

MicrobiomeAnalyst

**-- a web-based tool for comprehensive
exploratory analysis of microbiome data**

Jianguo (Jeff) Xia, Assistant Professor

McGill University, Quebec Canada

jeff.xia@mcgill.ca | www.xialab.ca

June 26, 2017



Tools @ XiaLab

(<http://www.xialab.ca/tools>)

Metabolomics

<http://metaboanalyst.ca>

Metabo-
Analyst

Systems transcriptomics

<http://networkanalyst.ca>

Network-
Analyst

Microbiome-
Analyst

Microbiome

<http://microbiomeanalyst.ca>

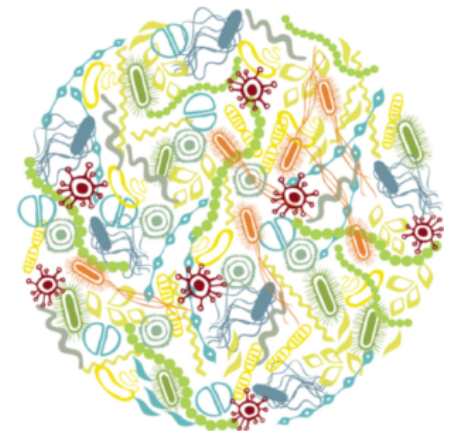
miRNet

MicroRNAs

<http://mirnet.ca>

Microbiome (I)

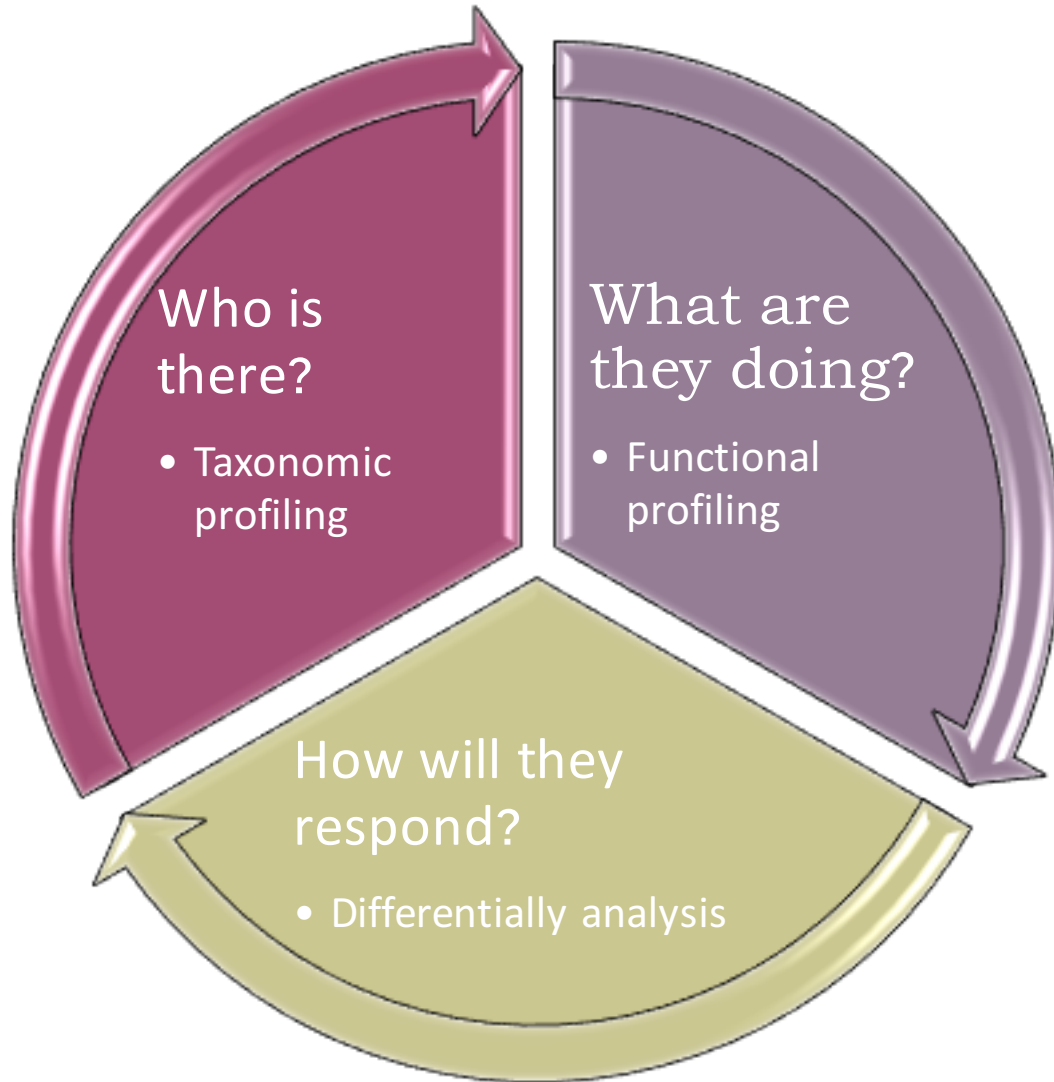
- All the microbes in a community (same as microbiota)
- The Human Microbiome is the collection of all the microorganisms living in association with the human body (source: Human Microbiome Project)
 - Gut microbiome
 - Oral microbiome
 - Vaginal microbiome
- Significant roles in health, disease and aging:
 - Obesity
 - Diabetes
 - Cardiovascular diseases
 - IBD



Microbiome (II)

