Quantifying Functional Microbiomes
An Integrated, Quantitative Metaproteomics Approach

Caleb Easterly & Pratik Jagtap
Outline

1) Introduction
   a) Current state of metaproteomic analysis

2) Methods
   a) Ontologies
   b) Taxonomy in metaproteomics
   c) Function in metaproteomics

3) metaQuant case studies:
   a) Lake sediment microbiome data
   b) Oral microbiome data

4) Conclusions and Next Steps
Multiple studies have shown correlation of microbial composition with physiological conditions.

**Metagenomics:** DNA Sequencing identifies species present within complex community (16S rRNA and Whole Genome Sequencing).

**Metatranscriptomics:** RNA Sequencing identifies species present and possible functions within complex communities (RNASeq).

**Metaproteomics:** The large-scale characterization of the entire protein complement of environmental microbiota at a given point in time. Potential to unravel the mechanistic details of microbial interactions with host / environment by analyzing the functional dynamics of the microbiome.

Metaproteomics approach goes beyond studies of compositional dynamics and sheds light on mechanistic details of microbial interactions with host / environment.
Microbiome: mass spectrometry

https://doi.org/10.1146/annurev-biodatasci-080917-013516
Metaproteomics: identification

**Database Generation**
- FASTQ
- Protein / Peptide FASTA

**Database Search & Strategies**
- Search Algorithm
- Spectra
- Peptides

**Functional Analysis**
- Known Function
- Proteins
- Unique Peptides

**Taxonomy Analysis**
Metaproteomics: identification

**Database Generation**
- FASTQ
  - Protein / Peptide FASTA
  - Spectra

**Database Search & Strategies**
- Search Algorithm
  - Peptides

**Functional Analysis**
- Known Function
- Hypothetical Function
- Unknown Function
- Proteins

**Taxonomy Analysis**
- Shared Taxonomy
- Unassigned Taxonomy
- Unique Peptides
- TAXONOMY ANALYSIS
Metaproteomics: quantification

**DATABASE GENERATION**
- FASTQ
- Protein / Peptide FASTA
- Spectra

**DATABASE SEARCH & STRATEGIES**
- Search Algorithm

**QUANTITATIVE ANALYSIS**
- Intensity data

**FUNCTIONAL ANALYSIS**
- Peptides
- Unique Peptides TAXONOMY ANALYSIS
- Proteins
Metaproteomics: quantification

**DATABASE GENERATION**
- FASTQ

**DATABASE SEARCH & STRATEGIES**
- Protein / Peptide FASTA
- Search Algorithm

**QUANTITATIVE ANALYSIS**
- Spectral counts or intensity data

**FUNCTIONAL ANALYSIS**
- Known Function
- Proteins
- Hypothetical Function
- Unknown Function
- Shared Taxonomy
- Unassigned Taxonomy
metaQuant: metaproteomic data analysis

Function
Taxonomy
Function × Taxonomy

Condition 1

Condition 2

Protein taxonomy

Example: *Streptococcus* are common in the oral microbiome.

To calculate total peptide intensity assigned to each taxon, propagate intensity values up the hierarchy (Riffle, et. al, Proteomes 2017).
Protein function

Gene Ontology annotation of DNA polymerase III subunit alpha (P10443)
Protein function


- Hierarchical, ~50,000 terms
- “Slim” version: smaller set of higher-level terms

**High level**: Clusters of orthologous groups (COG) categories (Galperin, et al. *Nuc Acid Res* 2015)

- 26 categories
  - **J**: Translation, ribosomal structure and biogenesis
metaQuant analysis

1) Taxonomy mode
2) Function mode
metaQuant analysis

1) Taxonomy mode
2) Function mode
3) Function/taxonomy interaction mode
Case Study 1: Anoxic Lake Sediment

- Sediment samples from Lake Matano, Indonesia, were incubated in three bioreactors.
- The search database was generated using metagenomic data from the bioreactors.

*Unpublished Data (Szeinbaum et al, Georgia Tech, Atlanta, GA)*
Anoxic lake: Galaxy workflow
Anoxic lake: results

All peptides assigned to *Betaproteobacteria* class, by COG category.
Case Study 2: Sucrose-induced oral dysbiosis

Rudney et al., BMC Microbiome DOI: 10.1186/s40168-015-0136-z

• Mass spectral data was acquired from plaque samples from five subjects at high risk for dental caries grown in biofilm reactor in the presence (With Sucrose, or WS) and absence of sucrose (No Sucrose, or NS) (5 in each group, 10 total samples)

• Mass spectra were searched against the Human Oral Microbiome database (HOMD) to identify microbial peptides.

• Quantitation, functional annotation, and taxonomic assignment was performed in Galaxy; metaQuant was used to analyze the results.
Oral dysbiosis results: taxonomy

metaQuant currently produces a tabular output, so R was used to create figures.

Volcano plot: differentially abundant taxa. *Proteobacteria*, *Escherichia* and *treponema* are higher in NS.
Oral dysbiosis results: function

Results of the ‘slim’ GO analysis

Volcano plot: Differentially abundant functional terms

Heatmap: clustered by functional term and sample.
Oral dysbiosis results: function-taxonomy

Peptides assigned to *Fusobacteria*, part of the ‘orange complex’ in periodontal disease (Costalonga and Herzberg, Imm Lett 2014).

Cog N: cell mobility.

Peptides assigned to carbohydrate metabolism (COG category G)
Conclusions and next steps

- **metaQuant** offers methods for quantitative metaproteomics analysis, providing a perspective on the interaction between taxonomy and function in microbiomes
- Source on Github ([github.com/caleb-easterly/metaquant](https://github.com/caleb-easterly/metaquant)), distributed as alpha on PyPI, Bioconda, and [Galaxy Test ToolShed](https://galaxyproject.org/).
Conclusions and next steps

Next steps:

- Develop interactive visualizations for metaQuant outputs
- Investigate peptides/proteins of unknown function/taxonomy

Differential expression analysis: proteins of known (L) and unknown (R) function
Conclusions and next steps

Next steps:

- Analyse more datasets (clinical and environmental)
- Integrate the metaproteomics workflow with the metatranscriptomics quantitative analysis and visualization pipeline (ASaiM) within Galaxy.
- We anticipate that integration and visualization of data from multi-omic approaches will aid in understanding the functional dynamics of the microbiome.
ACKNOWLEDGMENTS