Genomic Metadata Standards pertinent to

Fungi in the built environment - next steps

Lynn M. Schriml
University of Maryland, School of Medicine
Institute for Genome Sciences
lschriml@som.umaryland.edu

ALFRED P. SLOAN FOUNDATION

Important issues in to studying fungi indoors

define key metadata
utilize a common set of term
standards metadata in study design and collection
avoid post-collect data mapping
produce comparable datasets
reduce variability in metadata within public repositories
comparing community composition

Samples

Genomic Sequence

Indoor metadata

What ?
Where ?

Data Standards

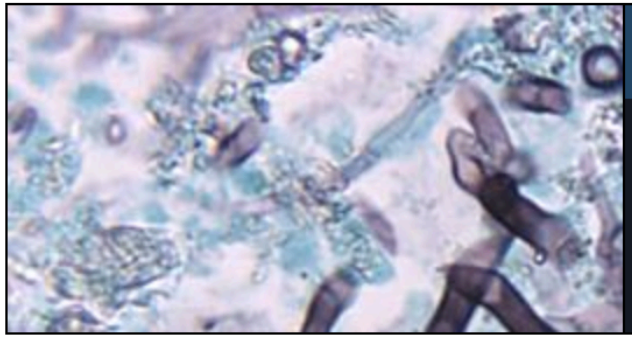
Data Integration

FungiDB
SGD
QIIME
EMBL/EBI
NCBI

NCBI Resources How To Ischriml My NCBI Sign Out

BioSample BioSample Search

Advanced Help



BioSample

The BioSample database contains descriptions of biological source materials used in experimental assays.

Query	# records
feces	22,592
faeces	1,750
ordure	2
dung	19
manure	154
excreta	153
stool	22,756

Pitfalls: multiple terms used for same biological concept -- community agreement
data standards utilized after data was collected -- post-collection data mapping
data provides submit the same depth or granularity of metadata

- ecological metagenomes
  - activated carbon metagenome
  - activated sludge metagenome
  - air metagenome
  - anaerobic digester metagenome
  - ant fungus garden metagenome
  - aquatic metagenome
  - beach sand metagenome
  - biocathode metagenome
  - biofilm metagenome
  - biofilter metagenome
  - biogas fermenter metagenome
  - bioreactor metagenome
  - bioreactor sludge metagenome
  - clinical metagenome
  - coal metagenome
  - compost metagenome
  - dust metagenome
  - fermentation metagenome
  - food fermentation metagenome
  - food metagenome
  - freshwater metagenome
  - freshwater sediment metagenome
  - groundwater metagenome
  - halite metagenome
  - hot springs metagenome
  - hydrocarbon metagenome
  - hydrothermal vent metagenome
  - hypersaline lake metagenome
  - ice metagenome
  - indoor metagenome
  - industrial waste metagenome
  - mangrove metagenome

ORGANISM

BioSample: [organism]

dust metagenome : 1 sample

indoor metagenome : 1,520 samples
{Hand, Mouse of Charge Nurse, Bed, Identification Card (Nurse), Doctor's table, Mouse/Keyboard of Doctor}

indoor metagenome AND dust: 844 samples

air metagenome AND fungi: 25 samples

ATTRIBUTES {metadata} : indoor AND fungi: 92 (QIIME)

altitude="0.0"
assigned_from_geo="n"
collection date="12512"
common_name="indoor metagenome"
elevation="8.0"
ENA-CHECKLIST="ERC000011"

env_matter="ENVO:dust"
environment biome="ENVO:terrestrial biome"
environment feature="ENVO:anthropogenic feature"

geographic location="GAZ:United States of America"
latitude="37.8731"
longitude="-122.2648"
room="Balcony"
room="Office" room="LR"
sample size="0.1, gram"
sample_id="1000298"
sample_name="1"
season="Fall"

taxon_id="1256227"
unit="1"
year_constructed="1940"

ENVO environmental ontology **ENVO:00002008** **ENVO:00000446**

GAZ Geographic Location Vocabulary

room="Bathroom" room="Bedroom"
room="Kitchen" **room="Outside"**

season="Summer" **Add Spring**
season="Winter"

??? building unit number ??

