

## Gionfriddo-et-al-*hgcAB*-amplicon-tutorial

---

This repository contains files for a step-by-step tutorial for processing and classifying mercury methylation genes (*hgcAB*) from a mock community dataset ([Gionfriddo et al. 2020](#)). The full-set of paired-end fastq files from the published study can be downloaded from the NCBI SRA database under BioProject: [PRJNA608965](#). The tutorial dataset contains *hgcAB* amplicon sequencing of two mock communities. Each community is a mix of three cultured Hg-methylator isolates from the *Deltaproteobacteria*, *Methanomicrobia* (*Euryarchaeota*), and *Firmicutes*. The dataset also includes a salt marsh sediment sample (ID: 1064), and the marsh sample spiked with the two mock communities (1064 + mock community 1, 1064 + mock community 2). The *hgcAB* genes were amplified with primers ORNL-HgcAB-uni-F, ORNL-HgcAB-uni-32R (Gionfriddo et al. 2020) and sequenced on Illumina Miseq 2x300 bp. The *hgcAB* amplicon is ~980 nt bp long, and therefore with short-read sequencing employed in this study, the forward and reverse reads do not overlap. Only the forward reads are used for downstream analyses. Please see the [paper](#) for a more detailed explanation of the methods. The trimmed (201 nt bp) forward read *hgcA* sequences are classified using the [ORNL compiled Hg-methylator database](#) reference package for short sequences, 'ORNL\_HgcA\_201.refpkg'.

### The contents of the repository include:

- **Gionfriddo-et-al-*hgcAB*-amplicon-tutorial-directions.pdf:** step-by-step instructions
- **The raw paired-end sequencing files for five samples:**
  - 1064\_Alt35\_S186\_R1\_001.fastq.zip
  - 1064\_Alt35\_S186\_R2\_001.fastq.zip
  - m1combo1\_Alt30\_S69\_R1\_001.fastq.zip
  - m1combo1\_Alt30\_S69\_R2\_001.fastq.zip
  - m2combo2\_Alt30\_S103\_R1\_001.fastq.zip
  - m2combo2\_Alt30\_S103\_R2\_001.fastq.zip
  - m1064\_1combo1\_Alt35\_S73\_R1\_001.fastq.zip
  - m1064\_1combo1\_Alt35\_S73\_R2\_001.fastq.zip
  - m1064\_2combo1\_Alt35\_S84\_R1\_001.fastq.zip
  - m1064\_2combo1\_Alt35\_S84\_R2\_001.fastq.zip
- **sample\_ids.txt:** text file of sample information/IDs for fastq sequencing files
- **ORNL\_HgcA\_201.refpkg:** the [ORNL compiled Hg-methylator database](#) reference package
- **HgcA\_201\_hmm:** hidden markov model of HgcA homology from **ORNL\_HgcA\_201.refpkg**
- **metadata.csv:** table of metadata for sample IDs to be used with R scripts
- **Gionfriddo-HgcA-tutorial-make-taxonomy-table.R:** custom R script for assigning taxonomy classifications from pplacer to OTUs
- **Gionfriddo-HgcA-tutorial-make-bar-plot.R:** custom R script for making a bar plot of relative abundances of OTUs

Gionfriddo, C.M., Wymore, A.M., Jones, D.S., Wilpiseski, R.L., Lynes, M.M., Christensen, G.A., Soren, A., Gilmour, C.C., Podar, M. and Elias, D., 2020. An improved *hgcAB* primer set and direct high throughput sequencing expand Hg-methylator diversity in nature. *BioRxiv*.