

# ImageGP 在微生物组可视化中的应用

## Visualization of microbiome data using ImageGP

陈同<sup>1, #, \*</sup>, 刘永鑫<sup>2, #, \*</sup>

<sup>1</sup>中国中医科学院, 中药资源中心, 北京; <sup>2</sup>中国科学院遗传与发育生物学研究所, 植物基因组学国家重点实验室, 北京;

\*通讯作者邮箱: [chent@nrc.ac.cn](mailto:chent@nrc.ac.cn), [yxliu@genetics.ac.cn](mailto:yxliu@genetics.ac.cn)

#共同第一作者/同等贡献

**摘要:** 高通量测序技术的发展极大加速了微生物组领域的研究, 微生物组分析也已应用于人、其它动物、植物、环境中微生物结构和微生物自身的研究, 也是当前火热的研究领域之一。大量数据的生成促成了很多分析工具和流程的开发, 经过多步生物信息分析获得很多结果表格。如何基于这些结果表进行进一步的可视化展示要求研究者需要熟悉数据格式的转换和至少一种绘图语言。ImageGP 工具基于常见的生信数据结果表格形式, 提供了定制化的参数, 可以快速、可定制的进行微生物数据的可视化, 如物种组成堆积柱状图、alpha 多样性箱线图、beta 多样性 PCoA 分析、LEfSe、PICRUSt、BugBase 和 Faprotax 分析等, 降低了微生物数据可视化的操作难度。

**关键词:** 微生物组, 可视化, 云平台

### 仪器设备

1. 个人电脑安装了上网浏览器如火狐、谷歌浏览器、Safari 等联网即可使用。

### 操作步骤

1. 在火狐或谷歌浏览器的地址栏输入网址 [http://www.ehbio.com/Cloud\\_Platform/front/](http://www.ehbio.com/Cloud_Platform/front/), 回车即可打开 ImageGP 绘图平台。首页可看到当前支持的图形类型和分析功能 (图 1 ImageGP 支持绘制的部分图形和分析功能)。



## 2. 微生物组样品内 Alpha 多样性结果的可视化与统计检验

2.1 Alpha 多样性是反应样品内物种的丰富度或均匀度的指标，可以比较样品组之间物种的构成是否存在差异。通常基于抽平后的 OTU 表计算，其指标包括 Richness、Chao1、Shannon\_e、Shannon\_2、Shannon\_10、Berger\_parker、Buzas\_gibson、Dominance、Equitability、Jost、Jost1、Reads、Robbins、Simpson 等。并可通过箱线图进行可视化和使用方差分析等检验 alpha 多样性指数的分布在组间是否有显著差异。常用的 alpha 多样性指数为以下几种：（1）Richness-Observed OTU 结果为整数，但只有物种种类信息，没有丰度信息，数值范围一般为几百至几千不等，范围很大，与研究对象有关；（2）Shannon index 数值为 1-10 左右的小数，数值越大，多样性越高，是综合物种数量和丰度两个层面的结果；（3）Chao1 是根据出现 1 次和 2 次的 OTU 来估算总体的多样性；详细计算方法见：Alpha diversity measures (<http://scikit-bio.org/docs/latest/generated/skbio.diversity.alpha.html>)。具体图像解读见扩增子图表解读 1 箱线图：Alpha 多样性 (<https://blog.csdn.net/woodcorpse/article/details/73810572>)。进一步了解 Alpha 多样性计算原理和不同多样性值代表的意义 ([http://www.drive5.com/usearch/manual/alpha\\_metrics.html](http://www.drive5.com/usearch/manual/alpha_metrics.html))<sup>1</sup>。

