

MicrobiomeAnalyst

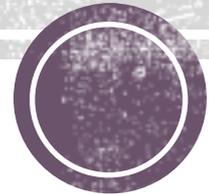
**-- a web-based tool for comprehensive
exploratory analysis of microbiome data**

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

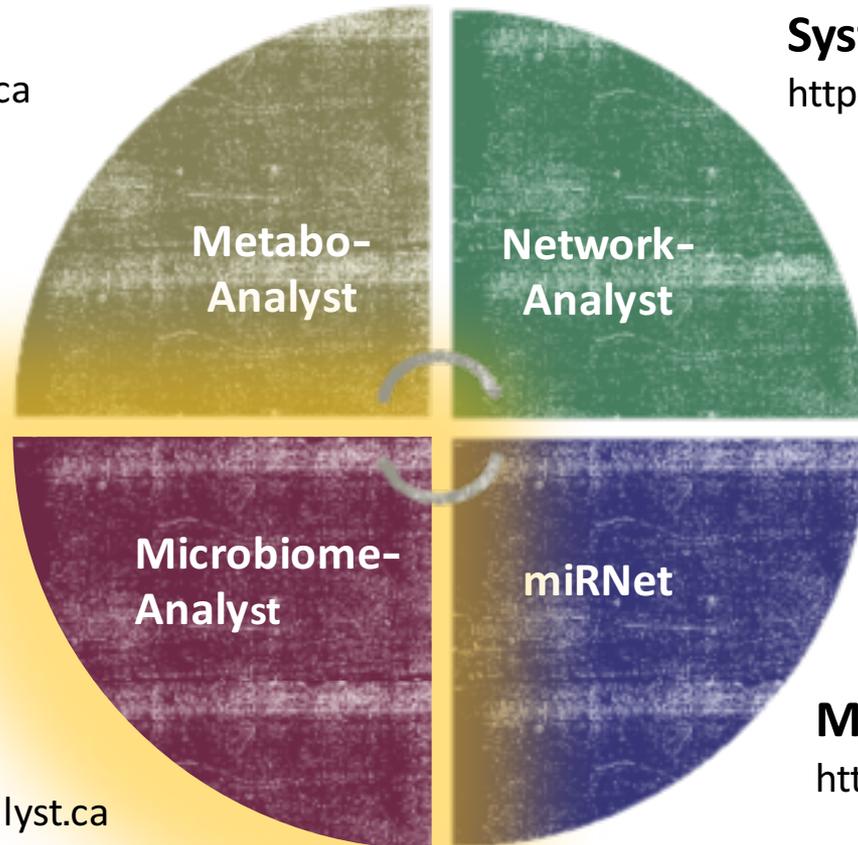
Microbiome

<http://microbiomeanalyst.ca>

miRNet

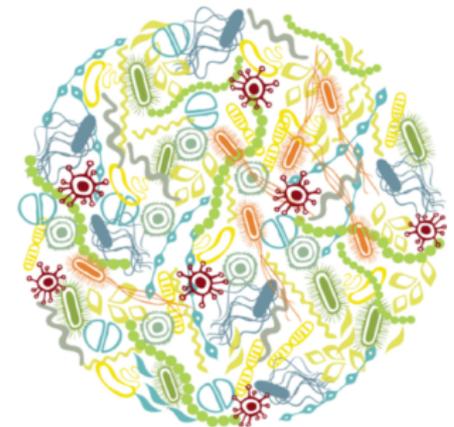
MicroRNAs

<http://mirnet.ca>



Microbiome (I)

