Oliveira_et_al_2024_V_Demo:

This repository pertains to the manuscript "Statistical Design of a Synthetic Bacterial Community that Clears a multi-drug Resistant gut pathogen."

Demonstration:

There are two different Demo:

1. Demo_on_Real_Data [Runtime 5 minutes]

This folder contains a demonstration of the construction of Strain and metabolite landscapes using the PCA projections of the real dataset. These diagrams can also be found in our manuscript.

2. Demo_on_Synthetic_Data: [Runtime 5 minutes]

This folder contains a demonstration of PCA Analysis, Metabolite RF modeling, and PCA landscape construction on the synthetic datasets. These synthetic datasets are of the same size and similar nature as the real dataset. The results here are solely for pedagogical demonstration of our code. No part of these datasets and results generated have been used in our analysis and the manuscript.

Installation:

1. In your terminal type and execute the following: [https://github.com/aramanlab/Oliveira_et_al_2024_V_Demo.git]

2. Alternatively, go to the link (https://github.com/aramanlab/Oliveira_et_al_2024_V_Demo.git (https://github.com/aramanlab/Oliveira_et_al_2024_V_Demo.git)) And download the zip file.

Installation/Download Time: 1 minute.

Datasets:

This repository contains the following datasets:

---

1. PCA_coordinates_Metabolite_space_original_81_actual.csv -------
   [Metabolite_Space_Principal_component_landscape.ipynb => Output]
2. PCA_coordinates_Strain_Presence_Absence_Space_Original_96_actual.csv -------
   [Strain_Space_Principal_component_landscape.ipynb => Output]

Upon Publication, the repository will also contain:

Strain presence-absence vs KP CFU for Original and out-of-sample experiments:

1. consortia_taxa_presence_and_KpCFUs.csv-------[Iput => RF_Model_Metabolites.ipynb, Strain_Space_Principal_component_landscape.ipynb]
2. consortia_taxa_presence_and_KPCFUs_OOS.csv

Metabolite Z score vs KP CFU for Original and out-of-sample experiments:

1. Metabolite_120_Hr_dataset.csv-------[Input => RF_Model_Metabolites.ipynb, Metabolite_Space_Principal_component_landscape.ipynb]
2. Metabolite_120_Hrs_OOS_set_log10_CFU.csv-------[Input => RF_Model_Metabolites.ipynb]

Python/R Notebooks and Scripts:

This repository contains the following notebooks:

Author: Bipul Pandey: [Mac OS 14.4 Python 3.9.12]

Major Package Detail: scikit-learn[1.0.2] , scipy[1.7.3], Matplotlib[3.5.1]

1. Jupyter Notebook to construct Metabolite space landscape of the original experiments [Metabolite_Space_Prinincipal_component_landscape.ipynb] [Runtime 5 minutes]
2. Jupyter Notebook to construct Strain space landscape of the original experiments [Strain_Space_Prinincipal_component_landscape.ipynb] [Runtime 5 minutes]
3. Jupyter Notebook to construct a Random Forest Model on Metabolites to predict KP suppression [RF_Model_Metabolites.ipynb] [Runtime 5 minutes]
   - Trained on original 96 experiments
   - Tested on 60 Out-of-sample experiments

Pseudocodes:

Pseudocodes describing the overarching steps in the calculations can be found in the Pseudocodes folder and the folders with notebooks. They also contain Run times for the codes they describe.