

Tutorial #1a

I would like to match you two up for the first tutorial that you will do on April 25th. It is expected to be at least 30 minutes (so each of you should spend 15 minutes each).

Taha/Richard – I will want you two to split the teaching duties and go over the basics – 1) how to do a t-test, 2) ANOVA (one-way, two-way), 3) MANOVA vs. ANOVA, and 4) multiple comparison corrections (Benjamin-Hochberg vs. Family-wide error rate). **Please see if the mouse dataset that Zhengqiao will use next week has enough classes in the OTU (species vs. sample) table to show most of the methods.**

Tutorial #1b

You are assigned Tutorial 1b (which I don't have defined in the document). You will present on this topic on April 25th. Please pair up and get started on the tutorial NOW!

This is what I want you to review: the Microbiome R package, specifically these functions:

Alpha Diversity: <https://microbiome.github.io/microbiome/Diversity.html>
(please review how the various metrics are calculated like Chao diversity, simpson/shannon, etc.)

Beta Diversity: <http://microbiome.github.io/microbiome/Betadiversity.html>

Univariate comparison of a taxon from microbial communities using linear mixed model:
<http://microbiome.github.io/microbiome/Mixedmodels.html>

Multivariate community-level differences:
<http://microbiome.github.io/microbiome/PERMANOVA.html>

Present on some of the datasets they use, so you can understand the results
[:http://microbiome.github.io/microbiome/Data.html](http://microbiome.github.io/microbiome/Data.html)

Tutorial #3

I have assigned you both to the METABAT tutorial for Tutorial 3 to present on **May 2nd**. You will have to install the package.

You will review the method
<https://peerj.com/articles/1165/>

and be able to fully explain what is going on in this tutorial:

<https://bitbucket.org/berkeleylab/metabat/wiki/Best%20Binning%20Practices>

I also will require that you schedule a meeting with Saeed Keshani (copied on this email) to meet once to go over the MetaBat tutorial to clear up any questions that you may have:

<https://bitbucket.org/berkeleylab/metabat/wiki/Best%20Binning%20Practices>

Tutorial #4

You two will team up for Tutorial 4 and will install Kraken. You will do this tutorial but only for Kraken2 (not Bracken or Centrifuge or the rest):

<https://genomics.sschmeier.com/ngs-taxonomic-investigation/index.html>

Instead of using the sorted-unmapped files, just use the raw/trimmed version of the R1/R2 files which you can get here:

type "wget <http://compbio.massey.ac.nz/data/203341/trimmed.tar.gz>"

Present on the technique used in Kraken:

<https://genomebiology.biomedcentral.com/articles/10.1186/gb-2014-15-3-r46>

The tutorial must be at least 30 minutes.

You will team up to do Tutorial 5: Kaiju classification. Your presentation will be on **May 9**. It must be at least 30 minutes, and talk about how the method works and a demonstration of using the webserver and visualizing the results. The paper is here:

<https://www.nature.com/articles/ncomms11257>

Tutorial #5

You will benchmark the Kaiju webserver.

<http://kaiju.binf.ku.dk/server>

You will compare "Reference Databases": 'RefSeq Genomes' vs. 'NCBI BLAST nr +euk'. You will also compare MEM and Greedy "Run Mode". In the end, you will have four comparisons -- showing how the taxonomic content changes in these 4 combinations.

Get DATA here:

"wget <http://compbio.massey.ac.nz/data/203341/trimmed.tar.gz>"

USE evolved-6-R1.fastq.trimmed.gz

Tutorial #6

Port it to work with Proteus and make a Markdown similar to their Github but that people can run on Proteus.

Do this tutorial: https://github.com/LangilleLab/microbiome_helper/wiki

Port it to work with Proteus and make a Markdown similar to their Github but that people can run on your compute cluster.

Tutorial #8

Tutorial 8: Alignments: Students will align H. Influenzae 16S genes using Cipres (<http://www.phylo.org/>). Students will compare Muscle and MAFFT alignments and comparing time that it takes.

In this tutorial, the students will be expected to dedicate a large portion of the class to how the underlying algorithms work for how the alignment algorithms work. An analysis of the pros and cons of each of the 2 methods mentioned above are expected. Please work closely with the TA on this assignment.

The sequences for the tutorial-

16S for Haemophilus influenzae:

/mnt/HA/groups/nsftuesGrp/data/Haemophilus_influenzae_16S.fasta

Tutorial #9

Tutorial #9: Tree Inference:

Students will align H. Influenzae 16S genes using Cipres (<http://www.phylo.org/>) using the **MAFFT** algorithm. Then students will build the trees using RaxML and FastTree. The students will comment on differences in the trees and compare the RaxML and FastTree algorithms.

In these tutorials, the students will be expected to dedicate a large portion of the class to how the underlying algorithms work for tree inference. An analysis of the pros and cons of each of the 2 methods mentioned above are expected. Please correspond with the TA.

Dataset:

16S for Haemophilus influenzae:

/mnt/HA/groups/nsftuesGrp/data/Haemophilus_influenzae_16S.fasta

Tutorial #10

Group last minute changed to Gene Set Enrichment Analysis