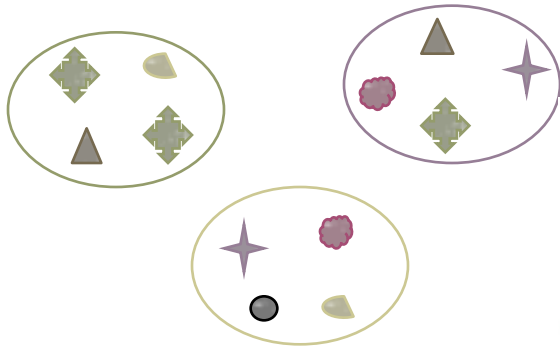
- The composition and structure of the microbiome is more fluid than human cells and organs
 - Host factors (bile acids, immunity, age)
 - Diet
 - Environment exposures
 - Medications
- Potential applications
 - Therapeutic targets
 - Diagnostic biomarkers
 - Precision medicine

Characterizing microbiome

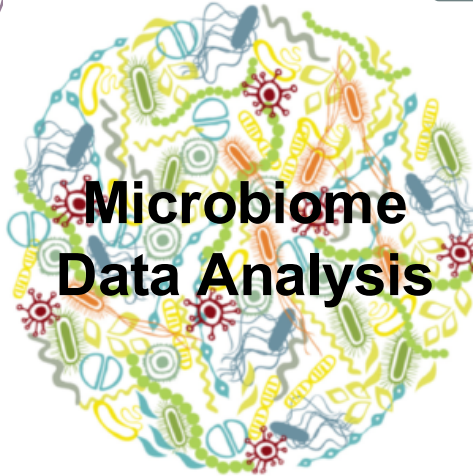
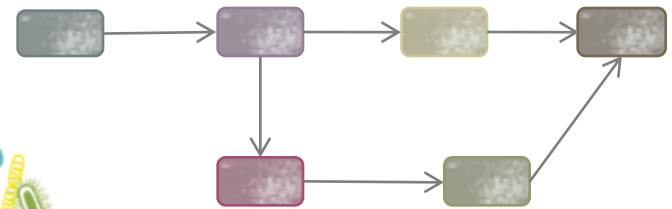