2.2 alpha 多样性计算后一般存储为表格格式，每个样本一行，每一列对应一种 Alpha 多样性指数，列数不限。每一列之间用 TAB 键分割 (图 2)。

Sample	berger_parker	buzas_gibson	chao1	dominance	equitability	jost	jost1	reads	richness	robbins	simpson	shannon_e	shannon_2	shannon_10
KO1	0.0572	0.0151	2383.3	0.99	0.787	191.6	456	30257	2382	0.259	0.00961	6.12	8.83	2.66
KO2	0.0523	0.0159	2394.3	0.992	0.794	220.1	480.4	30235	2393	0.251	0.00789	6.17	8.91	2.68
KO3	0.0599	0.0113	2050.5	0.99	0.764	157.9	339.8	30149	2050	0.255	0.0104	5.83	8.41	2.53
KO4	0.0576	0.0113	2101.4	0.99	0.762	158.4	341.1	30143	2101	0.261	0.0104	5.83	8.41	2.53
KO5	0.0738	0.0087	1893.6	0.986	0.738	116.7	262.5	30170	1893	0.291	0.0143	5.57	8.04	2.42
KO6	0.0486	0.0138	2257.4	0.992	0.781	199.9	415.7	30069	2257	0.256	0.00815	6.03	8.7	2.62
OE1	0.0366	0.0215	2628.8	0.994	0.823	301.2	652	30286	2628	0.197	0.00588	6.48	9.35	2.81
OE2	0.0454	0.0183	2554.7	0.993	0.805	243.5	554.2	30257	2554	0.208	0.0073	6.32	9.11	2.74
OE3	0.0426	0.0143	2459.8	0.991	0.777	183.1	432.1	30177	2459	0.228	0.00933	6.07	8.76	2.64
OE4	0.0379	0.0163	2447	0.993	0.795	231.3	492.4	30300	2446	0.249	0.00711	6.2	8.94	2.69
OE5	0.0333	0.0173	2516	0.993	0.799	247.4	523.1	30306	2515	0.242	0.00662	6.26	9.03	2.72
OE6	0.0428	0.0205	2608.9	0.994	0.818	283.5	620.9	30281	2608	0.202	0.00624	6.43	9.28	2.79
WT1	0.0583	0.012	2405.3	0.989	0.757	151.3	361.3	30231	2405	0.23	0.0112	5.89	8.5	2.56
WT2	0.0622	0.00985	2195.3	0.986	0.74	123.9	297.8	30251	2195	0.246	0.0136	5.7	8.22	2.47
WT3	0.0446	0.0171	2568.3	0.993	0.795	234.6	514.5	30107	2568	0.221	0.00726	6.24	9.01	2.71
WT4	0.052	0.0146	2427.4	0.991	0.781	198.5	440.4	30088	2427	0.247	0.00863	6.09	8.78	2.64
WT5	0.04	0.0108	2250.4	0.989	0.75	144.3	327.4	30178	2250	0.244	0.0111	5.79	8.35	2.52
WT6	0.0407	0.0161	2511.4	0.993	0.79	225.1	486.1	30115	2511	0.234	0.00735	6.19	8.93	2.69

图 2 Alpha 多样性数据表格展示

2.3 通常在可视化或比较样本组之间的 Alpha 多样性时，还需要提供 **metadata** 文件，指定样本的分组信息 (表 1)。这个文件可以与上面的 Alpha 多样性表格合并为一个文件（手动合并或使用 ImageGP 平台的 Merge matrixes 功能），或者作为单独的文件上传到 ImageGP 平台用于图形绘制。

SampleID	Group
KO1	KO
KO2	KO
KO3	KO
KO4	KO
KO5	KO
KO6	KO
OE1	OE
OE2	OE
OE3	OE
OE4	OE
OE5	OE
OE6	OE
WT1	WT
WT2	WT
WT3	WT
WT4	WT
WT5	WT
WT6	WT

**表 1 Metadata 示例表格**

第一列为样本名字，与 Alpha 多样性表格第一列一致；第二列为样本分组信息。可以有更多列信息。

2.4 数据准备好之后，可以直接粘贴到“Paste main data to text area”和“Input metadata data”处，然后点击“Check Data”按钮，效验数据的格式是否满足输入要求 (图 3)。

图 3 输入数据界面展示。右上角有 5 个 demo，可用于查看示例操作。修改过的参数会黄色高亮显示。

2.5 数据格式效验成功后，下面的参数转为可选或可填状态。只有 2 个必选参数，指定“X-axis variable”为“Group”，指定“Y-axis variable”为“Chao1”，点击“Submit”即可绘制“Chao1”水平的 Alpha 多样性图谱分布 (图 4 和 图 5)。

