**微生物群落定量宏基因组学和定量宏转录组学方法**

**Quantitative metagenomics and metatranscriptomics methods of microbial community**

黄昕瑜1, 2，张璐1, 2，袁凌1, 2，鞠峰1, 2 \*

1  浙江省海岸带环境与资源研究重点实验室，工学院, 西湖大学, 杭州, 浙江；

2 前沿技术研究所, 浙江西湖高等研究院, 杭州, 浙江

\*通讯作者邮箱: [jufeng@westlake.edu.cn](mailto:jufeng@westlake.edu.cn)

# 摘要

定量宏基因组学和定量宏转录组学技术是通过在样本中添加已知拷贝数的核酸内标片段，经建库测序后通过计算测序数据中内标序列的回收率，再结合原始环境样品的定量信息，估算样品中赋存的各基因和转录本的绝对丰度。对于不同类型的环境样品或受环境扰动前后生物量发生显著变化的样品，基因的相对丰度无法准确反映其绝对丰度的差异性与时空变化。因此，相较于基于序列相对丰度的常规定量方法，定量宏基因组与定量宏转录组能够在绝对定量的框架下更准确地比较样品之间的生物学差异，该定量宏组学方法的应用实例与技术优势见参考文献1前沿部分。该方法适用范围广，包括工程系统、天然水体、土壤、沉积物等不同类型的环境样品，都可以通过不同的绝对定量信息与方式（体积、生物量、质量、细胞数），结合内标的回收率来推算其中任一基因或转录本的拷贝数。本文在参考文献2的基础上定义了一系列实现微生物组内任一目标基因绝对转录本与绝对基因量的计算公式和相关参数。

**关键词:** 定量宏基因组、定量宏转录组、内标、绝对定量、微生物群落

# 材料与试剂

**耗材：**

0.22 μm注射过滤器单元：MILLEX 0.22 μm filter

50 mL无菌注射器

QubitTM assay tubes（Thermo Invitrogen, 货号：Q32856）

**试剂：**

十二烷基硫酸钠（SDS）（生工，货号：S0227-500g）

醋酸钠（国药，货号：10018818）

乙酸（国药，货号：10000208）

异丙醇（Acros, 货号：447080010）

无水乙醇（Sigma, BioUltra, for molecular biology, 货号：51976）

柠檬酸盐饱和苯酚（Sigma, BioReagent, for molecular biology, 货号：P4682-100mL）

氯仿（国药，货号：10006818）

无核酸酶纯水（Thermo UltraPure™ DNase/RNase-Free Distilled Water, 货号：10977023）

限制性核酸内切酶：Thermo FastDigest enzyme Mss I（Thermo Scientific™, 货号：FD1344）

胶回收试剂盒：Promega Wizard® SV Gel and PCR Clean-Up System（Promega, 货号：A9282）

proteinase K（Thermo Invitrogen, 货号：AM2546）

体外转录试剂盒：MEGAscript™ T7 Transcription Kit（Thermo Invitrogen, 货号：AM1333）

QubitTM 1X dsDNA HS Assay Kits（Thermo Invitrogen, 货号：Q33231）

Qubit™ RNA HS Assay Kit（Thermo Invitrogen, 货号：Q32852）

Q5超保真2 X 预混液（NEB Q5® High-Fidelity 2X Master Mix, 货号：M0492）

# 仪器设备

Qubit荧光计（Thermo QubitTM 4 Fluorometer）

全自动凝胶成像系统 （Tanon 2500）

水平电泳仪（Tanon EPS300+HE120）

高通量NGS片段分析系统：AATI Fragment analyzer（Agilent GENE-QC006）

低速离心机（杭州奥盛 Mini-6KS）

台式高速冷冻离心机（Eppendorf centrifuge 5427R）

PCR仪（Biometra TRIO 48）

振荡型恒温金属浴仪（杭州奥盛 MSC-100）

千分位天平（Precisa XJ220A SCS）

Nanodrop OneC（Thermo）

涡旋混合器（杭州奥盛 MTV-1）

# 实验步骤

## 1. 内标序列和PCR引物的设计与合成

本流程通过人工合成的DNA和RNA内标物投加来实现微生物群落的定量宏基因组学和定量宏转录组学分析，内标序列在设计时需要满足以下几个条件：

1）设计一段长度在1000个碱基左右的核酸内标序列（本文采用的内标序列见附录），GC 含量接近50%；起始端添加T7 promoter的序列（5’ -TAATACGACTCACTATAGG），以保证合成的质粒能够进行体外转录实验；在内标序列的末端嵌入一个酶切后产生平末端的限制性核酸内切酶酶切位点——Mss I（5’ -TTT↓AAA），用于质粒的线性化，且该酶切位点在内标序列中的其他位置不存在。