Microbiome data analysis

1. Taxonomic Profiling

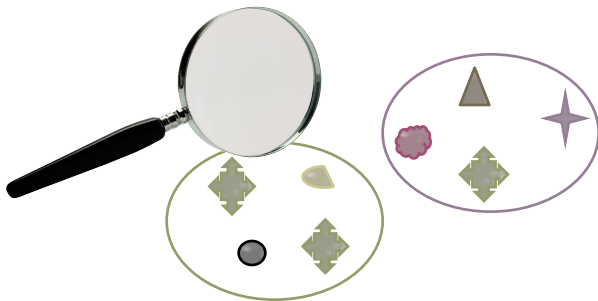


2. Functional Profiling

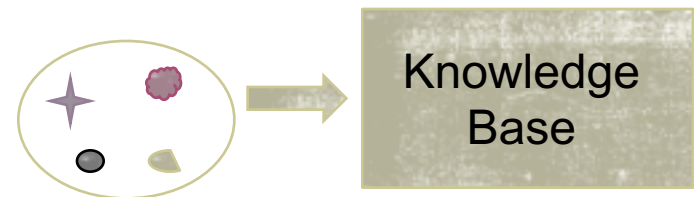


**Microbiome
Data Analysis**

3. Differential Analysis



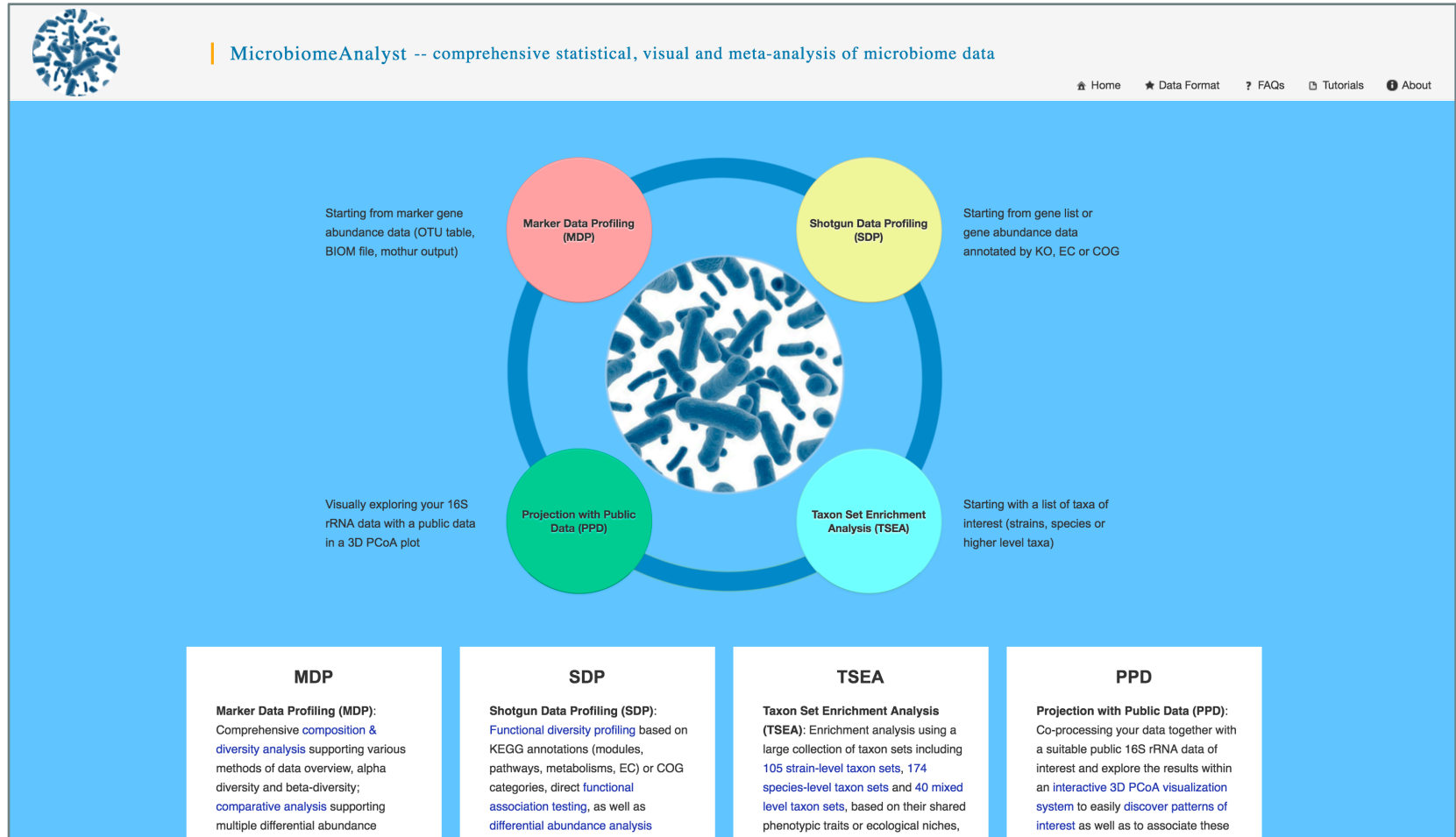
4. Meta-Analysis



Data analysis challenges

- Sparse data (too many zeros)
 - True absence of taxa across samples or under-sampling
- Vast difference in sequence depths across samples
 - Rescale read counts → issues with compositional data
 - Rarefying reads to equal sample depths → information loss
- Over-dispersion
 - Large variance in distribution

Introducing MicrobiomeAnalyst

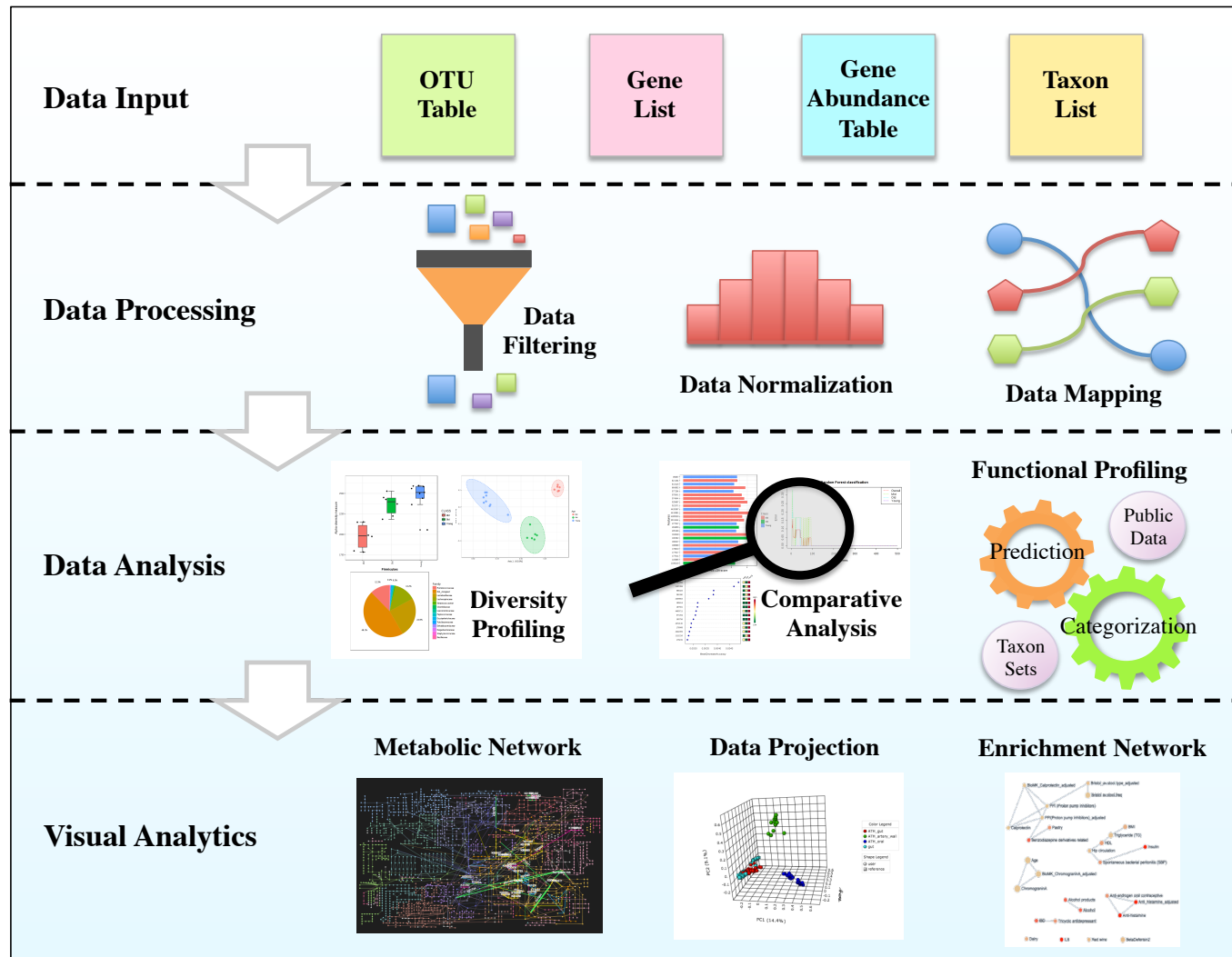


- Dhariwal, A., Chong, J., Habib, S., King, I., Agellon, LB., and Xia, J. (2017) "MicrobiomeAnalyst - a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data" Nucleic Acids Research (web server issue)

Feature overview

- Marker Data Profiling
 - 16S rRNA marker gene survey data
 - Plain text files or outputs from mothur, QIIME, BIOM files
- Shotgun Data Profiling
 - Metagenomics or meta-transcriptomics
- Projection with Public Data (PPD)
 - Marker gene data only
- Taxon Set Enrichment Analysis (TSEA)