图 4 Alpha 多样性绘图参数选择。根据用户选择的参数情况严格控制参数的可选状态，减少用户的选择茫然症

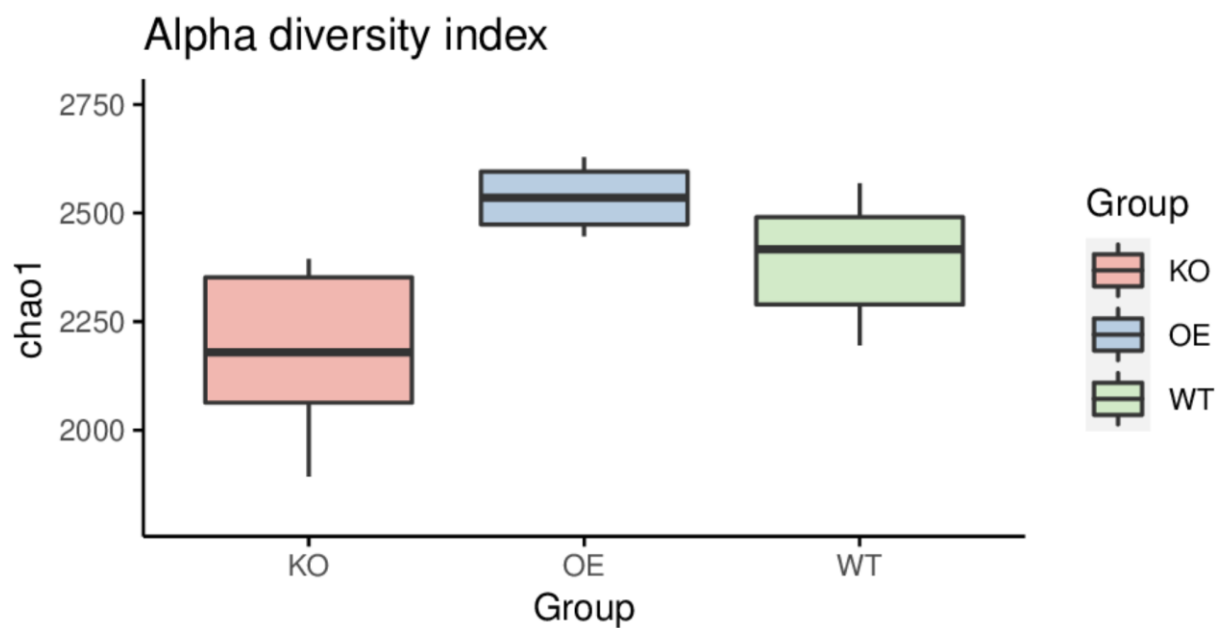


图 5 Alpha 最简多样性分布箱线图

2.6 设置“X-axis variable order”多选下拉的选择项为[“WT”, “OE”, “KO”] 调整各个箱体的顺序 (图 6 和图 7)。

Essential parameters

\* X-axis variable

Group

\* Y-axis variable

chao1

Legend variable

E.g. Group

X-axis variable order

WT OE KO

Legend variable order

图 6 设置箱体的出现顺序

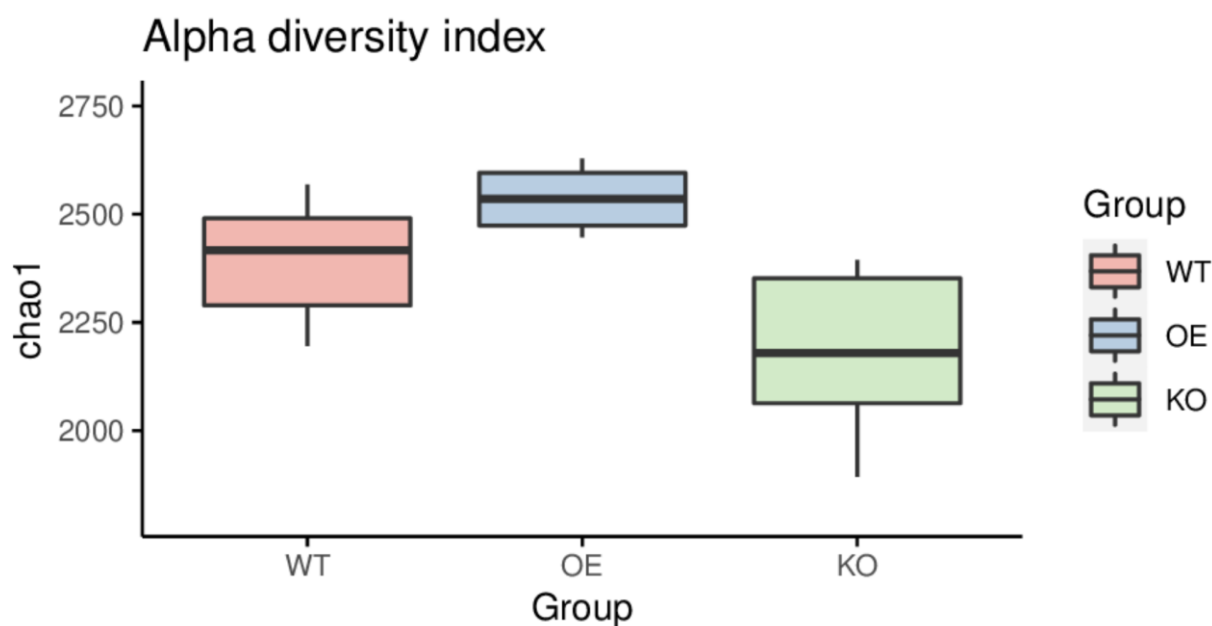


图 7 调整顺序后的结果图

2.7 设置“Plot type”为“Box+Violin+Jitter”，更详细的展示各个样品具体的 Alpha 多样性值和整体 Alpha 多样性值的密度分布 (图 8 和图 9)。