2）将序列提交到NCBI Blast比对nr数据库时，没有显著相似性参考序列被比对上；也没有任何bit-score得分大于50分的比对结果。

3）按照常规PCR的原则设计并合成PCR引物对质粒上人工合成的内标序列进行扩增，扩增产物经过回收纯化后作为后续定量宏基因组实验中的DNA内标使用。

**关于内标序列和PCR引物的设计的基本原则和注意事项请参考附录 1**

## 2. RNA内标（RNA Internal Standard, RIS）的制作

### 2.1 试剂的配制与准备

1）10% SDS buffer：称取5 g SDS固体粉末，并最终定容于50 mL无核酸酶纯水中，然后用无菌注射器和MILLEX 0.22 μm过滤器将溶液过滤到洁净的无菌容器中。

2）预冷的3 M醋酸钠溶液，pH 5.0：称取12.30 g醋酸钠，溶解于10-15 mL无菌超纯水中，用醋酸溶液调节pH到5.0，然后定容至50 mL，再用无菌注射器和MILLEX 0.22 μm过滤器将溶液过滤到洁净的无菌容器中。配制完成后适当分装，并置于-20 °C冰箱储存备用。

3）预冷的异丙醇（-20 °C冰箱）

4）预冷的70%乙醇（-20 °C冰箱）

5）预冷的无水乙醇（-20 °C冰箱）

6）柠檬酸盐饱和苯酚和氯仿1:1混合溶液（10 mL + 10 mL）

### 2.2 质粒的线性化

1）按表1准备反应体系，将所有组分按表1从上至下的顺序添加到0.2 mL 离心管中，轻弹混匀，再用低速离心机离心3-5秒以将反应混合物收集到离心管底部，对于新手建议每个质粒酶切5个反应，即5个20 μL的酶切体系（总计100 μL），以增加线性化质粒的回收后的产量，方便下游实验的进行；

表1.质粒酶切反应体系

|  |  |
| --- | --- |
|  | 体积（μL） |
| Nuclease-free water | 13.5（总体系20 μL，水的体积可按需调整） |
| 10×Buffer \* | 2 |
| Plasmid DNA | 3.5（约 1 μg） |
| FastDigest enzyme | 1 |
| \*限制性内切酶试剂盒自带的缓冲液。 | |

2）按以下条件设置金属浴仪程序（酶切时间需参考限制性内切酶的说明书，本示例使用的Thermo Scientific™ Mss I 的酶切条件为37 °C，5 min）：

金属浴仪：37 °C，5 min；

3）反应终止之后将样品取出，置于冰上，取2 μL样品进行DNA凝胶电泳检查（电泳时需要marker，及未酶切的质粒作为对照）根据电泳结果判断酶切是否完全：如果酶切完全，则可以进行下一步的实验，如若酶切不完全，可以继续酶切反应，直至质粒酶切充分；

4）确认酶切完全后，对限制性内切酶进行热失活处理（热失活处理的反应条件主要参考限制性内切酶说明书进行，本示例中使用的Thermo Scientific™ Mss I 的热失活条件为：65 °C，10 min）：

金属浴仪：65 °C，10 min；

5）蛋白酶处理：将5个相同的反应体系混合到一个1.5 mL的离心管中， 加入1 μL proteinase K（20 mg/mL）和5 μL 10% SDS buffer, 置于金属浴仪中于50°C反应30分钟。

### 2.3 线性化质粒的回收和纯化

线性化质粒的回收和纯化有苯酚氯仿纯化回收，使用一些公司生产的reaction clean up 的试剂盒，以及切胶回收，出于对产物纯度及后续实验稳定性的考虑，建议选择切胶回收来进行线性化质粒的回收和纯化。

本实验选用Promega的胶回收试剂盒：Promega Wizard® SV Gel and PCR Clean-Up System（Promega A9282）进行切胶回收，具体操作方法参见附录2。

切胶回收后，对产物——线性化的质粒进行质控：取2 μL胶回收产物进行琼脂糖凝胶电泳，确认回收片段的大小，并用Nanodrop OneC测定 其纯度；用Qubit荧光计测定回收线形质粒的准确浓度。