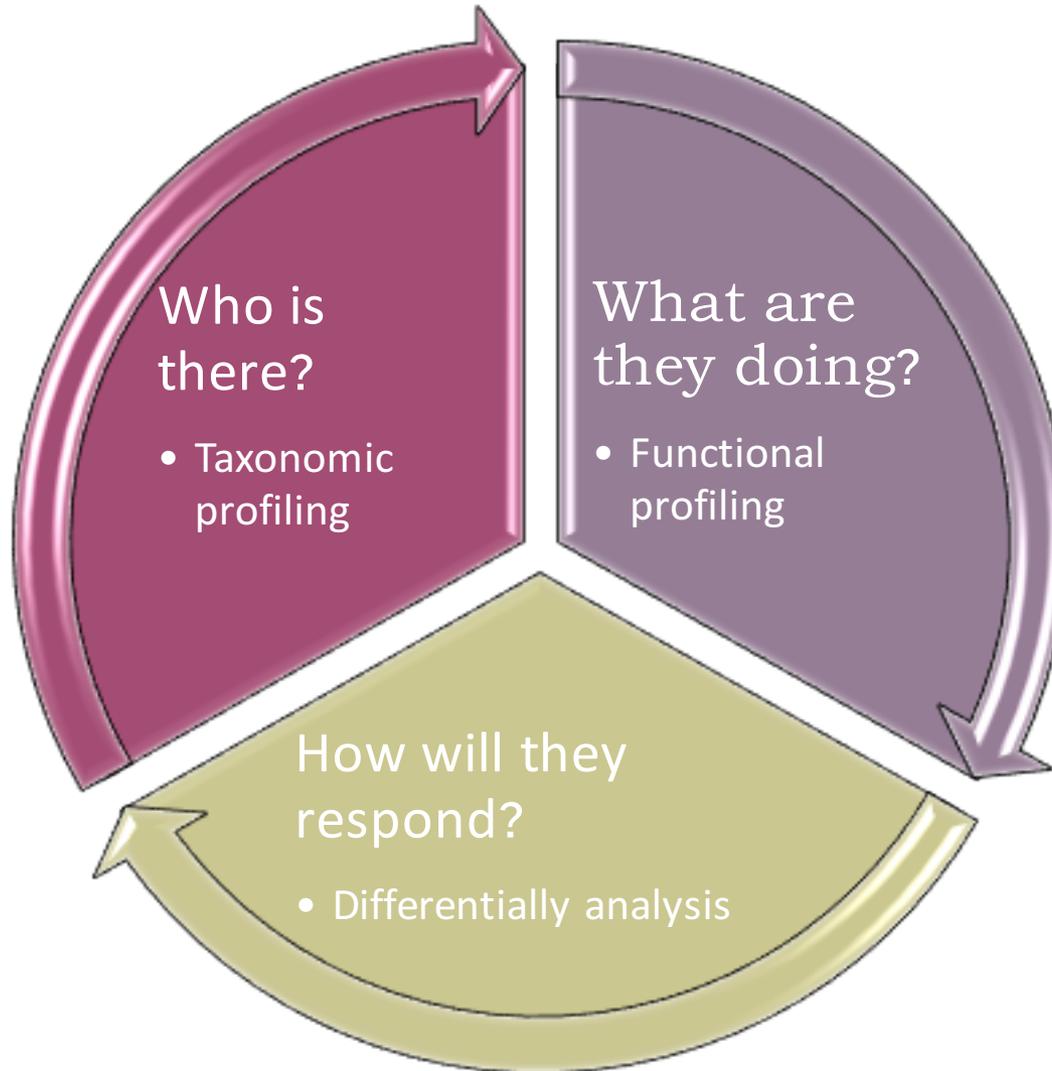
- All the microbes in a community (same as microbiota)
- The Human Microbiome is the collection of all the microorganisms living in association with the human body (source: Human Microbiome Project)
 - Gut microbiome
 - Oral microbiome
 - Vaginal microbiome
- Significant roles in health, disease and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



Microbiome (II)

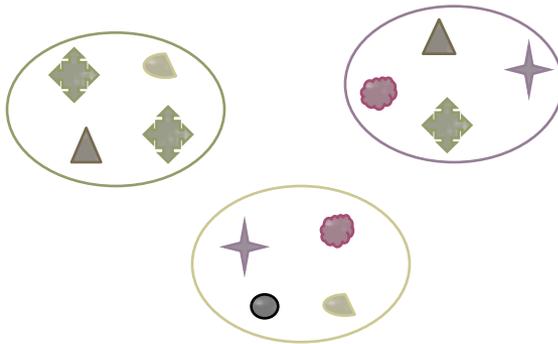
- The composition and structure of the microbiome is more fluid than human cells and organs
 - Host factors (bile acids, immunity, age)
 - Diet
 - Environment exposures
 - Medications
- Potential applications
 - Therapeutic targets
 - Diagnostic biomarkers
 - Precision medicine

Characterizing microbiome

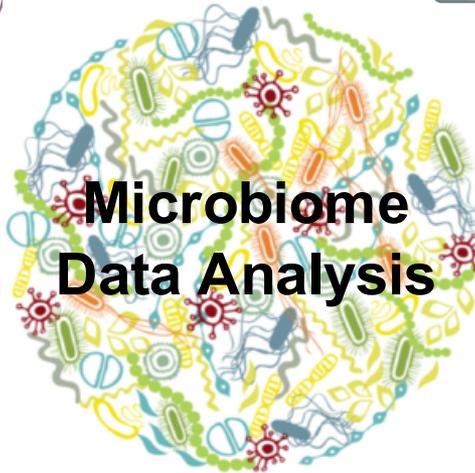
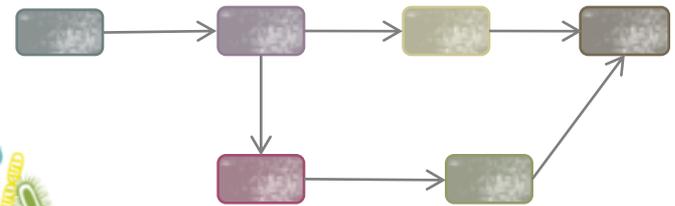


Microbiome data analysis

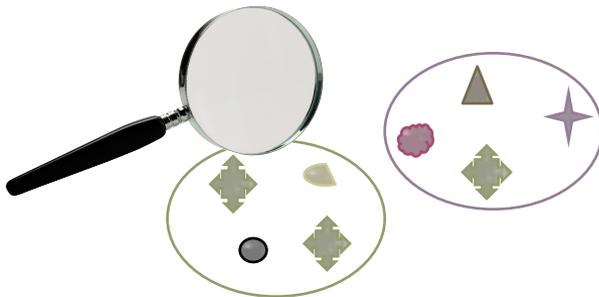
1. Taxonomic Profiling



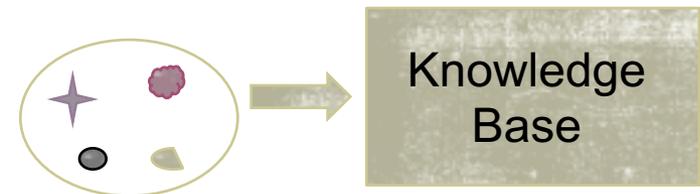
2. Functional Profiling



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

MicrobiomeAnalyst -- comprehensive statistical, visual and meta-analysis of microbiome data

