



Crowdsourcing the Advancement of Microbiome Research

Michalis Hadjithomas, Ph.D.
Microbiome Lead
DNAexus

What is Mosaic?

A collaborative cloud informatics platform which aims to:

Empower microbiome researchers

Facilitate the improvement of computational methods through **unbiased** benchmarking

Joining Mosaic provides access to:

~300 CPU hours to try out features

Viewing current activity in the Challenges

Free Challenge participation

Register for a free account at
mosaicbiome.com

Only name and email address are required

Features

Easy web-based user interface for:

Tool development

Tool execution on the cloud

The screenshot shows a web-based application interface for running bioinformatics tools. At the top, there is a vertical stack of tool icons and labels: Run GOTCHA, Run ConStrains, Run metaSPAdes version 3.11.1, Run Metaphlan2 v2.6.0, and Run IDBA-UD Assembler. Below this, a specific tool is highlighted with a blue border: Run CLARK-S for species level profiling (v 1.2.3.2). This tool has a yellow help bar at the top with the text "Need help? Learn more about running an app". The main interface is divided into sections: CONFIGURE, INPUTS, and OUTPUTS. The CONFIGURE section contains fields for Job Name (CLARK-S for species level profiling (v 1.2.3.2)), Workspace (Choose...), and Instance Type (High Mem 16). The INPUTS section contains fields for Input fastq file(s) (*.fastq), Paired end files? (True or False), Packaged CLARK DB (*.tar), Indicate Paired reads Identifier (Select...), and Concenate Reads (True or False). A note in the INPUTS section states: "This App Requires at least High Mem 16. Please visit the Readme section to learn more about the resources required." The OUTPUTS section is partially visible on the right.

Features

Facilitate collaboration:

Share:

Tools/Data

Tools executions

Jupyter notebooks

Groups

Discussions

The screenshot shows the mosaic platform interface. At the top, there's a navigation bar with links for Workspaces, Tools, Discussions, Groups, Challenges, and a user profile for Michalis. Below the navigation bar is a "Create New Notebook" button. The main area is divided into several sections: a left sidebar for "Antibiotic_resistance_Sumayah" showing a list of files like "antibiotic_resistance.csv", "al.stb", "Chdb.csv", etc.; a central workspace area with a title "antibiotic_resist X" and a subtitle "Analysis of antibiotic resistance in the premature infant gut microbiome" by Sumayyah F. Rahman 2017; and a right-hand terminal window titled "In [9]" containing Python code for data analysis and machine learning. The terminal code includes imports for pandas, numpy, sciPy, scikit-learn, and various stats modules, followed by specific commands for data processing and classification.

```
import pandas as pd
import numpy as np
import sciPy
from sciPy.stats import chi2_contingency
import matplotlib.pyplot as plt
import matplotlib.patches as mpatches
import seaborn as sns
import statsmodels
from statsmodels import sandbox
from statsmodels.sandbox import stats
from statsmodels.sandbox.stats import multicomp
import statsmodels.stats
from sklearn import preprocessing
from sklearn.decomposition import PCA
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier
from sklearn.metrics import precision_recall_fscore_support as score
import collections
from pylab import plot, show, savefig, xlabel, figure, hold, ylim, legend, boxplot, setp, axes
import warnings; warnings.simplefilter('ignore')
%load_ext rpy2.ipython
%load_ext ivertions
```

