**Sep, 2016**

**Introduction to Metagenome binning**

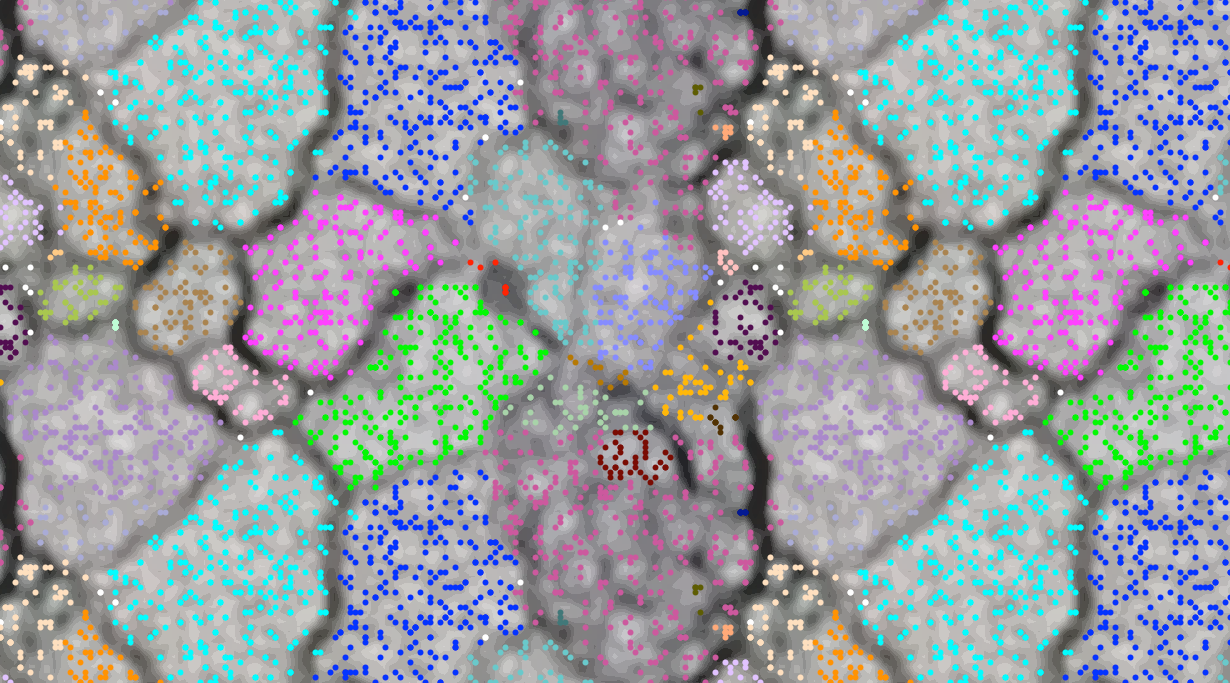
We will use an IMG available metagenome to bin contigs using MetaBAT and ESOM (based on tetranucleotide frequency similarity).

(1) Environment to analyze “Marine sediment microbial communities from the Hellenic Volcanic Arc” (Tax\_ID: 3300002231)

Scaffolds used in this exercise > 50 kb

Export scaffolds: Fasta file

Save scaffolds



**Guide to Use/Download ESOM**

ESOM website: <http://databionic-esom.sourceforge.net/>

1. Download (esom-1.1-installer.jar)
2. Tutorial to cover: in the website
3. ESOM input files: “.lrn” , “.cls”, “.names”

(4) Script: *prepare\_esom\_files.pl* (provided, courtesy of Dr. Itai Sharon)

perl prepare\_esom\_files.pl <flags> <output> <input.fa>

Flags: -w (Window size: default 3 k)

- k (K-mer size: default 4)

\*Be sure to have Installed the “Bio::SeqIO” module

modules installation: <http://triopter.com/archive/how-to-install-perl-modules-on-mac-os-x-in-4-easy-steps/>

1. Opening ESOM:
   1. Mac Applications (default saving location)
   2. ESOM/bin/esomana
   3. To add more memory locally edit –Xmx 5000 (esomstart) ../../../Applications/ESOM/bin/esomstart
   4. Upload “.lrn” file // “.cls” and “.names”
   5. Training
      1. Algorithm: k-batch
      2. #epochs (keep it low to save time)
      3. columns / rows (5-6 X neurons than datapoints)
      4. rs (24-50…)
   6. Upload “.wts” file
2. Visualizing / UI
   1. View data
   2. Create classes (save classes)

NOTE\* : All these files and the script© are provided to save the computing time needed.

**Guide to Use/Download MetaBAT**

https://bitbucket.org/berkeleylab/metabat/overview

Requirements:

* boost >= 1.55.0
* python >= 2.7
* scons >= 2.1.0
* g++ >= 4.8
* zlib >= 1.2.4
* binutils >= 2.2.2

(samtools 1.2 and igraph 0.7.1 are downloaded and installed automatically)

#stable release version

wget <https://bitbucket.org/berkeleylab/metabat/get/master.tar.gz>

tar xzvf master.tar.gz cd berkeleylab-metabat-\*

#run the installation script

scons install PREFIX=$HOME [BOOST\_ROOT=$BOOST\_ROOT] [SAMTOOLS\_DIR=$SAMTOOLS\_DIR] [HTSLIB\_DIR=$HTSLIB\_DIR]