



# Improving Genome-Scale Metabolic Models of the Human Gut Microbiome through Phylogenetic Protein Analysis

Zhonghan Chen<sup>1,2</sup>, Xinyi Ma<sup>1,3</sup>, Isha Sharma<sup>1,3,4</sup>, Yasamin Zarghami<sup>1</sup>, Andrew Hanzhuo Zhang<sup>1</sup>, Dafni Giannari<sup>5</sup>

1. Department of Computer Science, University of Toronto, Toronto, ON Canada

2. Department of Mathematics, University of Toronto, Toronto, ON Canada

3. Department of Statistical Sciences, University of Toronto, ON Canada

4. Department of Cell and Systems Biology, University of Toronto, Toronto, ON Canada

5. Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, ON, Canada

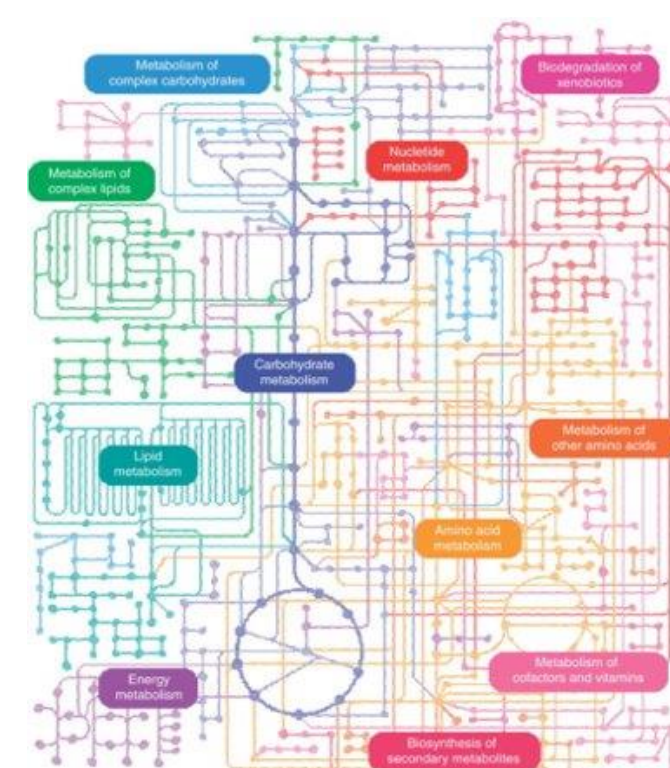


Laboratory for Metabolic System Engineering  
www.lmse.utoronto.ca

## Introduction

### Genome-Scale Metabolic Models (GSMMs)

- Mathematical models of the metabolism of living species
- Include a multitude of data
- Synthesized from functional genome annotations
- Predict effects of gene deletions on an organism

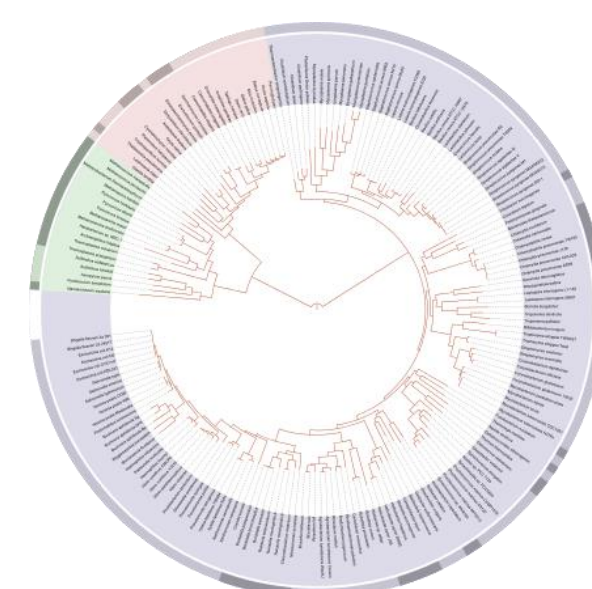


### The Human Gut Microbiome

- Complex ecosystem of bacteria, archaea, protozoa, and other organisms
- Plays an important role in human health and disease

### Phylogenetic Analysis

- Based on phylogenetic trees
- Used in calculating organism-distance values based on metabolism
- Promising for improving GSMMs



## Hypothesis & Aim

### Hypothesis

There is a significant enough correlation between phylogenetic information and metabolic models such that phylogenetic analysis can lead to meaningful improvements in constructing metabolic models for the human gut microbiome.

### Aim

Assess the quality of GSMMs of the Human Gut Microbiome with:

- the use of phylogenetic analysis for prediction of protein functions,
- the systematic curation of their metabolic reactions.