Features

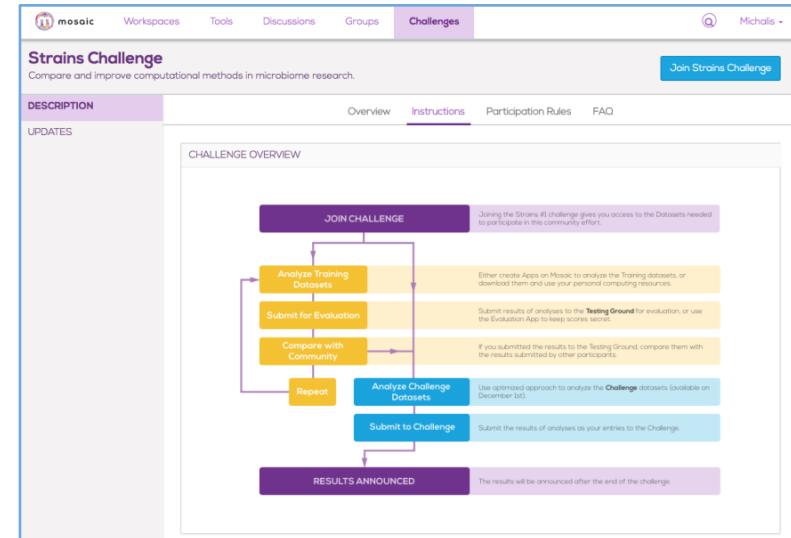
A space for hosting challenges

Community challenges engage the scientific community and enable the **unbiased** benchmarking of tools while promoting **innovation**

Interactive and immediate participation

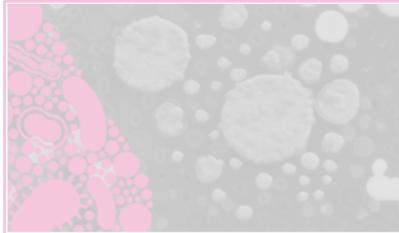
Unique datasets

Sponsored by hosts



Mosaic Community Challenges

Crowdsource the
benchmarking and
advancement of
microbiome methods

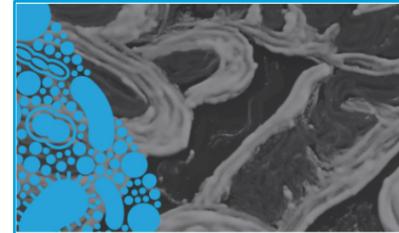


Mosaic Community Challenge:

Strains #1

Concluded

Benchmark and improve computational methods in microbiome characterization at the strain level



Mosaic Community Challenge:

Clinical Strain Detection

Opens: February 12, 2018

Closes: June 25, 2018

Benchmark and advance computational methods for targeted strain detection, tracking the presence of certain known strains in a sample.



Mosaic Community Challenge:

Standards

Coming Soon

Capture the diversity of protocols for NGS-based microbiome profiling methods and analyses as a first step in understanding their impact on the variability of the taxonomic results.

Hosts & Scientific Partners



Powered by **DNA**nexus

<http://mosaicbiome.com/challenges/>

Clinical Strain Detection

Objective:

Benchmark the ability of metagenomic tools in **detecting** the presence of certain known strains in a sample.



mosaic

Fundamental Question

IS  IN  ?

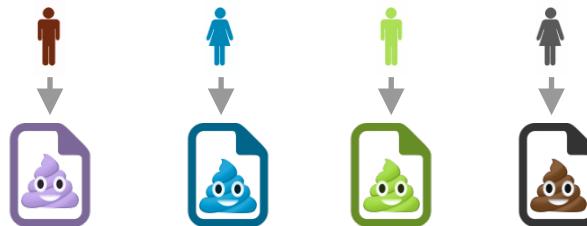
Given the sequence of a single bacterial genome and the metagenomic data from one microbiome, can we accurately ascertain the presence or absence of the bacterium in that microbiome?