MicrobiomeAnalyst Workflow



Statistical exploratory analysis

Community profiling

[Alpha diversity profiling & significance testing](#)

[Beta diversity profiling & significance testing](#)

[Core microbiome analysis](#)

Visual exploration & clustering analysis

[Sample-wise abundance \[stacked bar/area chart\]](#)

[Group-wise abundance \[interactive pie chart\]](#)

[Heatmaps & clustering analysis](#)

[Dendrograms & clustering analysis](#)

Differential abundance analysis

[Parametric & non-parametric univariate analysis](#)

[Statistical analysis for sparse high-throughput sequencing data \(metagenomeSeq\)](#)

[RNASeq methods - EdgeR & DESeq2](#)

Biomarker discovery & classification

[Correlation analysis](#)

[Linear Discriminant Analysis Effect Size \(LEfSe\)](#)

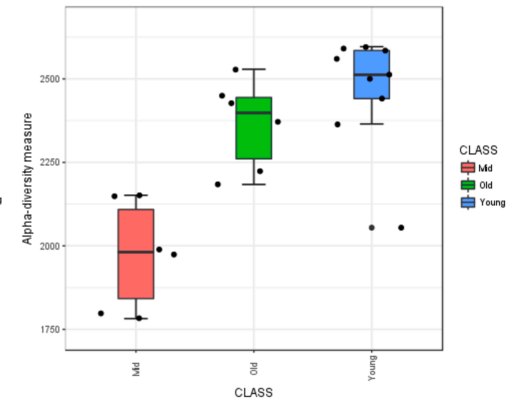
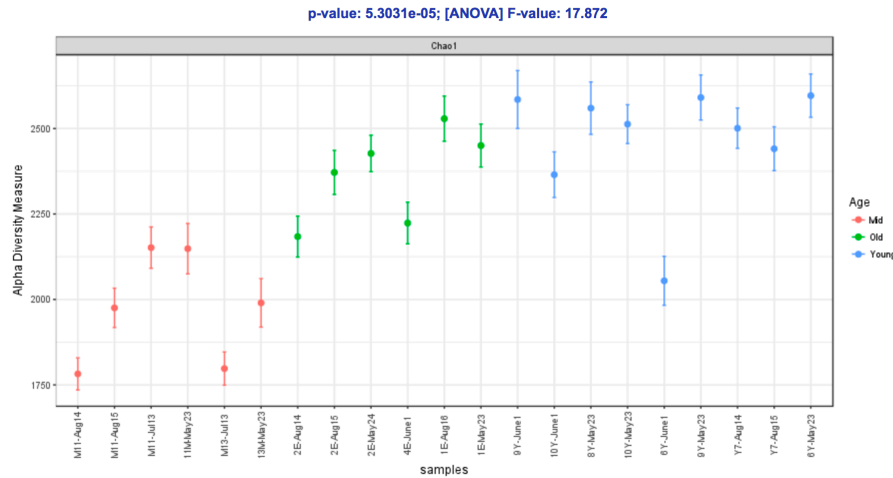
[Random Forests](#)

Prediction of Metabolic Potential

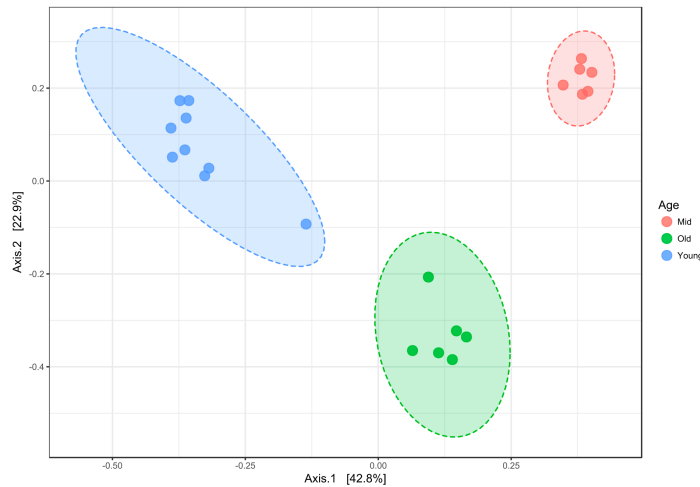
[Prediction for Greengenes Annotated OTUs \(PICRUSt\)](#)

[Prediction for SILVA Annotated OTUs \(Tax4Fun\)](#)

