Perelman
School of Medicine
UNIVERSITY of PENNSYLVANIA

Ontology Application to Microbiome data integration, exploration and analysis

Jie Zheng[1]
Daniel Beiting[2]

[1]Department of Genetics, [2]Department of Pathobiology
University of Pennsylvania

CTSOG 2017 Ann Arbor, MI
October 25, 2017

MicrobiomeDB

- Website: http://microbiomedb.org/
- Provides a platform for integrating, mining and analyzing microbiome data
- The platform allows exploring data based on sample details to examine the microbial community diversity:
  - Which sites show the lowest alpha diversity across the human body?
  - How does community composition in the gut change with age?

Microbiome Data

- Microbial census data contains around 14,000 samples collected from human, animal and environmental sources provided by multiple investigators
- Over 9,000 samples from healthy human subjects in the Human Microbiome Project (http://portal.ihmpdcc.org/)
- Minimum Information about any (x) Sequence (MIxS) is used to guide the data collection

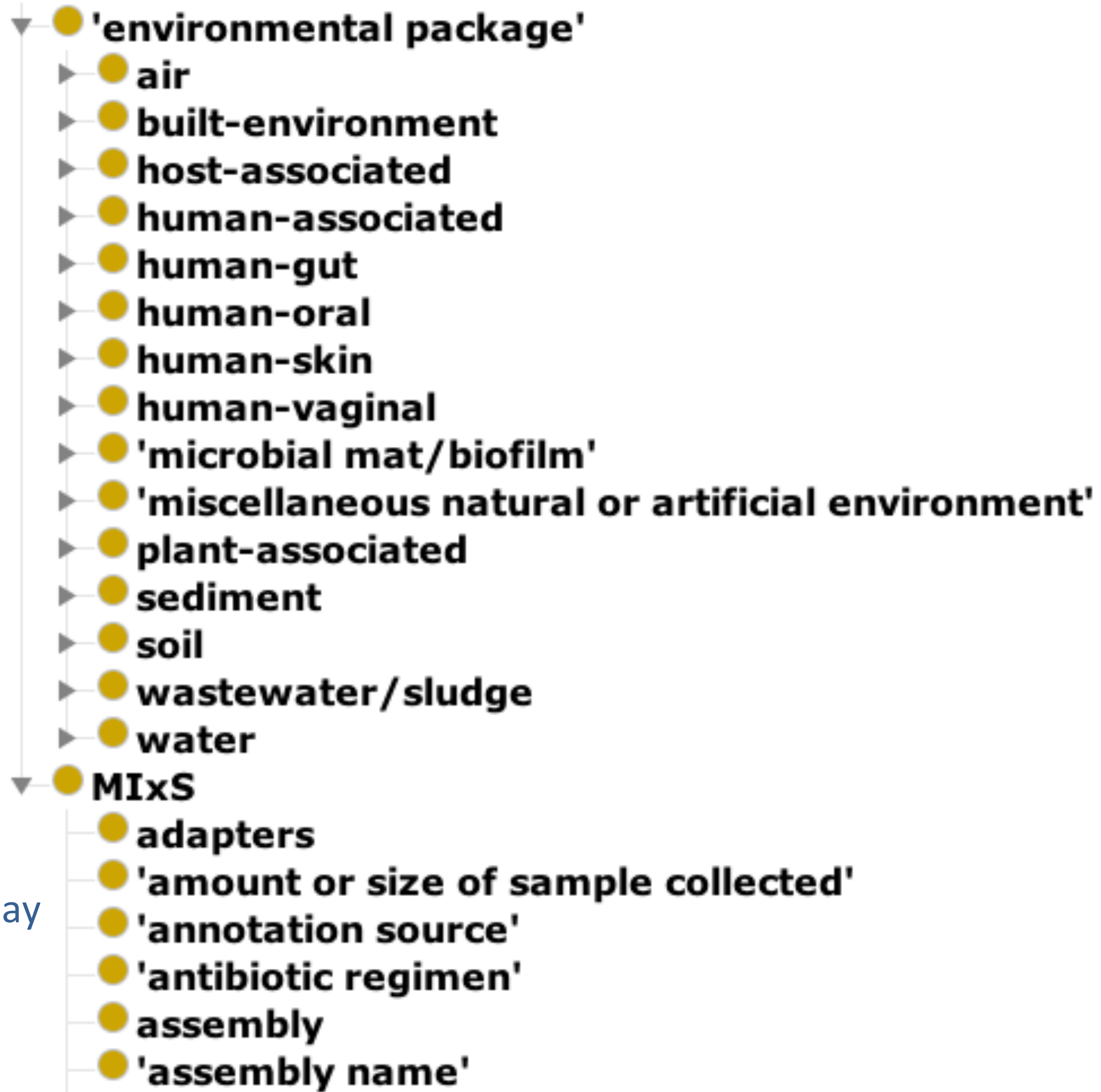
MIxS checklist

15 environmental packages

Cover over 400 different kinds of data that can be collected associated with samples used for sequencing

MIxS sequencing checklist

Minimal information required to collect for a sequencing assay



- 'environmental package'
  - air
  - built-environment
  - host-associated
  - human-associated
  - human-gut
  - human-oral
  - human-skin
  - human-vaginal
  - 'microbial mat/biofilm'
  - 'miscellaneous natural or artificial environment'
  - plant-associated
  - sediment
  - soil
  - wastewater/sludge
  - water
- MIxS
  - adapters
  - 'amount or size of sample collected'
  - 'annotation source'
  - 'antibiotic regimen'
  - assembly
  - 'assembly name'

Issues Using MIxS Checklist In Data Integration and Exploration

- Not good enough as consistent representation
