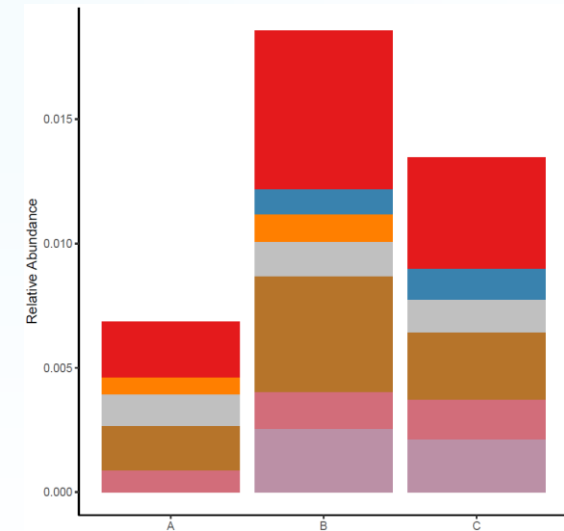




## 31表型预测BUGBASE

易生信  
2022年1月9日

易生信，毕生缘；培训版权所有。



# BugBase —— 基于16S OTU表预测细菌表型数据库

- 官网: <https://bugbase.cs.umn.edu/>
- 此文于17年5月2日发布在预印本杂志bioRxiv上面, 还没有正式发表。目前已被引用132次。

## BugBase predicts organism-level microbiome phenotypes

[T Ward, J Larson, J Meulemans, B Hillmann, J Lynch...](#) - BioRxiv, 2017 - biorxiv.org

Shotgun metagenomics and marker gene amplicon sequencing can be used to directly measure or predict the functional repertoire of the microbiota en masse, but current methods do not readily estimate the functional capability of individual microorganisms. Here we present BugBase, an algorithm that predicts organism-level coverage of functional pathways as well as biologically interpretable phenotypes such as oxygen tolerance, Gram staining and pathogenic potential, within complex microbiomes using either whole-genome shotgun ...

☆ Save Cite Cited by 132 Related articles All 4 versions

- 中文教程详见: [16S预测细菌表型-bugbase: 革兰氏阴阳、生物膜、致病力、移动元件、氧气消耗等](#)



