



phylogenetic tree insertion methods. Provides the tools for the creation of custom gene packages for the analysis of sequence data.

- **SingleM:** Provides highly resolved community composition from metagenomes using conserved single copy marker genes. By not heavily relying on reference databases of sequenced genomes, SingleM can be used to accurately profile communities that contain novel lineages. It can also be used to determine how representative a set of population genomes is of a community.
- **GroopM:** Recovers population genomes from large metagenomic datasets using differences in population abundance across metagenomic samples (differential abundance binning).
- **CheckM:** Assesses the quality of isolate, single cell, and population genomes using lineage specific single copy marker gene sets. Includes utilities for comparing genomes and exploring features such as GC content, sequence length, and tetranucleotide signatures.
- **RefineM:** Refines isolate, single cell or population genomes using qualitative and quantitative features such as GC content, coverage and coding density.
- **OrfM:** Rapidly predicts ORFs in raw metagenomic reads.

#### **iVirus:**

- **vContact/vContact-PCs:** Generates Protein Clusters (PCs) using a Markov clustering algorithm and incorporates metadata annotations. Then assigns contigs to taxonomic groups using the presence or absence of shared PCs along the length of the contig.
- **PCpipe:** Compares ORFs from user-defined datasets to existing viral PCs as a means to organize viral sequence space into functional units that can serve as (i) a universal functional diversity metric for viruses, (ii) a scaffold for iterative functional annotations, and (iii) input for ecological comparisons.
- **Fizkin:** Performs Bayesian network analyses based on the amount of shared sequence content in viromes and contextual data about the sample's environment.
- **VirSorter:** Identifies viral sequences in microbial genomes and metagenomic datasets

*This work is supported by the U.S. Department of Energy under funding opportunity announcement number DE-FOA-0000866. Pathways to carbon liberation: a systems approach to understanding carbon transformations and losses from thawing permafrost and iVirus is supported by the US Department of Energy Office of Biological and Environmental Research under the Genomic Science program (Award DE- SC0010580).*