

Inputs

- Input genome sequences (.fna)
- Merged phage + plasmid protein sequence fasta file (.faa)
 - This file will be used to create the diamond database
- Metadata file which contains database labels for distinguishing between phage vs plasmid

ORF Prediction

Step1:

- Predict open reading frames with Prodigal
 - Uses metagenomic mode ``-p meta``
 - Output is a protein fasta file ``-a _prodigal.faa``

Diamond Alignment

Steps 2 & 3:

- Create a reference database for diamond
- Using Diamond to perform a local alignment to identify hits.
- Hits ==> (e-value < 10^{-4} , identity $\geq 35\%$, coverage $\leq 50\%$)

Calculate gPPQ

Step 4

- Calculate average phage-plasmid quotient (gPPQ) scores
 1. Calculate PPQ score for each genome with $10 \leq$ protein sequences.
 2. Calculate average PPQ score per genome.