

# Microbiome Analysis

Research Day 2012

Ranjit Kumar



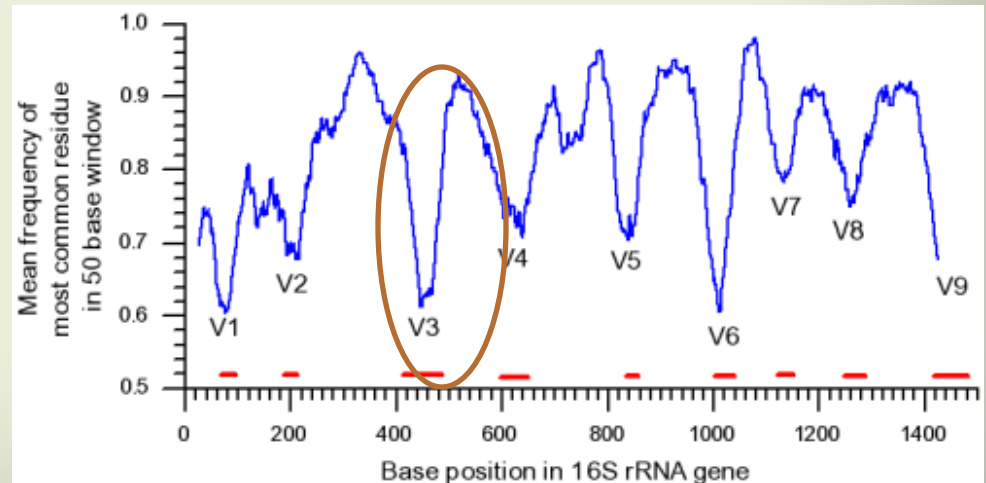
# Human Microbiome

- Microorganisms – Bad or good ?
- Human colon contains up to 100 trillion bacteria.
- Human microbiome - The community of bacteria that live on/in the human host (mucosal surfaces, skin, tooth surface, etc.) Beneficial/neutral/adversarial.
- Numerous studies have suggested that shifts in the populations of microbial communities may be associated with acute and chronic diseases: inflammatory bowel disease, obesity, cardiovascular disease, eczema and other skin diseases, vaginal infections.



# Studying microbiome : 16S rRNA gene sequencing

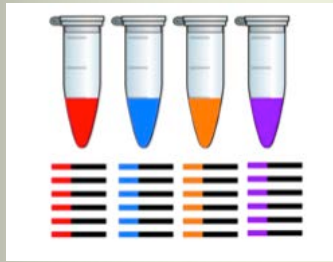
- 16S rRNA gene is found in all bacterial species, contains regions which are highly conserved and highly variable sequence.
- Variable sequence can be thought of as a molecular “fingerprint”. Can be used to identify bacterial genera and species.
- Large public databases available for comparison.
  - RDP Ribosomal Database
  - Greengenes
  - ARB-Silva



# Next Generation Sequencing

- ✓ Culture independent study
- ✓ Low quantity of sample needed (20-50ng DNA)
- ✓ High sequencing depth (identification of rare microbes).
- ✓ Multiplexing of many different samples in one run using indexes (up to 40 barcodes).
- Large scale sequencing of 16S rDNA gene amplicons. Different technologies have different read length and cost factors.
  - 454 : 450K – 500K reads per run
  - Illumina GAIIx : 10M – 40M reads/lane\*
  - Illumina HiSeq : 350M reads/lane\*
  - Illumina MiSeq (new) : 8M reads/lane\*

# Microbiome Analysis in Nutshell



Extract DNA and amplify 16S gene with barcoded primers



Next Generation sequencing using GAIIX



```
>GCACCTGAGGACAGGCATGAGGAA...
>GCACCTGAGGACAGGGGAGGAGGA...
>TCACATGAACCTAGGCAGGACGAA...
>CTACCGGAGGACAGGCATGAGGAT...
>TCACATGAACCTAGGCAGGAGGAA...
>GCACCTGAGGACAGGCAGGACGAC...
>CTACCGGAGGACAGGCAGGAGGAA...
>CTACCGGAGGACACACAGGAGGAA...
>GAACCTTCACATAGGCAGGAGGAT...
>TCACATGAACCTAGGGGCAAGGAA...
>GCACCTGAGGACAGGCAGGAGGAA...
```