### 2.4 体外转录实验

1）解冻冷冻的试剂：在冰上解冻10X Reaction Buffer和4种核糖核苷酸溶液（ATP，CTP，GTP和UTP），直到完全溶解，并适当涡旋混匀。解冻后，将核糖核苷酸溶液置于冰上，10X Reaction Buffer在室温条件下待用；

2）计算反应体系：本实验选用的体外转录试剂盒为：MEGAscript™ T7 Transcription Kit（Thermo Invitrogen AM1333）

标准反应体系为20 μL：

表2. 体外转录实验的标准反应体系

|  |  |
| --- | --- |
|  | 体积（μL） |
| Nuclease-free water | 6（总体系20 μL，水的体积可按需调整） |
| ATP Solution | 2 |
| CTP Solution | 2 |
| GTP Solution | 2 |
| UTP Solution | 2 |
| 10X Reaction Buffer | 2 |
| 线性化的DNA模板 | 2（约1 μg） |
| Enzyme Mix | 2 |

请根据Qubit测出线性化DNA模板的准确浓度计算添加体积和Nuclease-free water的体积。

3）室温下在0.2 mL PCR管中配制反应体系，按照表2从上到下顺序添加试剂；

***注意***：如果在冰上配制反应体系，10X Reaction Buffer中的亚精胺可能会与模板DNA发生共沉淀；在加入水和核糖核苷酸溶液后，才能添加10X Reaction Buffer；标准单个反应体系为20 μL，如果需要更多的产物，可以同时进行多个20 μL 的反应。

4）轻弹离心管或用移液枪轻轻吹打以使反应体系彻底混匀，再用低速离心机离心3-5秒，将反应混合物收集到离心管底部；

5）在37°C下孵育2-4小时；

6）降解模板DNA：向每一个标准体系（20 μL）中加入1 μL TURBO DNase，充分混匀，继续在37 °C下孵育15 min；

7）将台式高速冷冻离心机预冷到4 °C，进行下一步的实验操作：2.5 RNA内标的回收、纯化和质检。

### 2.5 RNA内标的回收、纯化和质检

2.5.1 RNA内标的回收和纯化

1）向每个混合体系（21 μL）中添加179 μL 无核酸酶纯水，再加入等体积（200 μL）苯酚/氯仿溶液（柠檬酸盐饱和苯酚和氯仿1:1混合溶液），将混合物涡旋1分钟，然后在4 °C下以17000 × g离心2分钟；

***注意***：如果在步骤2.4中进行了多个反应，即可将反应混合，并加入相应体积的无核酸酶纯水至200 μL。

2）将上层水相小心转移至干净的1.5 mL离心管中（约190 μL），并加入等体积的苯酚/氯仿溶液，将混合物涡旋1分钟，然后在4 °C下以17000 × g 离心2分钟；

3）将上层水相小心转移至干净的1.5 mL离心管中（约180 μL），加入0.1倍体积预冷的3 M乙酸钠（约18 μL）和0.9倍体积预冷的异丙醇（约162 μL），将混合物涡旋1分钟后置于-20 °C冰箱30分钟，然后在4 °C下以17000 × g离心10分钟；

4）小心丢弃上清液，并用200 μL -20 °C预冷的70％乙醇洗涤沉淀；

5）在4 °C 17000 × g下离心5分钟，小心除去上清液，注意在此过程中不要将RNA沉淀吹打上来；

6）风干沉淀物，直到没有乙醇残留，将干燥的沉淀物重悬于50μL 无核酸酶纯水中。

2.5.2 RNA内标的质检：

1）取1μL RNA产物进行水平琼脂糖凝胶电泳（0.7% Agarose），确认片段的大小；

2）用Nanodrop OneC测定 纯度；

3）用Qubit荧光计测定准确浓度。

*＃可选＃* 取1μL RNA 产物，用Fragment analyzer检测RNA长度分布情况。

2.5.3 RNA内标的分装与保存：

将获得的RNA内标分装并保存于-80 °C冰箱。

## 3. DNA内标（DNA Internal Standard, DIS）的制作

上述用于RNA内标制作的质粒同样可以作为定量宏基因组的内标使用。

### 3.1 内标序列的扩增

选用高保真的PCR酶对内标片段进行扩增，以NEB Q5® High-Fidelity 2X Master Mix为例，按照表4配制反应体系。整个PCR反应体系的配制应在冰上进行，配制完成之后充分混匀所有组分，然后将反应体系迅速转移至预热至变性温度（98 °C）的PCR中。