Strains #2 Setup

Bacteria were isolated from human fecal samples and sequenced

Genome and metagenome sequences are provided to the participants

Goal: match each organism to source microbiome

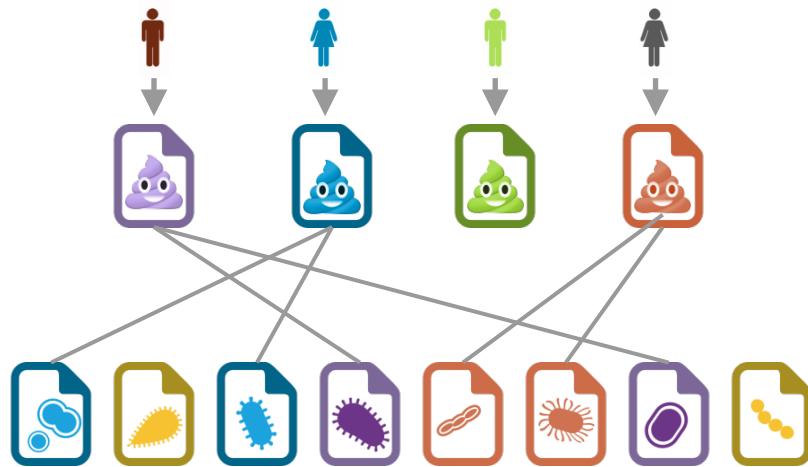


Strains #2 Setup

Bacteria were isolated from human fecal samples and sequenced

Genome and metagenome sequences are provided to the participants

Goal: match each organism to source microbiome



Need to Know

Cash prize for top 3 performers

- First place: \$10,000
- Second place: \$5,000
- Third place: \$2,000

Challenge ends: June 25, 2018

Standards

Objective:

Capture the **diversity** of protocols for NGS-based microbiome profiling methods.



mosaic

Why are Standards important?

Need for accurate and reproducible NGS-based microbiome profiling

Variability in results can come from:

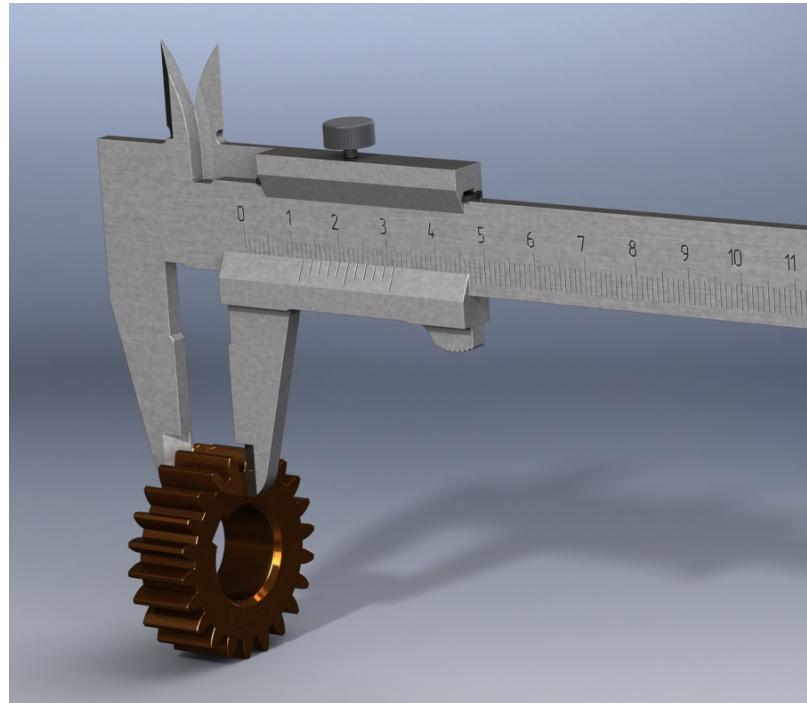
- Sample collection and storage

- Nucleotide extraction protocol

- Sequencing technology

- Bioinformatics

Need to capture the extent of **diversity** and **variability** across protocols using actual complex samples



Standards Setup

Crowdsource sequencing and bioinformatics

Anyone can participate; only requirement is a Mosaic account

Fecal samples and DNA mixtures provided



Samples, human feces and DNA mixtures,
created by partners

Participants request
and sequence
samples, and upload
files/metadata to
Mosaic

Uploaded metagenomes
are analyzed using
Mosaic's "house" pipelines

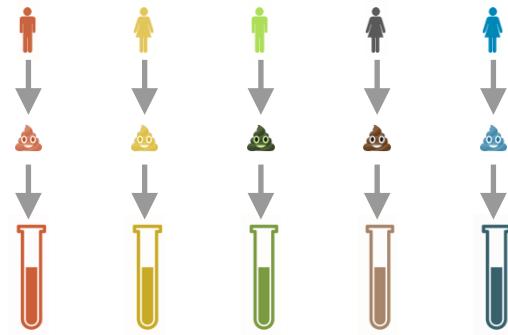
Results are presented to the
community through an
interactive experience



