Actinobacteriophage Genome Annotation Submission Cover Sheet

This Cover Sheet will accompany each genome’s annotation file(s) submission and succinctly describe the work that your students and you have done. This document ensures that the work done was as complete and thorough as it could be. Most important to the QC reviewer, denote where the trouble spots were in your annotation and how they were resolved.

**Note: we think Zippen and Chlochlo should be QC’ed by the same person.**

Phage Name. **Zippen**

Your Name. **Kieran Furlong**

Your Institution. **University of Ottawa**

Your email. **kfurl048@uottawa.ca**

Additional emails. (for correspondence). [**arudner@uottawa.ca**](mailto:arudner@uottawa.ca)

Describe any issues or specific genes that you would like to highlight for the QC reviewer. This includes any genes that you had questions about or received help with or that warrant further inspection in the QC review process. Include those genes that you deliberated on and/or want to strongly advocate for. If you contacted SMART, a workshop facilitator, or a buddy school for help, please document.

**Paste in any specific notes for the most troublesome/controversial calls.**

## Gene 30 (27285-28076)

#### Challenge:

* Most phages (61 of 100) call “Minor tail protein”. We call it a “Minor tail protein”.

#### Evidence for minor tail protein:

In the correct region of the genome for a third minor tail protein.

The protein sequence is very glycine rich.

#### Evidence against minor tail protein:

HHPRED: Several hits > 90% probability. However, they are to Homo Sapiens or Gallus. Also, there are no collagen nor glycine-rich protein hits as required by the Functional Assignment list.

## Gene 43 (32785-33276)

#### Challenge:

* All pham members (66 of 66) call NKF. We are choosing “nucleotide pyrophosphohydrolase”

#### Evidence for nucelotide pyrophosphohydrolase

HHPRED: 100% probability and ~60% coverage with dATP/dGTP diphosphohydrolase, N-terminal. This protein catalyzes the hydrolysis of dATP/dGTP to dNMPs and is part of the dZTP biosynthetic pathway. Strong match to the DUF550 domain described in <https://www.science.org/doi/epdf/10.1126/science.abe4882>, which was characterized in vitro for this activity. Although there is no evidence for the other genes in the dZTP pathway, the evidence for the presence of this function is very strong.

## Gene 56 (37,678 – 37,860)

Very weak coding potential, but some. Gene has -1 overlap with the downstream gene. This gap is present in comparator genomes and in some was called as a gene.

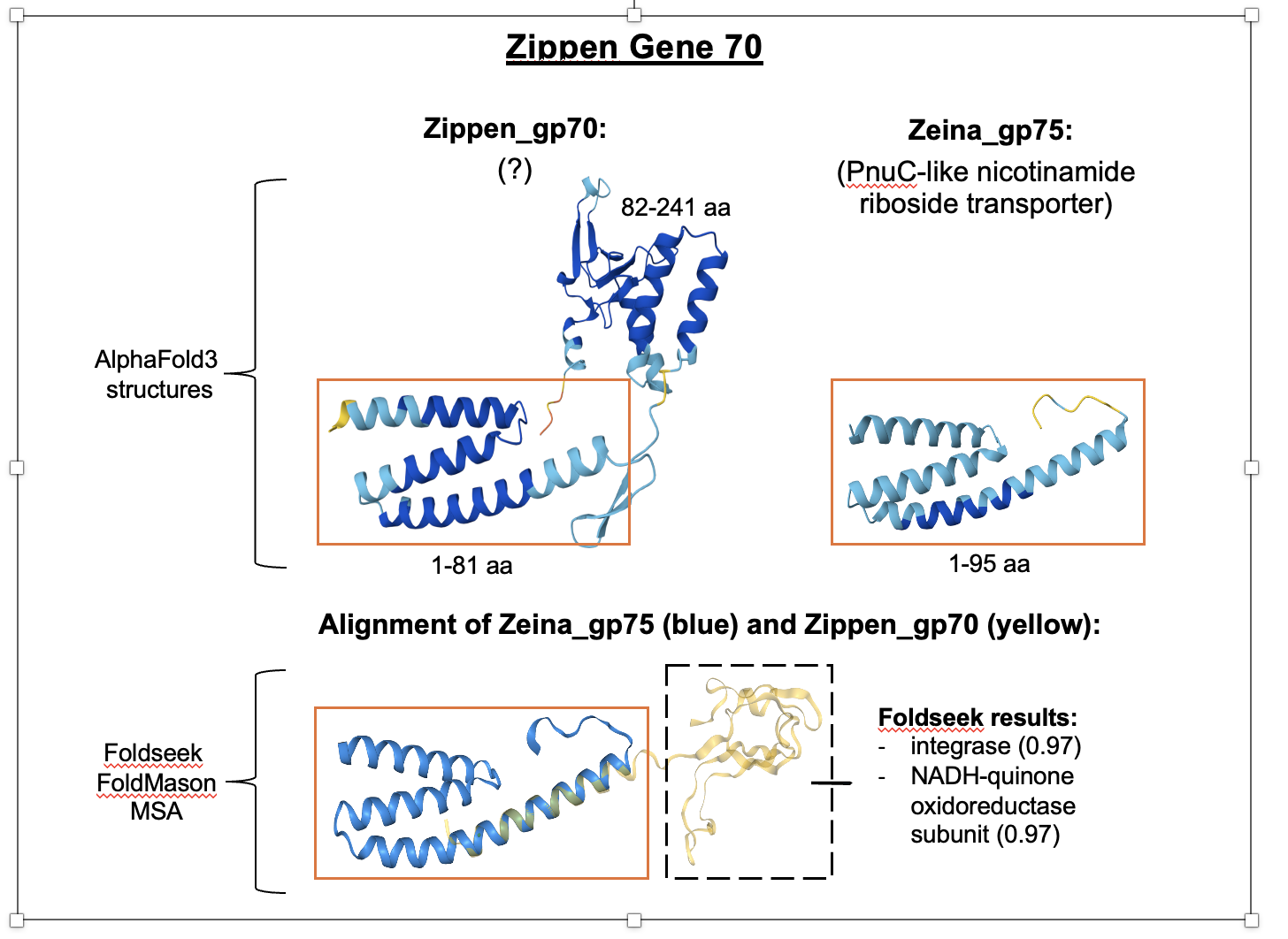
## Gene 61 (43,340 – 43,552)

