

# Supplemental material

## Summary of anvi'o workflows used in the study:

### # Genome-resolved metagenomics

#### Anvi'o programs (in order within the workflow):

```
# anvi-script-reformat-fasta
# anvi-gen-contigs-database
# anvi-run-hmms
# anvi-init-bam
# anvi-profile
# anvi-merge
# anvi-cluster-contigs
# anvi-interactive
# anvi-refine
# anvi-summarize
```

For the genome-resolved metagenomic survey, we followed the anvi'o metagenomic workflow (see <https://merenlab.org/2016/06/22/anvio-tutorial-v2/>). Here is a summary of the main steps we completed for each of the 11 regions:

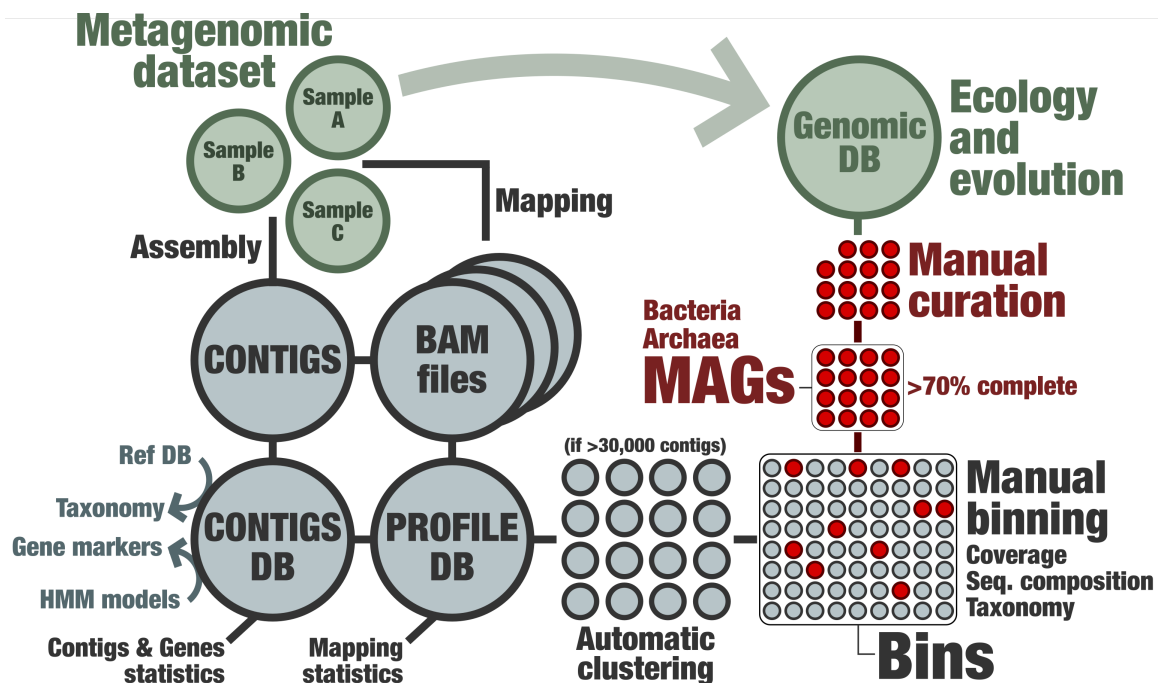


Figure 1: The manual genome-resolved metagenomic framework of anvi'o dedicated to Bacteria and Archaea. This workflow is to be applied to each assembly outcome.

We followed the workflow outlined in the figure 1 for each of the 11 metagenomic co-assemblies (e.g., for the Arctic Ocean). We used the sequence composition of contigs and their differential coverage across metagenomes to perform a first automatic binning step with CONCOCT<sup>3</sup>. Note that CONCOCT is used because the interactive interface of anvi'o cannot work efficiently when loading >25k contigs. For each of the CONCOCT clusters, we then used the anvi'o interactive interface to manually identify and curate the bacterial and archaeal MAGs. This interactive interface takes advantage of the sequence composition of contigs, their differential coverage across metagenomes, taxonomic signal using a reference database, and HMM models for single copy core gene collections (Bacteria, Archaea, Eukarya). When selecting a cluster of contigs corresponding to a MAG in the interface, anvi'o identified its domain affiliation in real time using random forest, and displayed its completion and redundancy values accordingly.

## **# Functional annotations in CONTIG databases**

### **Anvi'o programs (in order within the workflow):**

- # anvi-gen-contigs**
- # anvi-run-kegg-kofams**
- # anvi-run-ncbi-cogs**
- # anvi-export-functions**

For functional inferences, we created a anvi'o CONTIG database containing all diazotrophic MAGs using "anvi-gen-contigs", which identified genes. We then ran "anvi-run-kegg-kofams" and "anvi-run-ncbi-cogs" to identify functions in the CONTIG database. We finally used "anvi-export-functions" to generate a matrix table describing functional annotations stored in the CONTIG database.