Home Data Format FAQs Tutorials About

Starting from marker gene abundance data (OTU table, BIOM file, mothur output)

Marker Data Profiling (MDP)

Starting from gene list or gene abundance data annotated by KO, EC or COG

Shotgun Data Profiling (SDP)

Visually exploring your 16S rRNA data with a public data in a 3D PCoA plot

Projection with Public Data (PPD)

Starting with a list of taxa of interest (strains, species or higher level taxa)

Taxon Set Enrichment Analysis (TSEA)

MDP
Marker Data Profiling (MDP):
Comprehensive [composition & diversity analysis](#) supporting various methods of data overview, alpha diversity and beta-diversity; [comparative analysis](#) supporting multiple differential abundance

SDP
Shotgun Data Profiling (SDP):
[Functional diversity profiling](#) based on KEGG annotations (modules, pathways, metabolisms, EC) or COG categories, [direct functional association testing](#), as well as [differential abundance analysis](#)

TSEA
Taxon Set Enrichment Analysis (TSEA): Enrichment analysis using a large collection of taxon sets including [105 strain-level taxon sets](#), [174 species-level taxon sets](#) and [40 mixed level taxon sets](#), based on their shared phenotypic traits or ecological niches.

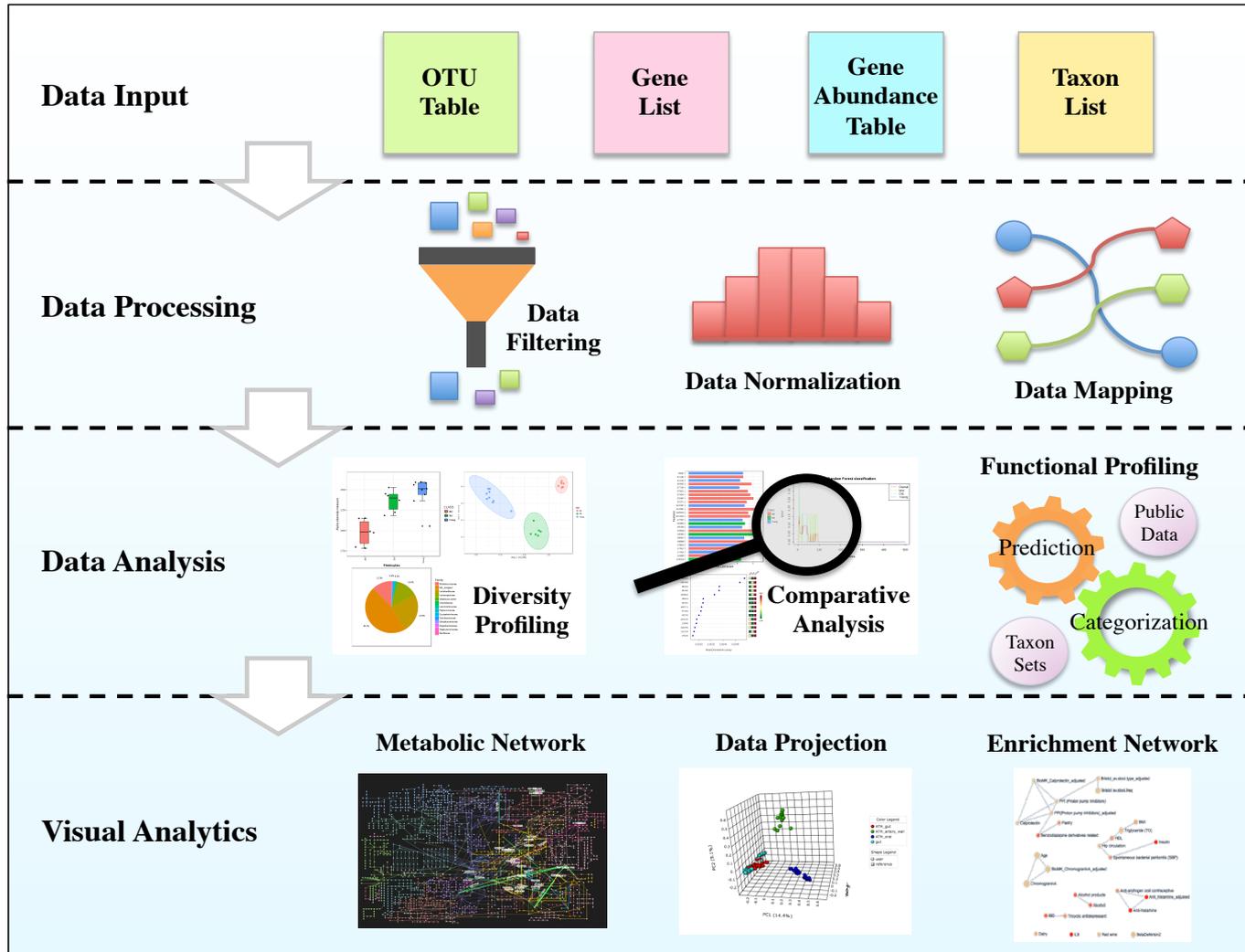
PPD
Projection with Public Data (PPD):
Co-processing your data together with a suitable public 16S rRNA data of interest and explore the results within an [interactive 3D PCoA visualization system](#) to easily [discover patterns of interest](#) as well as to associate these