- BugBase是一款分析微生物组样品表型的工具，此网站可以基于OTU表和Mapping files，预测大量信息和比较，包括以下七方面：

革兰氏阳性 Gram Positive

革兰氏阴性 Gram Negative

生物膜形成 Biofilm Forming

致病潜力 Pathogenic Potential

移动元件含量 Mobile Element Containing

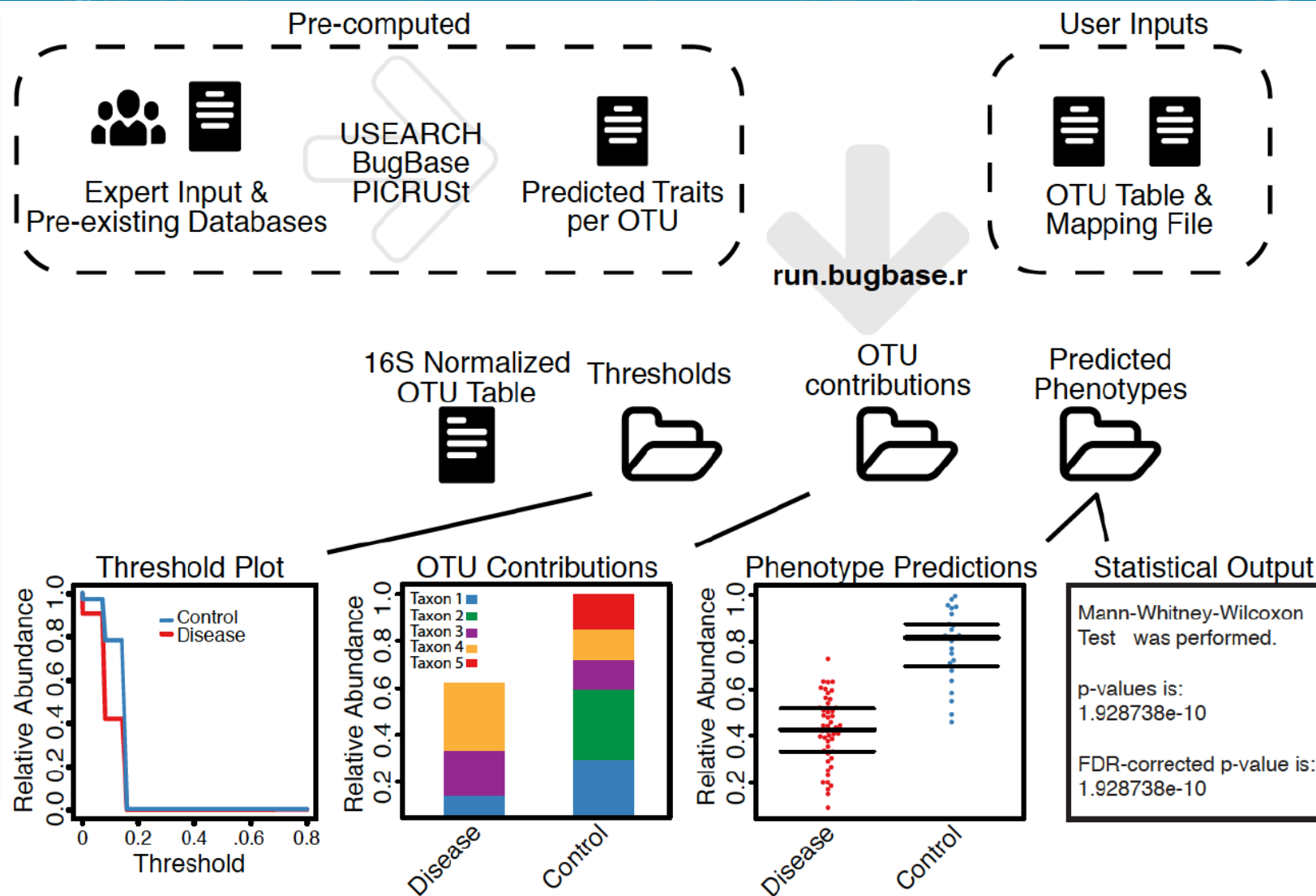
氧的利用 Oxygen Utilizing

氧化胁迫耐受 Oxidative Stress Tolerant

易生信  
生信宝典  
宏基因组



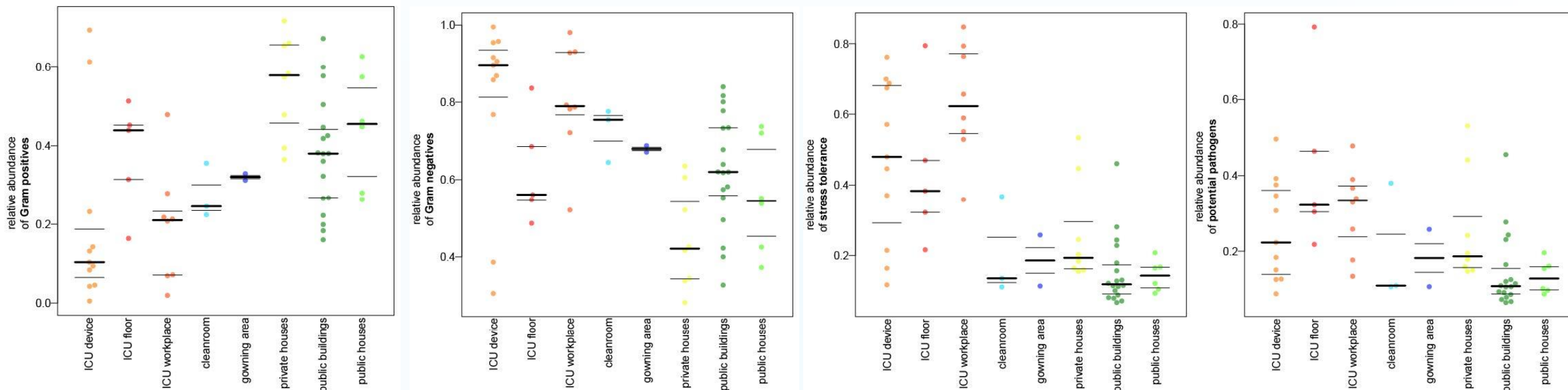
# BugBase工作流程图



宏基因组



# 使用实例——自然子刊: 洁净环境中耐药性比例增加



使用BugBase基于16S预测样品中革兰氏阳/阴性、胁迫和致病力菌的相对丰度。Phenotype prediction of Gram positive/negative、Potential stress、Potential pathogens bacteria based on 16S rRNA gene amplicon analysis.

毒力、疾病 ( $P = 0.008$ )、防御 ( $P = 5.2 \times 10^{-5}$ ) 和抵抗 ( $P = 0.08$ ) 有关的革兰氏阴性细菌在CB中富集 (P值由Kruskal–Wallis检验计算; 附图8—11)

# 输入文件要求

## OTU表

BIOM 1.0格式；16S以GreenGenes 13.5 为参数数据库(Reference)；宏基因组以IMG为参考；在线提交要求小于15 MB (本地版无限制)

## Mapping File

制表符分隔；第一行必须以#SampleID起始；第一行全为列标题；第一列必须为SampleID；只允许使用字母、数字、下划线和连字符；不允许包含空格，逗号、引号、括号；不要包含机密信息

SampleID	Group	Site	Date	BarcodeSequence	LinkerPrimerSequence	ReversePrimer
KO1	KO	Beijing	2017/6/30	ACGCTCGACA	AACMGGATTAGATACCCCKG	ACGTCATCCCCACCTTCC
KO2	KO	Beijing	2017/6/30	ATCAGACACG	AACMGGATTAGATACCCCKG	ACGTCATCCCCACCTTCC
KO3	KO	Sanya	2017/7/2	ATATCGCGAG	AACMGGATTAGATACCCCKG	ACGTCATCCCCACCTTCC
