Phage Name: **BENtherdunthat**

Rebecca Bortz, University of Pittsburgh, rlb6@pitt.edu

* Yes. Genome sequence contains same number of bases and is the same as on phagesdb.org.
* Yes. All genes are valid.
* Yes. All genes are numbered sequentially.
* Yes. All old BLAST hits were cleared for genes that were re-BLASTed.
* Yes. The locus tags are BENtherdunthat.
* Yes. Documentation has been recreated.
* Yes. tRNA information was adjusted according to tRNAscan SE.
* Yes. All the specified requirements are met for the final annotation file.
* Yes. All the specified requirements are met for the minimalistic annotation file.

Areas of Genome That Warrant Further Inspection:

* A single tRNA was identified that was not originally called by Aragorn. It was located in a gap using tRNAscan SE.
* Though there is an integrase gene (51) present in the genome, none of the BLAST information or HHpred data seems to suggest the presence of an immunity repressor. Gene 57 was found to code for an Antirepressor.
* Genes 37 and 41 appear to be abnormally short in length, although data suggests that these are probably legitimate genes.
* There are some areas of the genome with relatively large gaps, like between genes 52-53, 55-56, and 95-96, however none of these gaps indicate strong coding potential.
* Gene 90 is a reverse-transcribed gene among several forward-transcribed genes. It is not called by Glimmer, and its inclusion creates a large area of overlap with Gene 89.
* Gene 18 is called as the Tape Measure Protein, but Genes 15 and 16 only have weak hits to Tail Assembly chaperones and without analogs were not annotated as such. No translational frameshift was detected.
* Gene 17 is a reverse-transcribed gene among many forward-transcribed genes. The current ORF does not capture all coding potential, but changing the start site did not improve alignment and BLAST hits, so the current start site was retained.
* The following functional calls were designated but they are not currently in the list of approved nomenclature:
  1. Gene 17 - TAT Pathway Signaling Sequence
  2. Gene 23 – Lysin A, Glycosyl Hydrolase Domain
  3. Gene 42 - Procyclic Acidic Repetitive protein
  4. Gene 50 - nitrile hydratase
  5. Gene 54 - Metallopeptidase Domain Protein
  6. Gene 55 and 56 – Helix-turn-helix XRE protein
  7. Gene 66 - Monogalactosyldiacylglycerol (MGDG) synthase
  8. Gene 85 - Leucine-zipper of insertion element IS481
  9. Gene 88 - ATPase wrnip1
  10. Gene 89 - N-terminal phage replisome organiser
  11. Gene 91 - Loader and inhibitor of phage G40P
* The following genes were called by auto-annotation but were deleted.

Gene 25 (20241-20819): Lone reverse transcribed gene with no coding potential identified by GeneMark. Minimal base pair gap between Genes 24 and 26 suggesting that no gene should exist between them.

Gene 38 (28355-28486): No coding potential detected and no BLAST data

Gene 43 (29120-29410 Rev): Large overlap with Gene 42 and no coding potential identified by GeneMark. Dr. Welkin Pope reviewed section of genome and suggested deletion.

Gene 55 (33490-33642 Rev): No coding potential detected and no BLAST data

Gene 62 (36409-36579 Rev): Lone reverse transcribed gene with no coding potential identified by GeneMark. Extremely large overlap with previous Gene 61.

Gene 69 (39374-39703): Lone reverse transcribed gene with no coding potential identified by GeneMark. Overlaps with forward transcribed gene 68.

Gene 78 (41788-42288 Rev): No coding potential detected and no BLAST data

Gene 108 (53470-53943 Rev): There are two forward genes that have coding potential and overlap this reverse gene. GeneMark does not show coding potential for this reverse gene.

Gene 111 (54490-54627 Rev): No coding potential detected and no BLAST data

* The following functional calls were designated but they are not currently in the list of approved nomenclature:
  1. Gene 17 - TAT Pathway Signaling Sequence
  2. Gene 23 – Lysin A, Glycosyl Hydrolase Domain
  3. Gene 42 - Procyclic Acidic Repetitive protein
  4. Gene 50 - nitrile hydratase
  5. Gene 54 - Metallopeptidase Domain Protein
  6. Gene 55 and 56 – Helix-turn-helix XRE protein
  7. Gene 66 - Monogalactosyldiacylglycerol (MGDG) synthase
  8. Gene 85 - Leucine-zipper of insertion element IS481
  9. Gene 88 - ATPase wrnip1
  10. Gene 89 - N-terminal phage replisome organiser
  11. Gene 91 - Loader and inhibitor of phage G40P