<https://platform.mosaicbiome.com/challenges/8>

DNAexus

Standards Sample Creation



Fecal matter



Samples, human feces and DNA mixtures,
created by partners

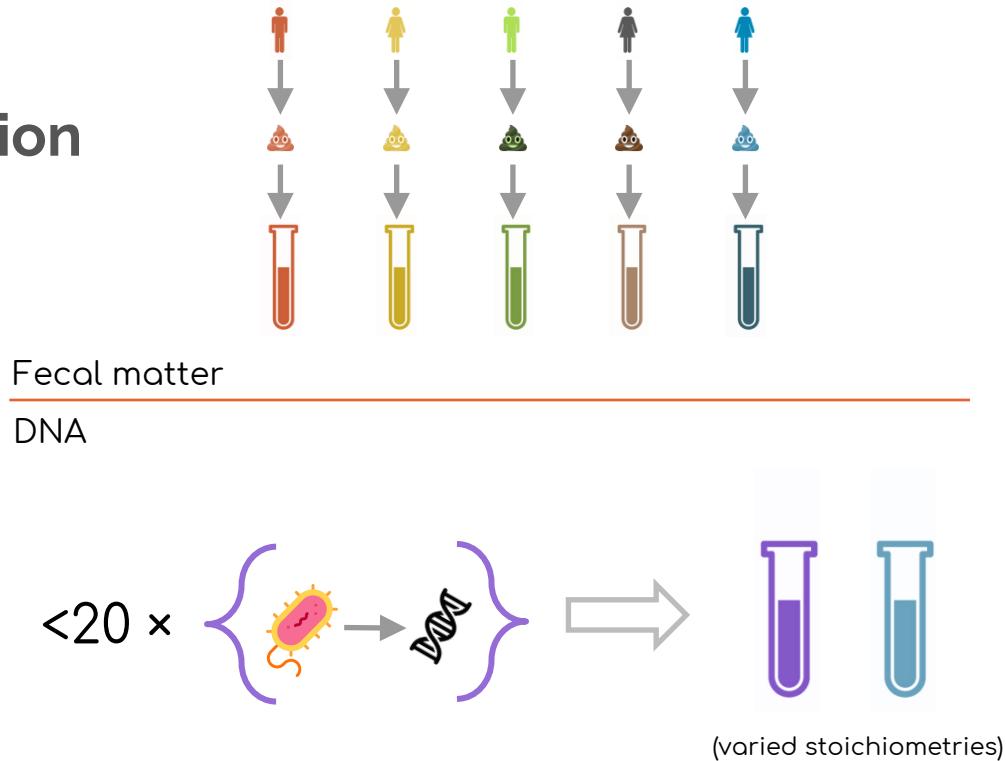
Standards Sample Creation



Samples, human feces and DNA mixtures,
created by partners



<https://platform.mosaicbiome.com/challenges/8>



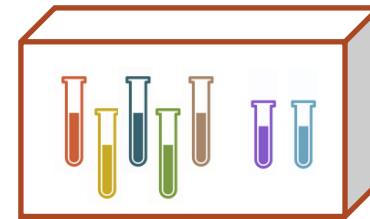
DNAexus

Standards Sample Creation



Samples, human feces and DNA mixtures,
created by partners

700 ×



1. No pre-approval needed
2. Up to 3 kits per user
3. Shipped **free** globally (if legal)
4. Commit to sequence and upload data

Standards Setup



Samples, human feces and DNA mixtures,
created by partners

Participants request
and sequence
samples, and upload
files/metadata to
Mosaic

Uploaded metagenomes
are analyzed using
Mosaic's "house" pipelines

Standards Challenge Sample Submission

Metadata Submission

Cancel

Save Changes

Submit

Focus on User Experience

Around 100 metadata values

Lower the hurdles of sharing metadata.



