Identifying Mycobacteriophage: DNA Primer Design for Phage Clusters

Lauren Esposito, Lauren Mordukhaev, Tara Sooknauth, Swati Gupta and John J. Dennehy
Biology Department, Queens College of the City University of New York, Flushing, NY

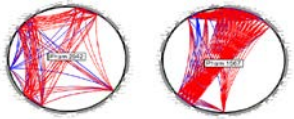
Abstract

To date, 363 mycobacteriophages have been fully sequenced and assigned to one of 40 clusters and subclusters. However, greater than 2,400 collected phages remain unidentified. In order to identify and categorize unknown mycobacteriophages, DNA primers specific for genes or groups of genes unique to each cluster and subcluster of mycobacteriophage were designed. These DNA primers were tested to confirm their ability to produce bands visible by gel electrophoresis of the PCR product produced by PCR template DNA of the corresponding cluster and subcluster of interest. Working primers were then tested against the template phage DNA from all other clusters and subclusters to ensure their specificity. Our results show that most of the DNA primers created for each cluster suitably identified the corresponding template phage DNA. Once the cluster of the template phage is determined, the DNA primers specific for the subclusters within that cluster can be tested on that template phage DNA in order to determine the subcluster to which the template phage DNA belongs to. These primers provide a simple, efficient, and low cost system to categorize the 2,400+ mycobacteriophages isolated from the wild by the HHMI Phage Hunters program, allowing the prioritization of unknown phages for full genome sequencing. Ultimately the long term goal is to discover genes or gene families that could be useful in the treatment of *Mycobacterium tuberculosis.*

Mycobacteriophage Diversity

Mycobacteriophages are viruses that infect members of the *Mycobacterium* genus, including *M. smegmatis* and *M. tuberculosis*. Mycobacteriophages are highly genetically diverse, perhaps due to extensive gene shuffling among the phages in this group. 326 mycobacteriophage genomes have been fully sequenced. A comparative analysis of these genomes has identified 2,345 unique Phams (genes or gene families).

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).



The mycobacteriophages are further assigned to clusters and subclusters based upon their Average Nucleotide Identity (ANI). To date, 31 subclusters have been identified.

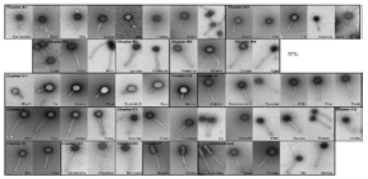
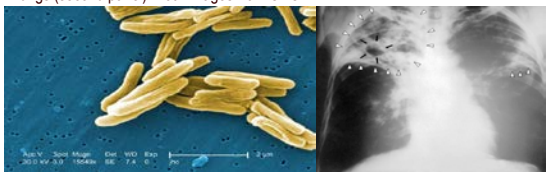


Fig. 2 - Electron micrograph images of representative members of all 31 subclusters. Images from PhagesDB.com.

Mycobacterium tuberculosis

Mycobacterium tuberculosis is the causative agent of tuberculosis (TB), one of the world's deadliest diseases. One third of the world's population is infected with TB. TB is considered to be one of the world's deadliest diseases due to its heightened capability to become antibiotic resistant. Extreme drug-resistant TB is a critical issue in many developing countries, and a totally drug-resistant TB, resistant to all currently used drugs, has emerged in India. The long term objective of our mycobacteriophage research is to ID *M. smegmatis* phage genes that can be used in the treatment of tuberculosis.

Fig. 3 – *Mycobacterium tuberculosis* (first panel). X-ray image of tuberculosis infection in lungs (second panel). Both images from CDC.



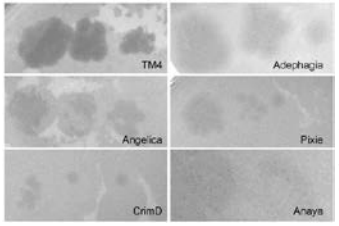


Fig. 4 – Mycobacteriophage of K cluster can form plaques on *M. tuberculosis*. Members of subclusters A2, B2, B3, C1, C2, D, F1, H1, J, L2 and M are also able to infect *M. tuberculosis.* Mycobacteriophage may hold the key for treating extreme drug resistant forms of TB.

