

mGPS Interface Tutorial

1. Introduction

mGPS interface is a web program based on the mGPS application created by Shiny. It can build a microbial origin prediction model and predict the origin of microbes. To learn more about mGPS, please visit [mGPS](#).

2. Function and Output

2.1 Function 1: Build a new prediction model using mGPS

2.1.1 Function 1 introduction

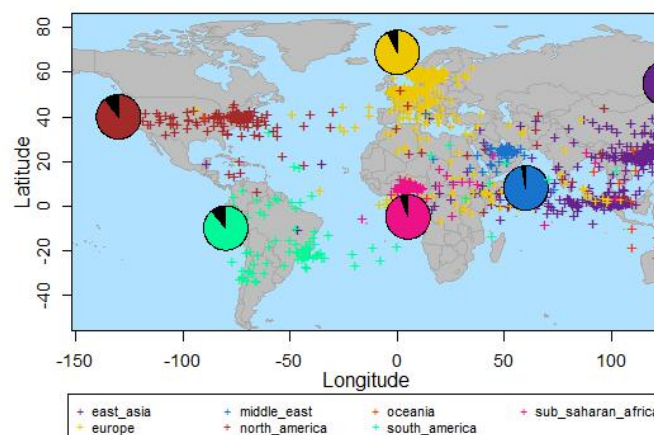
This mode can use the mGPS tool to build a microbial source prediction model based on the microbial abundance data and coordinates data uploaded by the user.

2.1.2 Output plots

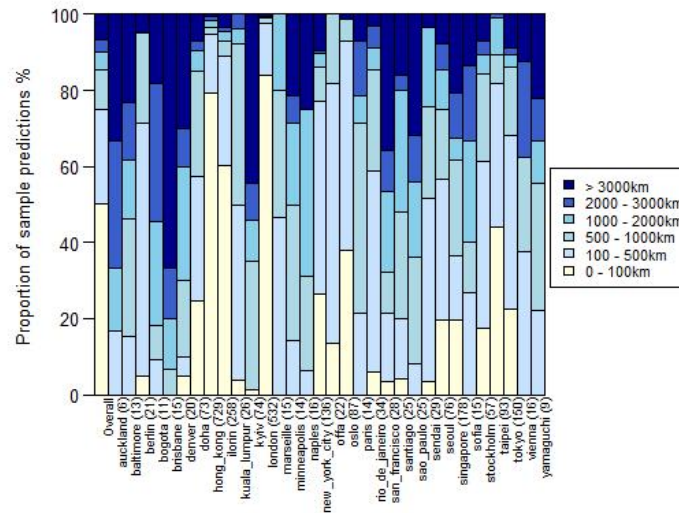
The "Result Plot" of this mode will show the accuracy of the prediction model trained by the mGPS tool and based on the reference microbial database uploaded by the user. The original database will be divided into 5 folds, and mGPS will use 4 of these folds to train the model, and the resulting model will be used to predict the microbial source of the remaining fold. Iteratively obtain the prediction result of the original database and compare it with the actual location of the microorganism.

Example output plot:

a. World map: samples' prediction origins are plotted on the world map. By adjusting the range of longitude and latitude, users can select the plotting area of the figure on the map. In addition, users can choose whether to pull the predicted point into the continent or the ocean. The predicted origin of original samples will be mapped onto the world map. The pie charts with colors divided by geographic area represents the proportion of samples in the area whose origin region is predicted to be the same as the true location.



b. Accuracy bar plot: model built by mGPS accuracy per site for the original reference dataset. mGPS accuracy is shown per-site as the distances between the predicted and true sampling site for the reference samples. The average prediction accuracy across all samples with each population given equal weight is shown on the left.