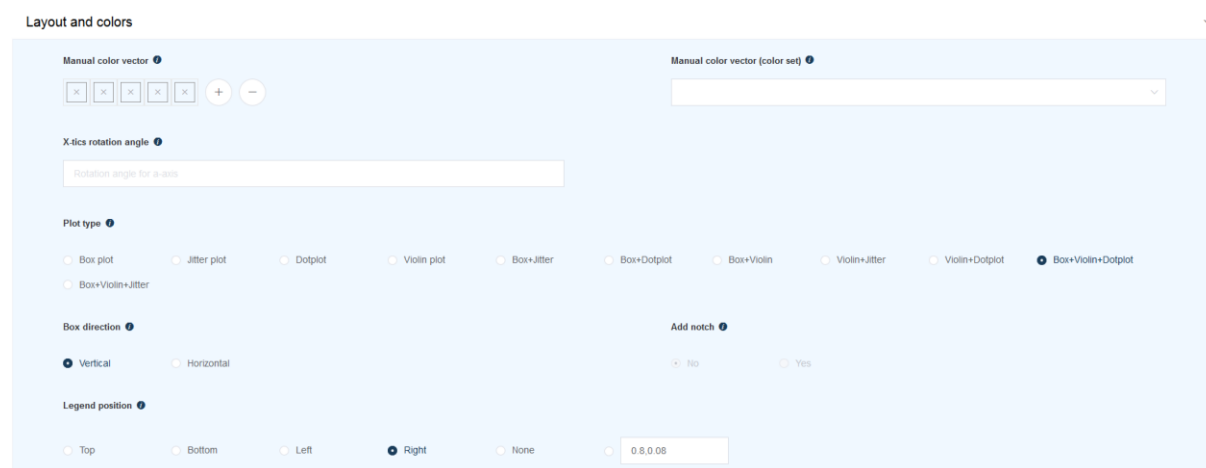


图 8 修改图形绘制类型

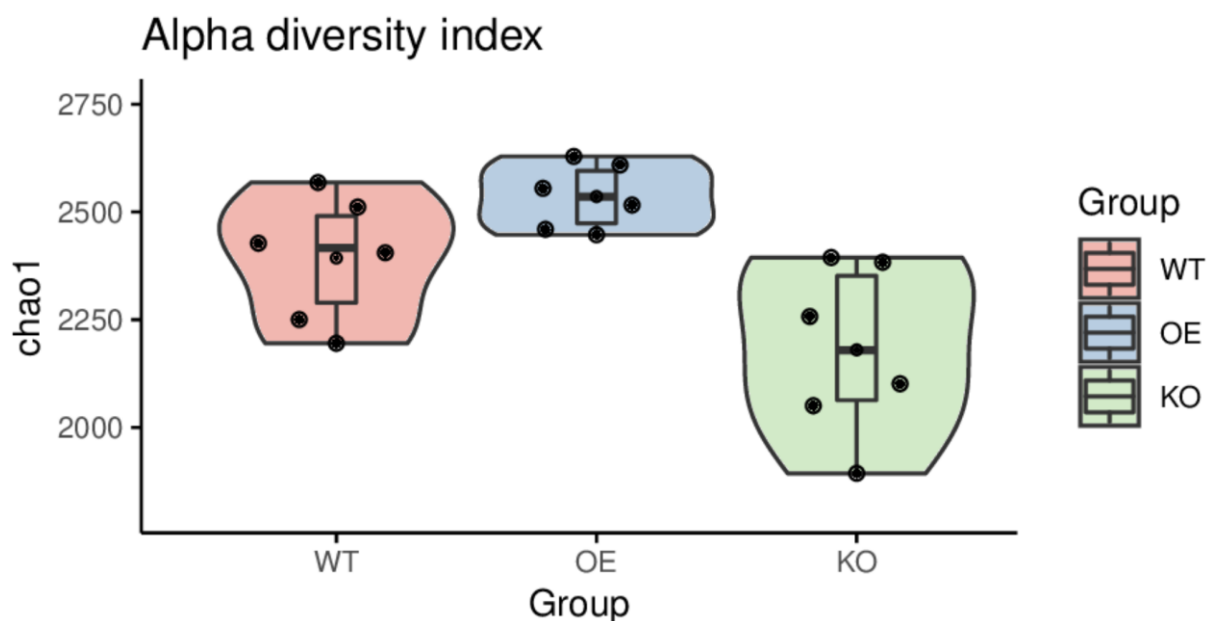


图 9 同时绘制箱线图、小提琴图和样本点

2.8 设置“Do statistics significance test for each group”选项为“yes”，若数据满足正态分布和方差齐性，则采用 ANOVA 方差分析检验不同样品组之间 Alpha 多样性水平是否存在显著差异，并进行标记。若数据不满足正态分布或方差齐性，则采用 Kruskal-Wallis Test（功能完善中）。