**使用Meta-Apo对16S扩增子的微生物组功能信息进行校正**

**Calibration of 16S-amplicon-based microbiome function by Meta-Apo**

张明乾1，张文科1，荆功超2，苏晓泉1 $ \*

1计算机科学技术学院，青岛大学，青岛市，山东省；

2单细胞中心，中国科学院青岛生物能源与过程研究所，青岛市，山东省；

$现工作单位：计算机科学技术学院，青岛大学，青岛市，山东省

\*通讯作者邮箱：suxq@qdu.edu.cn

摘要：微生物组的功能谱（functional profile）在宿主疾病诊断、生态健康检测等方面具有重要的研究和应用价值。目前功能谱可通过鸟枪法宏基因组测序（Shotgun Metagenomic Whole Genome Sequencing；以下简称WGS）数据直接解析；或基于16S rRNA基因扩增子（以下简称16S扩增子）测序数据，根据其参照基因组的关联进行预测。16S扩增子测序在实验和计算上的成本比WGS低得多，因此PICRUSt2等工具已广泛用于基于16S来预测微生物组的功能谱。然而，由于扩增子测序的PCR偏好性和16S rRNA基因-全基因组关联信息的不足，同一微生物组样本基于16S扩增子的功能谱与WGS产生的结果之间会存在偏差，从而导致相左的结论。为了解决以上问题，我们提出了Meta-Apo （Metagenomic Apochromat），它可以极大地减少甚至消除这种偏差。我们对来自4个身体部位超过5,000例人体微生物组的16S 扩增子样本进行测试发现，Meta-Apo仅使用15个WGS: 16S扩增子的配对样本来进行训练，就可以显著降低两种测序之间功能解析的差异。因此，利用Meta-Apo，可以让低成本的16S扩增子测序产生与WGS相近的、可靠的、高分辨率的微生物组功能图谱。Meta-Apo可以在https://github.com/qibebt-bioinfo/meta-apo下载。它以少数WGS：16S扩增子配对样本（例如，约15对配对样本）的功能谱作为训练集，可以对大量16S扩增子样本的功能信息进行校正。

关键字：微生物组，宏基因组，扩增子，功能预测，功能校正

# 仪器设备

Meta-Apo仅需要具有约1GB内存的标准计算机即可支持其安装与执行。目前Linux（如Ubuntu、CentOS、RedHat等）、Mac OS或Windows 10内置Linux子系统等操作系统均能够支持Meta-Apo。

**软件**

Meta-Apo软件最新版本为1.01。该软件主要由C++语言开发编写，所以软件的安装需要C++编译器（例如g++）。对于Linux操作系统，大多版本已经在系统中安装了g++。对于Mac OS，建议从App Store安装Xcode应用程序，即可完成编译器的安装与配置。

**实验步骤**

1. 安装Meta-Apo

我们建议选择步骤 1.1 中自动安装的方式来配置Meta-Apo软件。但如果自动安装程序失败，可以按照步骤 1.2 中的步骤手动安装Meta-Apo软件。

* 1. 自动安装（首选方案）

1. 下载安装包

git clone https://github.com/qibebt-bioinfo/meta-apo.git

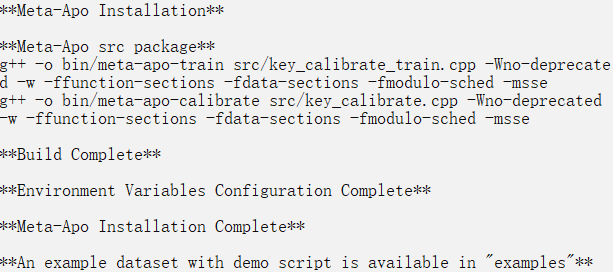
1. 安装

运行以下安装命令：

cd meta-apo

source install.sh

按照上述步骤操作，该软件包可以在1分钟内安装到计算机上，安装成功后提示信息如下（**图1**）所示：



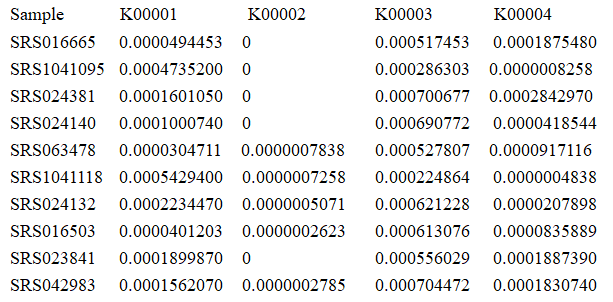
**图1. Meta-Apo安装成功提示信息**

示例数据集在安装包内“examples”文件夹下，可以查看 “examples/Read me”中的内容来获取演示运行的详细信息，或直接运行：

sh Readme

来自动演示示例数据集的处理运行。

该示例数据集包含有三个文件，其中，training.16s.ko.abd为训练建模所需的16S扩增子样本的相对丰度表（**图2**），training.wgs.ko.abd为训练建模所需的WGS样本的相对丰度表，16s.ko.abd为待校正16S扩增子样本的相对丰度表。相对丰度表的格式详见**表1**。