2.1.3 Output files

- Prediction data:** Records the predicted original coordinates of the reference samples. Also contains the distance between the predicted and true sampling site for the reference samples.
- Optimal features in prediction model:** Records the optimal features (microorganisms) used to predict the original coordinates that selected by mGPS algorithm.
- feature subsets accuracy in feature elimination:** Records the accuracy of prediction model with different features (microorganisms) subset size. The algorithm will use the subset size with highest accuracy.
- Model:** The origin prediction model built by mGPS algorithm saved in Rda format. The model can be downloaded to view the details of the model through load into r. Also, it can be upload to predict origin of new samples through function 3 Use existing model to predict new samples

2.2 Function 2: Build a new prediction model using mGPS and predict new samples

2.2.1 Function 2 introduction

This mode can train the microbial origin prediction model based on the reference data set uploaded by the user. The constructed prediction model will be used to predict the new sample to be tested provided by the user and report the prediction result of the sample source. (If user want

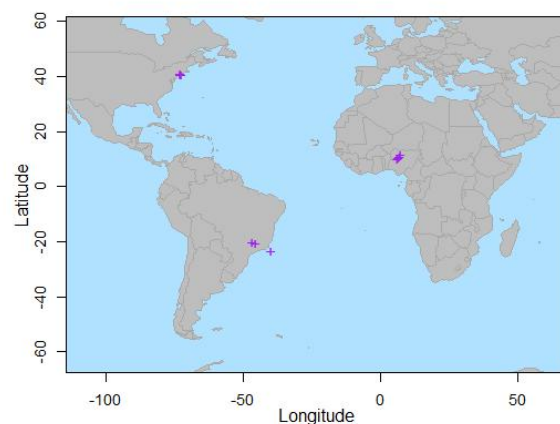
to visualize the accuracy of the model, please use function: Build a new prediction model using mGPS)

2.2.2 Output plot

The reference datasets will be used to construct a origin prediction model by mGPS. Then this model will be used to predict origin of new samples.

Example output plot:

a. World map: samples' prediction origins are plotted on the world map. By adjusting the range of longitude and latitude, users can select the plotting area of the figure on the map. In addition, users can choose whether to pull the predicted point into the continent or the ocean.



2.2.3 Output files

- a. Prediction data: Records the predicted original coordinates of the new samples.
- b. Optimal features in prediction model: Records the optimal features (microorganisms) used to predict the original coordinates that selected by mGPS algorithm.
- c. feature subsets accuracy in feature elimination: Records the accuracy of prediction model with different features (microorganisms) subset size. The algorithm will use the subset size with highest accuracy.
- d. Model: The origin prediction model built by mGPS algorithm saved in Rda format. The model can be downloaded to view the details of the model through load into r. Also, it can be upload to predict origin of new samples through function 3 Use existing model to predict new samples

2.3 Function 3: Use existing model to predict new samples

2.3.1 Function 1 introduction

This mode can predict new sample origin based on an existing prediction model. The

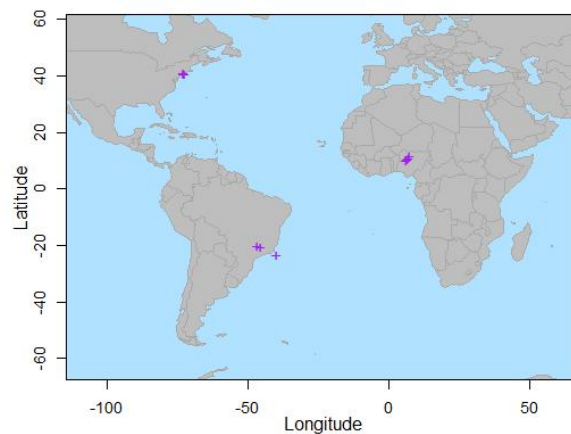
2.3.2 Output plot

This mode can predict new sample origin based on an existing prediction model. Model can be downloaded in Output tab of function: Build a new prediction model using mGPS or Build a new prediction model using mGPS and predict new samples

Example output plot:

a. World map:

new samples' prediction origins are plotted on the world map. By adjusting the range of longitude and latitude, users can select the plotting area of the figure on the map. In addition, users can choose whether to pull the predicted point into the continent or the ocean.



2.3.3 Output files

a. Prediction data: Records the predicted original coordinates of the new samples.