Course Description:
Microbiomes, microbes and their genetic material present in a host/environment, are linked to risk to disease in humans, animals and plants. Metagenomics, including 16S rRNA community survey methods and shotgun metagenomics, use high throughput sequencing technology to provide insight towards composition and potential function of microbiomes. This course provides hands on experience in using bioinformatics and statistical tools necessary to process and analyze the resulting large datasets.

Topics:
Week 1:
- An introduction to microbial ecology
- Highlight of data generation and high throughput methods
- The Galaxy interactive environment: An Orientation and basics of use

Week 2:
- Sequence Data + Quality view: the fastq and fasta files + FastQC and its output
- A Linux primer
- Preprocessing high throughput data: trimmomatic
- Metagenomic databases (JGI IMG, DACC, SILVA, GreenGene, RDP, …)

Week 3:
- Data processing and classification (or not): Mothur (16S)
- Data processing and classification (or not): Metagenomics

Week 4:
- An overview of R: R as a statistical and programming environment

Week 5:
- The normalization problem: Rarefying, quantile normalization, CCS
- Exploratory, multivariate data analysis (vegan): PCA, PcoA, NMDS, Clustering

Week 6:
- Comparative analysis: perMANOVA, RDA, metagenomeseq, DESeq2 and edgR

Week 7:
- Predictive approaches: Random forests, glmnet, …
- Network analysis (R): Bayesian, Association and Co-occurrence Networks
- Microbial Genome Alignment: Downloading and Installing Tychus

Week 8:
- Microbial SNP Calling and Consensus recovery and phylogenetic trees
- Microbial Genome de novo assembly and annotation: Tychus
- Highlights of Phylogenetics and phylogenomics

Exams and Homework:
We will have 4 homework assignments during weeks 1, 3, 5, and 7 each worth 25% of the total grade. These will be due Monday of week 3, 5 and 7 and last lecture on week 8, respectively. No late assignments will be accepted.