- 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 \(LEfSe\)](#)

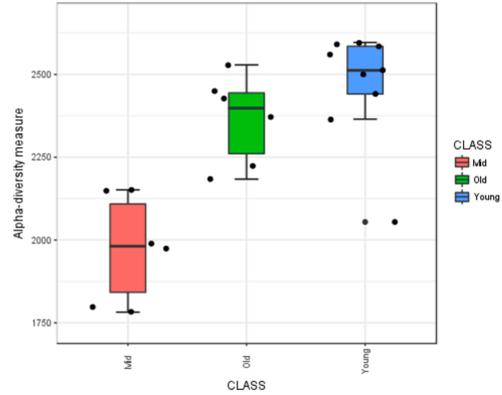
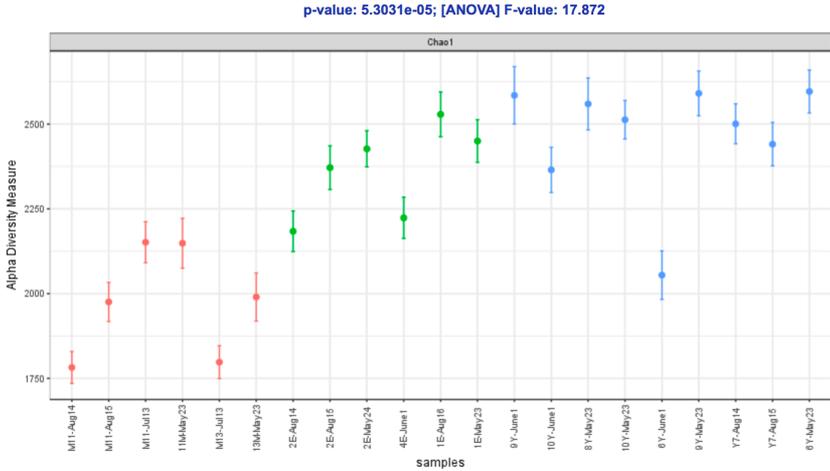
[Random Forests](#)

Prediction of Metabolic Potential

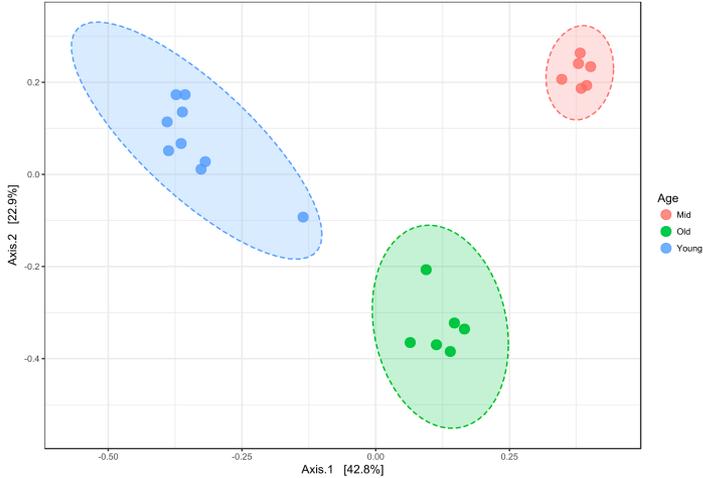
[Prediction for Greengenes Annotated OTUs \(PICRUSt\)](#)