图中可以看出，OE 组与 KO 组箱体上面标记的字母不同，表示其在 chao1 水平上存在显著的多样性差异，而 OE 与 WT，KO 与 WT 之间的差异则不具有统计显著性（图 10）。



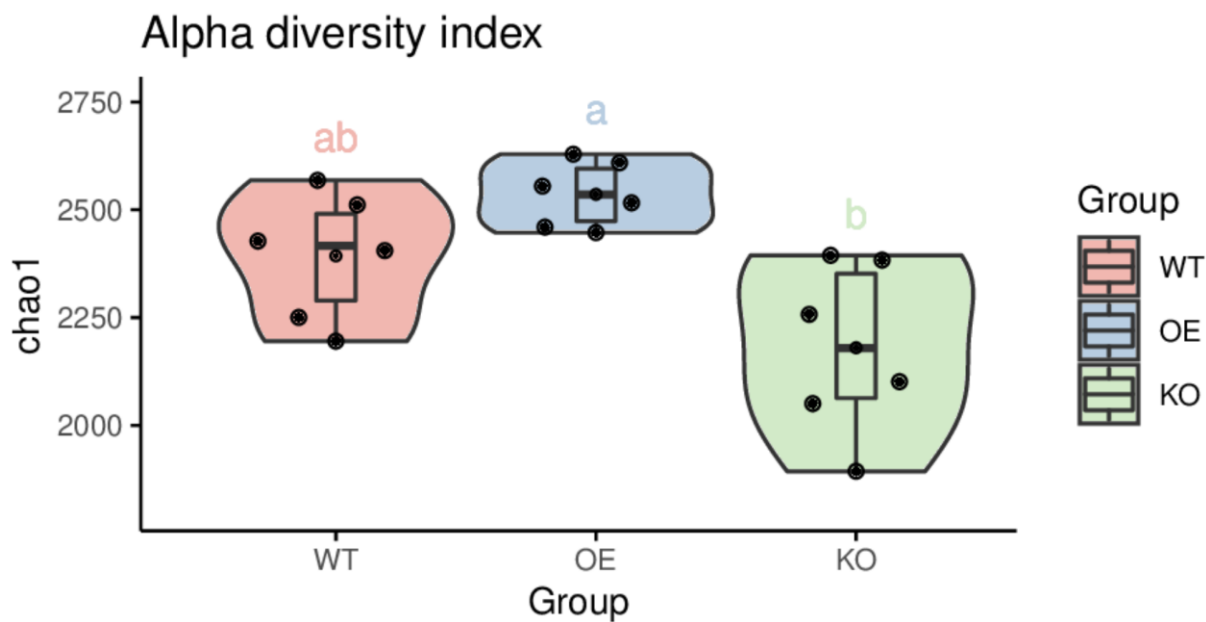


图 10 不同样本组 Alpha 多样性水平统计计算和标记

2.9 后续还可以继续调整图形的配色、长宽、方向等更多常用美化方式 (图 11)。每个参数旁都有一个感叹号图标，鼠标移上去之后会显示对应参数的描述。非单选类参数都有占位文本，以浅色文字形式呈现，描述了该参数的用途和注意事项。