## Methodology

### Retrieving GSMMs and Proteomes of bacterial species found in the Human Gut Microbiome

- Uniprot (fasta files: protein sequences of organisms)
- Virtual Metabolic Human Database (SBML models)



### Assess the Quality of the GSMMs

- Upload SBML models onto Memote platform and retrieve the reports
- Assess mass balance and charge balance of the models, which provides information of the reaction balances we need to fix

Input: SBML model

Online Platform: Memote

Output: Quality report

### Annotate Proteins with Phylogenetic Placement

- Use docker container to run TreeGrafter on each fasta file
- TreeGrafter determines for the protein sequences the best matching homologous family in its library of pre-annotated phylogenetic trees
- For establishment of evolutionary relationships, the Protein Analysis through Evolutionary Relationships (PANTHER) database is used.



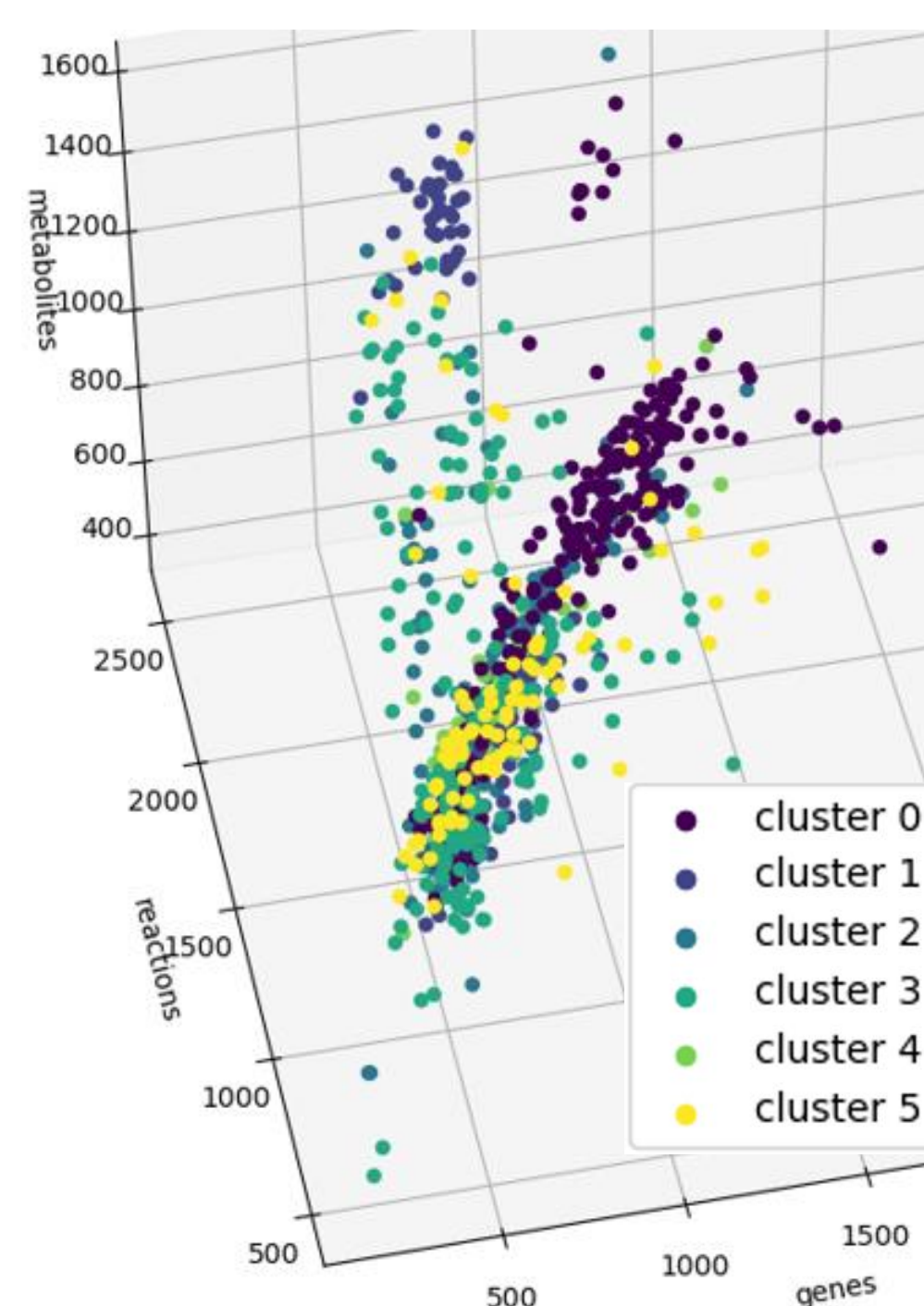
Input: fasta file

Program: TreeGrafter

Output: Protein Annotations

## Results

### Feasibility Analysis using K-modes Clustering

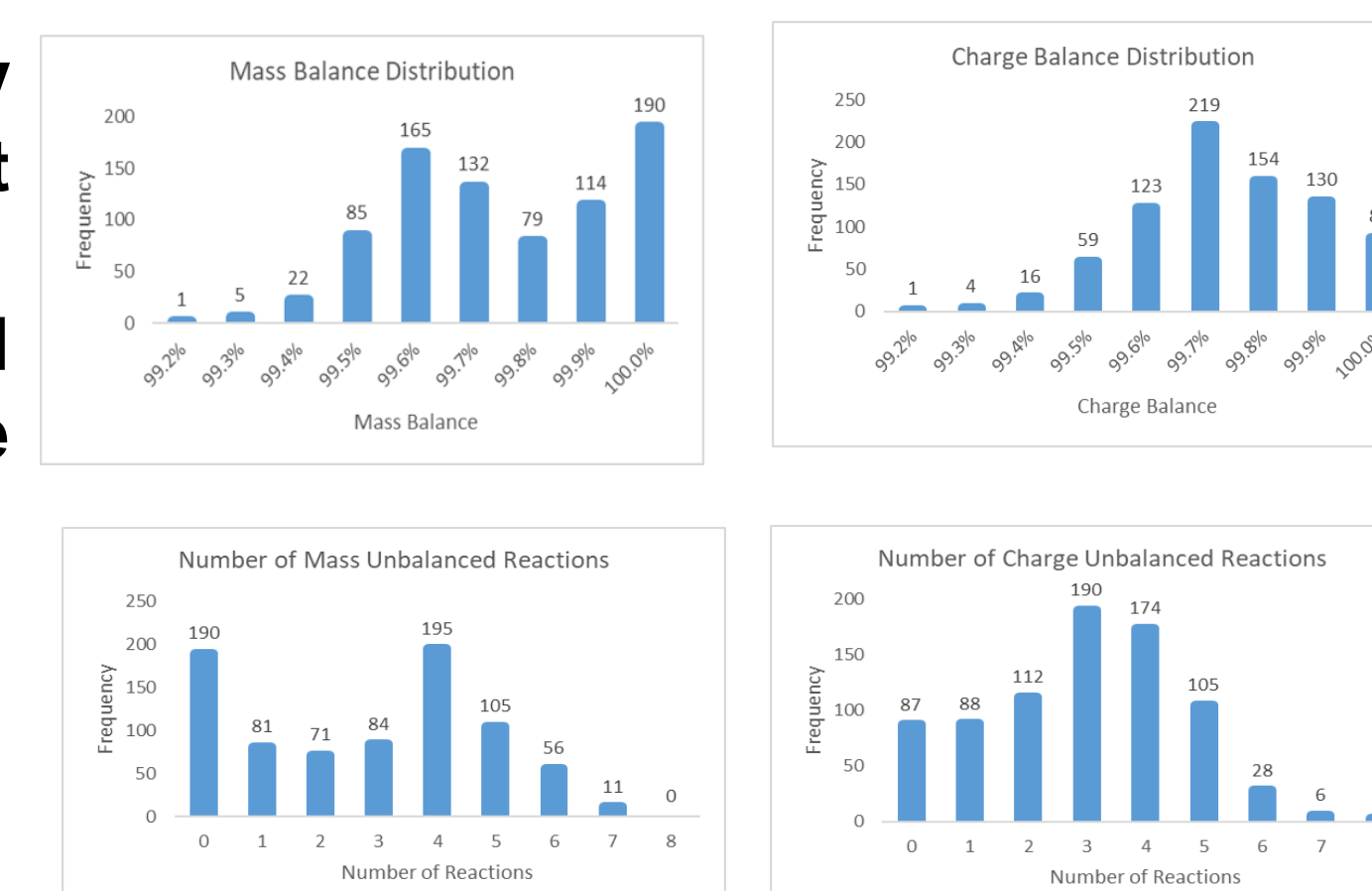


We implemented a K-modes clustering algorithm as a heuristic to verify that a significant enough correlation exists between phylogenetic information and metabolic models before we start our analysis.