  - Ambiguous use in data collection
    - MIxS: sample collection device or method
    - human-associated: urine/collection method
  - No guide on what terms should be used for a specific data type to collect
    - Host body site: name of body site where the sample was obtained from
      - Free text?
      - Controlled vocabulary?
      - Ontology term? UBERON anatomical
    - Host body site, host body product, host body habitat (tongue, feces, oral cavity)
  - Does not cover all the various data collected by investigators
- Cannot use for data exploration
  - Checklist, no hierarchy

Sample Details of Microbiome Data

- 6 data sets
- Around 100 different kinds of sample details collected
- Most sample details collected not in MIxS checklist
  - Baby’s delivery mode
  - Child’s diet
  - Number of child occupants

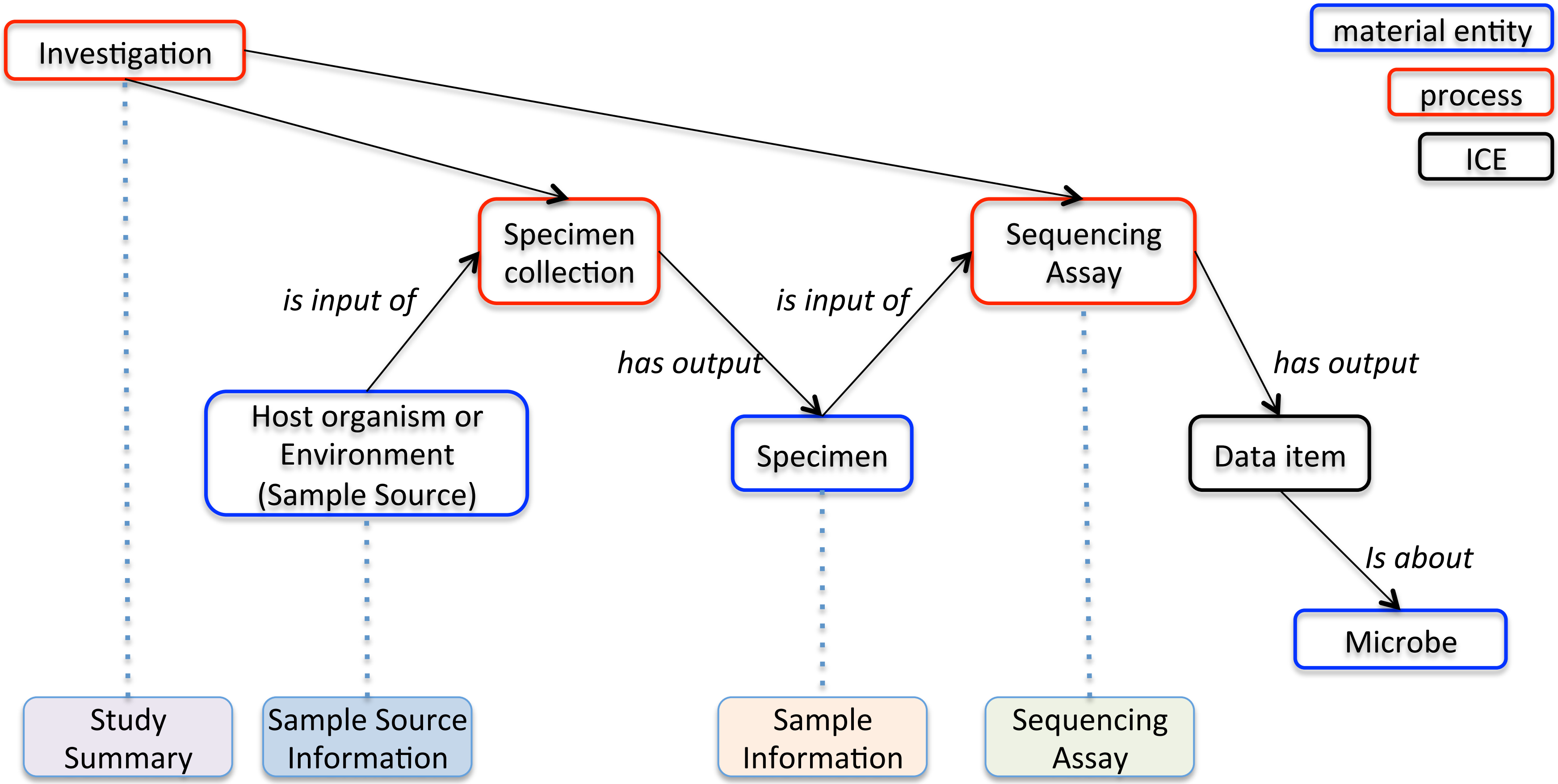
Ontology Application to Data Integration