KO4	KO	Sanya	2017/7/2	CACGAGACAG	AACMGGATTAGATACCCCKG	ACGTCATCCCCACCTTCC
KO5	KO	Harbin	2017/7/4	CTCGCGTGTC	AACMGGATTAGATACCCCKG	ACGTCATCCCCACCTTCC



# 方法1. Windows本地运行

- 代码和包进行重新基于3.6定制
- 在22Pipeline目录下，指定软件目录为bugbase变量，输入gg OTU表、元数据、指定分组列名和输出目录

```
bugbase=C:/public/script/BugBase
```

```
Rscript ${bugbase}/bin/run.bugbase.r -L ${bugbase} \  
-i result/gg/otutab.txt -m result/metadata.txt -c Group -o result/bugbase/
```

```
[1] "Loading Inputs..."
```

```
[1] "16S copy number normalizing OTU table..."
```

```
[1] "Predicting phenotypes..."
```

```
[1] "2035 OTUs from the input table matched the 203452 available database OTUs"
```

```
[1] "Plotting thresholds..."
```



```
[1] "Plotting predictions..."
```


```
[1] "Plotting OTU contributions..."
```


```
[1] "BugBase analysis complete"
```





# 方法2. ImageGP在线——BugBase一键分析


 


 Home


 Line plot


 GO Enrichment plot


 Pretty Heatmap

 Boxplot

 Scatter plot

 Bar plot

 Volcano plot

 Manhattan plot



Make plot be easy and exquisite

woodcorpse@163.com ▾

## BugBase



如果操作中遇到问题或想添加新功能，请扫码拉您入群

BugBase is a tool for measuring high-level phenotypes in your microbiome. Please refer to [BugBase](#) for input formats, results explanation and citation information.

