Figure S1 Write genomes that pass filters Download phage records from Produce MASH index for easy to new file (this is the final Genbank using "PHG" tag and comparison of new sequences database), and write individual length cutoffs against known genomes records to /GenomesDB/ Predict genes on individual Extract useful info from Download phage records for records in /GenomesDB/ using Genbank entries including known "megaphages" that may Prokka. This allows for phage taxonomy and host, and ordinarily be missed consistent gene calling write to summary table Manually inspect genomes that may be incomplete, and add to the exclusion list Filter genomes to only include Gather information on genome those >10kb, or their Use predicted genes to produce properties from Prokka output description matches strings input files for vConTACT2 and write to summary table "Complete" and "Genome" Exclude genomes using manually curated list of Produce annotation files for Produce annotation files for accession numbers for known vConTACT2 IToL

erroneous genomes