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

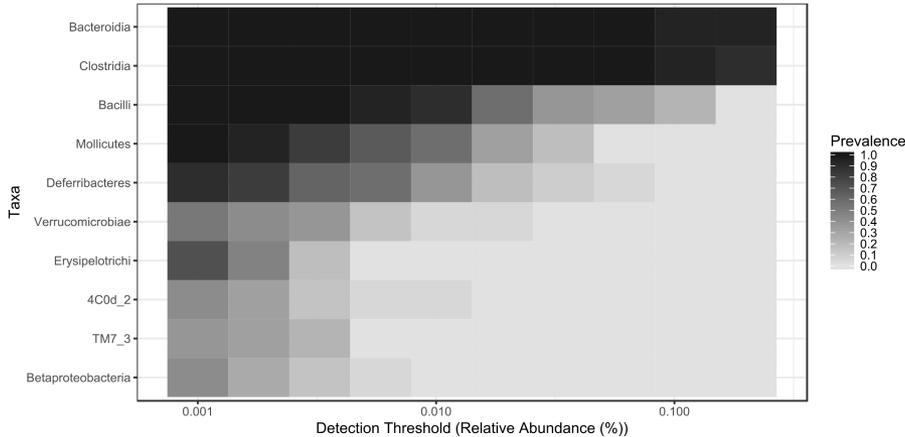
Taxonomic profiling



Alpha diversity



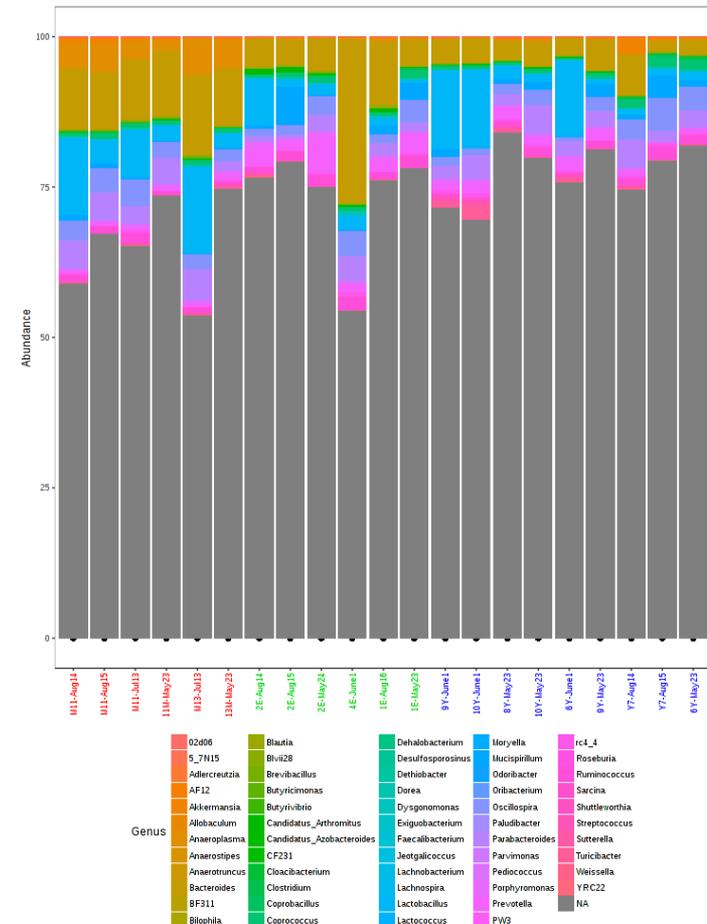
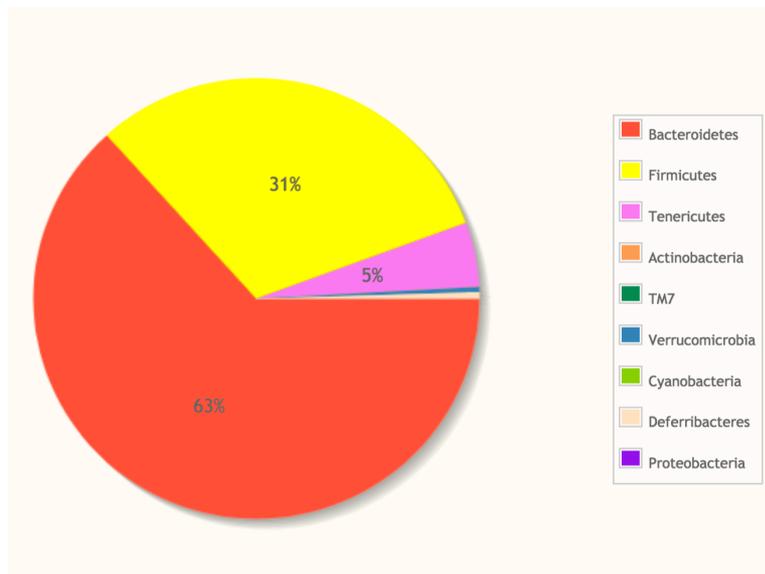
Beta diversity