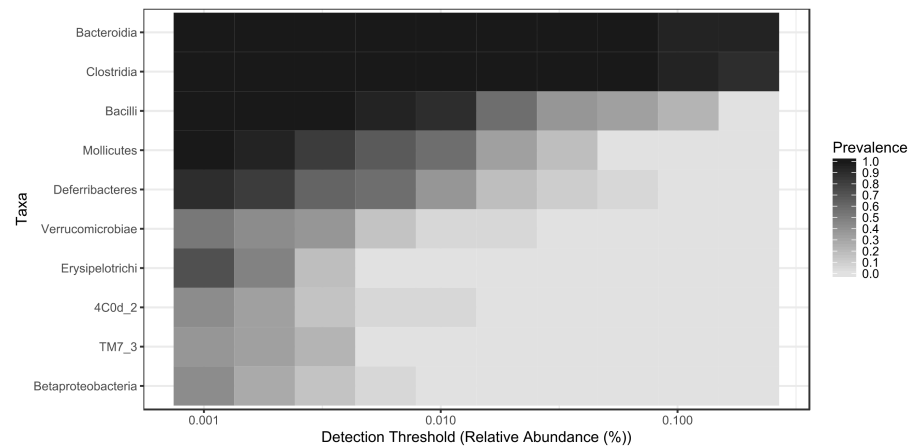
Taxonomic profiling



Alpha diversity



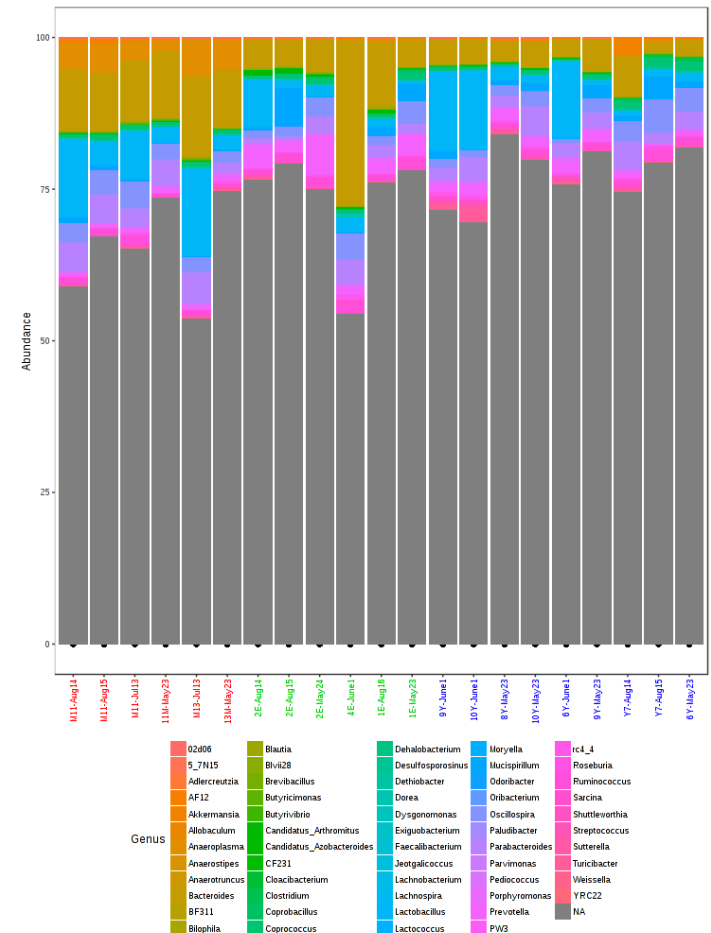
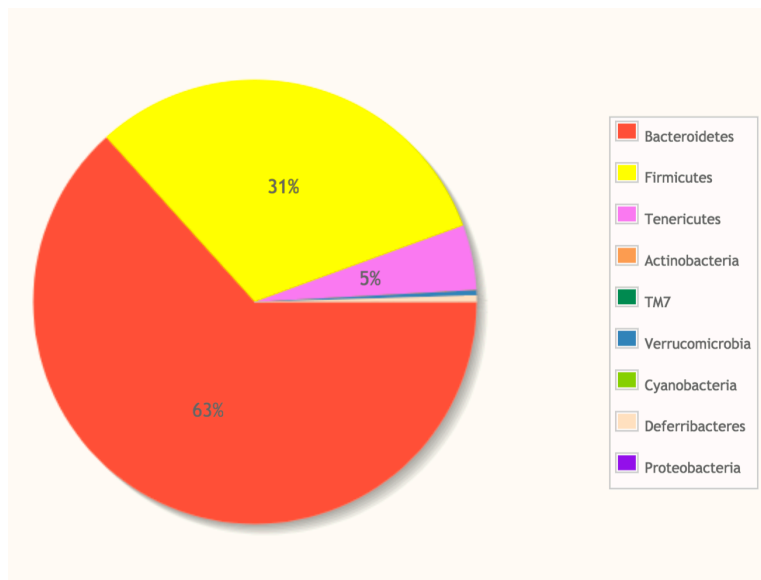
Beta diversity



Core microbiome

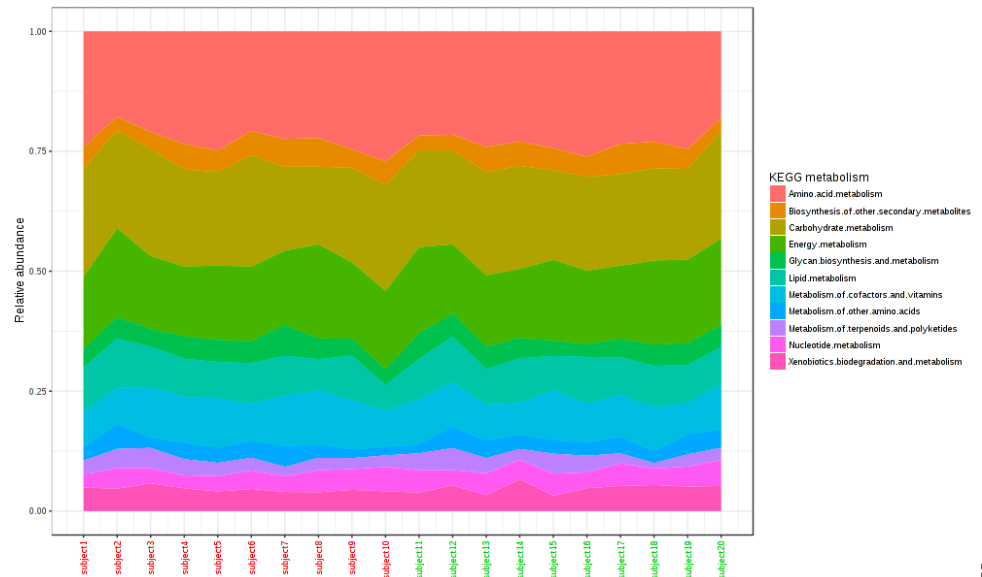
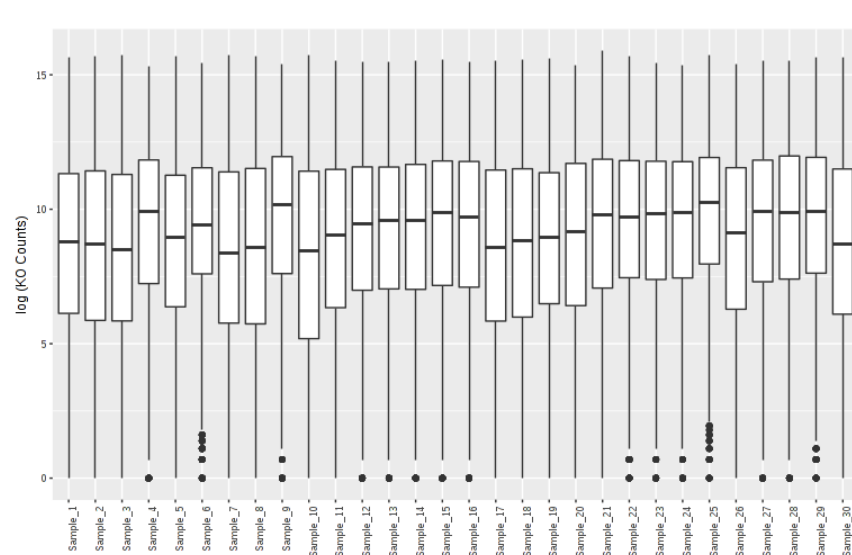
Flexible graphic generations

