**Introduction**

The mycobacteriophage population represents an incredibly large and diverse set of genomic sequences. Traditionally, phage genomes are compared using methods that require sequence alignment or gene annotation. However, these methods may be ineffective for populations with significant horizontal gene transfer and are relatively difficult to compute computationally. Mycobacteriophages also lack a common sequence or morphological feature (such as ribosomal RNA in bacteria) on which to compute phylogenetic relationships.

Alignment-free sequence analysis methods, such as measures that compute the usage of oligonucleotides (k-mers) in the genome, have the potential to infer relationships between significantly diverged sequences. We examined the usage of tetranucleotides in all 663 phage genomes available on phagesdb and found several interesting results. Tetranucleotide usage deviation (TUD) in a genome from a null model was highly conserved between subclusters and different between subclusters. A neighbor-joining phylogenetic tree constructed from a TUD-distance matrix accurately reconstructs the tree presented for 60 phage in Hatful et. al (2010). When a tree is constructed for all phage, genomes are almost always placed in a monophyletic clade with members of the same cluster.

Alignment-free methods can also be used to investigate horizontal gene transfer within the mycobacteriophage population. A segment of DNA transferred from another organism is likely to have an oligonucleotide usage signal more closely related to the original than to the target. We computed an index of genomic self-similarity in a sliding window across phage genomes called tetranucleotide difference index (TDI). The final portion of cluster L genomes are significantly different than the rest of the genome in terms of tetranucleotide usage, possibly indicating horizontal gene transfer.

Overall, we feel alignment- free sequence analysis is an excellent tool for preliminary investigation of phage genomes.