表4. DNA内标PCR扩增体系

|  |  |  |
| --- | --- | --- |
| 试剂 | 使用量（μL） | 终浓度 |
| Nuclease-free Water | 18（总体系50 μL，水的体积可按需调整） |  |
| Q5 High-Fidelity 2X Master Mix | 25 | 1X |
| Forward Primer | 2.5 | 0.2 μM |
| Reverse Primer | 2.5 | 0.2 μM |
| Template (Plasmid) | 2 | 1 pg - 10 ng |

表5. DNA内标PCR扩增程序（以Q5 High-Fidelity 2X Master Mix 为例）

|  |  |  |
| --- | --- | --- |
| 温度 | 时间 | 循环数 |
| 98 °C | 30 sec. |  |
| 98 °C | 5-10 sec. |  |
| 50-72 °C\*1 | 10-30 sec. | 35 cycles |
| 72 °C | 20-30 sec / kb |  |
| 72 °C | 2 min |  |
| 4-10 °C | ∞ |  |

***注意*：**\*1退火温度建议使用NEB官方提供的NEB Tm Calculator 进行计算（https://tmcalculator.neb.com/#!/main）。

### 3.2 内标序列的回收和纯化

PCR扩增结束后通过电泳鉴定产物，对与内标序列长度一致的片段进行切胶回收。本文的具体操作步骤可参见附录2。

## 4. 内标的定量与添加

### 4.1 内标物的浓度测定

每次实验开始之前都需要用Qubit对内标进行重新定量，测定其准确的浓度值（ng/μL），以便计算计算每个实验批次样品中所添加内标的绝对量（拷贝数）。

### 4.2 计算内标拷贝数

4.2.1 RNA内标的拷贝数计算公式：

：RNA内标的拷贝数

：内标的准确浓度（ng/μL）

：体积（μL）

：阿伏伽德罗常数，

：RNA内标的摩尔质量

：指RNA内标中对应的核苷酸数目

\*注：分子量额外加上“159”是为了将5’端三磷酸基团计算在内

4.2.2 DNA内标的拷贝数计算公式：

：DNA内标的拷贝数

：内标的准确浓度（ng/μL）

：体积（μL）

：阿伏伽德罗常数，

：DNA内标的摩尔质量（双链）

：指在内标对应的某一条多核苷酸单链中对应的核苷酸数目

### 4.3 样品的绝对定量信息

对于每一份拟开展定量宏基因组或宏转录组分析的样品，需要记录或测定反映原始样品的绝对定量信息：

1）样品体积或质量：对于水样，需要记录用于核酸提取的膜过滤水的总体积（L）。针对土壤等样本则需要记录其用于核酸提取的质量（g）。

2）样品生物量估测：对于环境样品的生物质总量，这里用来表示。例如：针对固定体积（L）的活性污泥和沉积物样本，其生物量可近似基于挥发性悬浮固体浓度（Volatile suspended solids，VSS；mg/L）来间接衡量，具体测定方法见附录3。

3）样品细胞数的测定：根据不同实验的定量需求和生物量相对较低的水样（例如自然水体），还可采用流式细胞仪对样品中的微生物的总细胞数进行定量。

### 4.4 内标的添加

1）内标添加量：内标添加量需要使其足以在序列数据集中进行有效的定量，但太高会浪费测序的通量和成本。实际实验操作时可以基于对同类型样品的经验，根据预期的核酸回收率估算出内标的添加量，根据我们的经验（Ju et al., 2018）和经典文献（ Satinsky et al., 2013）中的报道，添加最终核酸预期产量的0.5%的内标，最后获得的表征内标的reads数约占总reads数的0.1%～5%，这一数值受内标的添加时间影响 。

2）内标添加的时间：我们建议在提取环境样本的DNA和RNA时，在细胞裂解后立即加入已知拷贝数的DNA或RNA内标，然后继续进行提取。这个添加时间点考虑了破胞后DNA和RNA提取、建库过程中造成的核酸损失，可以用来计算基于样品体积、质量或生物量的基因或转录本绝对拷贝数，。