- Mapping sample details (including part of MIxS checklist) to OBO Foundry ontologies
  - Sequence Ontology (SO)
  - Ontology for Biomedical Investigations (OBI)
  - Environmental Ontology (EnVO)
  - Information Artifact Ontology (IAO)
  - Phenotype qualities (PATO)
  - Ontology for General Medical Science (OGMS)
  - Uberon multi-species anatomy ontology (UBERON)
  - Units of measurement (UO)
  - Eagle-I Resource Ontology (ERO)
  - Gazetteer (GAZ)

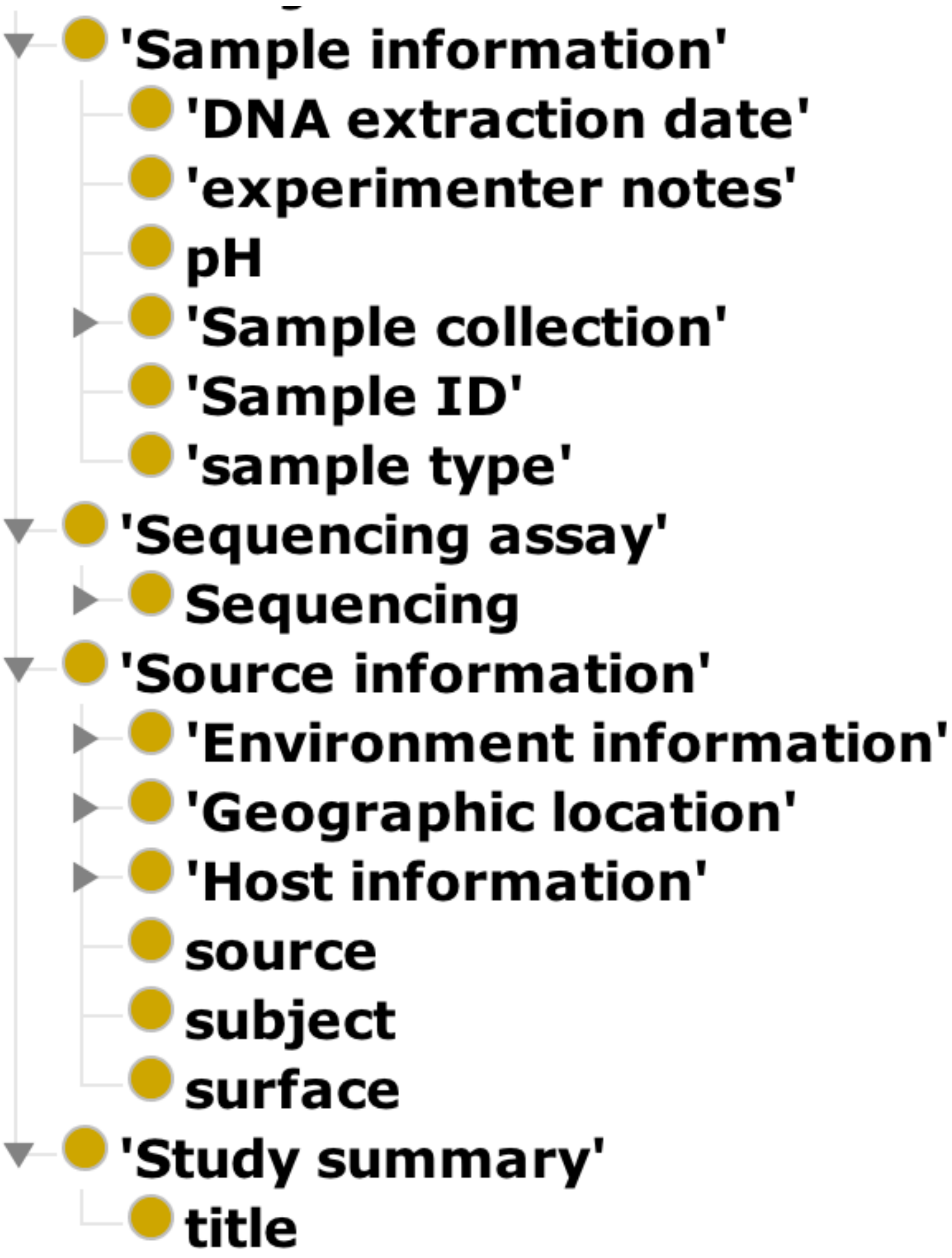
EuPath Ontology

- The EuPath ontology is an application ontology developed to support annotation of Eukaryotic Pathogen Database (http://eupathdb.org).
- The EuPath ontology was built based on the Ontology Biomedical Investigations (OBI).
- It uses the Basic Formal Ontology (BFO) as its upper level ontology.
- It imports terms from other OBO ontologies such as PATO, OGMS, DO, etc. when needed.
- It contains all ontology terms needed for Microbiome data annotations