GreenGenes based OTU abundance table for BugBase analysis

OTUID	KO1	KO2	KO3	KO4	KO5	KO6	OE1	OE2	OE3	OE4	OE5	OE6	WT1	WT2	WT3	WT4
57759	214	132	1225	366	1297	291	130	1206	833	503	192	219	798	918	546	214
579	856															
810167	1361	394	559	756	1310	515	520	586	461	616	670	432	1116	1463	555	593
1116	915															
1134692	968	764	981	1056	1732	1160	299	319	445	469	452	413	650	710	537	678
599	594															
546343	11	4	8	0	3	22	9	5	4	7	1	8	3	2	12	2
2	6															
48487	390	431	234	262	32	926	290	238	245	442	195	222	333	133	297	264
140	202															





# 方法3. Linux系统中安装和使用(选学)

- Linux本地分析1. 软件安装
- # 方法1. git下载，需要有git  
git clone <https://github.com/knights-lab/BugBase>
- # 安装依赖包  
cd BugBase  
export BUGBASE\_PATH=`pwd`  
export PATH=\$PATH: `pwd` /bin
- # 测试数据  
run.bugbase.r -i doc/data/HMP\_s15.txt -m doc/data/HMP\_map.txt -c  
HMPBODYSUBSITE -o output



# Linux本地分析2. 准备输入文件

```
cd ~/amplicon/result
```

- # 输入文件：基于greengene OTU表的biom格式(本地分析支持txt格式无需转换)和mapping file(design.txt首行添加#)
- # 上传实验设计+刚才生成的otutab\_gg.txt
- # 生成在线分析使用的biom1.0格式  

```
biom convert -i gg/otutab.txt -o otutab_gg.biom --table-type="OTU table" --to-json
```

```
sed '1 s/^/#/' metadata.txt > MappingFile.txt
```
- # 下载otutab\_gg.biom 和 MappingFile.txt用于在线分析



# Linux本地分析3. 设定参数和运行

- # 设计BUGBASE\_PATH变量为程序要求  
`export BUGBASE_PATH=`pwd``
- # 添加r脚本到环境变量  
`export PATH=$PATH:`pwd`/bin`
- # 基于OTU表和实验设计预测和组间比较  
`run.bugbase.r -i gg/otutab.txt -m MappingFile.txt -c Group -o phenotype/`
- # 结果位于phenotype目录

易生信 生信宝典 宏基因组



# 方法4. Bugbase官网在线分析

## Welcome to BugBase.



### About Bugbase

BugBase is a tool for measuring high-level phenotypes in your microbiome.



### Parse Data

Upload your microbiome data and run it through our pipeline.



### Downloads

Download BugBase and see sample analyses.

Visit the [documentation](https://bugbase.cs.umn.edu/) to get started!

<https://bugbase.cs.umn.edu/>



# 在线分析

- <https://bugbase.cs.umn.edu/>
- 选择Parse Data
- 选择本地biom OTU表和MappingFile
- Column Header填写group
- 点击最下方Parse Data运行
- 目前显示网络中断
- 但测试数据可成功

⚠ Error connecting to server -- please check your network connection. If this error persists, please contact us for assistance.

OTU File Upload ?

otutab\_gg.biom

📁 Browse

Whole genome sequencing (shotgun)? ? ☐

Output prediction table only? (no plots) ? ☐

Use KEGG modules? ? ☐

CLR transform data? ? ☐

Specific Traits or Modules (required for KEGG) ?

Module list is required for KEGG. Leave blank for default BugBase traits.

Use coefficient of variance? ? ☐

Threshold (optional) ?