**图2. 示例数据集中训练建模所需的16S扩增子样本的相对丰度表**

* 1. 手动安装（备选方案）

1. 下载安装包

git clone https://github.com/qibebt-bioinfo/meta-apo.git

2）配置环境变量

将以下内容写入环境变量配置文件(一般默认的文件是“~/.bashrc”)

export MetaApo=Path to MetaApo

export PATH="$PATH:$MetaApo/bin/"

并启用环境变量

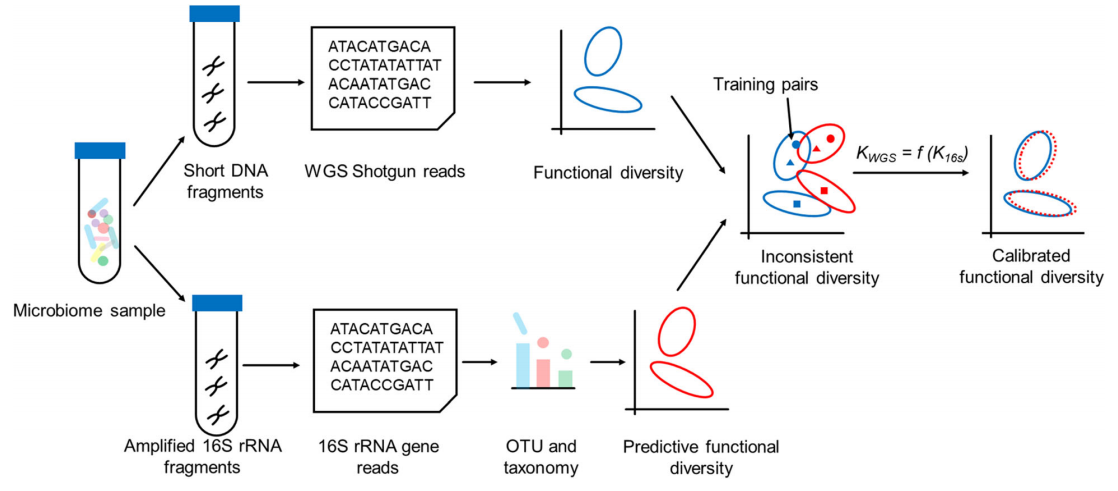
source ~/.bashrc

3）编译源代码

cd meta-apo

make

1. Meta-Apo校正原理



**图3. 通过对少量成对的WGS：16S扩增子样本进行训练来校正微生物组扩增子样本的预测功能图谱**

前期工作中，通过比较WGS和16S扩增子测序方法得出的功能谱，两种方法得到的WGS与16S扩增子之间距离高度相关(Jing*等，* 2021)。Meta-Apo仅使用少量的WGS：16S扩增子配对数据（即每一个样本都分别进行WGS和16S扩增子测序）用作训练集（如，15对训练样本），Meta-Apo就可以为大规模16S扩增子样本（如，数千例样本）的功能谱进行校正，使之结果与WGS更加一致（图3）。Meta-Apo主要包含两个部分：训练和校正。在训练部分中，Meta-Apo使用线性回归建模利用少量的WGS:16S配对样本来估算等式(1)中的*f*。在校正部分中，将WGS结果视为“黄金标准”，使用模型*f*校正16S扩增子样本的预测功能图谱。

*KWGS = f(K16S)*  （1）

1. 样本处理与输入格式

Meta-Apo仅使用少量的WGS：16S扩增子配对数据用作训练集。根据前期对来自4个身体部位超过5,000例人体微生物组的16S扩增子样本进行测试发现，Meta-Apo仅使用15个WGS: 16S扩增子配对样本来进行训练，就可以显著降低两种测序之间功能解析的差异(Jing*等，* 2021)(详见“**结果与分析**”)。因此我们建议训练集中包含10-20例WGS：16S扩增子配对数据即可。

训练集中每一个WGS：16S扩增子配对样本都分别进行WGS和16S扩增子测序，其功能谱信息需使用KEGG Orthology(Kanehisa*等，* 2011)（KO）来注释。其中，WGS样本我们建议用HUMAnN2(Franzosa*等，* 2018)进行功能分析，16S扩增子样本我们建议用PICRUSt2(Douglas*等，* 2020)进行功能预测。同时，待校正的16S扩增子样本，需按照与训练集中16S扩增子完全相同的测序流程和分析流程来处理。以上所有样本的输入文件中包含KO号和KO的丰度两类信息。目前Meta-Apo接受以下两种格式的输入文件格式（可任选其一）。

* 1. 丰度表

一个丰度表中可以包含多个样本的功能丰度信息。含有N个样本的丰度 表格式如**表1**所示。表中第一行为表头信息，接下来的N行为功能丰度信息。其中，第一列为样本的名称，其余列均为样本中所含有的KO功能的丰度。

