

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

Geographical origin prediction of microbiome

The screenshot displays the mGPS web interface. On the left is a sidebar with input controls, and on the right is the main content area with tabs for 'Welcome', 'HELP', 'Result Plot', and 'Output'. The 'HELP' tab is active.

Prediction program

Build a new prediction model using mGPS

Input file

☒ Merged metadata and abundance file
☐ Separate metadata and abundance file

Upload the reference merged dataset file

Browse... No file selected

Enter the main locality level

Enter the locality hierarchy

Column range of abundance data

☐ Locality sample size cut off (Optional)
☐ Remove values (Optional)
☐ Subsets in feature elimination (Optional)

Function description

This model 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.

Usage

In left side bar:

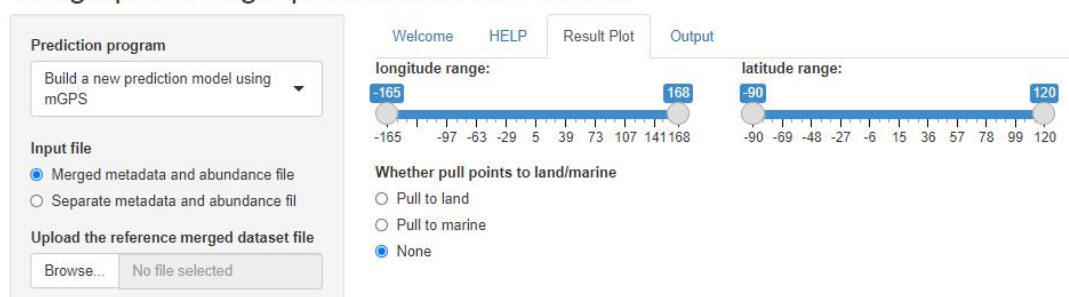
1. Select **Prediction program** as *Build a new prediction model using mGPS*
2. **Input file(s)**: Upload data file(s) (in .csv format) containing microbial abundance data and metadata. In metadata, at least one locality (eg. continent, city) and coordinates (necessary) data columns should be included. The metadata and abundance data of the sample can be merged into one file (*Merged metadata and abundance data*), or uploaded as two files (*Separate metadata and abundance data*)
- When *Separate metadata and abundance file* is selected, **Merge column name in metadata/abundance file**: Input the header name of column which is the merged column in two files.
3. **Enter the main locality level**: Input the main locality target. It should same as that column header. (eg. city)

2.1.1 Function 1 introduction

In this mode, you 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

Geographical origin prediction of microbiome

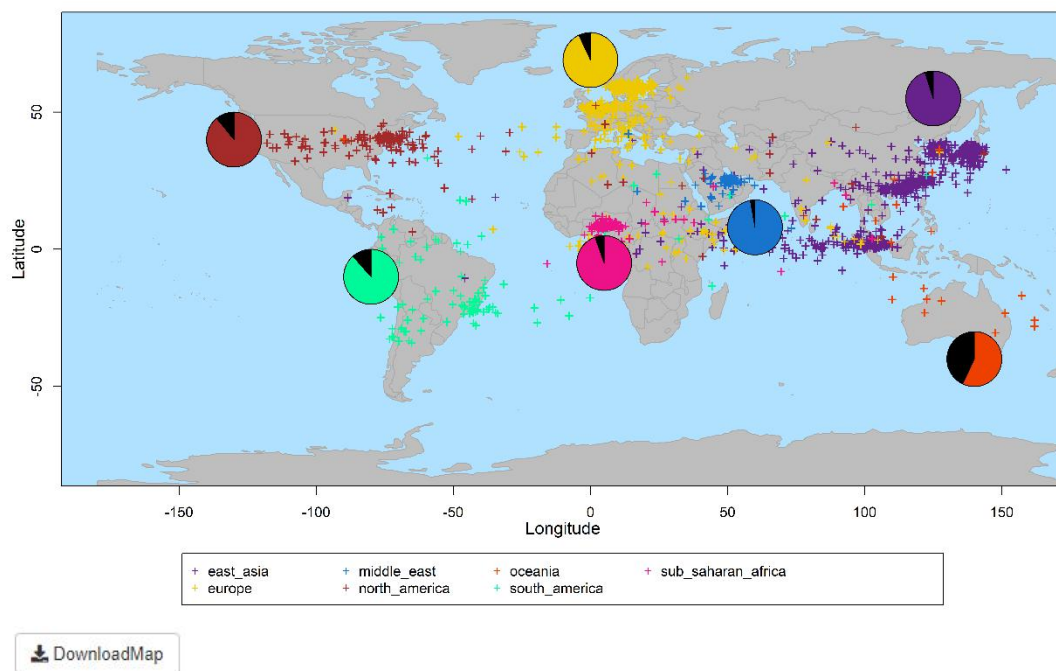


The "Result Plot" of this mode will show the accuracy of the prediction model trained by the mGPS tool and based on the reference microbial database you uploaded.

