

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。

1.1 自动安装（首选方案）

1) 下载安装包

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

2) 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：

```

**Meta-Apo Installation**

**Meta-Apo src package**
g++ -o bin/meta-apo-train src/key_calibrate_train.cpp -Wno-deprecated
d -w -ffunction-sections -fdata-sections -fmodulo-sched -msse
g++ -o bin/meta-apo-calibrate src/key_calibrate.cpp -Wno-deprecated
-w -ffunction-sections -fdata-sections -fmodulo-sched -msse

**Build Complete**

**Environment Variables Configuration Complete**

**Meta-Apo Installation Complete**

**An example dataset with demo script is available in "examples"**

```

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。

Sample	K00001	K00002	K00003	K00004
SRS016665	0.0000494453	0	0.000517453	0.0001875480
SRS1041095	0.0004735200	0	0.000286303	0.0000008258
SRS024381	0.0001601050	0	0.000700677	0.0002842970
SRS024140	0.0001000740	0	0.000690772	0.0000418544
SRS063478	0.0000304711	0.0000007838	0.000527807	0.0000917116
SRS1041118	0.0005429400	0.0000007258	0.000224864	0.0000004838
SRS024132	0.0002234470	0.0000005071	0.000621228	0.0000207898
SRS016503	0.0000401203	0.0000002623	0.000613076	0.0000835889
SRS023841	0.0001899870	0	0.000556029	0.0001887390
SRS042983	0.0001562070	0.0000002785	0.000704472	0.0001830740