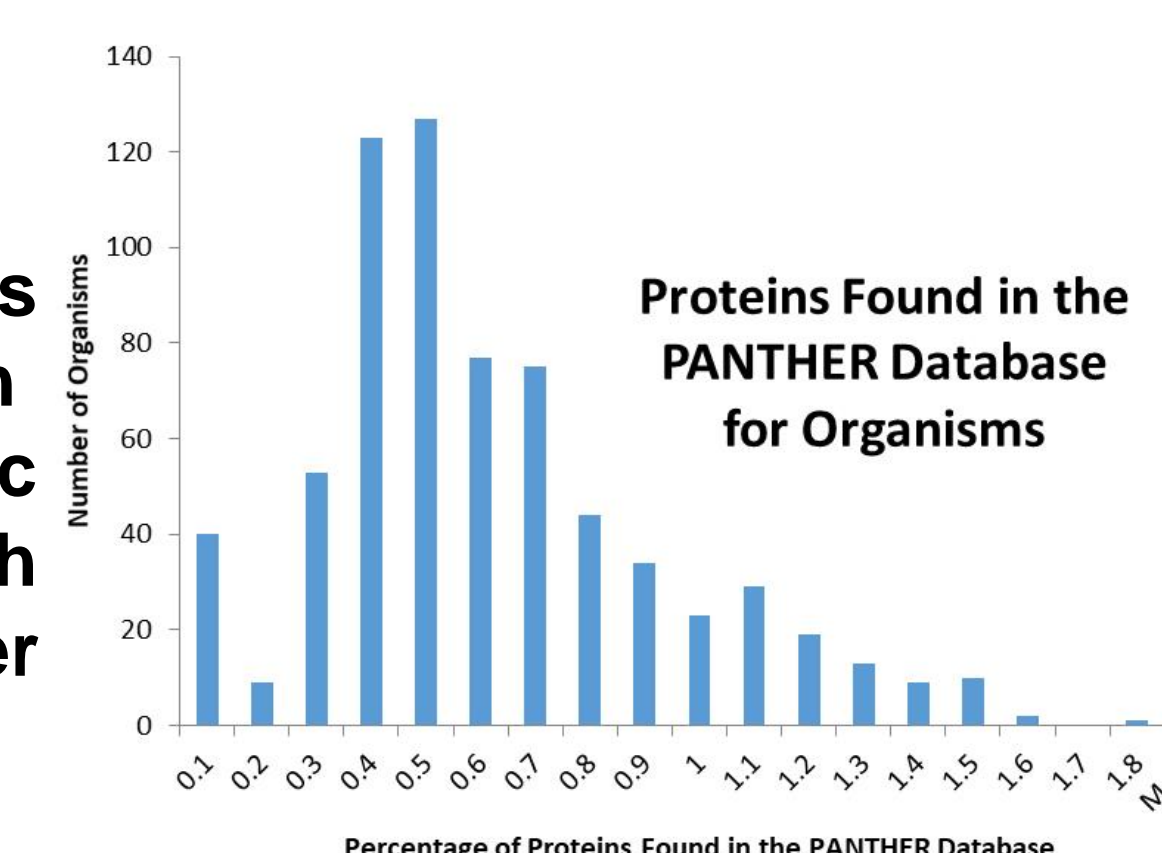
	phylum	oxygenstat	metabolism	mtype
0	Proteobacteria	Aerobe	Fermentative or respiratory	No Data
1	Bacteroidetes	Nanaerobe	Saccharolytic, fermentative	Commensal
2	Firmicutes	Facultative anaerobe	Saccharolytic, fermentative or respiratory	No Data
3	Firmicutes	Obligate anaerobe	Saccharolytic, fermentative	Commensal
4	Actinobacteria	Facultative anaerobe	Saccharolytic, fermentative	Opportunistic pathogen
5	Firmicutes	Obligate anaerobe	Saccharolytic, fermentative	No Data

Running said clustering algorithm on the following characteristics of the human gut microbiome species: **phylum**, **oxygen status**, **metabolism type**, and **role in the gut** (mtype), yielded 6 clusters as listed on the table above. When plotted against the number of **genes**, **reactions**, and **metabolites** as seen on the figure on the left, clear clusters are visible, and thus preliminary analysis shows that our approach is feasible.

### Quality Assessment of the GSMMs based on Memote



### Annotate Proteins with the Phylogenetic Placement with TreeGrafter



## Conclusions

- GSMMs from the Virtual Metabolic Human Database effectively represent bacterial species of the human gut microbiome.
- But they contain errors due to inconsistencies and low quality of functional genome annotations.
- The hypothesized approach to improve metabolic models is feasible, since significant enough correlation exists between phylogenetic information and metabolic models.

## Future Work

- Compare protein functions in the models with those we concluded.
- Comparison of predictions made by the initial models with those made by our final improved models.
- Use acquired protein annotations for improving existing GSMMs of human gut microbiome.
- Establish a pipeline that can be used for improving GSMMs