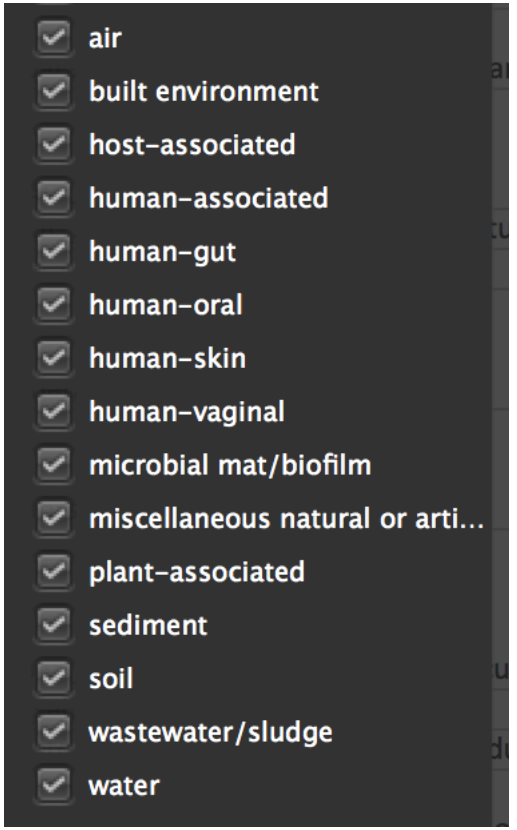
Organism: indoor metagenome

The MIxS Minimum Information family of standards

Capturing more Contextual Data

Geographic location of sample and environment
Extending the MIMS, MIGS and MIMARKS checklists

Extended by environmental packages: MIxS-BE



MIxS-BE:
A MIxS extension defining the Minimal Metadata for the Built Environment

Built Environment (BE) package. The BE package represents a minimal metadata description of the built environment to be collected and reported for each sequenced sample.

http://gensc.org/gc_wiki/index.php/MIxS_extensions

Minimal Information Checklist

MIxS Descriptors — project name, sample environment, sequencing specifications

MiXs – air environmental package — carbon dioxide, ventilation type, organism count

BE - core — surface material, temperature, humidity, occupancy

BE – building properties — room type, HVAC system, light source

BE – sample properties — sample processing, occupancy load

The ISME Journal advance online publication 24 October 2013; doi: 10.1038/ismej.2013.176

MIxS-BE: a MIxS extension defining a minimum information standard for sequence data from the built environment

Open

Elizabeth M Glass[1,1], Yekaterina Dribinsky[1], Pelin Yilmaz[2], Hal Levin[3], Robert Van Pelt[4], Doug Wendel[4], Andreas Wilke[1], Jonathan A Eisen[5], Sue Huse[6], Anna Shipanova[6], Mitch Sogin[6], Jason Stajich[7], Rob Knight[4,8], Folker Meyer[1,9] and Lynn M Schriml[10,1]

The need for metadata standards for microbe sampling in the built environment

NEW: MIxS-BE in Spring 2014 MIxS Release

MIxS-BE term
Carbon dioxide[a]
Ventilation type[a]
Organism count[a]
BE core
Surface material
Surface-air contaminant
Relative air humidity
Absolute air humidity
Surface humidity
Air temperature
Surface temperature
Surface moisture pH
Surface moisture
Dew point
Building occupancy type
BE Building properties
Indoor space (room type)
Indoor surface
Filter type
Heating and cooling system type
Substructure type
Building setting
Light type
BE Sample properties
Sample size sorting method
Space typical state
Typical occupant density
Occupancy at sampling
Occupant density at sampling