Essential parameters
Data preprocess
Layout and colors
Modify plot titles
Picture attributes

Submit

图 11 箱线图绘制参数概览

2.10 图形绘制完成后可以下载 png 格式 (鼠标右键另存) 和 PDF 格式, PDF 格式可导入 Adobe illustrator 等工具进行进一步的编辑或与其它图组合在一起 (图 12)。

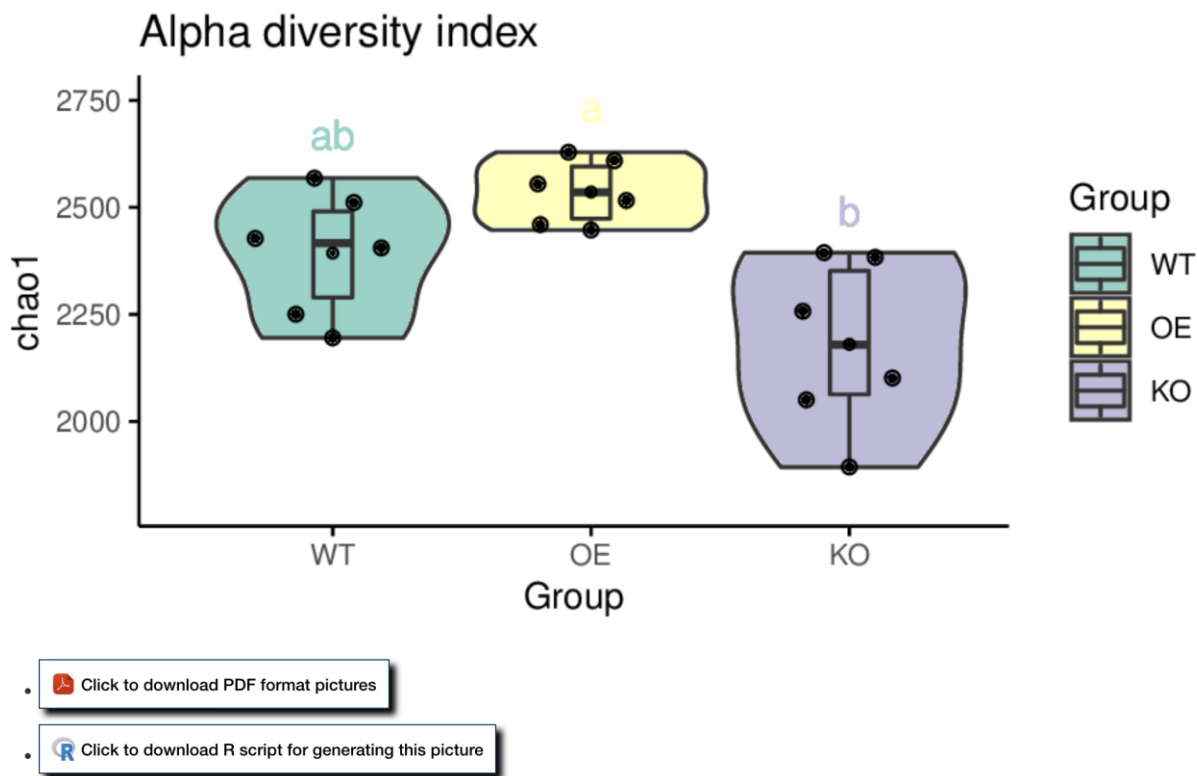


图 12 结果图和代码的下载和导出

2. 11 同时可以下载绘制此图的 R 代码 ([data](#), [metadata](#), [outputprefix](#) 按自己实际文件路径修改), 用于后续继续美化, 或发文时提交代码, 实现图形的可重复绘制<sup>2,3</sup> ([附件 1.zip](#))。