mosaic

ORDER ID 123986 [Edit Samples](#) [Share](#)

SAMPLE

*Mandatory Field

SAMPLE STORAGE

Were samples stored between receipt and processing? *

Storage period (days)

Storage temperature

Thawing procedure used

DNA EXTRACTION

Samples transported in lab *

Were samples spun down before processing? *

DNA extraction kit manufacturer

DNA extraction kit model

1
371

Library Type

-RDS-My-sample1.fq
DS-My-sample12.fq

ata from Poo #1

 yes
 #2
ID 97234

Unpaired Library Type

My-sample2.fq

Copy data from Poo #1

 no
 #3
ID 94121

Paired Library Type

FWD-RDS-My-sample3.fq
RV-RDS-My-sample3.2.fq

Copy data from Poo #1

 no
 #4
ID 7383

Interleaved Library Type

My-sample4.fq

Copy data from Poo #1

 no
 #5
ID 87234

Interleaved Library Type

Upload Data

Copy data from Poo #1

 yes
 #1
ID 87234

Interleaved Library Type

Upload Data

Copy data from Poo #1

 no
 

© 2017, Mosaic Project, All Rights Reserved

[Home](#) | [About](#) | [Terms of Service](#) | [Documentation](#) | [Contact Us](#)

Standards Setup

Analytical Pipelines
NINJA-OTU for 16S Data
SHOGUN for Metagenomic Data



Samples, human feces and DNA mixtures, created by partners

Participants request and sequence samples, and upload files/metadata to Mosaic

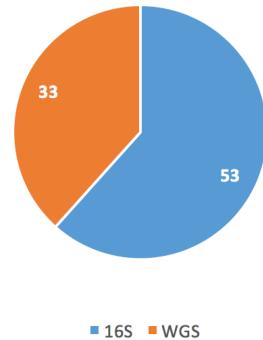
Uploaded metagenomes are analyzed using Mosaic's "house" pipelines

Results are presented to the community through an interactive experience

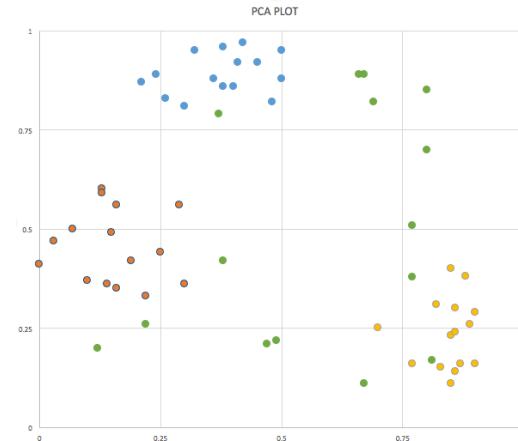
Standards Data Exploration

Metadata Summaries

Library Construction Strategy



Analysis Results



Standards Setup

Analytical Pipelines
NINJA-OTU for 16S Data
SHOGUN for Metagenomic Data



<https://platform.mosaicbiome.com/challenges/8>

DNAexus

Things to Know

1. Data Freezes
 - Data and Metadata will be made available for download at pre-defined milestones
2. Roundtable planned by Nature Publishing Group
 - Advance community-driven solutions to scientific hurdles in microbiome science and research

Why Participate in the Mosaic Standards effort?

1. Help the community by sequencing samples
2. See how your protocols and analyses of your data compare to the rest of the submissions
3. Be acknowledged on Mosaic (if you wish)
4. Participate in publishing the results (if you wish)
5. Have a chance to be invited to the Nature-hosted roundtable

Collaborators

Hosts

Janssen Research & Development, LLC

The BioCollective

DNAgentek

DNAexus

Scientific Partners

Scott Jackson, NIST (DNAmixtures)

Don Knights' group (Assistance with analytical pipelines)



mosaic

DNAexus



JOIN US



mosaic

Participate in Challenges

Share your ideas

Spread the word

mosaicbiome.com/challenges