Organize Sample Details for Data Exploration Based on Ontology



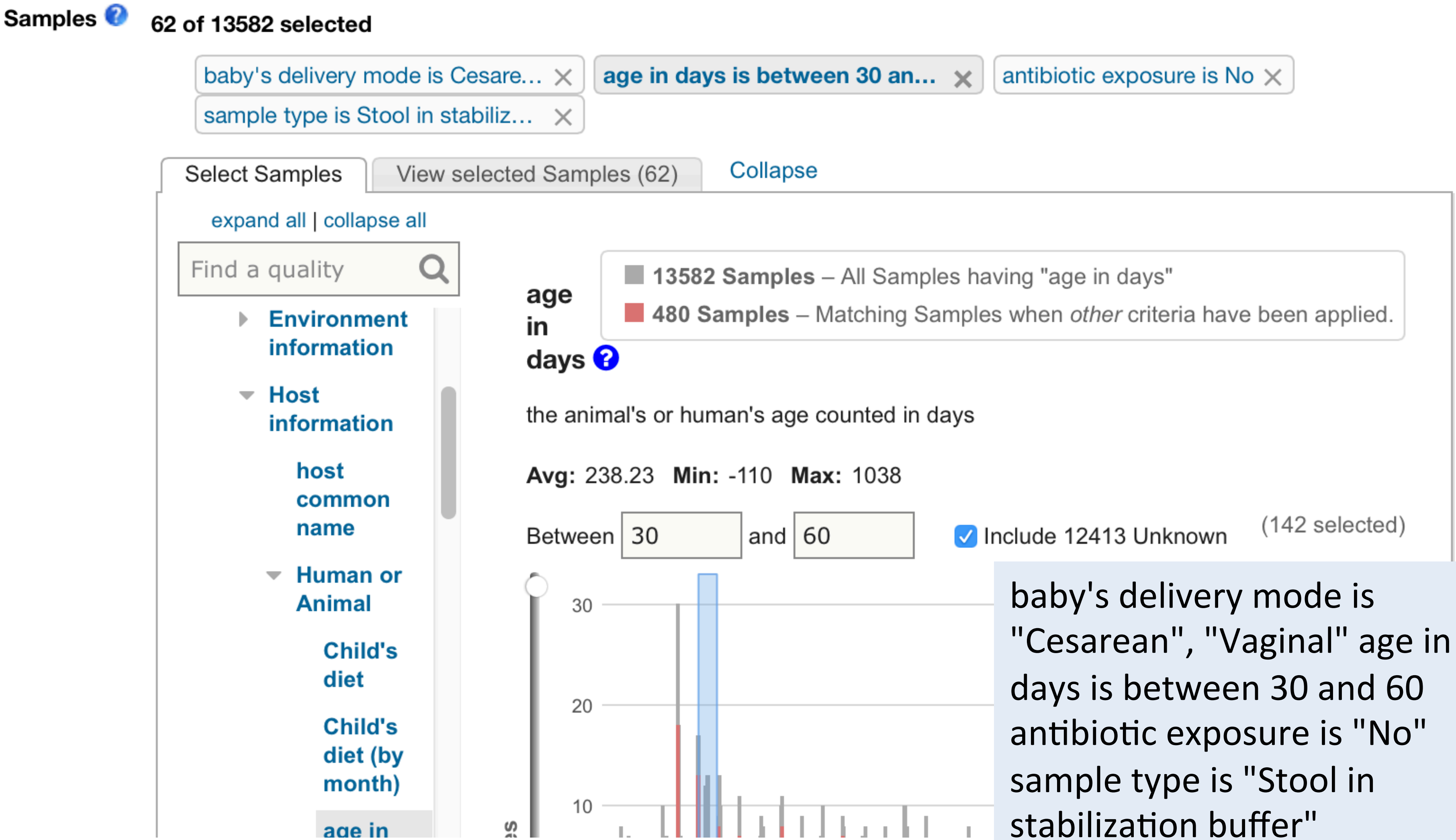
Rearrange sample details based on ontological model to guide search



- 'Sample information'
  - 'DNA extraction date'
  - 'experimenter notes'
  - pH
  - 'Sample collection'
  - 'Sample ID'
  - 'sample type'
- 'Sequencing assay'
  - Sequencing
- 'Source information'
  - 'Environment information'
  - 'Geographic location'
  - 'Host information'
  - source
  - subject
  - surface
- 'Study summary'
  - title

Using Sample Details To Explore and Analyze Microbiome Data

What is the impact of delivery mode and diet on the infant gut microbiome?



Samples 62 of 13582 selected

baby's delivery mode is Cesare… ✕ | **age in days is between 30 an…** ✕ | antibiotic exposure is No ✕ | sample type is Stool in stabiliz… ✕

Select Samples | View selected Samples (62) | Collapse

expand all | collapse all

Find a quality

age in days

■ **13582 Samples** – All Samples having "age in days"

🟥 **480 Samples** – Matching Samples when *other* criteria have been applied.