***注意：***在使用此内标添加方法时需要在正式实验之前进行预实验，预估核酸产量以确定内标的适宜添加量。

***备注：***以下另外2种内标添加方案在文献中均有报道，仅供用户参考与比较：1）选择在样品裂解之前加入内标（ Satinsky et al., 2013），可以用于评估物理破胞所造成的损失，但其存在的缺陷是内标序列不同于环境微生物，没有细胞结构包被，其受到的物理冲击可能是大于环境样品中的细菌基因组，给下游建库和目标片段回收过程带来不确定性影响；2）在DNA或RNA提取结束后再加入内标（Ju et al., 2018），此方法的优点是可以根据提取完成后DNA和RNA的实际产量，精准加入适量的内标，但此方法仅能反映出文库构建构成中的损失，无法反映核酸在提取过程的损失情况。

## 5. 文库构建与测序

使用双端测序（2 x 150）的测序策略在Illumina的Hiseq4000平台上对构建的DNA和cDNA文库进行测序（实际测序策略和测序平台的选择并不局限于本文示例，可根据实际实验需求对添加完内标的DNA和cDNA文库选择建库与测序的策略）。

## 6. 内标的回收与基础生物信息学分析

### 6.1数据前处理

### 6.1.1 定量宏转录组数据前处理

1）质控

用PRINSEQ对宏转录组的原始序列进行过滤，除去adapter，除去包含模糊碱基（ambiguous bases）数量>10％的序列（reads），除去平均质量值小于20的序列（reads），获得clean reads。

2）去除内标序列

用BlastN将宏转录组序列（reads）比对到内标序列，将比对得分（bit-score）高于50分的宏转录组reads计作属于内标序列的reads，记录属于内标序列的总reads数。然后在样品当中将这些reads全部剔除。使用Ribopicker （网址：http://ribopicker.sourceforge.net/）比对slr，ssr，rrnadb，humanrna，rfam5s和rfam58s等参考数据库，并使用其默认参数去除宏转录本数据中的rRNA序列，保留的所有clean reads作为有效cDNA reads，再用于后续的转录组数据分析。

### 6.1.2 定量宏基因组数据前处理

1）质控

用PRINSEQ对每个宏基因组的原始读数进行过滤，除去adapter，除去包含模糊碱基（ambiguous bases）数量>10％的reads，除去平均质量值小于20的reads，以获得clean reads。

2）去除内标序列

用BlastN比对内标和宏基因组数据，将比对得分高于50分的宏基因组reads全部计作属于内标序列的reads，记录属于内标序列的总reads数。然后在样品当中将这些reads全部剔除，保留的所有clean read作为有效宏基因组 reads，进行后续的数据分析。

## 6.2 绝对转录本量的计算

针对宏转录组中某一特定转录本的绝对转录本量的计算：

：Transcript copies per gram of biomass as VSS，每克生物量（以VSS计）的转录本拷贝数

：Transcript copies per liter，每升样品中的转录本拷贝数

：Transcript copies per gram of sample mass，每克样品中的转录本拷贝数

：Transcript copies per cell，每个细胞中的转录本拷贝数

：对应的样品中所添加的RNA内标的拷贝数

：对应的样品中总的生物量（以VSS计），单位g

：mapping到目标基因转录本上的reads数\*

：从对应样本宏转录组中回收到的属于RNA内标的reads数\*

：被mapping到目标基因转录本序列的长度，单位bp

：RNA内标序列长度，单位bp

：样品中的挥发性悬浮固体，单位mg/L

：样品的体积，单位L

：样品的质量，单位g

：样品中的细胞数量，单位个，该值可以通过流式细胞仪测定

\*注：通常质控后获得的reads长度一致，若不一致可相应采用目标基因和内标序列的覆盖度分别取代公式中的 *Ngene reads*/*Lgene*和*NRIS reads*/*LRIS*值即可。

## 6.3 绝对基因量的计算

针对宏基因组中某一特定序列的绝对拷贝数的计算：

：Gene copies per gram of biomass as VSS，每克生物量（以VSS计）的基因拷贝数

：Gene copies per litre，每升样品中的基因拷贝数

：Gene copies per gram of sample mass，每克样品中的基因拷贝数

：Gene copies per cell，每个细胞中的基因拷贝数

：对应的样品中所添加的DNA内标的拷贝数

：对应的样品中总的生物量（以VSS计），单位g

：mapping到目标基因上的reads数\*

：从对应样本中回收到的属于DNA内标的reads数\*

：目标基因序列的长度

：DNA内标序列长度，单位bp

：样品中的挥发性悬浮固体，单位mg/L

：样品的体积，单位L

：样品的质量，单位g

：样品中的细胞数量，单位个，该值可以通过流式细胞仪测定

\*注：通常质控后获得的reads长度一致，若不一致可相应采用目标基因和内标序列的覆盖度分别取代公式中的 *Ngene reads*/*Lgene*和*NRIS reads*/*LRIS*值即可。