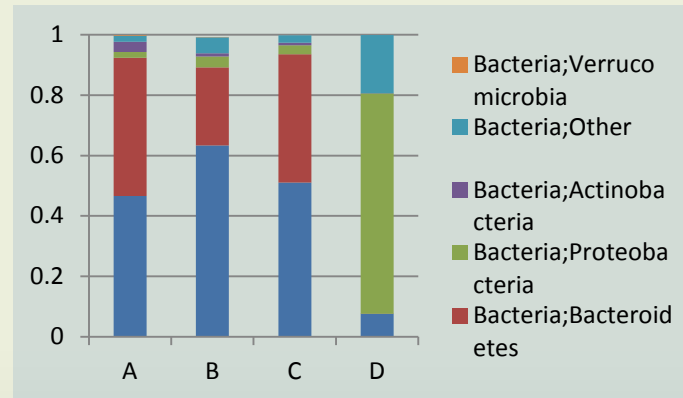
Assign reads to communities



- Sample Demultiplexing
- Quality Control
- Sequence clustering into OTUs (Operational Taxonomic Units)
- Pick representative sequences
- Assign Taxonomy
- phylogenetic tree



16S rRNA database



Bacterial community distribution

- ✓ Heatmap
- ✓ Phylogenetic tree
- ✓ Alpha diversity
- ✓ Beta diversity
- ✓ Statistical tests
- ✓ OTU correlation
- ✓ PCoA

# Microbiome Facilities at UAB

1. Sample preparation.
  - Processing DNA and 16S DNA PCR - Dr. Casey Morrow's lab.
2. Sequencing : Helfin genomics center
  - GAllx, HISEq, MISEq – Dr. Mike Crowley.
3. Data Analysis – Microbiome data analysis.
  - Data analysis by CCTS/BMI group : Dr. Elliot Lefkowitz
  - Software packages available QIIME\*, Mothur\*, R\*
4. Gnotobiotics facility