the animal's or human's age counted in days

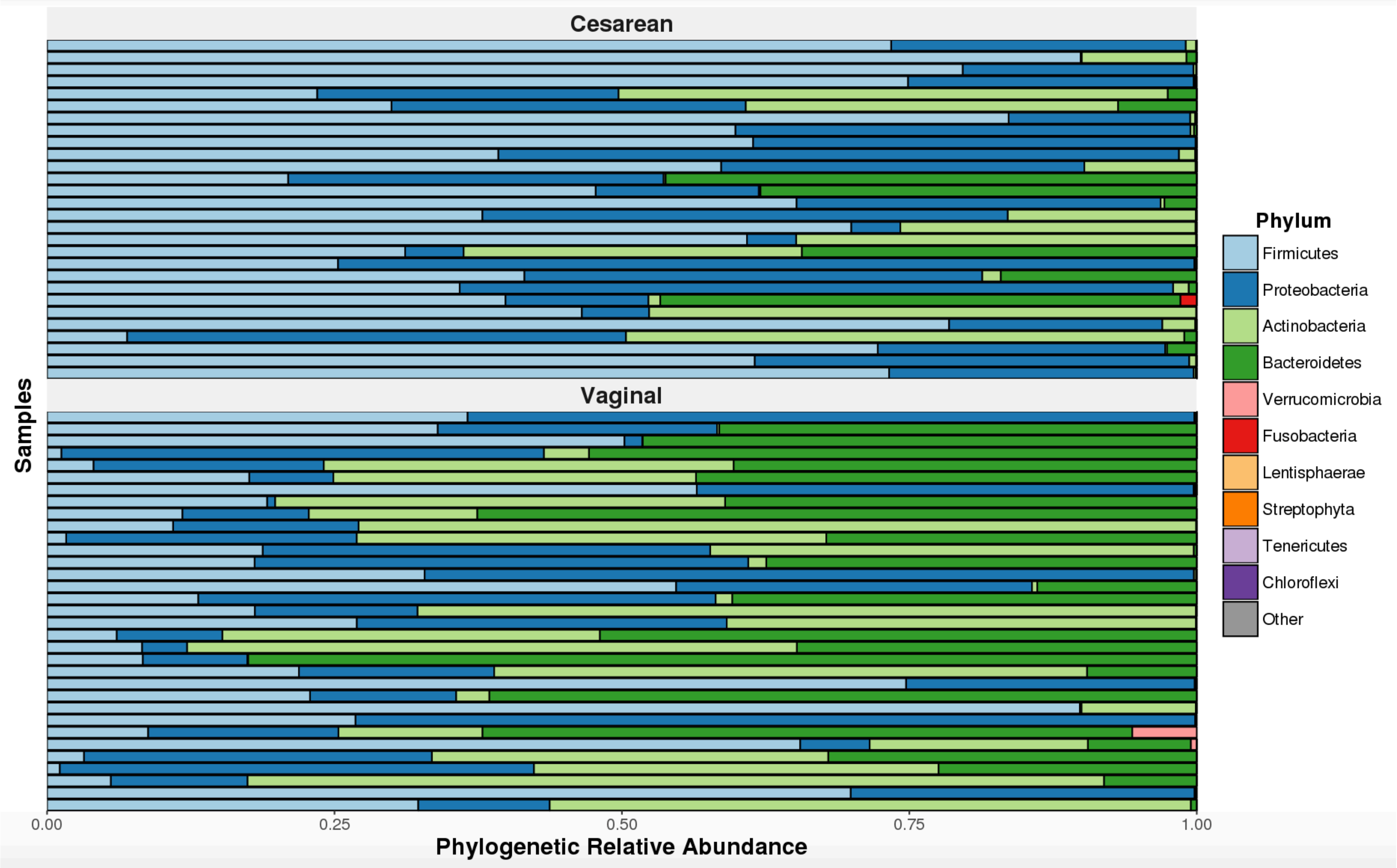
Avg: 238.23 **Min:** -110 **Max:** 1038

Between 30 and 60 ☑ Include 12413 Unknown (142 selected)

baby's delivery mode is "Cesarean", "Vaginal" age in days is between 30 and 60 antibiotic exposure is "No" sample type is "Stool in stabilization buffer"

Using Sample Details To Explore and Analyze Microbiome Data

What is the impact of delivery mode and diet on the infant gut microbiome?



EuPath Ontology Terms Submitted to OBO Foundry Ontologies

- Farm to fork food ontology (FOODON)
  - diet
  - feeding

https://github.com/FoodOntology/foodon/issues/28

- Ontology of Medically Related Social Entities (OMRSE)
  - family identifier
  - human traveling information

https://github.com/ufbmi/OMRSE/issues/62

OHMI: The Ontology of Host-Microbiome Interactions

- Aim: ontologically represent various entities and relations related to microbiomes, microbiome host organisms (e.g., human and mouse), and the interactions between the hosts and microbiomes at different conditions.
- Developed based on OBO principles.
- OHMI Github site: https://github.com/OHMI-ontology
- OHMI is available on:

Ontobee : http://www.ontobee.org/ontology/ohmi

BioPortal : http://bioportal.bioontology.org/ontologies/OHMI

OHMI is collaboratively Developed by Multiple Institutes

University of Michigan

Yongqun "Oliver" He, DVM, PhD

Haihe Wang, MD, PhD (visiting scholar from Harbin Medical University, China)

Yu Hong, MD, PhD (visiting scholar from Guizhou Univ. Medical School, China)

