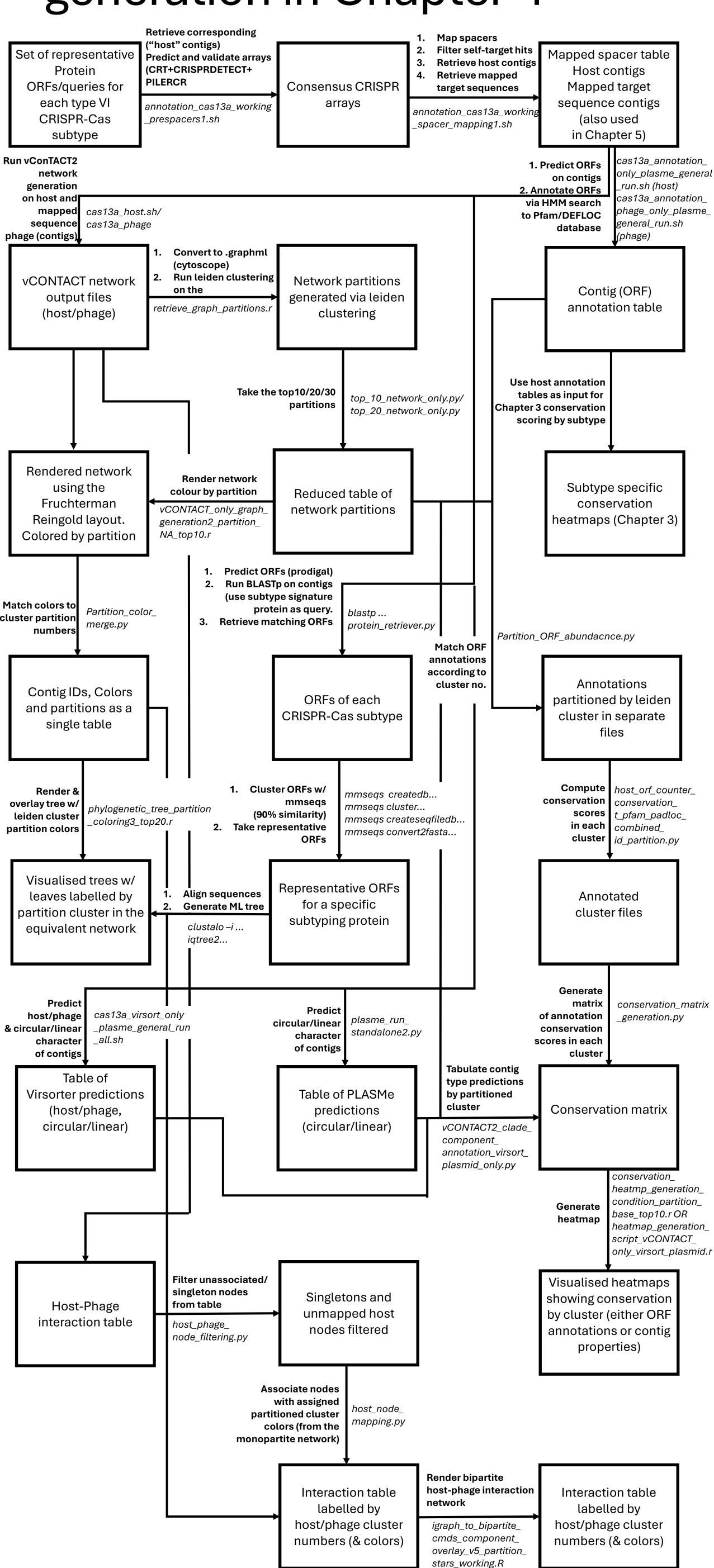
Workflow used for data generation in Chapter 4



Workflow to generate species/sample diversity heatmaps Table of all matched **JGI+NCBI** annotations (40kb "windows") Mapped spacer table Host contigs Mapped target sequence contigs cluster_specific_ Retrieve gold_annotation_ (also used contig local.py AND in Chapter 5) subtype cluster_specific_ specific ncbi_annotation annotations _local.py Network partitions JGI+NCBI tables of host-subtype generated via leiden specific annotations clustering Compute vCONTACT2_clade_ species/ component_ sample annotation_virsort_ conservation plasmid_only_species.py by cluster Cluster specific conservation matrix Visualise Heatmap_ species/ generation sample

script.r

Cluster-specific

species/sample

composition

heatmap

diversity by cluster