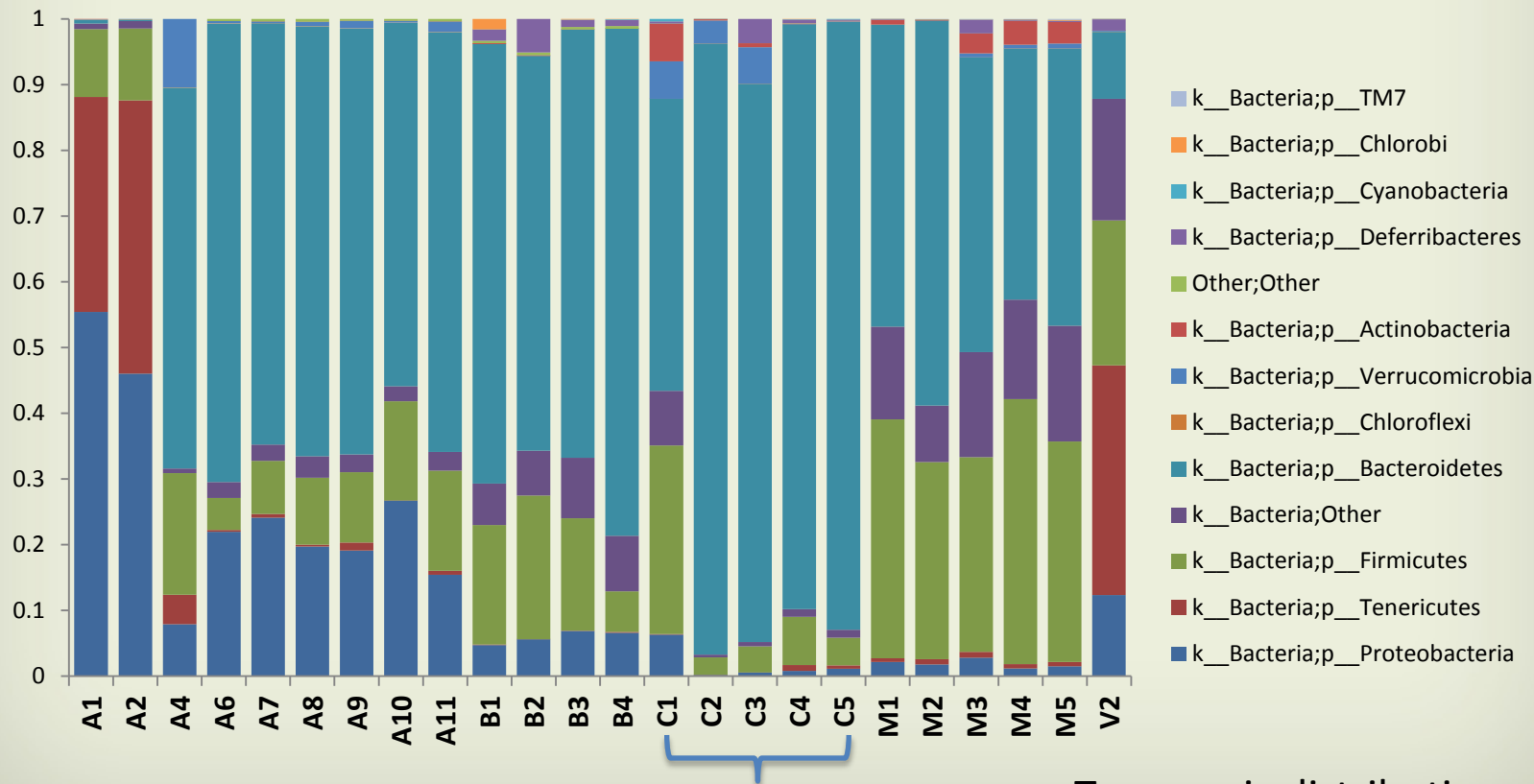
\* Programs installed at Cheaha Cluster

# Data Analysis and Cluster

- Recent microbiome run at UAB : GAIIX, 90 samples (8 projects), 7 lanes, 28GB ~ 120M reads.
- **Data Analysis Challenges : Sample grouping and de-multiplexing, Accuracy, Fast turnaround time, Storage space, parallel processing.**
- ✓ Cheaha cluster provides facilities for large scale data analysis.
- ✓ Programs installed QIIME, Mothur, R, FASTQC, FastX
- ✓ Developed automated scripts for parallel analysis.
- ✓ 3 day turnaround time for initial set of results.