**表1. 丰度表格式**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sample | K00001 | K00002 | K00003 | K00004 | K00005 |
| Sample1 | 0.1 | 0 | 0.3 | 0.1 | 0.1 |
| Sample2 | 0.3 | 0.1 | 0.1 | 0 | 0.1 |
| Sample3 | 0 | 0.2 | 0.1 | 0.3 | 0 |
| … |  |  |  |  |  |
| SampleN | 0 | 0.1 | 0.2 | 0.4 | 0 |

在训练集中，16S扩增子样本与WGS样本分别用单独的丰度表，两者格式相同，且要求待训练的WGS：16S配对样本在各自丰度表中样本的顺序一致。

* 1. 样本列表

一个样本列表中含有多个样本的功能丰度文件的地址路径，如**表2**所示：

**表2. 文件列表格式**

|  |  |
| --- | --- |
| Sample1 | /home/data/sample1.ko.out |
| Sample2 | /home/data/sample2.ko.out |
| Sample3 | /home/data/sample3.ko.out |
|  |  |
| SampleN | /home/data/sampleN.ko.out |

该文件有两列信息，其中，第一列为样本的名称，第二列表示每个样本单独的功能信息文件的路径。在训练集中，16S扩增子样本与WGS样本分别用单独的文件列表，两者格式相同，且要求待训练的WGS：16S配对样本在各自文件列表中样本的顺序一致。为了保证路径的合法性，我们强烈建议使用绝对地址（即包含完整的路径名称，如**表2**所示）。列表中每个样本单独的功能信息文件格式如**表3**所示：

**表3.样本的功能信息文件**

|  |  |
| --- | --- |
| #KO | Count |
| K00001 | 0.1 |
| K00003 | 0.3 |
| K00004 | 0.1 |
| K00005 | 0.1 |
| K00006 | 0.1 |
| K00010 | 0.2 |
| K00012 | 0.1 |

其中，第一列为KO号，第二列为样本中该KO功能的丰度。

1. 训练与校正
   1. 以丰度表为输入输出

训练建模由Meta-Apo程序包中的meta-apo-train程序提供。以WGS样本的KO相对丰度表training.wgs.ko.abd（由“-T”指定）和其配对的16S扩增子样本的KO相对丰度表training.16s.ko.abd（由“-t”指定）为例，训练过程如下：

meta-apo-train -T training.wgs.ko.abd -t training.16s.ko.abd -o meta-apo.model

训练过程所输出的模型文件为meta-apo.model。

接下来，就用生成的模型文件来校正大量的16S扩增子样本的功能信息。校正由Meta-Apo程序包中meta-apo-calibrate程序提供。在该程序中，模型文件由“-m”来指定，以待校正的16S扩增子样本KO相对丰度表16s.ko.abd（由“-t”指定）为例，校正过程如下：

meta-apo-calibrate -t 16s.ko.abd -m meta-apo.model -o 16s.ko.calibrated.abd

输出的文件16s.ko.calibrated.abd（由“-o”指定）是校正后的KO丰度表，其格式与输入文件16s.ko.abd一致（格式参考**表**1）。

* 1. 以样本列表为输入输出

训练建模由Meta-Apo程序包中meta-apo-train程序提供。以WGS样本列表training.wgs.list（由“- L”指定）和其配对的16S扩增样本列表training.16s.list（由“-l”指定）为例，训练过程如下：

meta-apo-train -L training.wgs.list -l training.16s.list -o meta-apo.model

训练过程所输出的模型文件为meta-apo.model。

接下来，就用生成的模型文件来校正大量的16S扩增子样本的功能信息。校正由Meta-Apo程序包中meta-apo-calibrate程序提供。在该程序中，模型文件由“-m”来指定，以待校正的16S扩增子样本列表16s.ko.list（由“-l”指定）为例，校正过程如下：

meta-apo-calibrate -l 16s.ko.list -m meta-apo.model -o 16s.ko.calibrated.out

输出的文件夹16s.ko.calibrated.out（由“-o”指定）包含每个输入样本校准后的功能信息文件（格式参考**表3**），同时校准后样本的文件列表也输出到16s.ko.calibrated.out.list中，格式与输入的16s.ko.list列表一致（格式参考**表2**）。

需注意的是，采用4.1和4.2两种不同格式的输入输出，训练程序meta-apo-train所生成的模型是通用的，可以被用不同输入格式的校正程序，无需重新训练。

1. Meta-Apo的计算过程
   1. 训练

每个微生物群落的功能由一系列代谢功能（例如KEGG Orthology；KO）及其相对丰度组成，如等式（2）

*Kmicrobiome = {kfunction 1, kfunction 2,…, kfunction i}*（2）

其中*kfunction i*代表功能*i*的相对丰度，由于WGS和扩增子之间的功能分布存在强线性相关性，对于每个功能*k*，我们可以在两种办法之间建立联系，如等式（3）

*kWGS = f(k16S) =θ0k16S + θ1*  （3）

在等式（3）中Meta-Apo使用N（例如N=15）例WGS：16S扩增子配对样本进行训练来计算模型f，通过优化等式（3）中的θ0和θ1，尽可能地降低*k16S* 和*KWGS*的差异，即等式（4）中的总差值E最小。