## 更多应用

1. ImageGP 目前支持 Boxplot、Flower plot、Line plot、Volcano plot、Heatmap、VennDiagram、Enrichment plot、Euler plot、Scatterplot、UpsetView plot、Bar plot、Histogram plot、PCA、PCoA、cPCoA 等图形绘制。其它功能如 WGCNA 分析、差异菌群分析、菌群相关性网络的绘制、LEfSe, PICRUST, FAPROTAX、BugBase 等正在逐步更新中 (表 2)。

支持的可视化类型	应用场景	相关软件
Boxplot	Alpha 多样性展示, OTU 丰度展示, Beta 多样性距离展示	R
Flower plot	共有个特有的高丰度 OTU、物种等的展示	R
Line plot	稀释曲线, OTU 丰度变化曲线	R

Volcano plot	差异 OTU、差异物种展示	R
Heatmap	OTU 丰度信息展示	R
VennDiagram	共有个特有的高丰度 OTU、物种等的展示	R
Enrichment plot	宏基因组中基因的功能富集分析展示	R
Euler plot	共有个特有的高丰度 OTU、物种等的展示	R
Scatterplot	通用型展示如 OTU 在两个样品中的丰度比较等	R
UpsetView plot	共有个特有的高丰度 OTU、物种等的展示	R
Bar plot	物种组成展示	R
Histogram plot	OTU 丰度分布展示	R
PCA	样品聚类	R
PCoA	基于 Beta 多样性距离的样品聚类展示	R
cPCoA	基于 Beta 多样性距离的样品聚类展示	R
LEfSe	差异 OTU 分析	LEfSe
PICRUSt	基于 16S 扩增子数据预测潜在功能通路	PICRUSt
FAPROTAX	基于 16S 扩增子预测群落功能信息和参与的元素循环通路	FAPROTAX
BugBase	预测微生物组样品的表型信息如革兰氏阳性阴性、病原性等	BugBase

表 2 ImageGP 当前支持的数据可视化类型和描述。

2. 每个工具都提供了 Demo 示例，并用轮播图展示了输入数据、输入参数和所能获得的图的样式 (图 13)。



图 13 图形输入数据、参数示例和 Demo 按钮

3. ImageGP 使用基础视频已上传至 <https://www.bilibili.com/video/BV1Zh411974X>；进阶视频已上传至 <https://www.bilibili.com/video/BV17D4y1o7y4>。更多使用视频录制后会持续上传。

### 失败经验

ImageGP 绘图时遇到的问题多为输入数据格式问题。在数据输入或选择后, 点击`Check Data`按钮, 会进行数据格式效验, 提示数据中存在的问题。

1. 绘图数据与 **metadata** 数据不匹配。通常是绘图数据中的样本未包含在 **metadata** 数据中时会出现如下错误提示:

Error: The first column of *Paste main data to textarea* is not equal to the first column of *Input metadata data*.

错误: *Paste main data to textarea* 的第一列不等于 *Input metadata data* 的第一列。

2. 宽矩阵格式 (常见的表达矩阵、**OTU** 丰度矩阵等) 第一列通常会作为行名字, 是不允许有重复的。如果不符合, 会弹出如下错误提示

Error: No duplicate names are allowed in the *first* column of the data in *Paste main data to text area* when *Matrix format* is *Wide*.

错误: 当 *Matrix format* 是 *Wide* 时, *Paste main data to text area* 数据第一列不允许有重复值。

3. 长宽矩阵数据格式混用。用户初始不了解长宽矩阵, 可能出现混用。或者数据在 **Excel** 等操作时引入了非数字时会出现此错误。若存在此问题, 会弹出如下错误提示

Error: All value in *Paste main data to text area* should be *numbers* except for the first row and the first column when *Matrix format* is *Wide*. *Column Group* contains non-numerical values.

错误: 类型错误, 当 *Matrix format* 是 *Wide* 时, *Paste main data to text area* 中的数据除第一行第一列外都是数值类型, 列 *Group* 包含非数字信息。

## 致谢

本项目由中国中医科学院基本科研业务费优秀青年科技人才 (创新类) 培养专项 (编号: ZZ13-YQ-095 和 ZZXT201708) 支持。感谢易汉博基因科技 (北京) 有限公司的同仁对网站建设的技术支持。感谢 20 万网友的使用和持续反馈。目前 ImageGP 已经有 119 次正式引用。

Google Scholar

www.ehbio.com/ImageGP

文章

找到约 119 条结果 (用时0.04秒)

时间不限

2020以来

2019以来

2016以来

自定义范围...