- At different taxonomic levels
- PNG, PDF, SVG
- Underlying tables

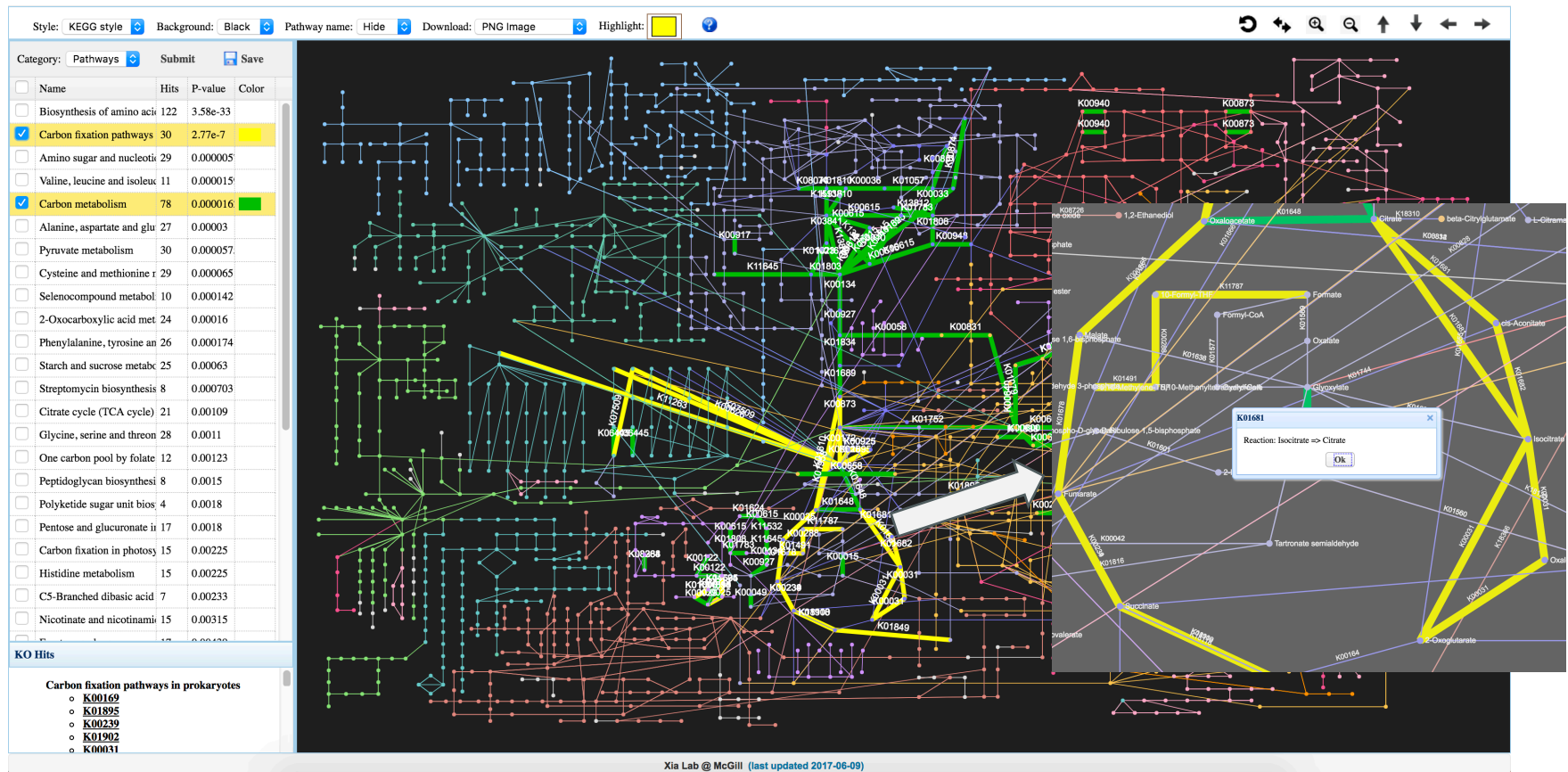


Predicting functional potentials

- Inference of metabolic profiles based on 16S rRNA
 - PICRUSt
 - Tax4Fun
- Functional diversity profiling
 - Annotation to COG and KEGG