Value 0 to 1. Leave blank for automatic.

Proceed without mapping file? ? ☐

Mapping File Upload ?

MappingFile.txt

📁 Browse

Column Header ?

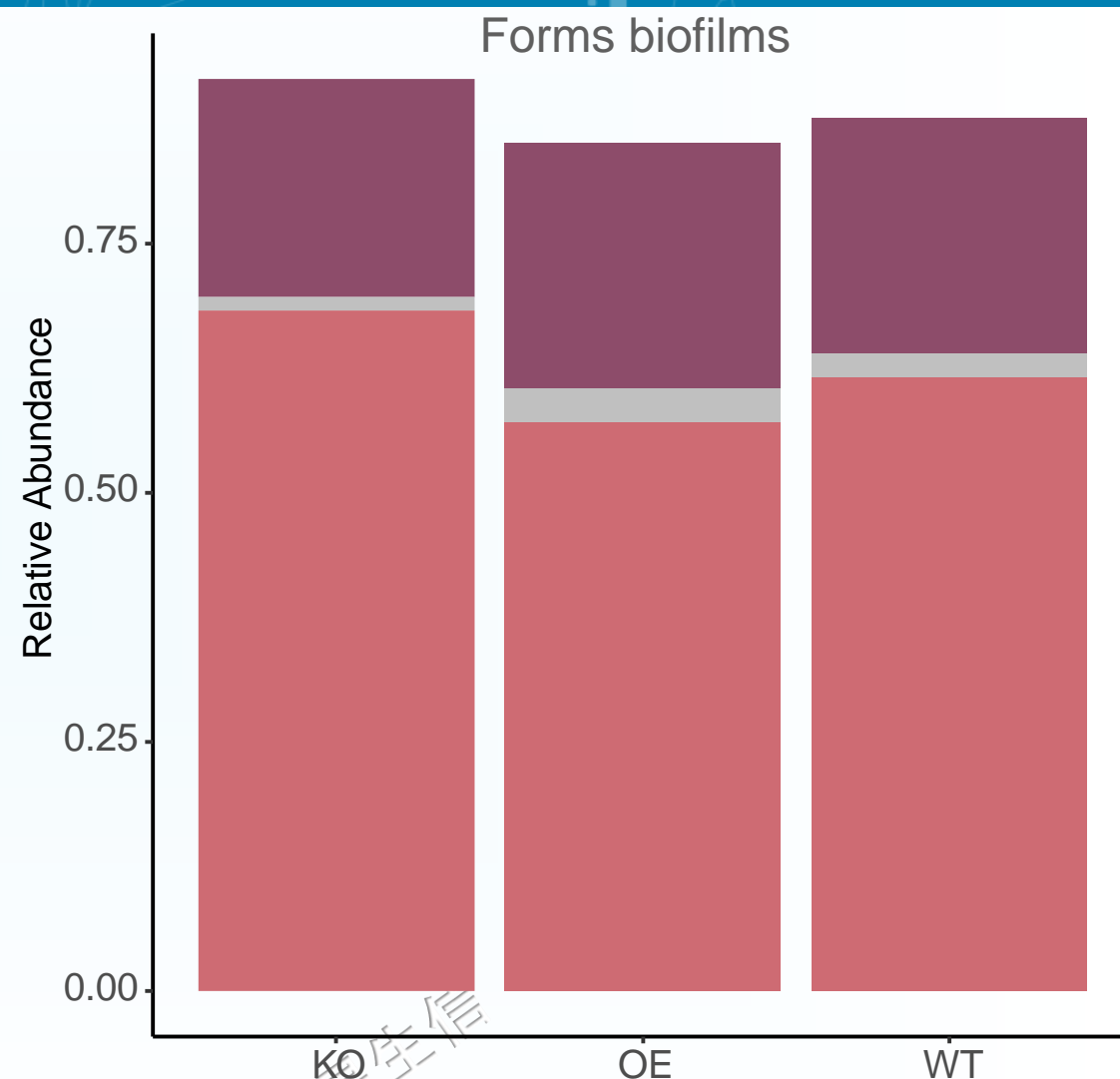
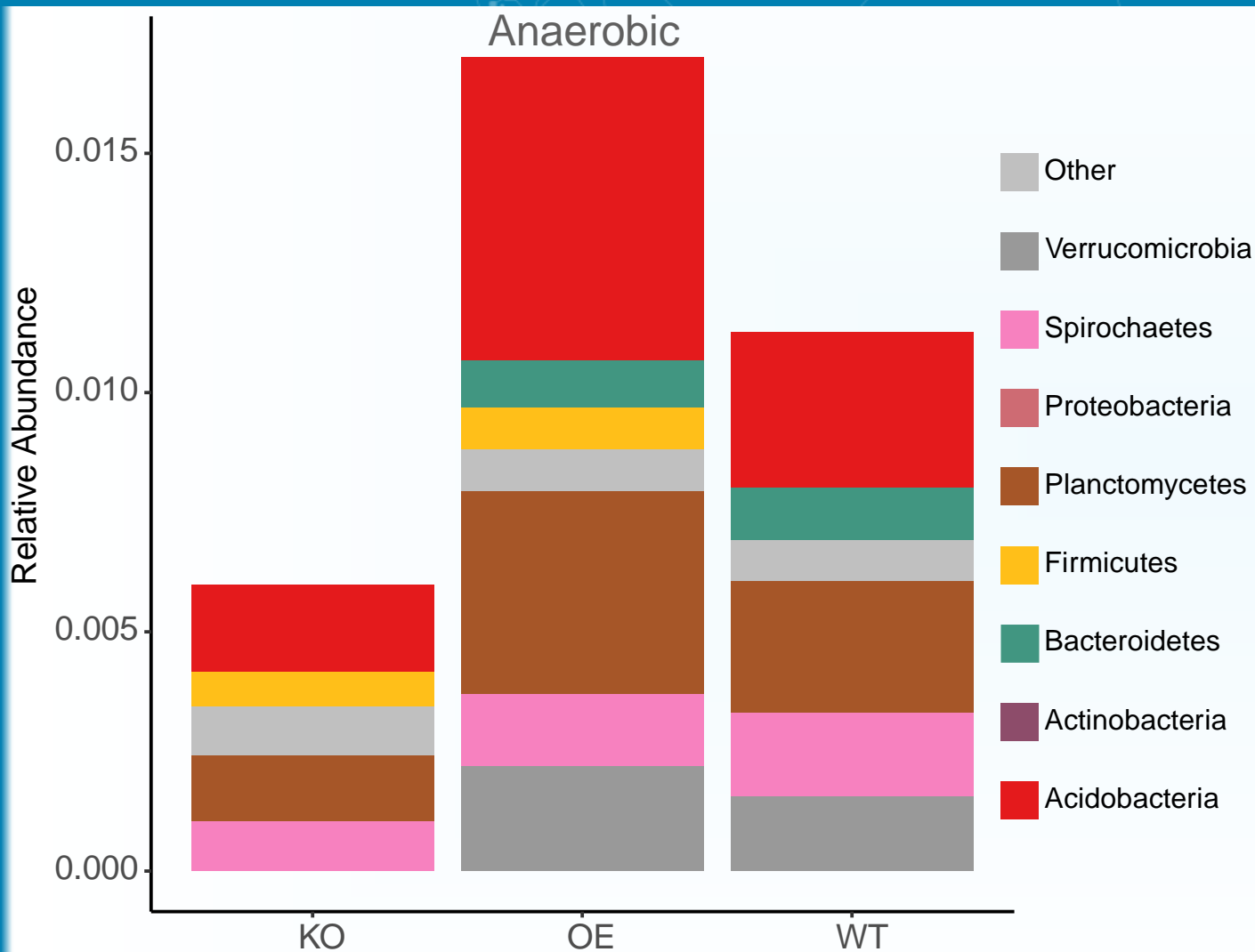
group