Core microbiome

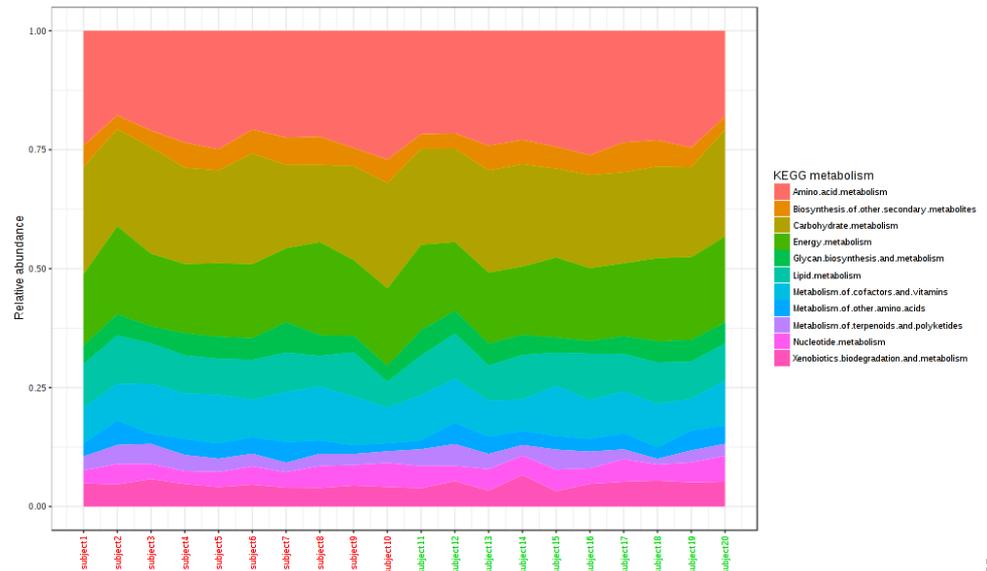
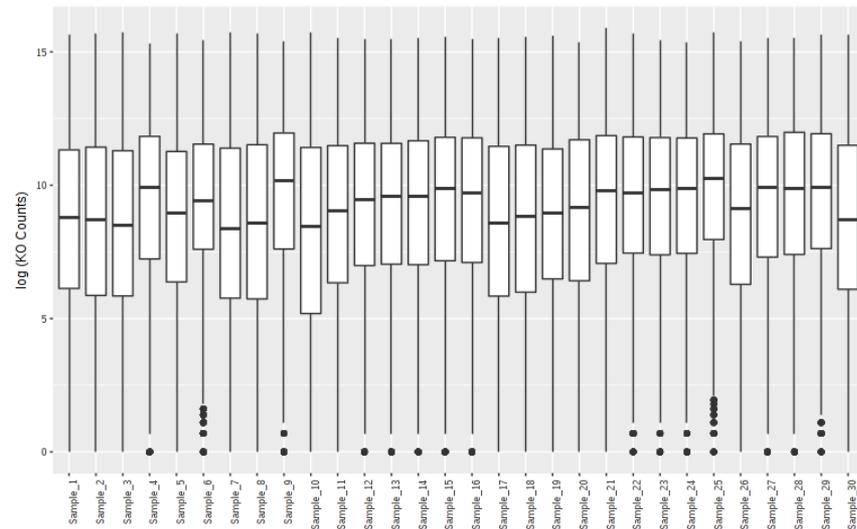
Flexible graphic generations

- 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

Style: KEGG style Background: Black Pathway name: Hide Download: PNG image Highlight:

Category: Pathways Submit Save

Name	Hits	P-value	Color
<input type="checkbox"/> Biosynthesis of amino aci	122	3.58e-33	
<input checked="" type="checkbox"/> Carbon fixation pathways	30	2.77e-7	
<input type="checkbox"/> Amino sugar and nucleoti	29	0.000005	
<input type="checkbox"/> Valine, leucine and isoleu	11	0.000015	
<input checked="" type="checkbox"/> Carbon metabolism	78	0.000016	
<input type="checkbox"/> Alanine, aspartate and glu	27	0.00003	
<input type="checkbox"/> Pyruvate metabolism	30	0.000057	
<input type="checkbox"/> Cysteine and methionine r	29	0.000065	
<input type="checkbox"/> Selenocompound metabol.	10	0.000142	
<input type="checkbox"/> 2-Oxocarboxylic acid met	24	0.00016	
<input type="checkbox"/> Phenylalanine, tyrosine an	26	0.000174	
<input type="checkbox"/> Starch and sucrose metab	25	0.00063	
<input type="checkbox"/> Streptomycin biosynthesis	8	0.000703	
<input type="checkbox"/> Citrate cycle (TCA cycle)	21	0.00109	
<input type="checkbox"/> Glycine, serine and threon	28	0.0011	
<input type="checkbox"/> One carbon pool by folate	12	0.00123	
<input type="checkbox"/> Peptidoglycan biosynthesi	8	0.0015	
<input type="checkbox"/> Polyketide sugar unit bios	4	0.0018	
<input type="checkbox"/> Pentose and glucuronate ii	17	0.0018	
<input type="checkbox"/> Carbon fixation in photosy	15	0.00225	
<input type="checkbox"/> Histidine metabolism	15	0.00225	
<input type="checkbox"/> C5-Branched dibasic acid	7	0.00233	
<input type="checkbox"/> Nicotinate and nicotinami	15	0.00315	

KO Hits

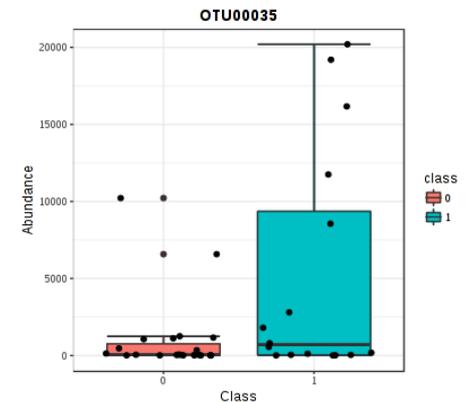
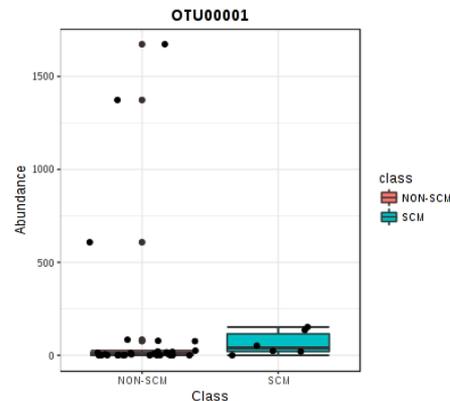
Carbon fixation pathways in prokaryotes