PHAGE	CLUSTER/ SUB-CLUSTER	PHAM/ GENE AMPLIFIED	EXPECTED AMPLICON SIZE (bp's)	OBSERVED AMPLICON SIZE (bp's)
Bethlehem	**A1 Cluster**	Pham 2967	180	~200
D29	**A2 Cluster**	D29_gp27	655	~700
Bxz2	**A3 Cluster**	TBD	TBD	TBD
LHTSCC	**A4 Cluster**	Pham 6000	314	~300
George	**A5 Cluster**	Pham 7429	147	~150
DaVinci	**A6 Cluster**	DaVinci_gp39	159	~200
Chah	**B1 Cluster**	Cha_gp2	696	~700
Qyrzula	**B2 Cluster**	Pham 2332	525	~500
Pipefish	**B3 Cluster**	Pham 3084	680	~700
Cooper	**B4 Cluster**	Pham 3023	257	~300
Arcadian	**B5 Cluster**	Pham 5956	366	~400
Cali	**C1 Cluster**	Bxz1_gp29	402	~400
Myrna	**C2 Cluster**	TBD	TBD	TBD
Gumball	**D1 Cluster**	gp_11/ gp_13	725	~700
CJW1	**E1 Cluster**	Pham 244	802	~800
Boomer	**F1 Cluster**	Boomer_gp8 /gp12	336	~300-400
Che9D	**F2 Cluster**	Pham 510	402	~400
BPs	**G1 Cluster**	TBD	TBD	TBD
Predator	**H1 Cluster**	Pham 3585	195	~200
Barnyard	**H2 Cluster**	TBD	TBD	TBD
Brujita	**I1 Cluster**	Brujita_gp17	622	~600
Che9C	**I2 Cluster**	Pham 2165	1280	~1200
Omega	**J1 Cluster**	Pham 2842	261	~250
CrimD	**K1 Cluster**	TBD	TBD	TBD
TM4	**K2 Cluster**	Pham 1367	307	~300
Pixie	**K3 Cluster**	Pham 2597	236	~200
JoeDirt	**L1 Cluster**	JoeDirt_gp84	204	~200
Faith1	**L2 Cluster**	Pham 3930	276	~300
Rey	**M1 Cluster**	Pham 2689	428	~450
Corndog	**O1 Cluster**	Pham 496	450	~450

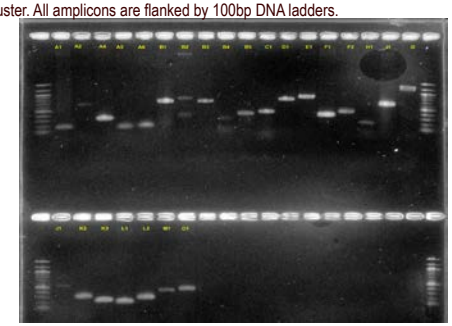
Mycobacteriophage Distribution

Mycobacteriophages have been isolated from many locations across the US and around the world. No biogeographic structure has yet been identified to explain the distribution of various clusters and subclusters.

Fig. 5 – Sequenced phage isolation locations from GPS data. First panel shows distribution of isolation locations across the US; second panel shows close up of Eastern region. Letters represent phage cluster ID. Images from PhagesDB.com.

Results and Discussion

Fig. 6 – The bands in the gel electrophoresis represent PCR product produced by amplification of template DNA for the corresponding subcluster of interest. Amplicon size was used to test the accuracy of the DNA primers that were designed for each subcluster. All amplicons are flanked by 100bp DNA ladders.



As shown in the gel electrophoresis image above, DNA primers designed for each cluster or subcluster were able to amplify DNA from the unique pham region of the corresponding template phage. The location of a band measured against a 100bp DNA ladder indicates the amplicon size of the amplified pham region. For all the visible bands, the experimentally observed amplicon sizes match the actual, computed amplicon lengths. This testing confirms the functionality of nearly all the designed primers. The second phase of analysis includes negative cross-testing of each DNA primer set against non-corresponding phage template to confirm the specificity of each primer set to its corresponding cluster. Once the primers have been validated, the 2,400+ phages collected can be identified. To date, there are cluster-specific primers for the D, J, and M 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.

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 UM/RE 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, Brian Ford, Marc Tollis, Peter Novick, Rafael Ovalle, and all Phage Hunter students.