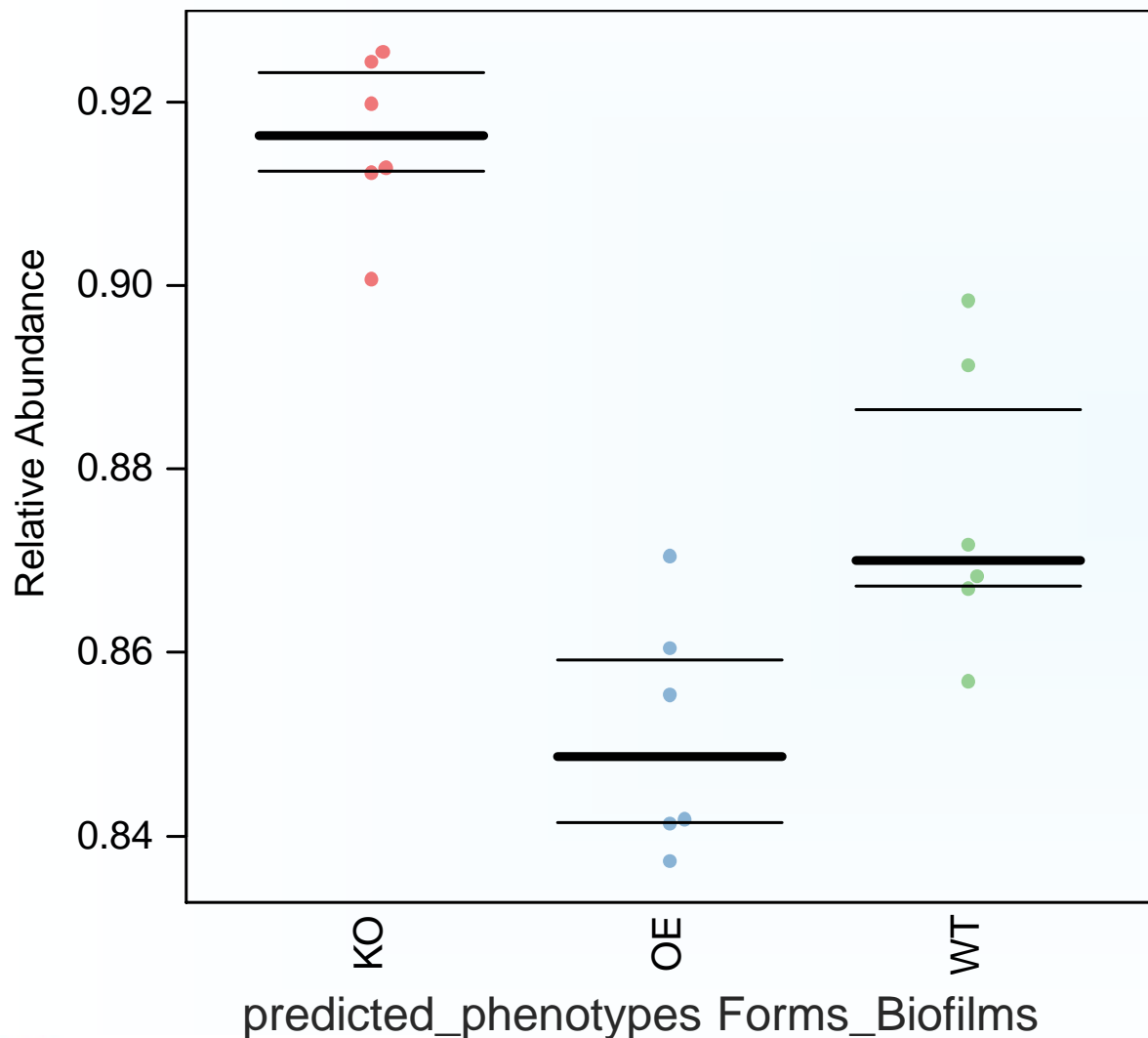


- **normalized\_otus** # 按16S拷贝数标准化的OTU表  
16s\_normalized\_otus.txt
- **otu\_contributions** # 功能由哪些OTU贡献  
Aerobic/Anaerobic/./Stress\_Tolerant.pdf # 每组功能门水平组成堆叠图和图例  
contributing\_otus.txt # 每个OTU对应每类功能的有无矩阵
- **predicted\_phenotypes** # 表型的相对丰度  
Aerobic/Anaerobic/./Stress\_Tolerant.pdf # 每组功能的箱线图  
Aerobic/Anaerobic/./Stress\_Tolerant\_stat.txt # 各组均值和组间比较P值
- **thresholds** # 不同阈值下可注释的丰度  
Aerobic/Anaerobic/./Stress\_Tolerant.pdf # 不同阈值的分类比例与相对丰度折线图

# 功能由哪些OTU贡献



# 表型的相对丰度和统计



Forms\_Biofilms

Number of samples in each treatment group:

KO OE WT

6 6 6

Proportion with phenotype (mean):

KO OE WT

0.9158879 0.8511340 0.8755880

Proportion with phenotype (median):

KO OE WT

0.9162921 0.8486079 0.8700309

Standard deviation:

KO OE WT

0.009292643 0.013062136 0.015858715

Pairwise Mann-Whitney-Wilcoxon Tests were performed.

Pairwise p-values are:

KO\_vs\_OE KO\_vs\_WT OE\_vs\_WT

**0.0021645 0.0021645 0.0259740**

FDR-corrected pairwise p-values are:

KO\_vs\_OE KO\_vs\_WT OE\_vs\_WT

0.003246753 0.003246753 0.025974026

Kruskal-Wallis Test was performed.

Group p-value is:0.001080416

Forms\_Biofilms\_stats.txt

- BUGBASE是著名的Dan Knight实验室出品，仅在预印本发布，他们还发布过微生物来源分析软件SourceTracker、微生物组机器学习数据库等。
- BUGBASE有网络版，但要求biom格式，而且我用自己数据没成功；
- BUGBASE本地版较大，依赖包较多，但支持txt格式且容易运行成功；
- 推荐使用方法：基于易生信定制的Windows本地版或ImageGP在线版；
- 不推荐使用：Linux安装或官方在线版。



# 进一步学习

- 0概述：根据16S预测微生物群落功能最全攻略，对以下4种常用的功能预测工具进行简介、实例说明和点评，必读；
- 1KO通路PICRUST：本地和在线分析，及统计分析可视化指导；
- 2元素循环FAPROTAX：元素循环相关菌代谢预测，只有本地版python脚本；识别菌种名称；
- 3表型bugbase：表型预测、革兰氏、氧气利用等；有在线和本地版，基于Greengene数据库；
- 4KO通路Tax4Fun：R包基于作者整理的Sliva123数据，准确率不可知，输出结果为百分比；最新版Tax4Fun2
- PICRUST2：OTU/ASV等16S序列随意预测宏基因组，参考数据库增大10倍







扫码关注生信宝典，学习更多生信知识



扫码关注宏基因组，获取专业学习资料

# 易生信，没有难学的生信知识



# 常见问题：找不到biom包

- Error in library(biom) : 不存在叫'biom'这个名字的程辑包
- 在CRAN和Bioconductor上都没有biom包，因为已经在Bioconductor上面更新为biomformat包
- 参考主软件主原则同意安装最新版biomformat
- <http://www.bioconductor.org/packages/release/bioc/html/biomformat.html>
- 修改run.bugbase.r脚本中library(biom)为library(biomformat)