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

1.2 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
```

2. Meta-Apo 校正原理

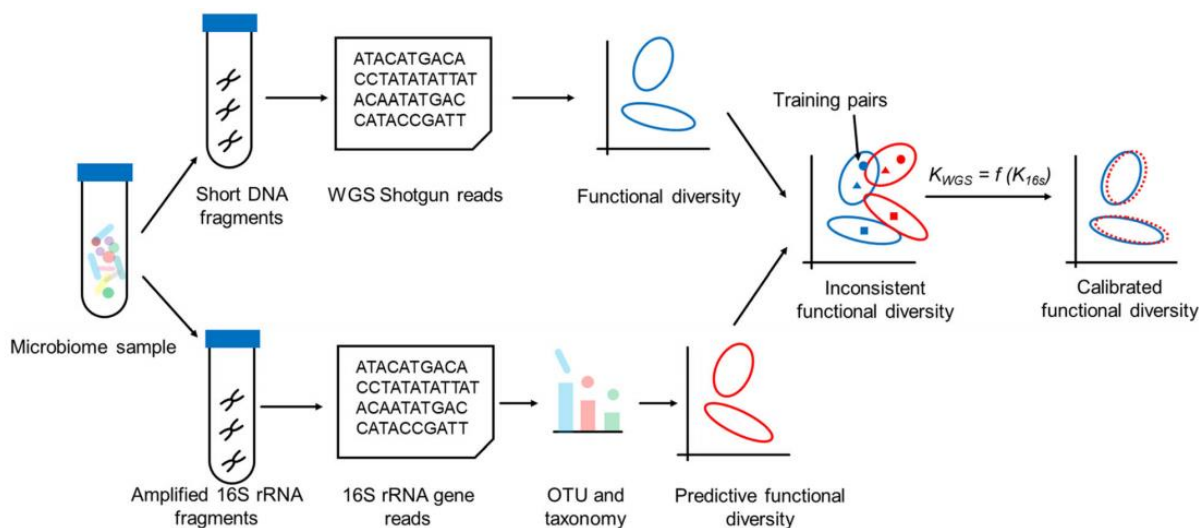


图 3. 通过对少量成对的 WGS: 16S 扩增子样本进行训练来校正微生物组扩增子样本的预测功能图谱

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。

$$K_{WGS} = f(K_{16S}) \quad (1)$$

3. 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）。

3.1 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。

3.2 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。

4. 训练与校正

4.1 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）。

4.2 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。

5. Meta-Apo 的计算过程

5.1 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）

$$K_{\text{microbiome}} = \{k_{\text{function } 1}, k_{\text{function } 2}, \dots, k_{\text{function } i}\} \quad (2)$$

其中 $k_{\text{function } i}$ 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）

$$k_{WGS} = f(k_{16S}) = \theta_0 k_{16S} + \theta_1 \quad (3)$$

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 θ_0 和 θ_1 , 尽可能地降低 k_{16S} 和 k_{WGS} 的差异, 即等式 (4) 中的总差值 E 最小。

$$E = \frac{1}{2} \sum_{i=1}^N (f(k_{16S}) - k_{WGS})^2 \quad (4)$$

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 θ_0 和 θ_1 , 如等式 (5) 和等式 (6) 所示:

$$\theta_0 = \frac{N \times \sum_{i=1}^N k_{16S} \times k_{WGS} - \sum_{i=1}^N k_{16S} \times \sum_{i=1}^N k_{WGS}}{N \times \sum_{i=1}^N k_{16S}^2 - (\sum_{i=1}^N k_{16S})^2} \quad (5)$$

