 11; and LHTSCC to Unknown 15. Understanding of why these differences have occurred is currently being investigated.

Discussion

A panel of 15 unknown mycobacteriophages was sent out by SEA, in order to determine an more efficient method for identifying the cluster of a phage. Here, an attempt was made using the banding patterns produced by a cohort of 15 PCR primers. Initially, these primers were designed to amplify genomic regions or phams that pertained to a particular cluster, but through multiple rounds of PCR analysis, the primers were found to amplify DNA across multiple clusters. Although various BlastP searches were performed by the researchers that created these primers (data not shown), the unique regions that the primers were designed off of were not as unique as previously theorized. It is reason enough to believe this is attributed to the architectural mosaicism of the genome of the mycobacteriophages. The attempt at designing PCR primers analysis that specifically pertain to one cluster further elucidates the complexity of the mosaic architecture of the genomes of the mycobacteriophages. It is problematic to find genomic regions that are specific to a particular cluster, especially when recombination and divergence from the recombination events has been the driving force behind the clusters.

References

Pope WH et al. (2011) Expanding the Diversity of Mycobacteriophages: Insights into Genome Architecture and Evolution. PLoS ONE 6(1): e16329. doi:10.1371/journal.pone.0016329.

Hatfull, GF et al. (2010) Comparative genomic analysis of 60 Mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. Journal of Molecular Biology. Vol 397:119-143.

Graham F Hatfull. "Mycobacteriophages: Genes and Genomes." Annual Review of Microbiology 64.1 (2010): 331-56. DOI: 10.1146/annurev.micro.112408.134233

Acknowledgements

This research was supported by funding from the Howard Hughes Medical Institute SEA PHAGES program, the National Science Foundation, the CUNY LSAMP program, the Queens College Office of the Provost, the Queens College Divisions of General Education and Math and Natural Sciences, the Queens College UMRE program and the Queens College Biology Department. This work was conducted in part with equipment in the Core Facility for Imaging, Cellular and Molecular Biology at Queens College. Many thanks to James Carpino, Marc Tollis, Peter Novick, Rafael Ovalle, researchers at the University of Pittsburgh, the University of Alabama at Birmingham, and all Phage Hunter students.