Meta-omics Workflows: microbiome Analysis using metagenomics, metatranscriptomics, Metaproteomics, Metabolomics...

Pratik Jagtap
University of Minnesota, Minneapolis
pjagtap@umn.edu
Microbiome: Genes, proteins, molecules expressed in response to stimulus

Multiple studies have shown correlation of microbial composition with physiological conditions.
Potential to unravel the mechanistic details of microbial interactions with host/environment by analyzing the functional dynamics of the microbiome.
Potential to unravel the mechanistic details of microbial interactions with host/environment by analyzing the functional dynamics of the microbiome.
Multiple studies have shown correlation of microbial composition with physiological conditions. Also used to study interaction with the environment.

Metagenomics: Identifies species present within complex community (16S rRNA and Whole Genome Sequencing) by studying DNA from the samples.

OTU = Operational Taxonomic Unit, a group of very similar 16S sequences.
**Metagenomics Software Tools**

Data Collection and Sequencing

- Pre-processing of sequencing reads

Analysis to profile taxonomy, genomic features and function

- High level analyses (alpha and beta diversity, random forests regression or source tracker)

- Read-based profiling (Kraken, MEGAN or HUMAnN)

- Assembly-based analysis (metaSPAdes or MEGAHIT)

Statistics and biological interpretation

Validation

Metagenomics Software Tools

Marker gene (16S rRNA, ITS or 18S rRNA)

Denoise to sOTUS (Deblur)

DATABASES: Greengenes, RDP, Silva

High level analyses (alpha and beta diversity, random forests regression or source tracker)

Gene content Prediction PiCRUST

QIIME Output: Emperor Plots

Mothur Output: Taxonomy Plot

**Shotgun Metagenomics Software Tools**

**READS-BASED ANALYSIS (MAPPING)**

Provides an *aggregate view* of community structure and function - based on reads mapping to *Reference databases*.

<table>
<thead>
<tr>
<th>Taxonomy</th>
<th><strong>DATABASES</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bowtie2</td>
<td><em>RefSeq</em></td>
</tr>
<tr>
<td>KRAKEN</td>
<td><em>PFAM</em></td>
</tr>
<tr>
<td>MetaPhlan</td>
<td><em>UniRef</em></td>
</tr>
<tr>
<td>TIPP</td>
<td><em>NCBI-nr</em></td>
</tr>
<tr>
<td><strong>PHASTER</strong></td>
<td></td>
</tr>
<tr>
<td><strong>FOAM</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUMaN2</td>
</tr>
</tbody>
</table>

**Taxonomy and Function**


**HUMaN2 Outputs:**
- *Gene Family Abundance*
- *Pathway Coverage*
- *Pathway Abundance*
Shotgun Metagenomics Software Tools

Assembly-based Analysis
Can construct multiple whole genomes – for organisms with enough coverage to be assembled and binned.

Assembly
MetaSPADES
MEGAHIT

Binning Contigs (Supervised & Unsupervised)
MaxBin2
CONCOCT

Quality Control
Visualization Tools such as Anvio, VizBin, ATLAS

Metagenomics Software Tools

**TAXONOMY**

- Marker gene (16S rRNA, ITS or 18S rRNA)
  - Denoise to sOTUS (Deblur)
    - High level analyses (alpha and beta diversity, random forests regression or source tracker)

**FUNCTION**

- Metagenomics
  - In silico removal of host DNA (bowtie 2)
    - Read-based profiling (Kraken, MEGAN or HUMAnN)
    - Assembly-based analysis (metaSPAdes or MEGAHIT)

The Importance of the MICROBIOME by the Numbers

- 90% of all disease can be traced to the gut and health of the microbiome.
- 10-100 trillion: Number of symbiotic microbial cells harbored by each person, primarily bacteria in the gut, that make up the human microbiota.
- >10,000: Number of different microbes species researchers have identified living in the human body.
- 100 to 1: The genes in our microbiome outnumber the genes in our genome by about 100 to 1.
- 3.3 million: Number of non-redundant genes in the human gut microbiome.
- 99.9%: Percentage individual humans are identical to one another in terms of host genome.
- 10X: There are 10 times as many outside organisms as there are human cells in the human body.
- 22,000: Additional number of genes in the human gene catalog.
- 80%-90%: Percentage individual humans are different from one another in terms of the microbiome.

http://draxe.com/microbiome/
Potential to unravel the mechanistic details of microbial interactions with host/environment by analyzing the functional dynamics of the microbiome.
**Metatranscriptomics Software**

- **RNAseq Data Preprocessing:**
  Quality Control (*FastQC*), Ribosomal RNA removal (*SortMeRNA, barrnap*), host RNA removal (*BMTagger*).
- **De Novo Assembly:**
  *Trinity, MetaVelvet, Oases, IDBA-MT, TAG*, etc.
- **Transcript Taxonomy:**
  *Kraken, GOTTCHA, MetaPhlan2*
- **Functional Annotation:**
  Annotation of assembled contigs are subjected to gene finding programs such as *FragGeneScan* followed by functional assignment using DIAMOND searches against *KEGG, NCBI RefSeq, UniProt*.