ALFRED P. SLOAN FOUNDATION

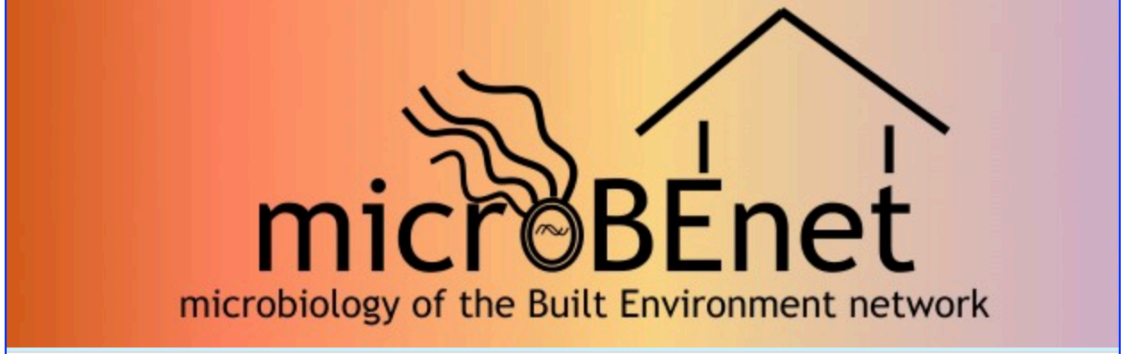
microBEnet

genomic STANDARDS consortium

Standards to support science http://gensc.org/

NATURAL ENVIRONMENT RESEARCH COUNCIL

NSF



Home Blog Resources Sloan Program microBEnet Projects Social Media Event

MoBE Metadata Resources

MoBE Metadata Resources

Metadata collected for MoBE studies can be organized and submitted to QIIME utilizing the MIxS-BE metadata standard. Templates for data submission and a Metadata Standard Submission Guide are below.

Please contact Lynn Schriml (lschriml@som.umaryland.edu) for questions regarding mapping your data to the MIxS metadata standard or Gail Ackermann (gail.ackermann@colorado.edu) at QIIME for data submissions.

MoBE_Metadata_Standard_Guide_2014

Templates:

MIxS-BE_metadata_template_2014

MIxS_metadata_template_2014

QIIME : sample_template_MoBE

QIIME: prep_template_MoBE

What is metadata?

MIxS-BE Metadata Standard

Prepare and load metadata files

QIIME Metadata Loader

http://microbe.net/

package item	definition	value syntax
surface material	Surface materials at the point of sampling.	[adobe|carpet|cinder blocks|concrete|glass|hay bales|metal|paint|plastic| stainless steel|stone|stucco|tile|vinyl|wood]
surface-air contaminant	Contaminant identified on surface.	[biocides|biological contaminants|dust|nutrients|organic matter|particulate matter|radon|volatile organic compounds]
relative air humidity	Partial vapor and air pressure, density of the vapor and air, or by the actual mass of the vapor and air.	{float} {unit} [%]
building occupancy type	The primary function for which a building or discrete part of a building is intended to be used	[airport|agricultural|commercial|educational|government|health care|high rise|industrial|low rise|market|military|office|parking|residential|restaurant|school|sports complex|storage|religious|transport|high rise|wood framed]
indoor space	A distinguishable space within a structure, the purpose for which discrete areas of a building is used.	[bathroom|bedroom|elevator|foyer|hallway|kitchen|locker room|office]
indoor surface	Type of indoor surface.	[cabinet|ceiling|counter top|door|wall|shelving|vent cover|window]
filter type	A device which removes solid particulates or airborne molecular contaminants.	[chemical air filter|electrostatic air treatment|gas-phase air treatment|HEPA filter|low-MERV pleated media|particulate air filter|ultraviolet air treatment]
heating and cooling system type	Methods of conditioning or heating a room or building.	[forced air system|heat pump|radiant system| steam forced heat|wood stove]

MIxS-BE: Building & Room Metadata Checklist (step 1)

building (general)	geographic location, region, city, state, country building setting, building type, building identifier, net area of building, building setting, building type (purpose).
	BE documentation: drawings, specifications, design, construction
architectural elements	building: # floors, # rooms, stairs, elevators room: ID/name, floor, volume, condition, # windows
	floors, walls, ceilings, room, windows, shading devices, doors & archways, furnishings:
	area, location, type, structure, material, condition, age, signs of water/mold
building systems	heating, cooling, ventilation: system ID, delivery method & locations
occupants and function	data types: manual, automated counts, video, estimated, average daily, houseplants, cleaning & cooking frequency
building environmental conditions	temperature (globe, variance), humidity, illuminations, solar radiation, air distribution

