**Phage Genome Annotation Group Quality Control Checklist**

**Phage Name:**

**Team Members:**

|  |  |  |
| --- | --- | --- |
| **Task** | **Team Members Responsible** | **Notes** |
| Examine all genes with a called function.  Be sure all lines of evidence are described for each gene (Starterator, GM Coding Potential, PhagesDB BlastP, NCBI BlastP, Conserved Domain Database, HHPred, TmHmm, Topcons, Synteny) with appropriate top informative match evidence boxes checked.  Are all of the assigned functions on the APPROVED function list?  *Difficulty level: low to medium, recommend 1-2 teammates*  *Time commitment: moderate* |  |  |
| Examine all genes assigned the function NKF.  Confirm that all genes with a designation of NKF have been checked by TMHMM and Topcons for transmembrane domains and have been correctly updated (1 TMH called by both programs OR 2+ TMH called by one)  *Difficulty level: low, recommend 1 teammate*  *Time commitment: low* |  | List genes called as membrane proteins |
| Confirm that the genome has been checked for tRNAs, and that the ends of any tRNAs have been adjusted with web-based Aragorn and/or tRNAScan-SE. Infernal scores from web-based tRNAScan-SE are included in tRNA notes.  *Difficulty level: low to medium, recommend 1-2 teammates*  *Time commitment: low to moderate* |  | How many tRNAs were found? List coordinates and details. |
| Frameshift in tail assembly chaperone genes has been correctly identified (when applicable)  *Difficulty level: medium to high, recommend 2 teammates*  *Time commitment: moderate to high* |  | Does genome have an identifiable frameshift? Yes/no  Which genes are the tail assembly chaperones?  What is the frameshift? +1 or -1  What is the coordinate of the frameshift? |
| Functions present in almost all genomes (<https://seaphagesbioinformatics.helpdocsonline.com/article-91>) have been identified in your genome  *Difficulty level: medium, recommend 1 teammate*  *Time commitment: low* |  | List any function on the list that you weren’t able to identify in your genome and indicate the priority level. |
| Gaps in the genome have been checked for missing genes:   * Check for ORFs with good coding potential that were missed in auto-annotation * Check ORFs longer than 120 bp with Blast and HHPred   *Difficulty level: moderate, recommend 2 teammates*  *Time commitment: moderate to high* |  | List new genes that need to be added, including start and stop coordinates and direction of transcription  List gaps that were investigated but were found to not contain any genes |
| All start sites have been updated as needed and appropriate databases are updated  *Difficulty level: low, recommend 1 teammate*  *Time commitment: low to moderate* |  | List all genes with start changes from original position |
| Additional review of difficult-to-annotate genes  *Difficulty level: moderate to high, recommend 2-4 teammates*  *Time commitment: moderate to high* |  | See below |

**In the space below, list all “difficult” genes in the genome that require additional review (be as detailed as you can regarding what is difficult about the gene-confusing functional assignment, ambiguous start site, etc.) and briefly describe final annotation decision regarding these genes:**