**Differential Expression:**
*EdgeR, DeSeq2 and limma*

**Metatranscriptomics Workflows**

**READS-BASED ANALYSIS**
- MetaTrans
- COMAN
- FMAP
- SAMSA2
- AsaiM

  - Quality Control
    - Assess Quality
    - Trim and Filter raw reads
    - Filter ribosomal RNA (rRNA)
  - Community profiling (Who?)
    - Determine composition of sample
    - Visualisation
  - Functional Analysis (What?)

**ASSEMBLY-BASED ANALYSIS**
- SqueezeMeta
- IMP
- MOSCA


Batut et al Gigascience. 2018 7(6) doi: 10.1093/gigascience/giy057
Potential to unravel the mechanistic details of microbial interactions with host/environment by analyzing the functional dynamics of the microbiome.
**Microbiome:** Microbial genetic potential and response

Multiple studies have shown correlation of microbial composition with physiological conditions. Also used to study interaction with environment.

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

DNA from samples. 16S rRNA (economical) or Shotgun sequencing (expensive).

Multiple studies that correlate taxonomy with observed phenotype.

**Metaproteomics:** The large-scale characterization of the entire protein complement of environmental microbiota at a given point in time.

Proteins from samples.

Potential to unravel the mechanistic details of microbial interactions with host / environment by analyzing the functional dynamics of the microbiome.
Metaproteomics

Bond and Wilmes 2004
“The large-scale characterization of the entire protein complement of environmental microbiota at a given point in time”
Environ. Microbiol. 6, 911–920.

Bond and Wilmes 2015
“Through the application of metaproteomics to different microbial consortia over the past decade, we have learnt much about key functional traits in the various environmental settings where they occur.”
Metaproteomics Workflow

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

**DATABASE SEARCH & STRATEGIES**
- Peptides

**FUNCTIONAL ANALYSIS**
- Known Function
- Proteins
- Hypothetical Function
- Unknown Function
- Shared Taxonomy
- Unassigned Taxonomy
- Unique Peptides

**TAXONOMY ANALYSIS**
Metaproteomics Workflow

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

**DATABASE SEARCH & STRATEGIES**
- Peptides

**QUANTITATIVE ANALYSIS**
- Spectral counts OR Intensity data

**FUNCTIONAL ANALYSIS**
- Known Function
  - Proteins
    - Hypothetical Function
    - Unknown Function
- Shared Taxonomy
  - Unassigned Taxonomy
- Unique Peptides
  - TAXONOMY ANALYSIS
Metaproteomics Workflow

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

**Database Search & Strategies**
- SearchGUI
- PeptideShaker

**Quantitative Analysis**
- moFF
- FlashLFQ
- metaQuantome

**Functional Analysis**
- UniPept
- Proteins
- Unique Peptides

**Taxonomy Analysis**
metaQuantome Workflow

**PEPTIDE IDENTIFICATION**

- Spectrum files
  - msconvert: Format conversion
  - SearchGUI: Database search
  - Peptide Shaker PSM

**PEPTIDE QUANTIFICATION**

- Peptide Quantitation and normalization
  - Peptide Quantitation: FlashLFQ
    - Limma: Peptides with Normalized Quant
  - Unipept 4.0: Peptides with functional assignments
    - Function and Taxonomy annotation
  - Unipept 4.0: Peptides with taxonomic assignments

**FUNCTION TAXONOMY**
**metaQuantome OUTPUTS**

*metaproteomic data input*

- **MS1 intensity**
- **function annotation**
- **taxonomic annotation**

---

**FUNCTION: VOLCANO PLOTS**

Fold-change: 33 hours versus 8 hours

---

**FUNCTION: HEATMAP**

---

Functional abundance values separate time point T1 (8 hr) from other time points (which are clustered together). Taxonomy abundance alone does not separate the time points, thus highlighting the importance of understanding functional state of the microbiome.
Potential to unravel the mechanistic details of microbial interactions with host/environment by analyzing the functional dynamics of the microbiome.
Metabolomics Software Tools

- MetFrag, MAGMa (combinatorial fragmenters)
- CFM-ID, CSI:FingerID (machine learning)
- MS-FINDER (rule-based fragmenters)

An overview of selected compound databases, EI-MS databases and soft ionization MS/MS databases.

“Rumsfeld Quadrants” showing the intersection of yes/no answers for whether analysts expect a compound to be identified in the sample (prior probability) and whether it was identified in a library search.

Metabolomics Analysis

DATA INTEGRATION: MULTI-OMICS / MICROBIOMICS

RECOMMENDED READING:


Accessing tools and Workflows

**METAGENOMICS:**
Toolshed: [z.umn.edu/metagenomics_toolshed](z.umn.edu/metagenomics_toolshed)
Galaxy Training Network: [https://training.galaxyproject.org/training-material/topics/metagenomics/](https://training.galaxyproject.org/training-material/topics/metagenomics/)

**METATRANSCRIPTOMICS:**
Workflow: [http://z.umn.edu/MTWF2020](http://z.umn.edu/MTWF2020)

**METAPROTEOMICS:**
Workflow: [z.umn.edu/MPWF2020](z.umn.edu/MPWF2020)
Galaxy Training Network: [http://z.umn.edu/gtn-metaproteomics](http://z.umn.edu/gtn-metaproteomics)

Also available on: [https://proteomics.usegalaxy.eu/](https://proteomics.usegalaxy.eu/) and Metaproteomics Gateway: [z.umn.edu/metaproteomicsgateway](z.umn.edu/metaproteomicsgateway)
REFERENCES (z.umn.edu/RefsforLorentz)