*E = f*(*k16S*)-*kWGS*)2  （4）

具体来讲，在训练步骤中，Meta-Apo采用最小二乘法（Least Square Method）计算参数θ0和θ1，如等式（5）和等式（6）所示：

（5）

（6）

* 1. 校正

Meta-Apo利用从训练中得到的模型*f*，可以利用等式（7）估算出16S扩增子样本中每个功能校正后的丰度。

*kexpected*= θ1*k16S*+θ0 ≈ *kWGS* （7）

由于已经优化了映射模型*f*，使得从16S rRNA基因预测的功能丰度和来自WGS的真实的功能丰度之间的差异最小化，因此Meta-Apo可以将16S扩增子样本的预测功能谱校准为WGS的水平。

**结果与分析**

为了验证Meta-Apo对于校准扩增样本功能丰度的可靠性、准确性，本工作采用了5个来自人类微生物组计划HMP(Huttenhower*等，* 2012)的数据集**（表4）**进行验证。

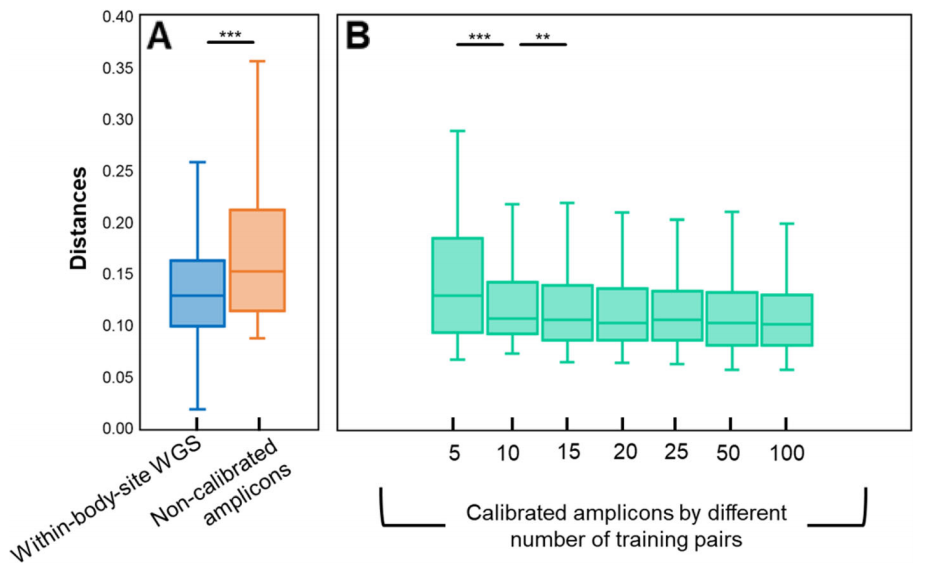
**表4.测试数据集**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dataset | # of WGS sample | # of amplicon samples | Amplicon type | Paired | Source study | Body |
| **数据集 1** | 622 | 622 | V3-V5 16S rRNA | Yes | HMP | Gut, Oral, Skin and Vaginal |
| **数据集 2** | 295 | 295 | V1-V3 16S rRNA | Yes | HMP | Gut, Oral and Vaginal |
| **数据集 3** | 2,354 | 5,350 | V3-V5 16S rRNA | No | HMP | Gut, Oral, Skin and Vaginal |
| **数据集 4** | 2,045 | 2,186 | V1-V3 16S rRNA | No | HMP | Gut, Oral and Vaginal |

