|  |  |
| --- | --- |
|  | **Hiro**Florida Gulf Coast UniversityMay 25, 2015 |

Phage Hiro was isolated as part of the *Rhodococcus erythropolis* pilot. Hiro is highly similar to RER2, a previously published *Rhodococcus* phage isolated in Australia {97% identity, 95% query cover, E-value=0.0}. RER2’s annotation did not appear to make use of the function calls of similar *Mycobacterium smegmatis* phage genes. Using NCBI Blastp, PhagesDB Blastp, and HHpred, we were able to assign many predicted functions. We are confident with all areas of Hiro’s annotation.

**tRNAs**

Aragorn and tRNAscan predicted 3 tRNAs in the 5’ end of the genome. The first is Asn (gtt anti-codon), the second has a single base deletion in the anti-codon loop and is predicted to be either Val (gac) or Arg (acg), and the third is Trp (cca).

**Head-to-tail connector proteins**

A group of 5 predicted proteins in the structural operon between the major capsid protein and the major tail subunit show similarities to *M. smegmatis* phage head-to-tail connector proteins. We are unsure of the biological significance of this cluster of proteins.

**Frameshifts**

There is a typical tail assembly chaperone frameshift with a -1 slippage site at 11634 bp (G).

There’s a putative DNA primase frameshift. The frameshift could either be with a –2 shift or a +1 shift in the slippery sequence UUUUG [in forward direction: 35537 bp - CAAAA - 35542 bp]. After consultation with Drs. Graham Hatfull and Welkin Pope, we kept the two DNA primase orfs separate until the frameshift can be verified experimentally. Our start site choice for the 2nd DNA primase orf was based on coding potential in the GeneMark graphical output using *R. erythropolis* PR4 preferences. This choice resulted in a very large overlap.

**Gaps**

There are several gaps equal to or greater than 100 bp in the genome. We did not find good coding potential or BLAST matches in these regions. The first major gap is 1,544 bp between gp68 at the 3’ end of the genome and gp1 at the 5’ end. Other gaps near the 5’ end of the genome are 343 bp between gp1 (NKF) and gp2 (NKF), 131 bp between gp2 (NKF) and gp3 (NKF), and 352 bp between gp3 and gp7 (the 3 tRNAs are found in this gap). A gap of 113 bp is found between gp18 (major tail subunit) and gp19/20 (tail assembly chaperone). A gap of 570 bp is found between gp29, the integrase (the first gene in the reverse direction) and gp30, the excisionase (which switches back to the forward direction). This gap probably contains transcriptional control sequences for the integrase and excisionase genes. Additional gaps were located as follows: 103 bp between gp38 (thymidylate synthase) and gp39 (NKF); 100 bp between gp54 (NKF) and gp55 (NKF); and 663 bp between gp58 (immunity repressor) and gp59 (NKF).