# 失败经验

1. 使用外转录试剂盒为：MEGAscript™ T7 Transcription Kit（Thermo Invitrogen, 货号：AM1333）时应在室温下配制反应体系，不能在冰上进行，在冰上配制反应体系温度过低会导致10X Reaction Buffer中的亚精胺可能会与模板DNA发生共沉淀。
2. 在使用苯酚/氯仿溶液对RNA内标进行回收时，样品与试剂混合再离心后，水相分布在上层，有机相分布在下层，吸取上层水相时应尽量小心避免吸到下层的有机相。
3. 每次使用RNA内标之前都需要对其重新进行定量，若重新定量时RNA内标的浓度低于初始浓度的90%，就不再建议使用，建议重新进行体外转录实验，获取新鲜的RNA内标。

# 结果与分析

## RNA内标的质检结果

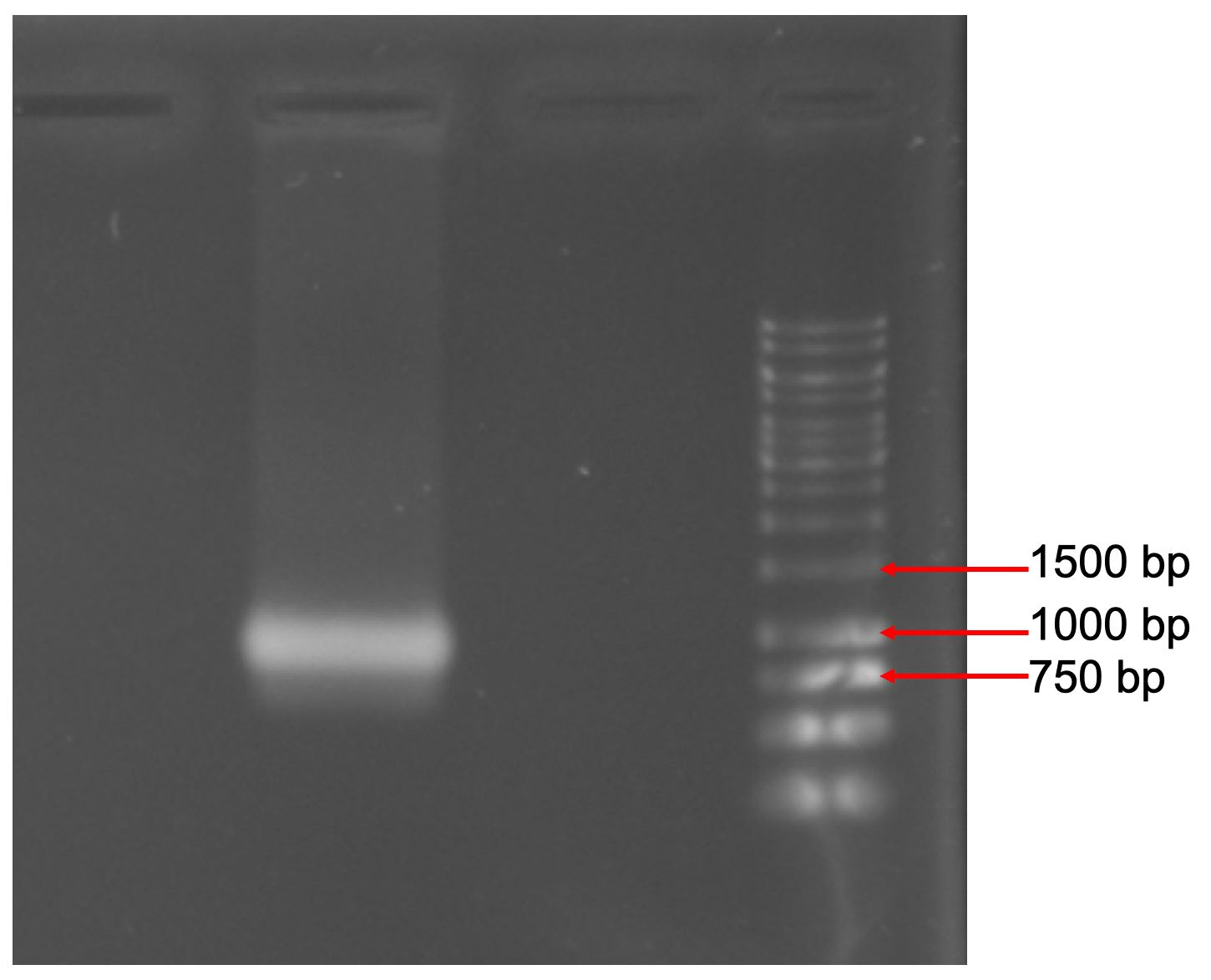


图1. RNA内标的水平琼脂糖凝胶（0.7%）电泳结果图

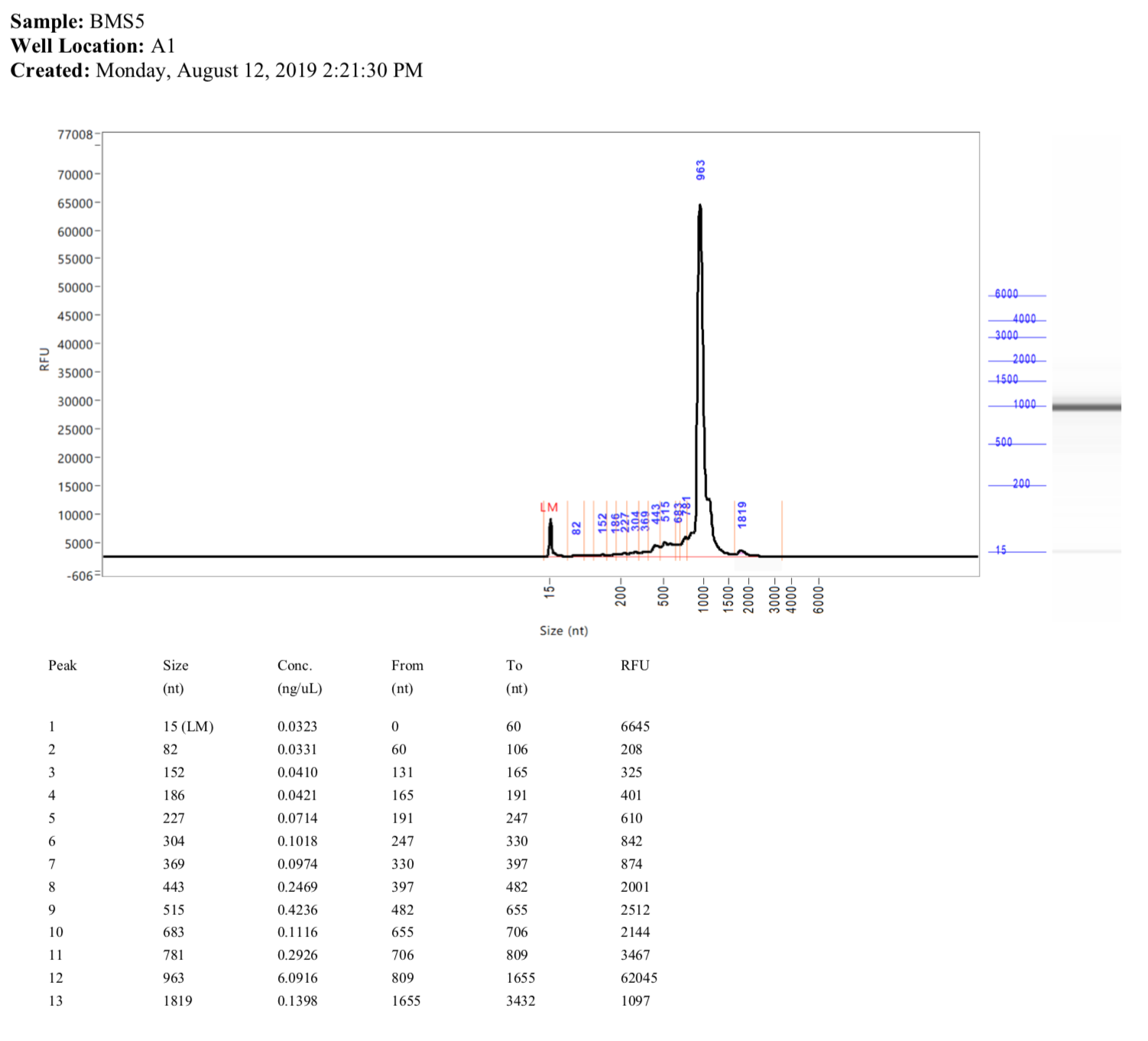


图2. RNA内标的Fragment analyzer结果图

\*Fragment analyzer的实验结果可能会与内标的真实长度有部分偏差（<10%）

# 附录

## 附录1. 内标序列设计的基本原则和注意事项

本流程通过人工合成的DNA和RNA内标物投加来实现微生物群落的定量宏基因组学和定量宏转录组学分析，内标序列在设计时需要满足以下几个条件：

1）该序列为受测环境样品或自然环境中不存在的序列；提交到NCBI Blast比对nr数据库时，没有显著相似性参考序列被比对上；也没有任何bit-score得分大于50分的比对结果，以保证所设计内标序列的独特性和有效性。