The protein has strong matches to Fin, an Anti-sigma-Factor. See forum discussion: <https://seaphages.org/forums/topic/5799/?page=1#post-11539>

We have called it an NKF, though we still have questions.

## Gene 70 (46,004 – 46,729)

This protein has high probability hits on HHpred to the “PnuC-like nicotinamide riboside transporter” present in AU6 phages (Zeina\_75). The C-terminus doesn’t match proteins in HHpred, but does have structural homology on Foldseek to a NADH-quinone oxidoreductase subunit. Although we don’t know the exact function of the C-terminus, we think the protein should be annotated as “PnuC-like nicotinamide riboside transporter”. No other pham members have made this call, but we think this is in error.



Please record yes/no for each of the questions below. If further explanation is needed, please add this item to the above box.

In the submitted DNA Master file (Yes/No):

YES 1. Does the genome sequence in your submitted DNA Master file match the nucleotide fasta file posted on phagesDB (same number of bases, no N bases, etc.)?

YES 2. Are all the genes ‘Valid” when you click the [Validation button](https://seaphagesbioinformatics.helpdocsonline.com/article-84)?

YES 3. Are the genes (and matching LocusTag numbers) [sequential](https://seaphagesbioinformatics.helpdocsonline.com/article-77), starting with #1, counting by 1s.

YES 4. Are the Locus Tags the “[SEA\_PHAGE NAME](https://seaphagesbioinformatics.helpdocsonline.com/article-77)” format?

YES 5. Has the [documentation been recreated](https://seaphagesbioinformatics.helpdocsonline.com/article-86) from the Feature Table to match the latest file version?

YES 6. Have tRNAs followed the [tRNA protocol](https://seaphagesbioinformatics.helpdocsonline.com/undefined), **COPYING** tRNA-AMINOACID type (DNA equivalent of the anti-codon) from Aragorn output - ﻿tRNA-Gln(ctg) - AND the ends been adjusted to match the Aragorn output?

N/A 7. Has the [frameshift in the tail assembly chaperone](https://seaphagesbioinformatics.helpdocsonline.com/article-54) been annotated correctly (if applicable)?

YES 8. Have you cleared your Draft\_Blast data and have you [re-Blasted](https://seaphagesbioinformatics.helpdocsonline.com/article-57) the submitted DNA Master file?

YES 9. Has every gene been [described and supported in your Supporting Data file](https://seaphagesbioinformatics.helpdocsonline.com/article-44)?

YES 10. Did you investigate ‘[gaps](https://seaphagesbioinformatics.helpdocsonline.com/article-31)’?

YES 11. Did you [delete the genes](https://seaphagesbioinformatics.helpdocsonline.com/article-65) that you meant to delete?

Now, [make a profile of the file](https://seaphagesbioinformatics.helpdocsonline.com/article-64) you plan to send. (And you can save this file for [Review to Improve!)](https://seaphagesbioinformatics.helpdocsonline.com/untitled-18)

YES 1. Have any duplicate genes been deleted?

YES 2. Has the Notes field been cleared (using the automated buttons)?

YES 3. Do the gene numbers and locus tags match?

YES 4. Are the correct Feature\_Types correctly selected (most will be ORFs, but check that tRNAs and tmRNAs are correctly labeled)?

YES 5. Do the function names in the Product field either match the official function list or say “Hypothetical Protein”?

YES 6. Has the Function field been cleared (using the automated buttons)?

How are you documenting your gene calls in class? Choose any/all that apply:

**X** PECAAN output into DNA Master Notes

      DNA Master shorthand (previously used format)

      Spreadsheet

      Powerpoint

**X** Word document (full notebooks available, but not submitted)

      Other: Describe.

What is the file type (sort) submitted for QC to document your gene calls? Choose only one.:

**X** PECAAN output to dnam complete and minimal files

      DNA Master shorthand (previously used format)

      Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

      Other: Describe.

**The QCer can look at the PECAAN file Zippen\_draft6**