# **KBS 2014 PRE SURVEY**

# Please indicate your level of ABILITY with each of the following concepts, tools, or approaches.

1. Navigating terminal
   * High Ability
   * Intermediate Ability
   * Low Ability
   * No Ability
2. Executing Python scripts
3. Using Amazon EC2
4. Installing and running mothur
5. Installing and running QIIME
6. Executing R scripts
7. Choosing ecological statistics to analyze microbial community data
8. Visualizing patterns in microbial community data
9. Interacting with microbial metagenomic databases
10. Assembling metagenomes and determining assembly quality

|  |
| --- |
|  |

## Please indicate your degree of agreement with each of the following statements.

I understand what “cd” means.

* Strongly Agree
* Agree
* Disagree
* Strongly Disagree

I understand what “.py” signifies.

I know how to process Illumina data.

I understand what per\_library\_stats.py does.

I know how to run R.

I know the main differences in analyses offered by QIIME and mothur.

I am familiar with .biom formatted files.

I can name at least two different microbial metagenomic databases.

I know what an “R package” is.

I understand the structure of an OTU table.

I know what a “k-mer” is.

I know how to start an EC2 instance.

I know the difference between alpha and beta diversity.

I know how to visualize microbial metagenomic data

I know how to use metadata to guide community analyses.

I know how to assemble shotgun metagenomic data.

## Please indicate your degree of agreement with each of the following statements.

I can independently execute the steps of microbial **shotgun** metagenomic analysis from raw data processing to statistics.

* Strongly Agree
* Agree
* Disagree
* Strongly Disagree

I can independently execute the steps of microbial **amplicon** metagenomic analysis from raw data processing to statistics.

I can make informed choices for analysis parameters (e.g., OTU clustering, tree building, denoising) for **amplicon** analyses in QIIME.

I can make informed choices for analysis parameters (e.g., OTU clustering, tree building, denoising) for **amplicon** analyses in mothur.

TWO STATEMENTS: I understand the differences between and limitations of unassembled and assembled shotgun metagenomic data.

TWO STATEMENTS: I understand what digital normalization is and why I might use it.

TWO STATEMENTS: I understand what a k-mer is and what it means in the context of assembly.

I can make informed choices about the programs to use for assembling **shotgun** metagenome data.

I can determine the quality of a shotgun metegenome assembly.

I can use an R package to analyze data.

BE SPECIFIC: I can move data out of one tool/program (e.g., mothur) and into another (e.g., R).

I can find and move files on a computer using terminal.

I can evaluate the quality of raw sequencing data.

I can deposit a sequencing dataset and make it publicly accessible.

I can search for public datasets that are relevant to my own project.

I CANNOT process and quality control **shotgun** metagenome data unless someone else helps me get started.

I CANNOT process and quality control **amplicon** metagenome data unless someone else helps me get started.

I can complete an analysis project if I have a lot of time to complete the project.

CHANGE “PI” TO “MENTOR” FOR PEOPLE WHO MAY NOT USE “PI”: I can complete an analysis project if I have just the help of my research PI for assistance.

I can find ways of overcoming the problem if I get stuck at a point during data analysis.

I CANNOT complete a data analysis project unless I can call someone for help if I get stuck.

I can find ways to motivate my data analysis, even if the research question is of no interest to me.

I can design an experiment collecting 16S marker gene data that someone else can comprehend.

I understand the technical limitations of metagenomic data.

## Please answer the following questions to the best of your ability in a few sentences.

Suppose that you are doing a 16S marker analysis of the gut microbial community from a non-model worm. This worm has a unique digestive tract evolved for the breakdown of a specialized diet. Discuss a sequencing strategy, a number of challenges you expect to face, and what kinds of additional resources are needed to help explore your data set.

Suppose that you are using the Illumina platform to sequence total DNA from soil samples. For this particular study, the soil samples come evenly distributed from two plot types on a single farm - one growing corn, the other dairy cow pasture. Discuss one or more approaches you would take to analyze the data, as well as your expected sensitivity and specificity to taxonomic and functional sequence identification. Include in your discussion the technical limitations of your analysis.

## Please circle the number that corresponds to your rating of each aspect of the workshop.

Day 1: Shell/Terminal

Day 2: Diversity/amplicon analyses / QIIME

Day 3, part a: EC2

Day 3, part b: amplicon analyses/ mothur

Day 4: Shotgun metagenomics

Day 5: Metagenome assembly

Day 6a: R

Day6b: Data visualization

Day 7: Databases

Day 8:

OVERALL WORKSHOP RATING

* Very Poor
* Poor
* Adequate
* Good
* Very Good

OVERALL INVITED SPEAKER RATING

Any comments you would like to add (particular speakers/topics that were especially distinctive)

HPCC

Box and Arrow Diagram:

Congratulations! You have discovered a new microbial community growing on Martian soil. You want to use the Illumina platform to evaluate both the phylogenetic diversity and functional potential of this community. If you succeed, you will be short-listed for a Nobel prize.

a. **Please draw a boxes-and-arrows process diagram** laying out the different steps involved in going from your raw sequencing data to your final "answer".

b. If you have time, **write some ideas** for what kinds of computational and experimental cross-validation and comparisons you can do, and where in the process.