按相关性排序

按日期排序

不限语言

中文网页

简体中文网页

☒ 包括专利

☒ 包含引用

☒ 创建快讯

[HTML] Survival associated alternative splicing events in diffuse large B-cell lymphoma

R Zhang, P Lin, X Yang, RQ He, HY Wu... - American journal of ..., 2018 - ncbi.nlm.nih.gov

... The percent spliced in (PSI) value was applied to quantify various AS events from zero to one. UpSet view and network construction. To analyze and visualize the intersected AS events, an UpSet plot was created using ImageGP (<http://www.ehbio.com/ImageGP/>) ...

☆ 99 被引用次数 : 10 相关文章 所有 5 个版本

[HTML] nih.gov

OmicsOne: Associate Omics Data with Phenotypes in One-Click

Y Hu, M Ao, H Zhang - bioRxiv, 2019 - biorxiv.org

... this era of omics data explosion, a simple, efficient bioinformatic analysis tool is critical to work with the large amount of data. Individual tools, such as GiaPronto7, PANDA-view8, ImageGP (<http://www.ehbio.com/ImageGP/>), offer select options for data analysis and visualization ...

☆ 99 被引用次数 : 1 相关文章 所有 3 个版本

[PDF] biorxiv.org

[HTML] Alterations in immune-related genes as potential marker of prognosis in breast cancer

B Li, R Geng, Q Wu, Q Yang, S Sun, S Zhu... - Frontiers in ..., 2020 - ncbi.nlm.nih.gov

... Additionally, we used ImageGP (<http://www.ehbio.com/ImageGP/index.php/Home/Index/index.html>) to construct a visual volcano plot to display the results of the dataset analysis and Morpheus online analysis software (<https://software.broadinstitute.org/morpheus/>) to construct a ...

☆ 99 被引用次数 : 3 相关文章 所有 3 个版本

[HTML] nih.gov

[HTML] Identification of differentially expressed key genes between glioblastoma and low-grade glioma by bioinformatics analysis

Y Xu, R Geng, Q Sun, B Liu, Q Chen - PeerJ, 2019 - peerj.com

... In addition, we used visual hierarchical cluster analysis to show the two groups by Morpheus online analysis software (<https://software.broadinstitute.org/morpheus/>) and volcano plot of two groups by ImageGP (<http://www.ehbio.com/ImageGP/index.php/Home/Index/index.html>) ...

☆ 99 被引用次数 : 6 相关文章 所有 8 个版本

[HTML] peerj.com

[HTML] Silencing TRIP13 inhibits cell growth and metastasis of hepatocellular carcinoma by activating of TGF-β1/smad3

J Yao, X Zhang, J Li, D Zhao, B Gao, H Zhou... - Cancer cell ..., 2018 - Springer

... The visual hierarchical cluster analysis was performed using volcano plot and heat map in ImageGP (<http://www.ehbio.com/ImageGP/index.php/Home/Index/index.html>) software. Gene set enrichment analysis (GSEA). According ...

☆ 99 被引用次数 : 16 相关文章 所有 8 个版本

[HTML] springer.com

[PDF] The identification of a common different gene expression signature in patients with colorectal cancer

ZW Zhao, XX Fan, LL Yang, JJ Song, SJ Fang... - Math Biosci ..., 2019 - aimspress.com

... criterion. Finally, we used the ImageGP (<http://www.ehbio.com/ImageGP/index.php/Home/Index/index.html>) to visualize the major biological processes (BP), molecular function (MF) and cell composition (CC) of these DEGs. 2.4 ...

☆ 99 被引用次数 : 9 相关文章 所有 4 个版本

[PDF] aimspress.com

## 参考文献

- Liu, Y. X., Qin y., Chen, T., Meiping Lu, M. P., Xubo Qian, X. B., Guo, X. X., Bai, Y. [A practical guide to amplicon and metagenomic analysis of microbiome data.](#) Protein Cell (2020) doi:10.1007/s13238-020-00724-8.
- Hadley Wickham. [ggplot2: Elegant Graphics for Data Analysis.](#) (Springer-Verlag New York, 2016).
- R Core Team. R: [A Language and Environment for Statistical Computing.](#) (2020).