

Known Knowns: Comprehensive Virus-Host Interaction Database Pipeline

Initial Data Acquisition

PhageDB.org: data -> download data -> Full Tab-Delimited Text Files: All sequenced phages

NCBIVirusDB: NCBI Virus Variation Resource database (November 6th, 2019)

TASKS:

1. Identify fields that are needed for the final database
2. Fill in missing information for PhageDB data

Final Database Format

virus tax id	virus name	virus lineage	virus accession #	host tax id	host name	host lineage	source
--------------	------------	---------------	-------------------	-------------	-----------	--------------	--------

What do we currently have in PhageDB?

VIRUS

virus tax id	virus name	virus lineage	virus accession #
--------------	------------	---------------	-------------------

✗

✓

✗

✓

[]

[]

1. Submit virus names to NCBI Taxonomy Browser

✓

✓

HOST

host tax id	host name	host lineage
-------------	-----------	--------------

✗

✓

✗

[]

[]

Submit host names to NCBI Taxonomy Browser

✓

✓

NCBI & PhagesDB databases were combined and duplicate interactions.

KnownInteractionsDB.csv

44,975 virus-host pairs

29, 847 unique viruses

7,974 unique hosts