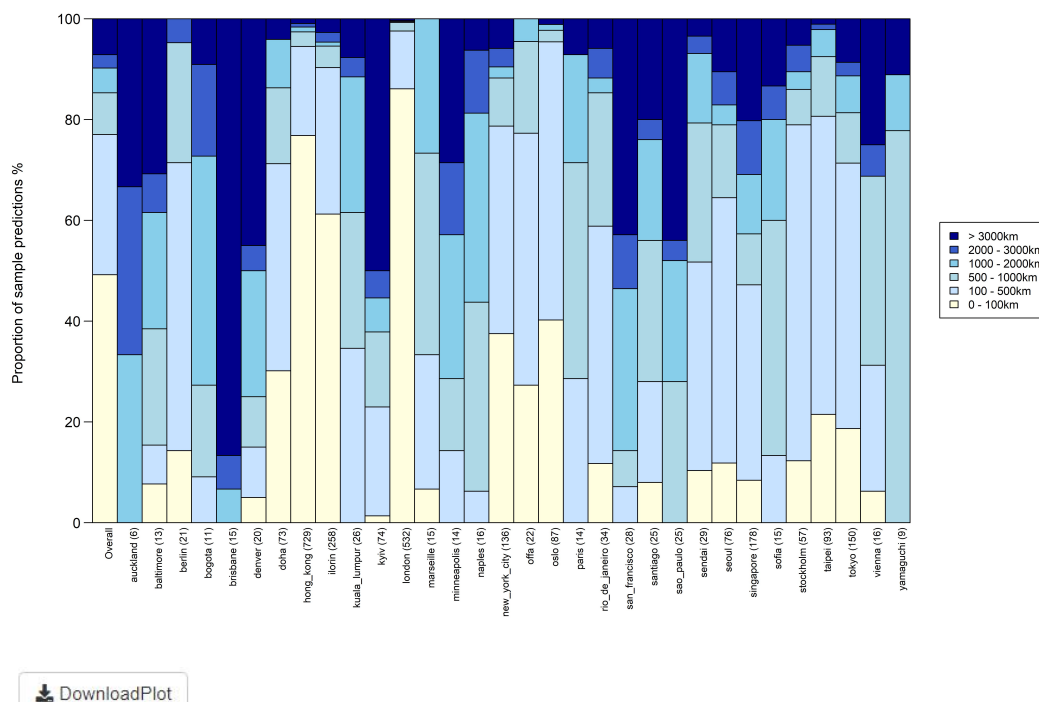
The original database will be divided into 5 folds, and mGPS will use 4 of these folds to train the model, and the resulting model will be used to predict the microbial source of the remaining fold. Iteratively obtain the prediction result of the original database and compare it with the actual location 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, you can select the plotting area of the figure on the map. In addition, you 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 represent the proportion of samples in the area whose origin region is predicted to be the same continent as the true location. You can download and save the png file of this map through the button 'DownloadMap' below the figure.



b. Prediction accuracy bar plot: The prediction accuracy of the model is shown per-site as the distances between the predicted and true sampling sites for the reference samples. The average prediction accuracy across all samples with each population given equal weight is shown on the left. You can download and save the png file of this plot through the button 'DownloadPlot' below the figure.



2.1.3 Output files

Geographical origin prediction of microbiome

Prediction program
 Build a new prediction model using mGPS

Input file
☒ Merged metadata and abundance file
☐ Separate metadata and abundance file

[Welcome](#)
[HELP](#)
[Result Plot](#)
[Output](#)

Predicted original coordinates of samples can be downloaded. Prediction model in RDA format can be downloaded

[Download prediction data](#)
[Download optimal features in prediction model](#)
[Download feature subsets accuracy in feature elimination](#)
[DownloadModel](#)

a. Prediction data: Records the predicted original coordinates of the reference samples (columns: LatPred and LongPred). Also contains the distance between the predicted and true sampling site for the reference samples (column: Distance_from_origin). The original metadata and abundance data will be merged.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		uuid	continent	city	latitude	longitude	taxal_abundance	taxan_abundance	cityPred	latPred	longPred	Distance_from_origin	
2	1	1	africa	offa	8.1548483	4.7263633	0.00017	0	offa	7.881690979	11.71926498	771.4253212	
3	2	2	africa	offa	8.15485	4.726465	9.00E-05	0	offa	8.323236465	-0.264049619	550.1233564	
4	3	3	africa	offa	8.1556609	4.7244244	8.00E-05	0	offa	7.446023464	4.620732784	79.81995773	
5	4	4	america	new_york_city	40.70670433	-74.0111377	7.00E-05	2.00E-05	new_york_city	40.67137527	-72.10720062	160.7544161	
6	5	5	america	new_york_city	40.67953148	-73.9035273	0.00016	0	new_york_city	18.32634163	-43.18954086	3843.356172	

b. Optimal features in prediction model: Records the optimal features (microorganism) used to predict the original coordinates that selected by the mGPS algorithm. The column “Overall” records the importance value of each feature. The higher the value, the more important the feature in the prediction algorithm.

a. Prediction data: Records the predicted original coordinates of the new samples. (columns: LatPred and LongPred).