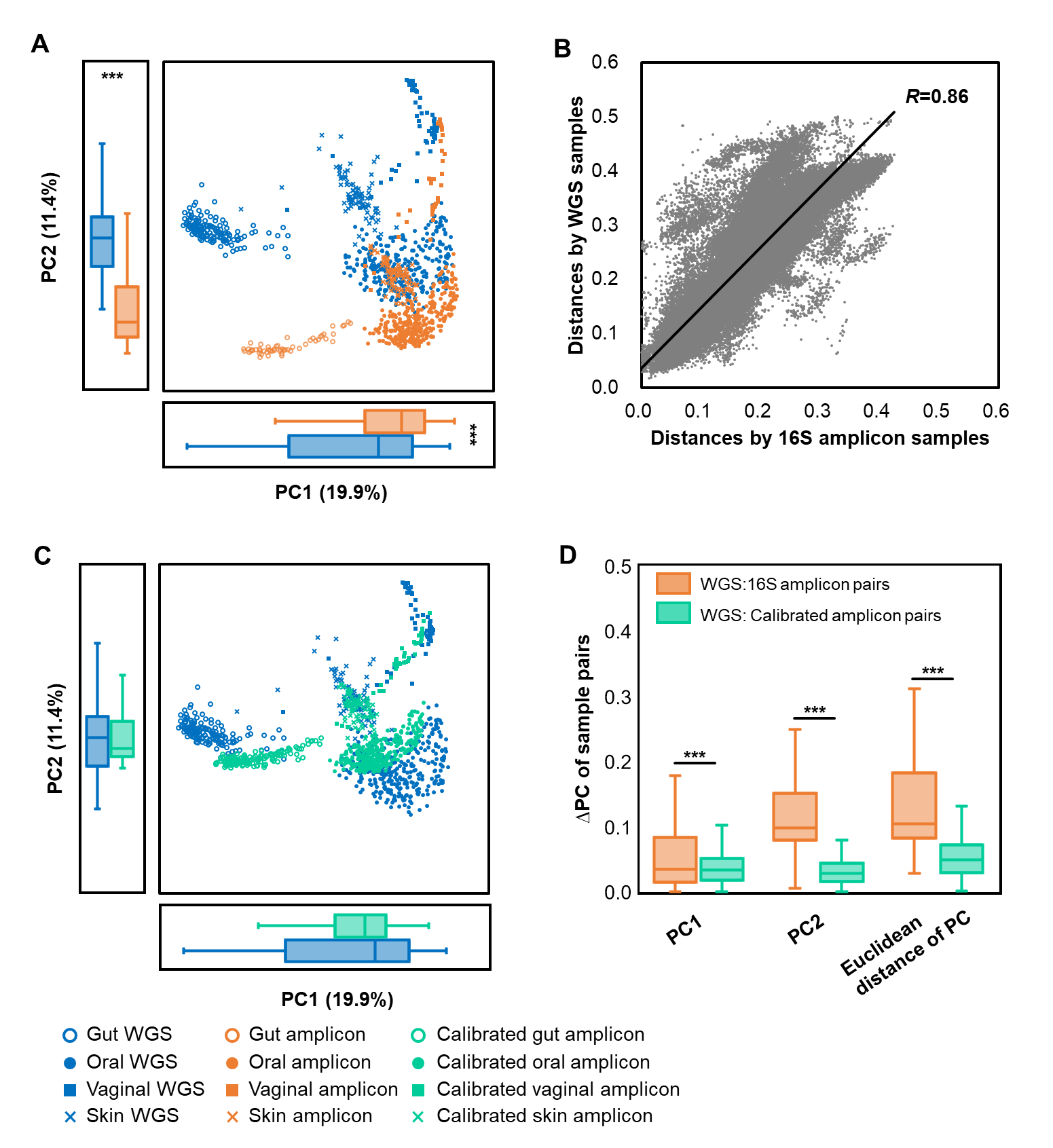
以上测试中所有的数据集均可在 Meta-Apo 软件下载页面的“Supplementary”部分中下载。

我们首先比较了622例配对的人体微生物组功能谱（数据集1；来自四个身体部位：肠道，皮肤，口腔和生殖道；**表4**）来评估两种测序策略之间的差异程度。每个样本都通过WGS和V3-V5区16S rRNA扩增子进行测序。WGS的功能谱由HUMAnN2(Franzosa*等，* 2018)分析生成。16S扩增子则使用PICRUSt2(Douglas*等，* 2020)预测得出，均使用KEGG Orthology(Kanehisa*等，* 2011)（KO）注释。通过比较从两种测序方法得出的功能谱，我们发现配对的WGS：16S扩增子之间差异显著高于WGS的内部差异（即来自同一部位的WGS样本之间的距离；**图4A**）。两种策略之间的差异十分显著，β多样性也表现出非常不同的模式（**图5A**；PC1 双尾配对Wilcox秩和检验*p* < 0.01；PC2双尾配对Wilcox秩和检验*p* < 0.01）并导致了一些错误的分类。例如，一些皮肤的16S扩增子的功能谱与口腔的WGS的功能谱被错误的分成一类。然而，这两种方法得到的WGS与16S扩增子之间距离高度相关（**图5B**；Pearson相关性*R* = 0.86，*p* < 0.01），而且其β多样性之间的总体形状相似（**图5A**；蒙特卡洛检验*p* < 0.01）。