- K00169
- K01895
- K00239
- K01902
- K00031

Xia Lab @ McGill (last updated 2017-06-09)

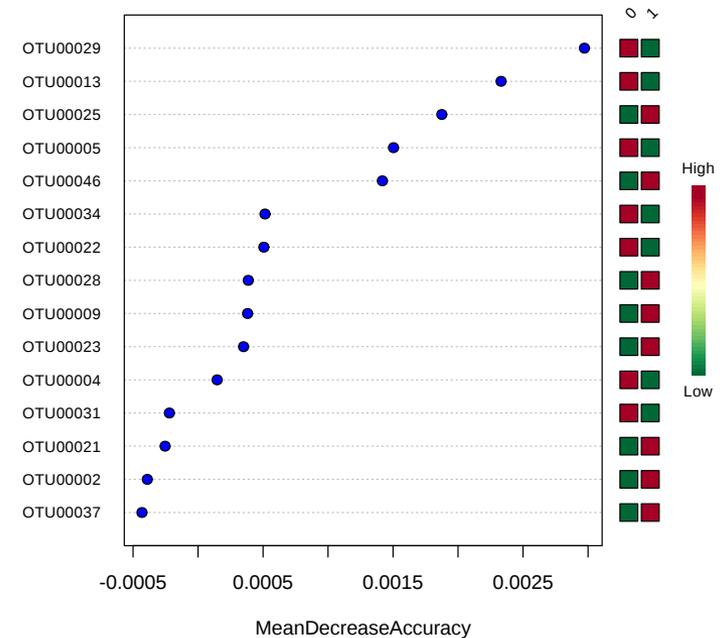
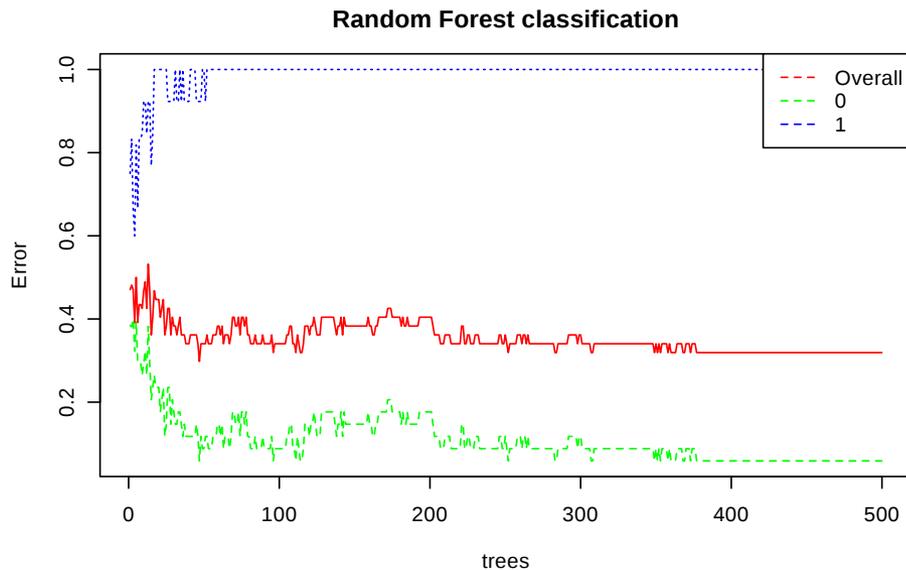
Differential abundance analysis