Duke University

Anna Maria Masci, PhD

University of Pennsylvania

Jie Zheng, PhD

Daniel P Beiting, PhD

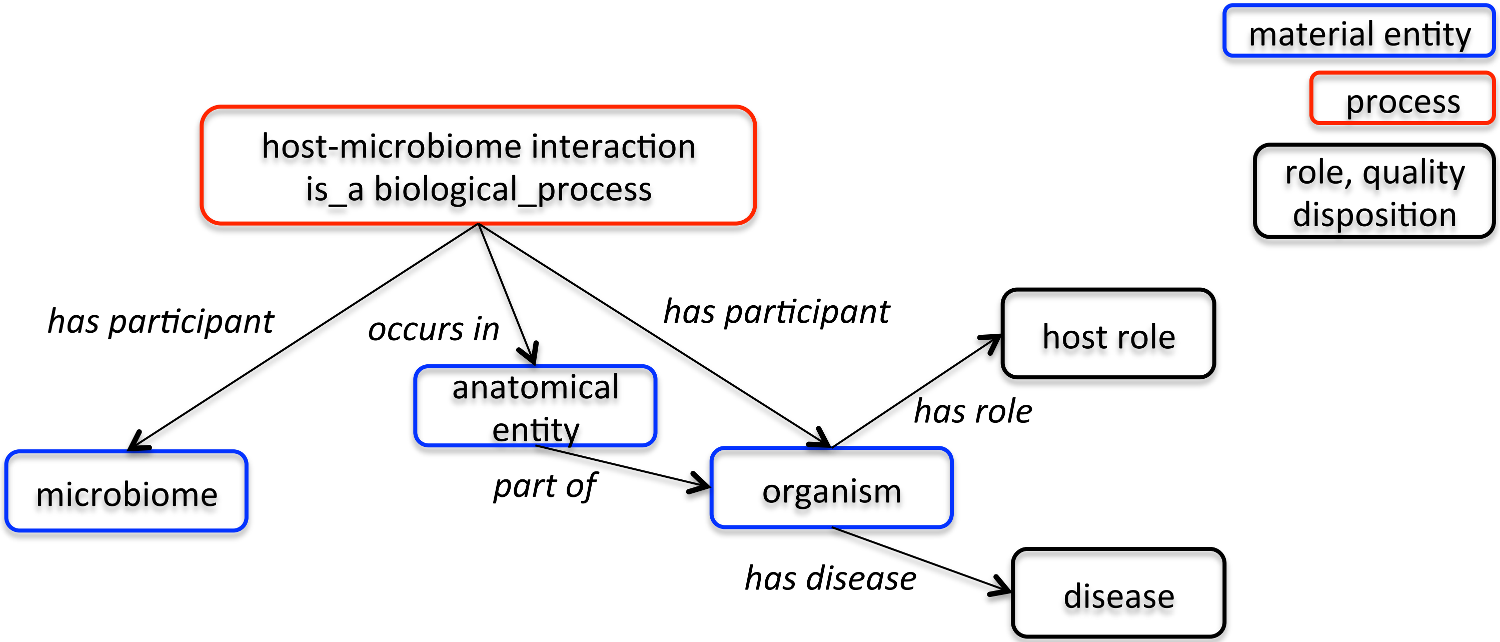
Medical University of South Carolina

Alexander V. Alekseyenko, PhD

Jihad S. Obeid, MD

OHMI

- 600 Classes
  - 279 OHMI classes
  - 321 imported classes (240 are NCBITaxon)
- Host-microbiome interaction model



Use of OHMI in MicrobiomeDB

- Data annotation
- Microbiome diversity in different conditions
  - Sample details ontology term enrichment analysis
- Microbiome diversity
  - Apply OHMI to relate microbiome diversity to host disease

Acknowledgement

EuPath Team
- John Brestelli
- John Iodice
- Shon Cade
- Steve Fischer
- Cristina Aurrecoechea
- Brian P. Brunk
- Jessie Kissinger
- Chris Stoeckert
- David Roos

MicrobiomeDB
- Dan Beiting (PI)
- Gabriel Fernandes (Co-I Belo)
- Francislon Oliveira

OHMI Ontology
- Alexander V. Alekseyenko
- Jihad S. Obeid
- Anna Maria Masci
- Yongqun "Oliver" He
- Haihe Wang
- Yu Hong

FOOD Ontology
- Damion Dooley

OMRSE Ontology
- Amanda Hicks