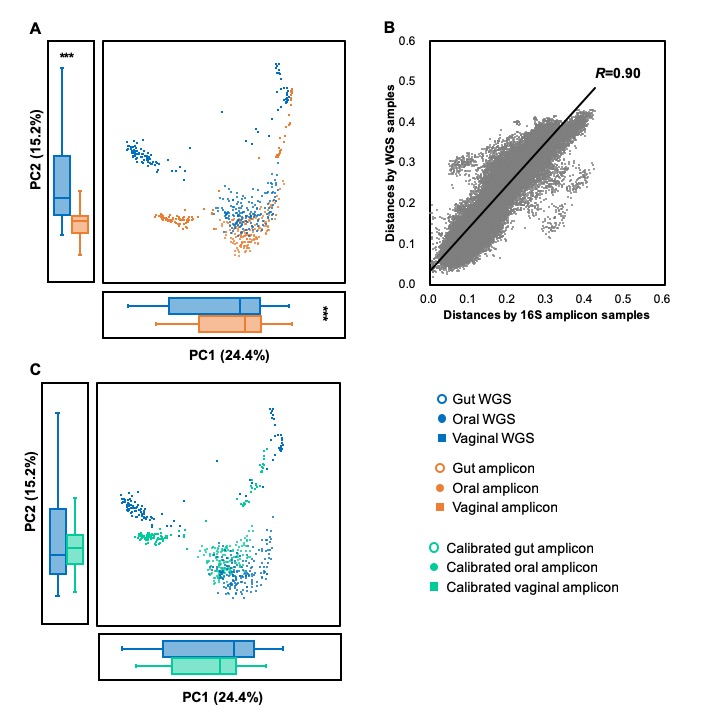
为了定量评估Meta-Apo的效果，我们分别从数据集1中随机选择了N = 5、10、15、20、50和100个WGS:16S扩增子配对样本作为训练集，并使用Meta-Apo校正该数据集中其他16S扩增子样本。当使用N = 15个训练对建立模型f时，Meta-Apo校正效果变得稳定，并且在增加更多训练对之后（最多100个；**图4B**），校正效果也不会明显增加。在校正后（即N = 15个训练对），配对的WGS：16S扩增子距离（0.121±0.055）显著低于WGS样本的组内距离（0.136±0.056）。经主坐标分析（PCoA）证实，Meta-Apo消除了两种测序策略产生的样本之间的总体功能分布差异（**图5C**；PC1双尾配对Wilcox秩和检验*p* = 0.30，PC2双尾配对Wilcox秩和检验*p* = 0.29；**图5D**。与此同时，Meta-Apo对于来自数据集2（**表4**）的V1-V3区16S rRNA序列也同样适用（**图6**）。



**图4. Meta-Apo显著减少了数据集1中WGS和16S扩增子配对样本之间的功能谱的距离。**A. WGS：16S扩增子配对样本之间的Bray-Curtis距离（未校正，橙色条）高于WGS体内位点距离（来自同一部位的WGS样本之间的距离，蓝色条）。B. 仅使用15个训练对，校正的16S扩增子样本与其配对的WGS样本之间的Bray-Curtis距离变得稳定，且显著低于WGS的组内距离。两个图像共用X轴。通过双尾Wilcox秩和检验计算*p*值，\*\*表示*p* < 0.05，\*\*\*表示*p* < 0.01。

****

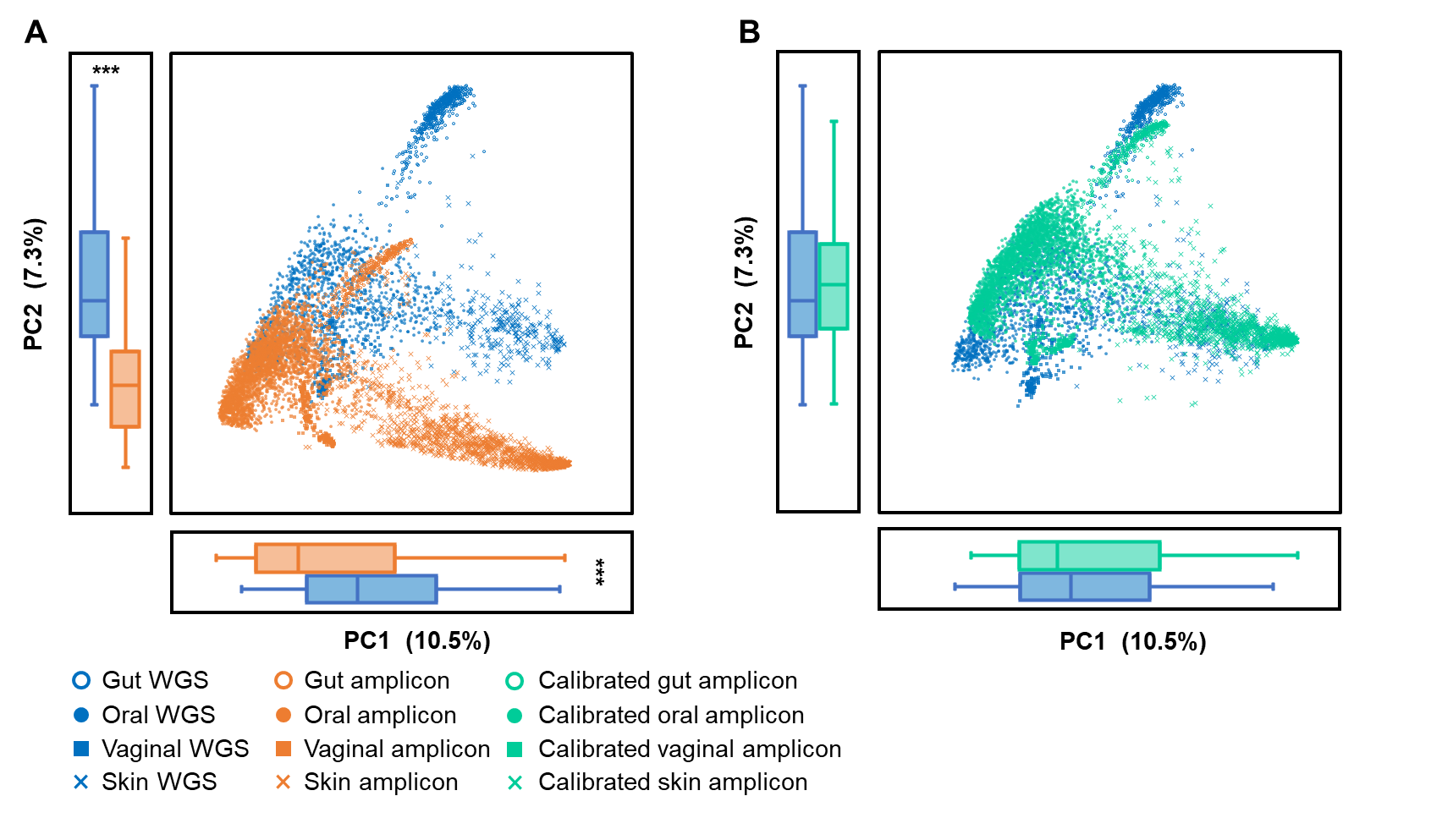
**图5. 数据集1的622个WGS：16S扩增子配对样本的beta多样性。**A. 16S扩增子和WGS方法的总体功能模式是同构的，但在PC1和PC2分布上存在明显差异。B. 由WGS和16S扩增子计算的Bray-Curtis距离高度相关（Pearson相关*R* = 0.86，*p* < 0.01）。C. Meta-Apo使用15个配对样本进行训练，将16S扩增子样本的预测功能谱与WGS样本的预测功能谱进行比对，从而使校正的功能谱的PC1和PC2比原始的未校正的16S扩增子样品更接近WGS样品。D. WGS：16S扩增子对的ΔPC显著降低。PCoA使用Bray-Curtis距离计算主坐标。通过双尾配对的Wilcox秩和检验计算*p*值，\*\*\*表示*p* < 0.01。

