**MarineDNA Hackathon Meeting**

**2/6/2023**

Nastassia Patin, Eric Archer, Kimani Kimbrough, Katie Pitz

**Big picture issues**

* How to determine “best” clustering algorithm(s) for eDNA data sets?
* All clustering will show structure; we need to determine which structure is most suitable for our questions, i.e. connectivity among trophic levels, or between sequences and environmental metadata
* Can we leverage computing power to quickly test multiple clustering approaches on each data set?
* NVP: One priority should be connectivity between 16S amplicons and higher-order taxa of interest, e.g. HAB phytoplankton species or other indicator taxa
* Jupyter notebooks need to be well-commented and commit notes provided on each GitHub push so all team members understand the application of each script and each section of code

**Hackathon mentor questions**

* GPU vs CPU for MCMC burn-in speed
* Benchmark timing for HMC burn-ins, effective sample sizes
* EA: HMC more efficient than MCMC