Lehigh University would like Mycobacteriophage Derek’s genome investigated further for the following features.

Many features in Derek’s genome were evaluated for functions that are called by HHpred but not by BLASTp for any phages. These functions are not currently annotated but we believe are valid functions.

Feature 88 has hhpred homology to Endonuclease; PI-PFUI, intein-encoded, hydrolase; 2.10A {Pyrococcus furiosus} with Probab=99.93 E-value=1.2e-25. There is BLASTp data from Pio\_94 to call as a DNA Polymerase subunit/ Intein with Q1:S1. This function is no longer one of the approved functions from the Functional Assignments 2015-16 list. More recently annotate phages have this called as a portal protein with Q384:S116 (Yucca\_90). There is HHpred data lower down the list to support a portal protein function { Portal protein; transport protein; 1.98A {Bacteriophage G20C}

Probab=99.07 E-value=2.8e-10}. No other portal protein is currently called in the Derek genome.

Feature 100 has HHpred data to support a capsid-like protein but there is no BLASTp data to support this. HHpred data to support {Head fiber protein; supercoiled triple repeating helix-turn-helix, viral protein; 1.52A {Bacillus phage PHI29} Probab=99.06 E-value=2.1e-10}.

Feature 121 has strong HHpred homology to ParB. HHpred data for 121 {PARB, chromosome partitioning protein PARB; nuclear protein, chromosome segregation, DNA-binding,

helix-turn-helix; 2.3A {Thermus thermophilus} SCOP: [a.4.14.1](http://scop.berkeley.edu/sccs=a.4.14.1) [d.268.1.1](http://scop.berkeley.edu/sccs=d.268.1.1) Probab=99.90 E-value=4.1e-25}. This is not called because there is no ParA called, however there is weak HHpred homology to ParA in feature 193. HHpred for 193 {Plasmid partition protein A; type IA, DNA binding, winged-HTH, DNA bindin; HET: ADP EPE; 2.05A {Escherichia coli} PDB:  [3ez6 \_A\*](http://pdb.rcsb.org/pdb/explore.do?structureId=3ez6)  [3ez7 \_A](http://pdb.rcsb.org/pdb/explore.do?structureId=3ez7) Probab=64.98 E-value=0.44}.

Feature 160 has HHpred homology to DNA endonuclease I-MSOI, linker, DNA endonuclease; protein-DNA complex, chloroplast, hydrolase-DNA complex; HET: DNA; 2.69A {Monomastix SP} SCOP: [d.95.2.1](http://scop.berkeley.edu/sccs=d.95.2.1) [d.95.2.1](http://scop.berkeley.edu/sccs=d.95.2.1) Probab=98.14 E-value=1.4e-05. There are no other C1 phages that have an endonuclease function called.

We have called functions for the following functions but there are some discrepancies based on BLASTp and other C1 phage functional annotation.

Feature 208 is currently called a RusA resolvase (endonuclease) based on the list of accepted function. This feature has homology to Yucca\_212, Gabrielle\_215, and Erdmann\_213 which call a Holliday junction resolvase. Unsure of which nomenclature to use.

Feature 237 is called as a glycosyltransferase. Yucca\_242, Zeenon\_243 call as glycosyltransferase. Wally\_240, Shrimp\_244 call a galactosyl transferase. HHpred data supports glycosyltransferase.

Some features were called based on HHpred data. Feature 193 was called as a ThyX thymidylate synthase based on HHpred homology and BLASTp data from Gabrielle. Feature 216 was called a hydrolase based on HHpred data, no BLASTp data supports this.

We would like the QC team to evaluate features 127 and 128 to ensure that these were called correctly. These features are a tail assembly chaperone caused by a translational frame shift.

Feature 90 has no BLASTp data in DNAmaster but has blast data from NCBI web based server.

Features 151, 152, 176, 181 are tRNA but have no COVE score because they were only called by Web based Aragorn not tRNA scan.

We have investigated all gaps and they are valid based on Phamerator homology to comparable C1 phages. Derek contains 229 ORF, 31 tRNAs, and 1tmRNA (feature 180).