****

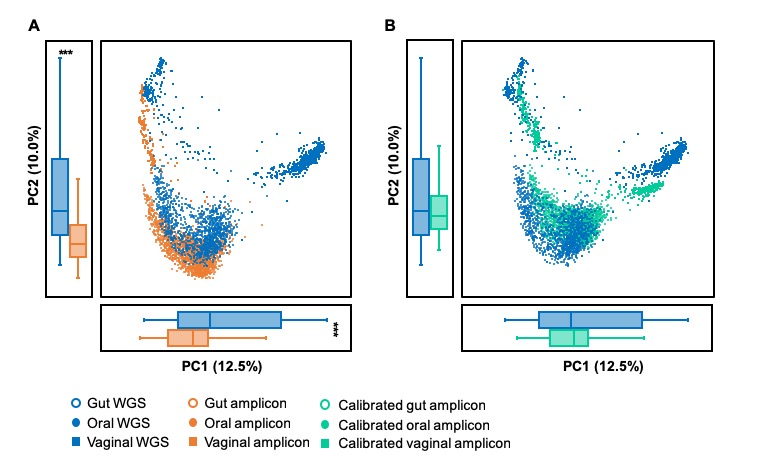
**图6. 数据集2的295个WGS：16S扩增子配对样本的beta多样性。**A. 16S扩增子和WGS方法的总体功能模式是同构的，但在PC1和PC2分布上存在明显差异。B. 由WGS和16S扩增子计算的Bray-Curtis距离高度相关（Pearson相关*R* = 0.90，*p* < 0.01）。C. Meta-Apo使用15个配对样本进行训练，将16S扩增子样本的预测功能谱与WGS样本的预测功能谱进行比对，从而使校正的功能谱的PC1和PC2比原始的未校正的16S扩增子样品更接近WGS样品。通过双尾配对的Wilcox秩和检验计算*p*值，\*\*\*表示*p* < 0.01。

我们进一步将Meta-Apo样本扩展至5,350 个V3-V5 16S rRNA扩增子样本和与2,354 个WGS样本（数据集3，同数据集1一样从四个身体部位收集，并使用相同的方法处理序列；**表4**），从而评估大规模16S扩增子功能图谱的校正性能。该数据集尽管是来自于相同的健康宿主队列，并由同一研究进行测序（HMP），但WGS和16S扩增子样品并未配对。另外我们发现，无论选择何种测序策略(Rausch*等，* 2019)，由WGS和16S扩增子所得出的物种结构组成是一致的，但在功能图谱上则有显著差异（**图7A**； PC1双尾Wilcox秩和检验*p* < 0.01；PC2双尾Wilcox秩和检验*p* < 0.01）。例如，在功能图谱上，肠道部位的16S扩增子与口腔中WGS聚类在一起，口腔等相同部位的样本会按照不同的测序策略分离，即身体部位在人类微生物组的功能格局中占主导地位(Turnbaugh*等，* 2009; Huttenhower*等，* 2012)。之后，我们使用Meta-Apo，利用数据集1的WGS：16S扩增子对做训练样本（训练样本*N* = 15）构建的模型，对所有扩增子样本的预测功能图谱进行校正。经β多样性的分析证明，Meta-Apo校正后的16S扩增子和WGS样本之间功能谱的偏差大大降低（**图7B**； PC1双尾Wilcox秩和检验*p* = 0.20；PC2双尾Wilcox秩和检验*p* = 0.03）。