- 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



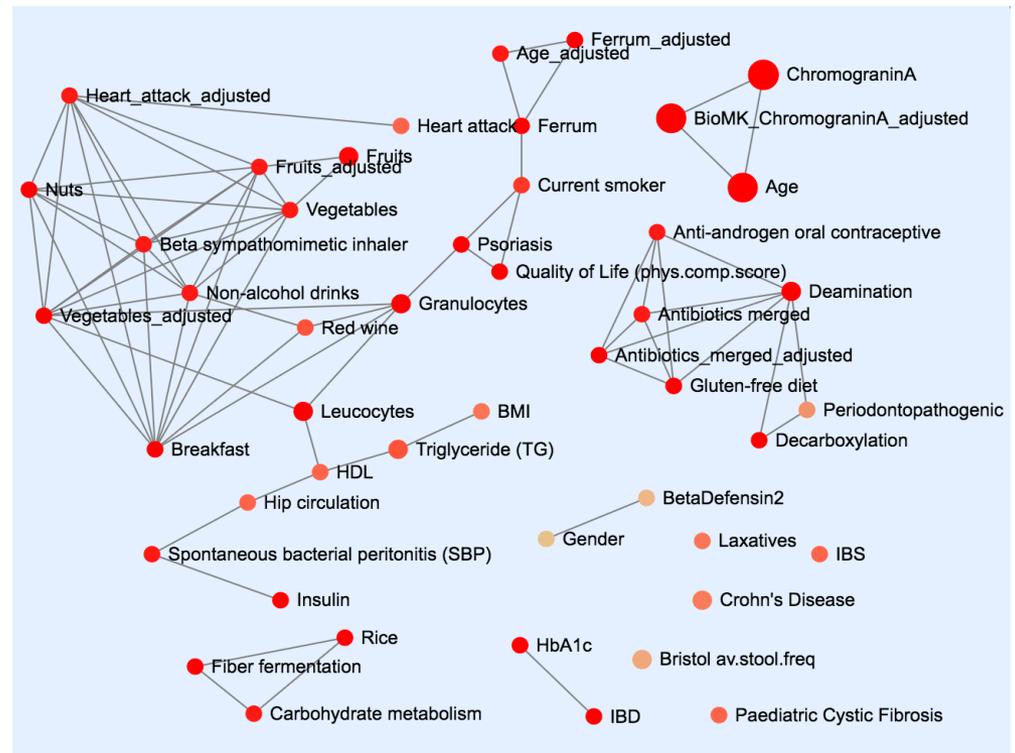
Biomarker analysis

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

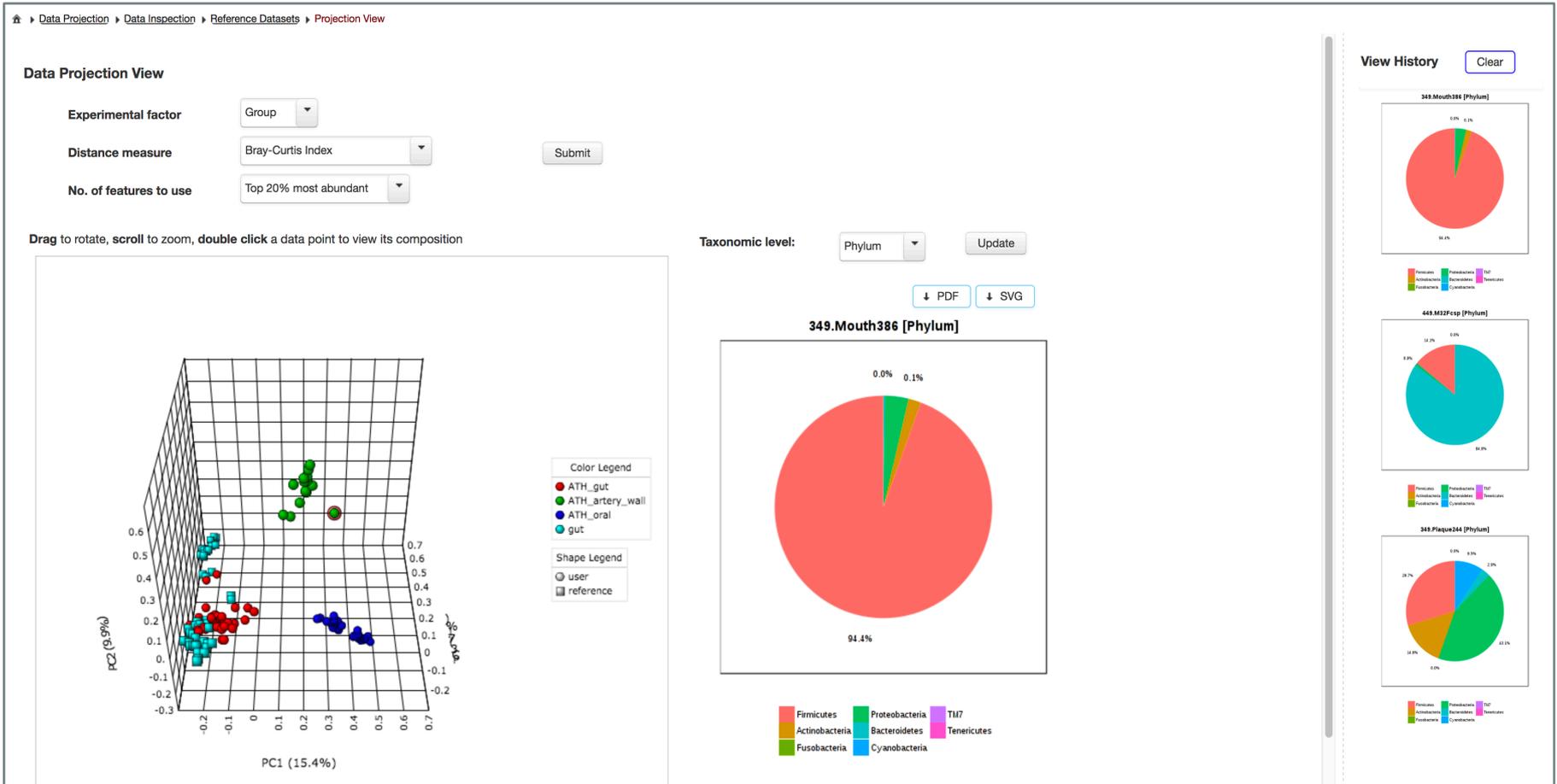


Taxon set enrichment analysis (TSEA)

- 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



GenomeCanada

■ Fundings

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

*Fonds de recherche
sur la nature
et les technologies*

Québec 