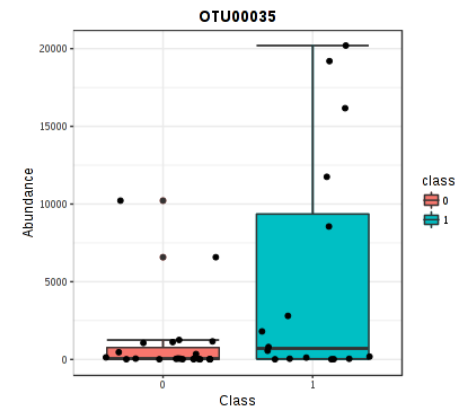
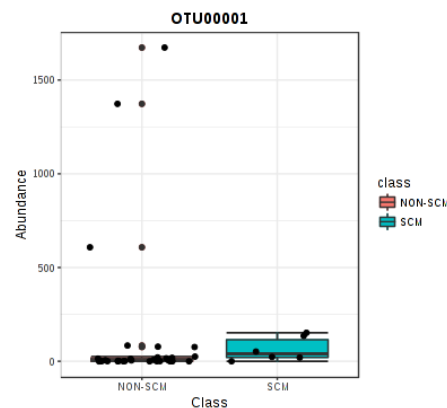


Explore metabolic potentials



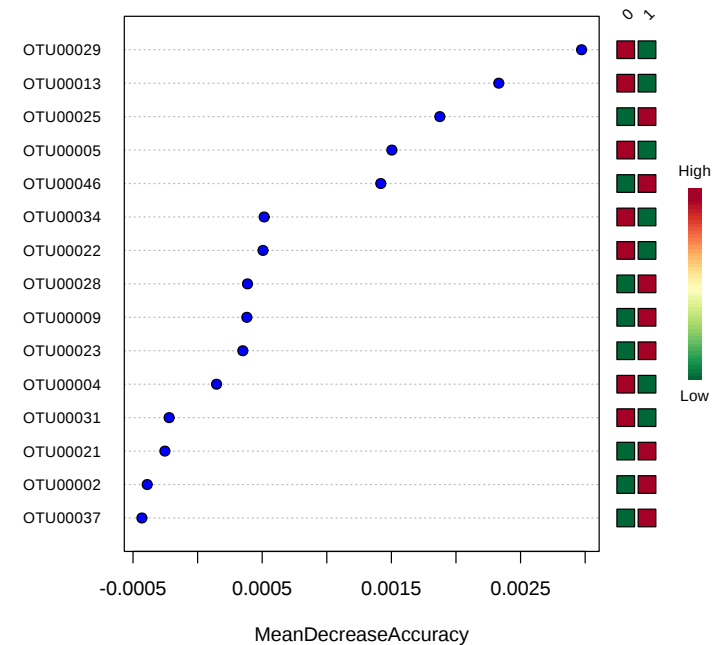
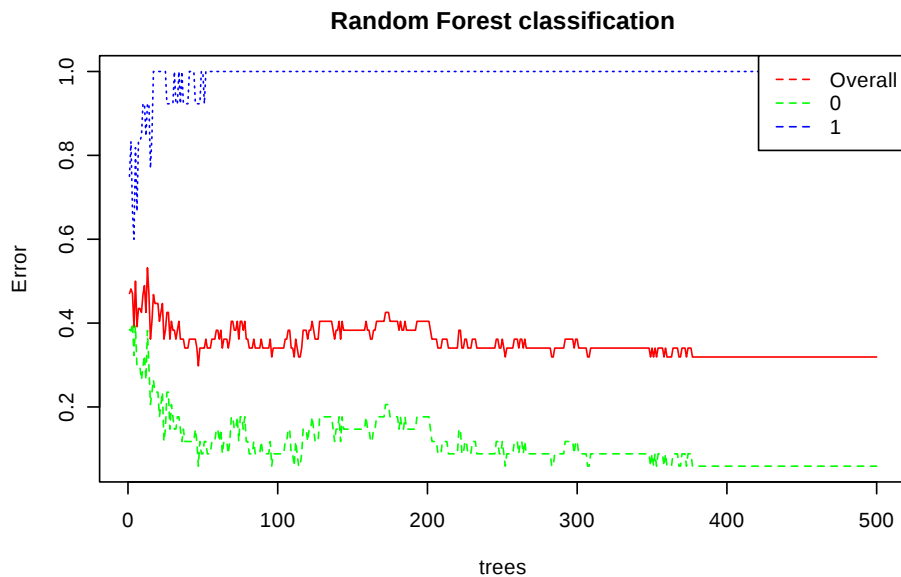
Differential abundance analysis

- Allows users to identify statistically significant features
 - At different taxonomy levels
 - Relating to different meta-data
- Implemented a suite of statistical methods
 - Common univariate analysis
 - Parametric
 - Non-parametric
 - metagenomeSeq
 - EdgeR
 - DEseq2