	A	B	C	D	E	F	G
1		taxa1_abundance	taxa2_abundance	taxan_abundance	cityPred	latPred	longPred
2	1	0.00017	0	0	offa	7.881690979	11.71926498
3	2	9.00E-05	0.00023	0	offa	8.323236465	-0.264049619
4	3	8.00E-05	0	0	offa	7.446023464	4.620732784
5	4	7.00E-05	2.00E-05	2.00E-05	new_york	40.67137527	-72.10720062
6	5	0.00016	1.00E+00	0	new_york	18.32634163	-43.18954086

b. Optimal features in prediction model: Same content as this file in function 1: Build a new prediction model using mGPS. [Jump](#)

c. feature subsets accuracy in feature elimination: Same content as this file in function 1: Build a new prediction model using mGPS. [Jump](#)

d. Model: Same content as this file in function 1: Build a new prediction model using mGPS. [Jump](#)

2.3 Function 3: Use an existing model to predict new samples

Geographical origin prediction of microbiome

Prediction program

Use existing model to predict new samples

Upload sample(s) abundance file

Browse... No file selected

Upload the prediction model (in .Rda format)

Browse... No file selected

Start

HELP Result Plot Output

Use existing model to predict new samples

Function description

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

Usage

In left side bar:

1. Select **Prediction program** as *Use existing model to predict new samples*
2. **Upload sample(s) abundance file:** Upload data file(s) (in .csv format) containing new microbial sample abundance data.
3. **Upload the prediction model:** Upload a constructed origin prediction model in .Rda format. 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*

In **Result Plot** tab:

4. **Change longitude/latitude range in output map** (Optional)
5. **Whether pull points to land/marine:** (Optional) If checked, predicted origin location will be pull to the nearest land/marine if predicted coordinates are out of the expected boarder.

2.3.1 Function 3 introduction

In this mode, you can predict new sample origin based on an existing prediction model. You can download the model in the Output tab of functions: Build a new prediction model using mGPS or Build a new prediction model using mGPS and predict new samples. [Jump to model](#)

2.3.2 Output plot

Geographical origin prediction of microbiome

Prediction program

Use existing model to predict new samples

Upload sample(s) abundance file

Browse... No file selected

HELP Result Plot Output

longitude range:

-165 168

latitude range:

-90 90

Whether pull points to land/marine

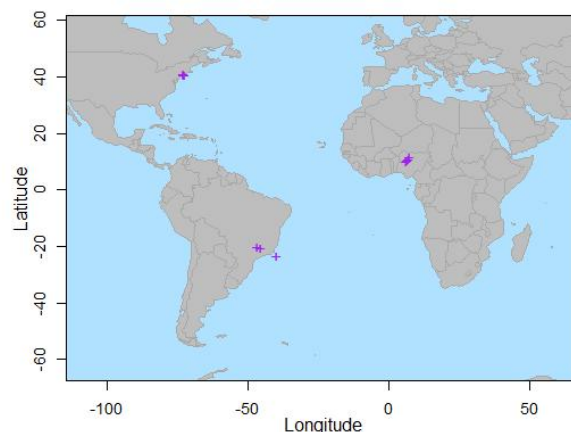
☐ Pull to land
☐ Pull to marine
☒ None

The model uploaded will be used to predict the origin of 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, you can select the plotting area of the figure on the map. In addition, you can choose whether to pull the predicted point into the continent or the ocean. You can download and save the png file of this map through the button 'DownloadMap' below the figure.



DownloadMap

2.3.3 Output files

Geographical origin prediction of microbiome

Prediction program

Use existing model to predict new samples

Upload sample(s) abundance file

Browse... No file selected

HELP Result Plot Output

Here you can download the predicted original coordinates of samples in your uploaded file. (Reference microbiome locations are based on MetaSub project microbiome database)

DownloadData

a. Prediction data: Records the predicted original coordinates of the new samples. (columns: LatPred and LongPred).

	A	B	C	D	E	F	G
1		taxa1_abundance	taxa2_abundance	taxan_abundance	cityPred	latPred	longPred
2	1	0.00017	0	0	offa	7.881690979	11.71926498
3	2	9.00E-05	0.00023	0	offa	8.323236465	-0.264049619
4	3	8.00E-05	0	0	offa	7.446023464	4.620732784
5	4	7.00E-05	2.00E-05	2.00E-05	new_york	40.67137527	-72.10720062
6	5	0.00016	1.00E+00	0	new_york	18.32634163	-43.18954086

3 Reference

3.1 If you meet error:

```
> runApp('mGPS_interface.r')
Warning in file(con, "w") :
cannot open file 'C:\Users\Temp\...': No such file or directory
Error in file(con, "w") : cannot open the connection
```

Try to restart the R (Rstudio). Sometimes this error is due to caching in Rstudio.

3.2 If you meet error:

```
Error in file(file, ifelse(append, "a", "w")) :
Cannot open the file:'.....': Permission denied
Warning: Error in file: cannot open the connection
```

Try to see if you have the file open in another program, so the file cannot be modified by the R application.