# Example

- 24 samples in five group(A,B,C,M,V)
- Sample prefixed with C are control

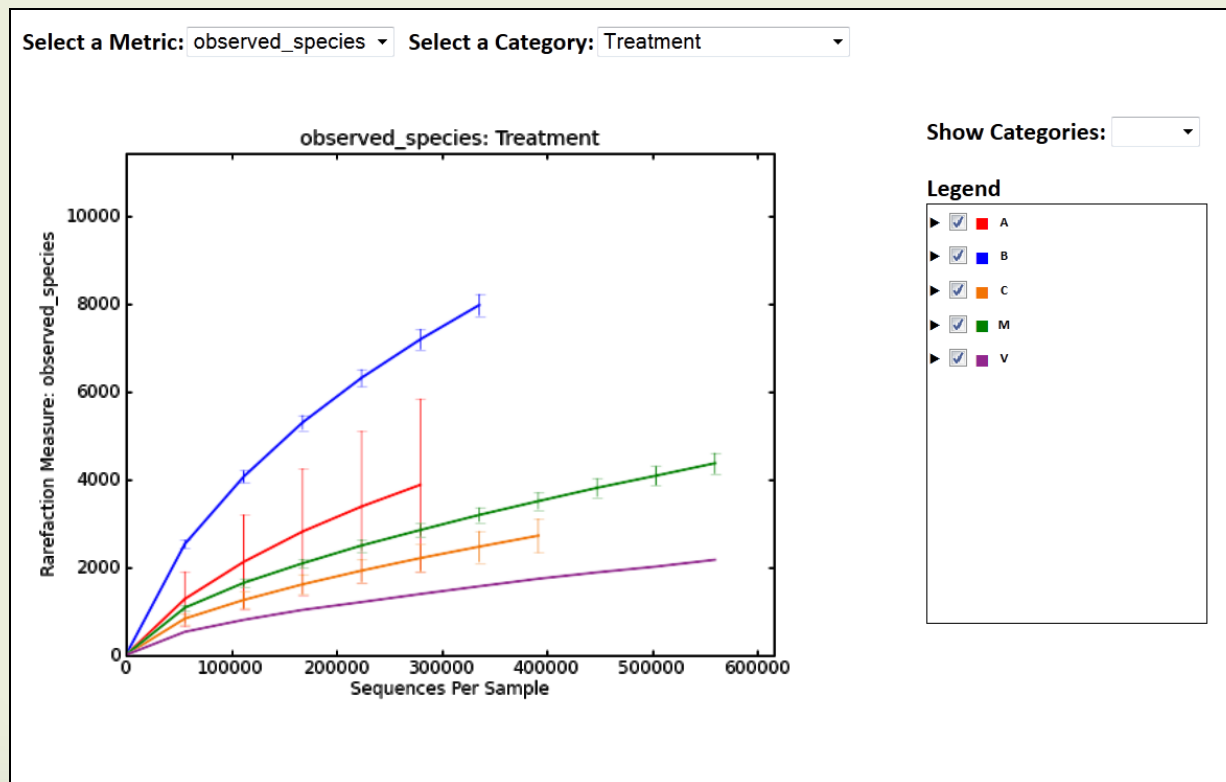


Taxonomic distribution



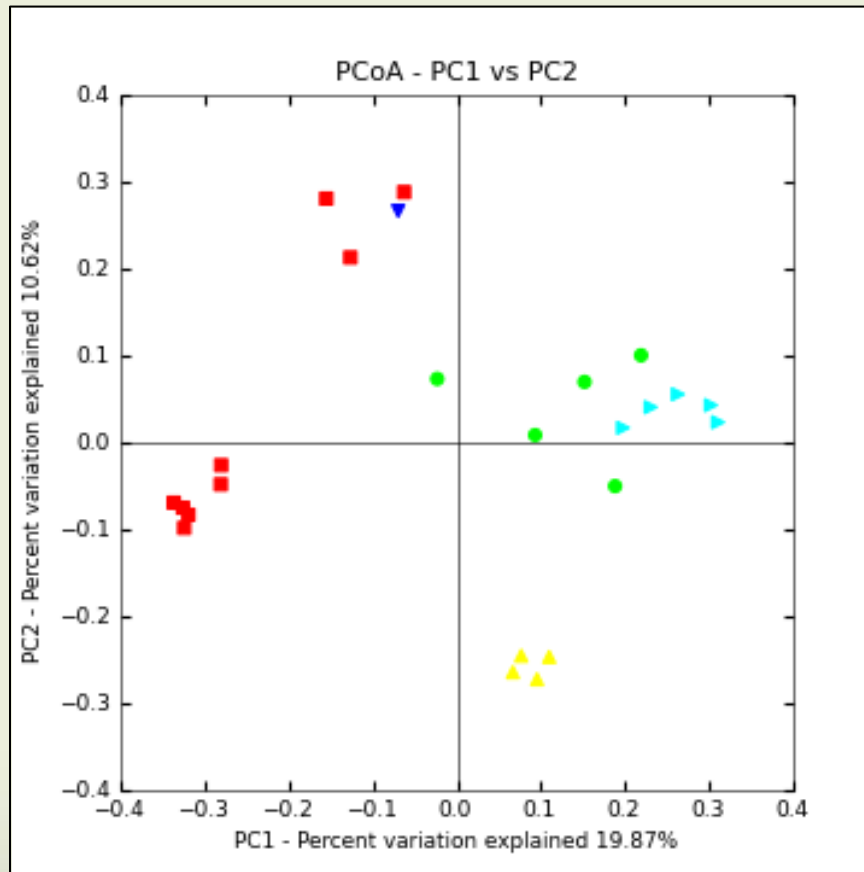
# Alpha diversity

Used to access the diversity within a sample



# Beta diversity

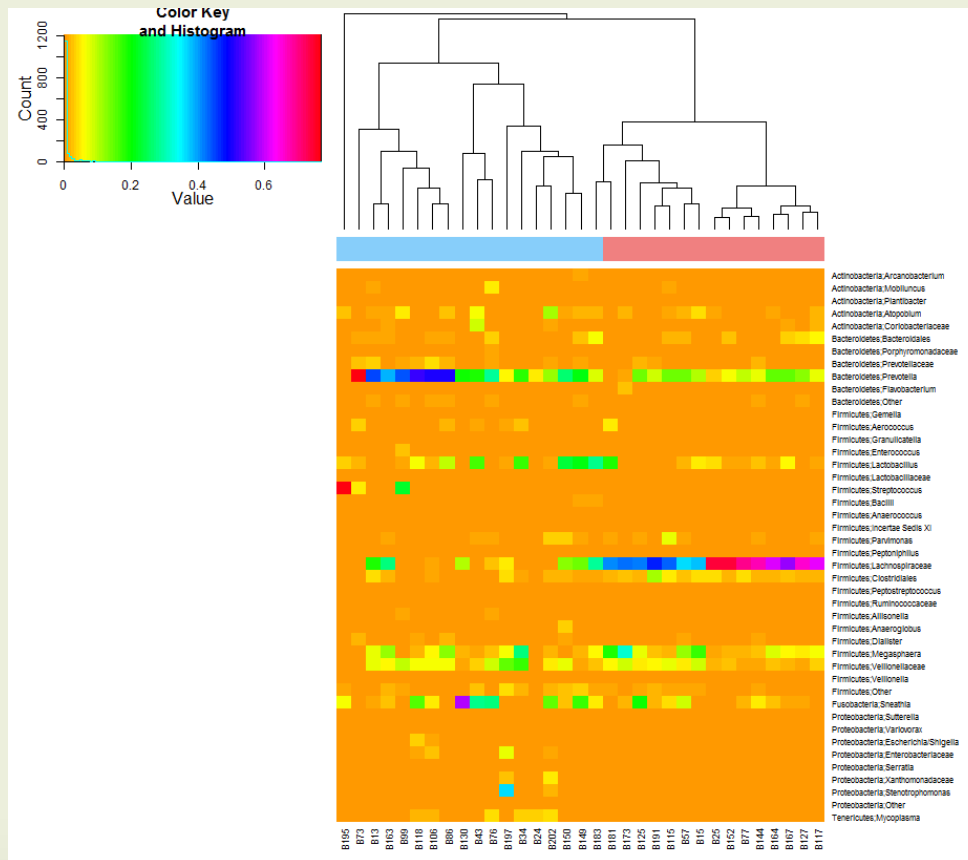
Used to assess the diversity between the samples



- A\*
- C\*
- B\*
- M\*
- V

PCoA plot

# Other plots : Heatmap



# UAB Biomedical Informatics Resources and Assistance

- **CCTS-BMI**
  - Elliot Lefkowitz, Ph.D.
    - [ElliotL@uab.edu](mailto:ElliotL@uab.edu)
    - (205) 934-1946
- **Heflin Center**
  - David Crossman, Ph.D.
    - [dkcrossm@uab.edu](mailto:dkcrossm@uab.edu)
    - (205) 996-4045
- **Microbiome Core Facility**
  - Casey Morrow, Ph.D.
    - [caseym@uab.edu](mailto:caseym@uab.edu)
    - (205) 934-5705
- **Section on Statistical Genetics (School of Public Health)**
  - Hemant Tiwari, Ph.D.
    - [Htiwari@soph.uab.edu](mailto:Htiwari@soph.uab.edu)
    - (205) 934-4907
- **Department of Pathology Division of Informatics**
  - Jonas Almeida, Ph.D.
    - [jalmeida@uab.edu](mailto:jalmeida@uab.edu)
    - (205) 975-3286
- **Comprehensive Cancer Center (CCC) Biostatistics and Bioinformatics**
  - Karan Singh, Ph.D.
    - [kpsingh@uab.edu](mailto:kpsingh@uab.edu)
    - (205) 996-6122