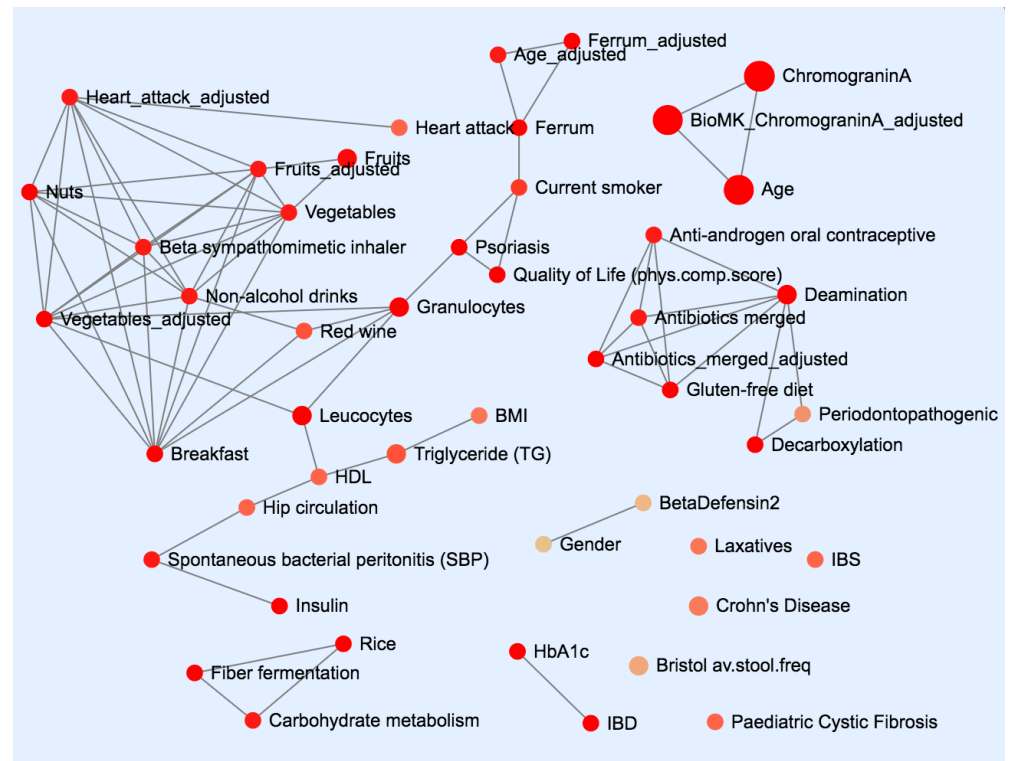
Biomarker analysis

- Linear Discriminant Analysis Effect Size (LEfSe)
- Random Forest

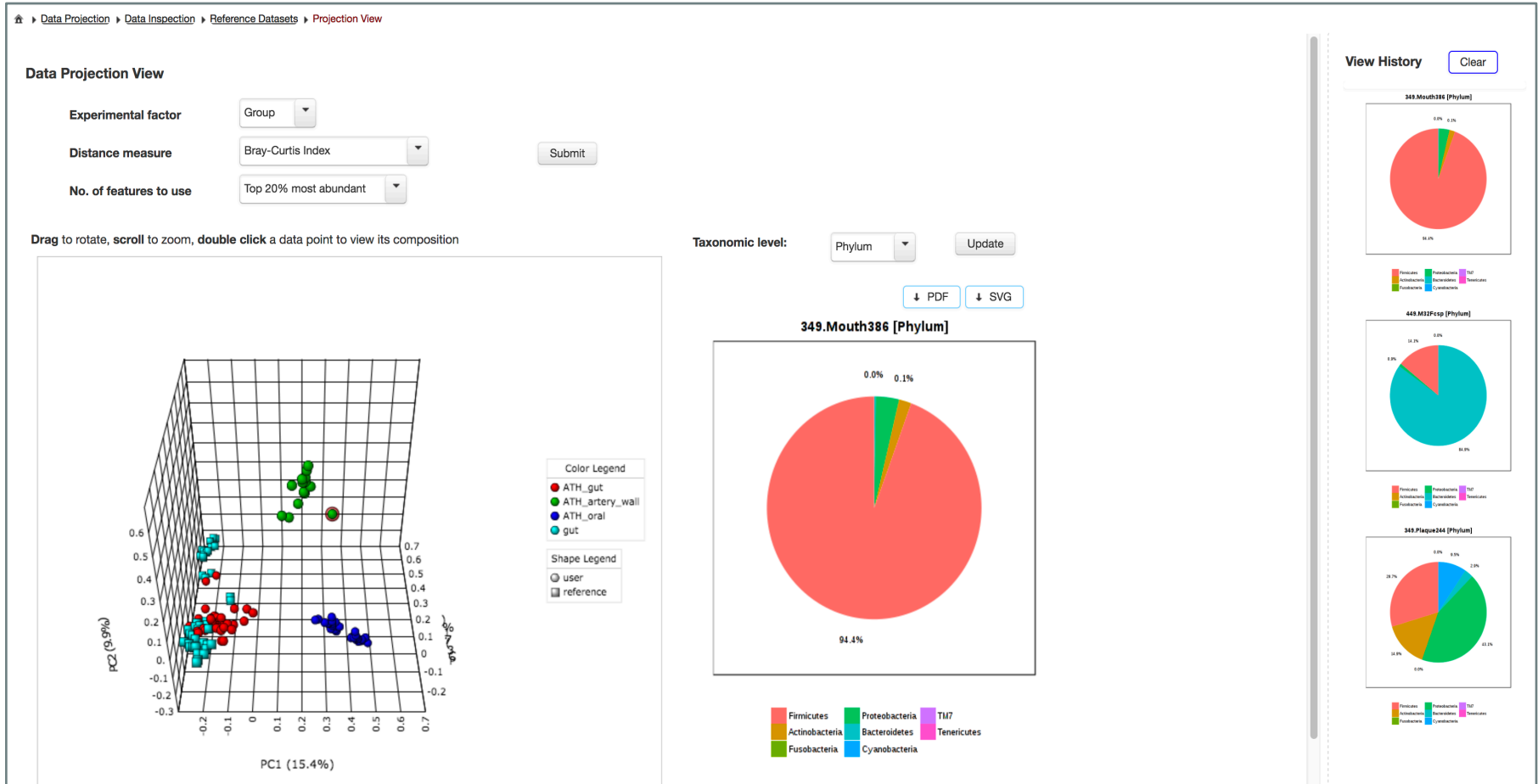


Taxon set enrichment analysis (TSEA)

- Taxon sets manually curated from literature and public databases
 - Phenotypic traits
 - Ecological niches
 - Disease states
 - Associated with host physiological states or lifestyles
- Enrichment analysis by hypergeometric tests
 - Strain level
 - Species level
 - Mixed level



Projection with public data (PPD)



Limitations & future work

1. Multiple-omics data integration
 - Metabolomics (current efforts)
2. To support multivariate association analysis
 - Different life-style factors
3. More formal & rigorous meta-analysis methods
 - Integrating effect sizes, p values, ranks, *etc.*
4. Support for reproducible research

Acknowledgements

■ McGill University

- Xia Lab
 - Achal Dhariwal
 - Jasmine Chong
- Agellon Lab
 - Salam Habib



■ Fundings

- McGill University Startup Fund
- FQRNT New Investigator Award
- NSERC
- Genome Canada