$$\theta_1 = \frac{\sum_{i=1}^N k_{16S}^2 \times \sum_{i=1}^N k_{WGS} - \sum_{i=1}^N (k_{16S} \times k_{WGS}) \times \sum_{i=1}^N k_{16S}}{N \times \sum_{i=1}^N k_{16S}^2 - (\sum_{i=1}^N k_{16S})^2} \quad (6)$$

5.2 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。

$$k_{expected} = \theta_1 k_{16S} + \theta_0 \approx k_{WGS} \quad (7)$$

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 Wilcoxon 秩和检验 $p = 0.30$, PC2 双尾配对 Wilcoxon 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)。

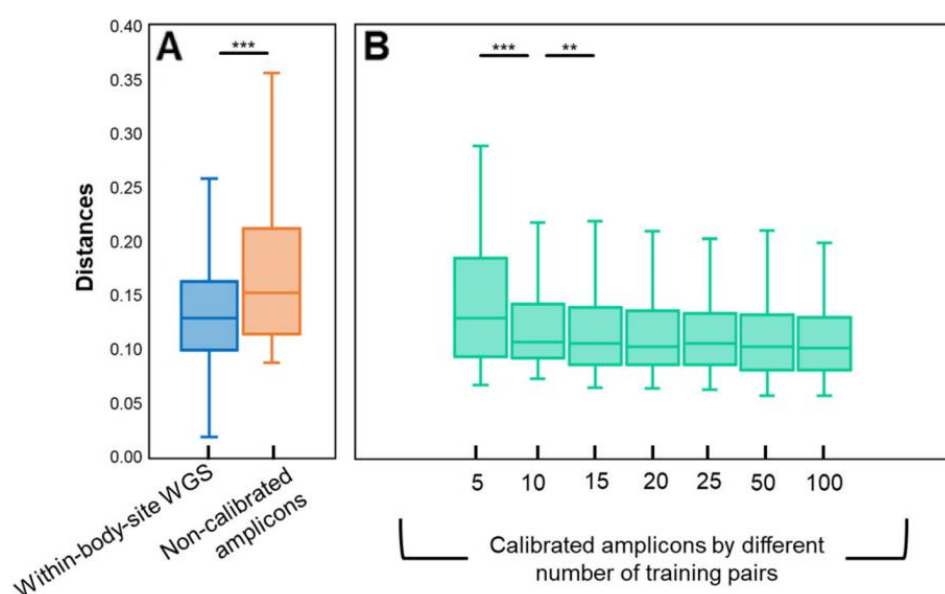


图 4. Meta-Apo 显著减少了数据集 1 中 WGS 和 16S 扩增子配对样本之间的功能谱的距离。 A. WGS: 16S 扩增子配对样本之间的 Bray-Curtis 距离 (未校正, 橙色条) 高于 WGS 体内位点距离 (来自同一部位的 WGS 样本之间的距离, 蓝色条)。B. 仅使用 15 个训练对, 校正的 16S 扩增子样本与其配对的 WGS 样本之间的 Bray-Curtis 距离变得稳定, 且显著低于 WGS 的组内距离。两个图像共用 X 轴。通过双尾 Wilcoxon 秩和检验计算 p 值, **表示 $p < 0.05$, ***表示 $p < 0.01$ 。

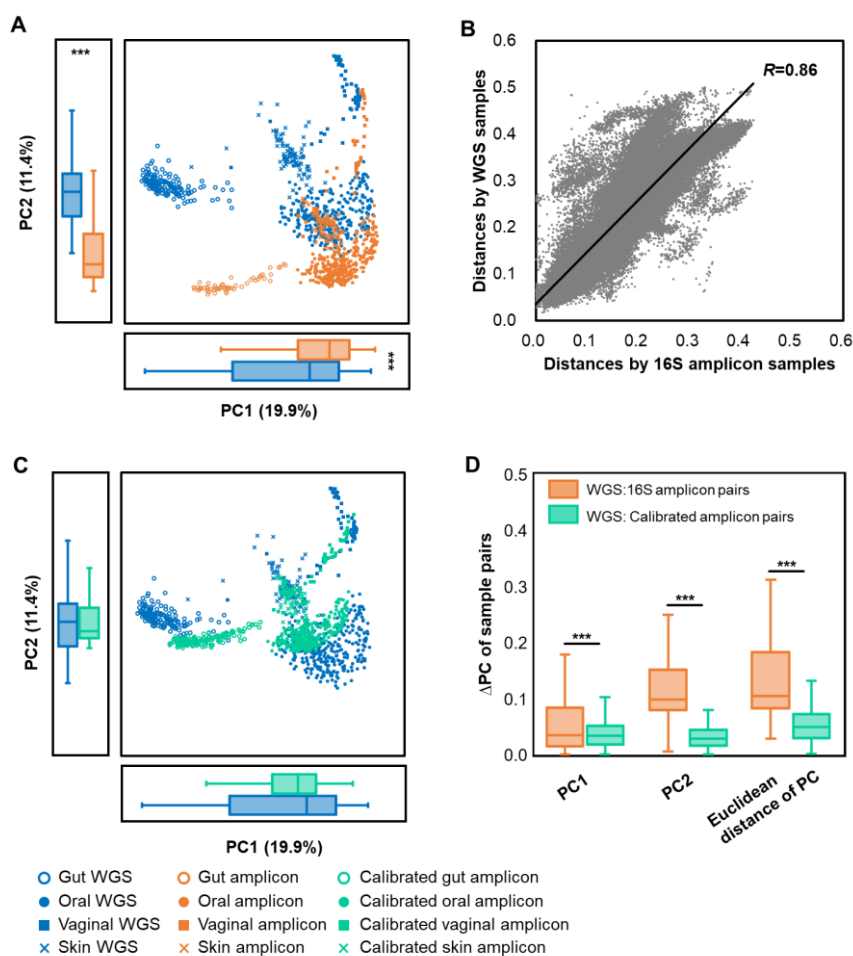


图 5. 数据集 1 的 622 个 WGS: 16S 扩增子配对样本的 beta 多样性。 A. 16S 扩增子和 WGS 方法的总体功能模式是同构的，但在 PC1 和 PC2 分布上存在明显差异。B. 由 WGS 和 16S 扩增子计算的 Bray-Curtis 距离高度相关 (Pearson 相关 $R = 0.86$, $p < 0.01$)。C. Meta-Apo 使用 15 个配对样本进行训练，将 16S 扩增子样本的预测功能谱与 WGS 样本的预测功能谱进行比对，从而使校正的功能谱的 PC1 和 PC2 比原始的未校正的 16S 扩增子样品更接近 WGS 样品。D. WGS: 16S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

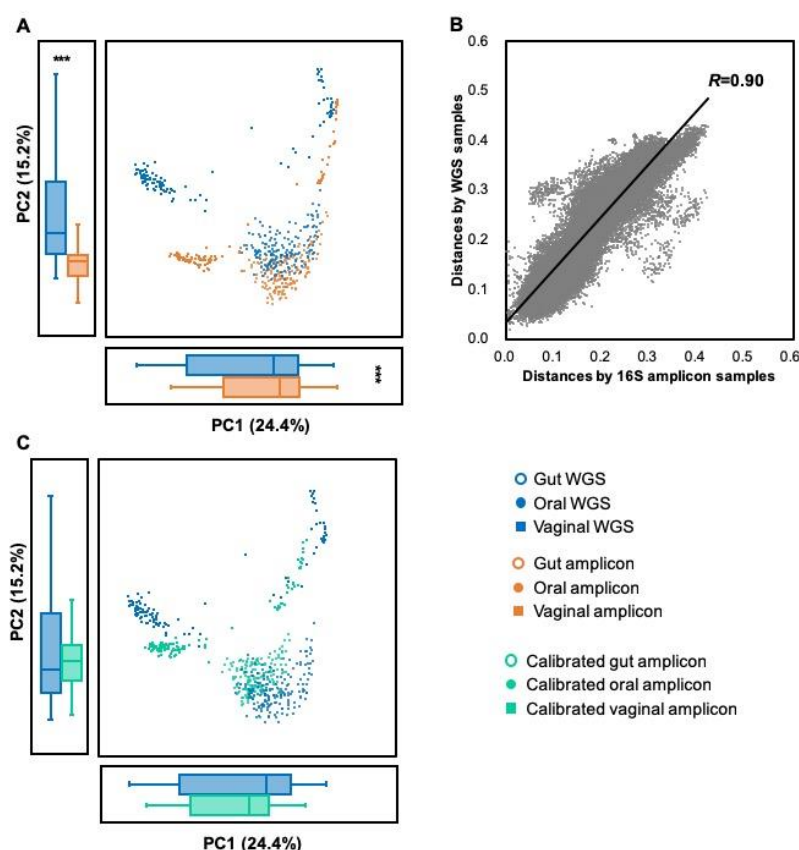