(Lynn Schriml & Jeff Kline)

MIxS-BE: Building & Room Data Gathering (step 2)

Scott Kelley, John Chase, Jeffrey Siegel, Greg Caporaso, Jack Gilbert, Kin Handley, Brent Stephens, Curtis Huttenhower, Regina Joice, Rachel Adams, Seema Bhangar, Jean Francis Ruiz Calderon, Atila Novoselec, Maria Gloria Dominguez-Bello, Humberto Cavallin Calanche, Jessica Green, James Meadow, Jeff Kline, Jordan Peccia, Susan Lynch, Chris Johnson, Al Levin

building (collection location)

Age of building
building type
University housing
location
longitude
Train stop collection location
train line
Station collection location
latitude
site

collection date

date
day
collection date
collection time
month
time
weekday
Season

occupancy

number of adults
number of children
Pets (# and kind)
Crowdedness
occupancy_sum

Number of infrared beam breaks at patient room doorways (a metric of human occupancy/activity)
House occupied during the day (1-2, 3-4, or 7)
Times/week a meal is cooked (1-2, 3-4, or 7)
sex
Houseplant (yes/no)
Usage
Indoor air CO2 concentration (patient rooms only)
Last time swapt/mopped/vacuumed
Indoor illumination levels

architectural elements

number of bathrooms
number of bedrooms
flooring type
flooring material: vinyl tiles, carpeting
Flooring (carpet, hard-floor, throw-rug)
exterior siding
exterior siding: fiber-cement (HardiPlank)
building exterior material
Slope of roof
wall
wall material: painted sheetrock
Interior wall construction (drywall/plaster)
interior wall type

humidity

average relative humidity
humidity at time of sampling
humidity on sampling day
Indoor air absolute humidity (humidity ratio)
relative_humidity
RelHumid_mean
Indoor air relative humidity
Chamber Temp/relative humidity

Standards Enable Large Scale Data Analysis

microbiome metadata rich samples annotated in compliance with
the MIxS metadata standard
(June 2014, the American Gut Project as sequenced 3,102 samples, 175 metadata variables)

Rachel Park (UMD SOM), Rob Knight, Gail Ackermann, Daniel McDonald
GitHub: https://github.com/biocore/American-Gut/tree/master/data)

Title

Environmental Drivers of gut microbiome community composition
in the American Gut Project

Study

Test if the taxonomic diversity of a person's gut flora
varies with the extent of their interaction with multiple environments

Cohort: 2,440 or 79% of the American Gut samples
data associated with fecal samples submitted by US residents over the age of 3.

Metadata (independent) Variables
(environmental microbiome exposure)

exercise frequency
exercise location
length of time in current state
drinking water source
recent travel

Microbiome associated
age
body mass index

Sample population descriptors
race
health status

Outcome Variables
(gut flora diversity)

unique observed operational taxonomic units (OTUs)
unique phyla, classes, orders, families and genera identified in each sample

Statistical Analysis
Multiple regression analysis
→ develop an OTU predictive model based on environmental microbiome exposure

Note: The square root of the OTU values was used for this analysis in order to mitigate for outliers in the OTU distribution.

Results

Multivariate Model ($p < 0.05$)

Exercise frequency	estimate	p value
less than 3x/week	-1.3325	**9.92 E-06**

increased gut microbiome diversity

→ people that exercised more than 3 times per week

OTUs decreased by 1.3325 in participants that exercised less than 3 times per week compared to those that exercised at least 3 times per week.

Simple Linear Regression

increased gut microbiome diversity

-- traveling outside the US within the past year

More recent travel increased gut diversity by 0.8238

decreased gut microbiome diversity

-- with increases in BMI

An increase in BMI decreased the square root of observed OTUs by 0.09093

metadata	estimate	p value
BMI	-0.09093	**< 2 E-16**
Outside US travel in past year	0.0838	0.00595


KEEP
CALM
AND
USE
METADATA