接下来，为了测试对不同可变区16S数据集的校正效果，我们也将Meta-Apo应用于表4中数据集4的2,186个V1-V3区16S扩增子样本。使用数据集2的WGS：16S扩增子对做训练样本（训练样本*N* = 15）来构建的模型，Meta-Apo也可以有效地提高16S扩增子的功能谱重建的准确性（**图8）**。因此，Meta-Apo普遍适用于16S rRNA基因的多个可变区域。



**图7. 来自数据集3的2,354个WGS样本和5,350个16S扩增子样本的功能beta多样性。**A.16S扩增子和WGS方法获得的功能模式在PC1和PC2分布上有显著差异。B. Meta-Apo使用15个配对样本进行训练，将扩增子样本的预测功能图谱与WGS样本的预测功能图进行比较，与原始的未经校正的扩增子样品相比，校正后的扩增子样本的功能谱的PC1和PC2更接近WGS样本。PCoA使用Bray-Curtis距离计算主坐标。通过双尾Wilcox秩和检验计算*p*值，\*\*\*表示*p* <0.01。



**图8. 来自数据集4的2,045个WGS样本和2,186个16S扩增子样本的功能beta多样性。**A.16S扩增子和WGS方法获得的功能模式在PC1和PC2分布上有显著差异。B. Meta-Apo使用15个配对样本进行训练，将扩增子样本的预测功能图谱与WGS样本的预测功能图进行比较，与原始的未经校正的扩增子样品相比，校正后的扩增子样本的功能谱的PC1和PC2更接近WGS样本。PCoA使用Bray-Curtis距离计算主坐标。通过双尾Wilcox秩和检验计算*p*值，\*\*\*表示*p* <0.01。

**失败经验**

问题1

安装提示：“make: g++: command not found”

问题原因：没有安装Meta-Apo所需要的g++编译器。

解决方法：根据不同的操作系统，利用相应的命令安装 g++，常见的操作系统：

Ubuntu Linux系统：sudo apt-get install g++

CentOS Linux系统：sudo yum install g++

Mac OS 系统：通过App Store安装Xcode应用程序

问题2

运行提示：“Please set the environment variable MetaApo to the directory”

问题原因：环境变量设置失败。

解决方法：请参考实验步骤 1.2.2 中手动配置环境变量的方法将 Meta-Apo 所需要的环境变量添加到配置文件中。

问题3

运行提示：“meta-apo-train: command not found”

问题原因：环境变量设置失败。

解决方法：请参考实验步骤 1.2.2 中手动配置环境变量的方法将 Meta-Apo 所需要的环境变量添加到配置文件中。

问题4

运行提示：“Error: Cannot open file: XXX”

问题原因：输入了错误的输入/输出文件路径。

解决方案：请检查正确的输入文件路径(可在输入时用Tab 键自动补全)，并确保用户在输出路径下有足够的写权限。

问题5

运行提示：“Argument #X Error : Arguments must start with -”

问题原因：运行命令中所有参数选项名称必须以“-”开头。

解决方法：请检查第 X 个参数并更正。

**致谢**

本项工作得到了国家自然科学基金**31771463**、**32070086**和**32000389**项目，以及山东省自然科学基**ZR201807060158**项目的资助。

**参考文献**

1. Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., Huttenhower, C. and Langille, M. G. I. (2020). [PICRUSt2 for prediction of metagenome functions](https://www.nature.com/articles/s41587-020-0548-6). *Nature Biotechnology*

2. Franzosa, E. A., Mciver, L. J., Rahnavard, G., Thompson, L. R., Schirmer, M., Weingart, G., Lipson, K. S., Knight, R., Caporaso, J. G. and Segata, N. (2018). [Species-level functional profiling of metagenomes and metatranscriptomes](https://www.nature.com/articles/s41592-018-0176-y). *Nature Methods* 15

3. Huttenhower, C., Gevers, D., Knight, R., Abubucker, S. and White, O. (2012). [The Human Microbiome Project (HMP) Consortium. Structure, function and diversity of the healthy human microbiome](https://www.nature.com/articles/nature11234). Nature 486: 207-214. *Nature* 486(7402): 207–214

4. Jing, G., Zhang, Y., Cui, W., Liu, L. and Su, X. (2021). [Meta-Apo improves accuracy of 16S-amplicon-based prediction of microbiome function.](https://europepmc.org/article/med/33407112) *BMC Genomics* 22(1)

5. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. and Tanabe, M. (2011). [KEGG for integration and interpretation of large-scale molecular data sets](https://pubmed.ncbi.nlm.nih.gov/22080510/). *Nucleic Acids Research* 40(D1): D109-D114.

6. Rausch, P., Rühlemann, M., Hermes, B. M., Doms, S. and Baines, J. F. (2019). [Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms.](https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-019-0743-1) *Microbiome* 7(1)

7. Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. A. and Affourtit, J. P. (2009). [A core gut microbiome in obese and lean twins](https://www.nature.com/articles/nature07540). *Nature* 457(7228): 480