图 6. 数据集 2 的 295 个 WGS: 16S 扩增子配对样本的 beta 多样性。 A. 16S 扩增子和 WGS 方法的总体功能模式是同构的，但在 PC1 和 PC2 分布上存在明显差异。B. 由 WGS 和 16S 扩增子计算的 Bray-Curtis 距离高度相关 (Pearson 相关 $R = 0.90$, $p < 0.01$)。C. Meta-Apo 使用 15 个配对样本进行训练，将 16S 扩增子样本的预测功能谱与 WGS 样本的预测功能谱进行比对，从而使校正的功能谱的 PC1 和 PC2 比原始的未校正的 16S 扩增子样品更接近 WGS 样品。通过双尾配对的 Wilcoxon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

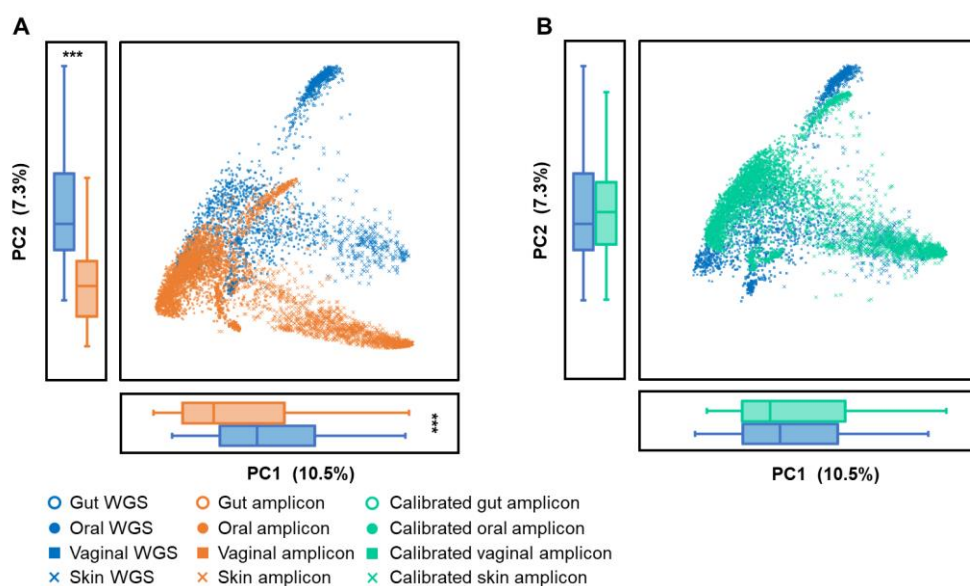


图 7. 来自数据集 3 的 2,354 个 WGS 样本和 5,350 个 16S 扩增子样本的功能 β 多样性。A.16S 扩增子和 WGS 方法获得的功能模式在 PC1 和 PC2 分布上有显著差异。B. Meta-Apo 使用 15 个配对样本进行训练,将扩增子样本的预测功能图谱与 WGS 样本的预测功能图进行比较,与原始的未经校正的扩增子样品相比,校正后的扩增子样本

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 Wilcoxon 秩和检验计算 p 值，***表示 $p < 0.01$ 。

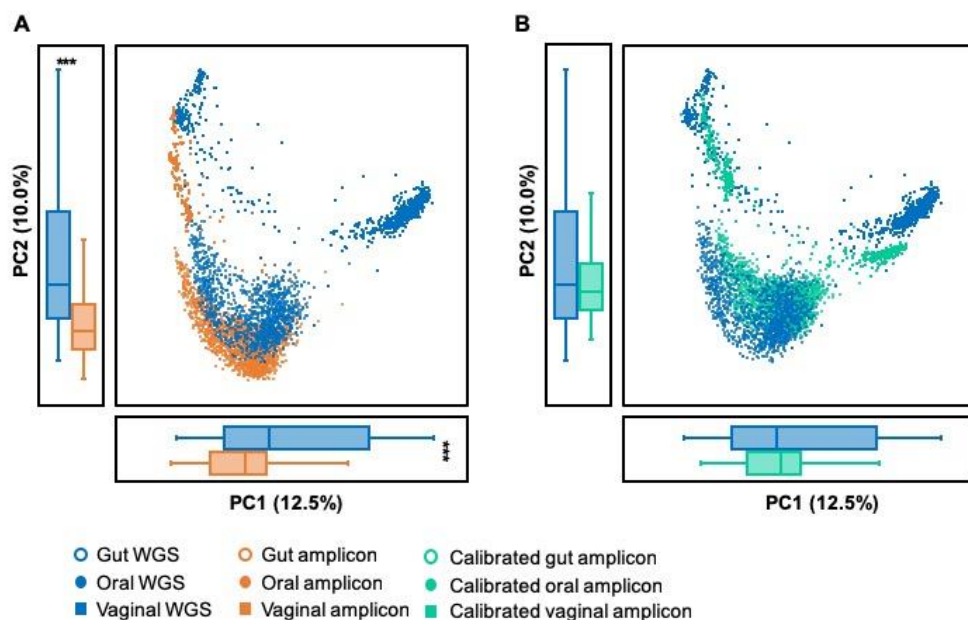


图 8. 来自数据集 4 的 2,045 个 WGS 样本和 2,186 个 16S 扩增子样本的功能 beta 多样性。A.16S 扩增子和 WGS 方法获得的功能模式在 PC1 和 PC2 分布上有显著差异。B. Meta-Apo 使用 15 个配对样本进行训练，将扩增子样本的预测功能图谱与 WGS 样本的预测功能图进行比较，与原始的未经校正的扩增子样品相比，校正后的扩增子样本的功能谱的 PC1 和 PC2 更接近 WGS 样本。PCoA 使用 Bray-Curtis 距离计算主坐标。通过双尾 Wilcoxon 秩和检验计算 p 值，*表示 $p < 0.01$ 。**

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：

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

285 CentOS Linux 系统: `sudo yum install g++`

286 Mac OS 系统: 通过 App Store 安装 Xcode 应用程序

287 问题 2

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

289 问题原因: 环境变量设置失败。

290 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
291 的环境变量添加到配置文件中。

292 问题 3

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

294 问题原因: 环境变量设置失败。

295 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
296 的环境变量添加到配置文件中。

297 问题 4

298 运行提示: “Error: Cannot open file: XXX”

299 问题原因: 输入了错误的输入/输出文件路径。

300 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
301 用户在输出路径下有足够的写权限。

302 问题 5

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

304 问题原因: 运行命令中所有参数选项名称必须以“-”开头。

305 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金 **ZR201807060158** 项目的资助。

参考文献

- 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](#). *Nature Biotechnology*
- 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](#). *Nature Methods* 15
- 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](#). *Nature* 486: 207-214. *Nature* 486(7402): 207–214
- Jing, G., Zhang, Y., Cui, W., Liu, L. and Su, X. (2021). [Meta-Apo improves accuracy of 16S-amplicon-based prediction of microbiome function](#). *BMC Genomics* 22(1)
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. and Tanabe, M. (2011). [KEGG for integration and interpretation of large-scale molecular data sets](#). *Nucleic Acids Research* 40(D1): D109-D114.
- 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](#). *Microbiome* 7(1)
- 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](#). *Nature* 457(7228): 480