2）细菌和古细菌基因的平均大小约为924 bp，考虑到不同的转录本长度在RNA的提取、文库构建、测序过程中的差异性，即转录本长度对最终测序数据回收率的影响，该内标序列的设计长度建议在1000个碱基左右。GC 含量接近50%，不存在一些高GC，或发卡结构的区域。

***注意***：如果研究的重点是小RNA或短转录本，则可以缩小内标的长度；高GC的核酸片段更容易形成二级结构，在常规建库过程中难以被扩增，难以被打断，其最终的测序丰度信息往往会受此影响而偏低。

3）在设计RNA内标时需要在序列的起始端添加T7 promoter的序列（5’ -TAATACGACTCACTATAGG），以保证合成的质粒能够进行体外转录实验。

4）在设计RNA内标时，需要在内标序列的末端嵌入一个酶切后产生平末端的限制性核酸内切酶酶切位点，用于质粒的线性化，且该酶切位点在内标序列中的其他位置不存在。

5）当实验计划用PCR对质粒上人工合成的内标序列进行扩增，将扩增产物作为定量宏基因组实验中使用的DNA内标时，在引物设计过程中需要注意不要将T7启动子序列包含进去，避免影响后续的内标序列回收。

6）内标参考序列BMS5 （ Satinsky et al., 2013）：

TAATACGACTCACTATAGGGTTCGGTGGTCTATACTACTACCTAAGTTGGATGTACTGGTGGAAGTGCTACCAACACAGTAATGCTGGTATAGGTAGGCAACACTACGACTTCAGGAAGAGTCTAACGAATGTATGCATAATACTAATGCCTTACATGTGGAAGCACCCTATAACGGACAGGATGAGGCACAGGCACATGTGCAGGCAAGCTTTCAAGTGGATGTGCGGTGAAAATAAGTTCTGGCTAGTAAGGGCTATGGAAAATCAACCTGACGAAAGGATACTAGCTCAAATGACGATAACGGACAGTGACTGGCAACCTGAAGAATGGTACAAGAAGAGGCACGACCCTGGTGAAAATGACGTAATAAGGTGCGTATACATAGGTATAGAAAATCTAGTAAATGCTACGTGCCCTGACATGGAAGACTACTACGCTATGACGGGTAATAAGCCTCTACTAGAACTAAATAGTATAGGTCCTTGCACGCAATGCACGGTACACAAGCTAGAAGGTGTACACTGCATATGGTGGATAGTAAGGAGGGACCACTTCCCTGTACCTATAATACAAATAGTAGACGTATTCAATCTATACAATTTCGCTAGTGGTACGGTACTATGCATACAACACGCTGCTCACCCTTGGGGTGACTGGATGTTCGACGTACAATACGAAAGTTGCAGGATGTACAGGTGGTGGATGACGAGGAATGACTGGAGTGGTCCTAATAAGTGGAGTGGTGCTCACAGTATATGCCAACCTCACTGCTGCAGGAGTGACGACAGTAGGGTAAGTAAGAGGATGGCTACGGTAACGAAGGAAGTAGTAGAAATGAGTCACATGGACCTAAAGAGGAGTGCTTACTGCAATAGGACGCAACTAGAAGAATACGACGCTTTCTACACGAGGTGGAAGTTCGTACCTTGGATGTACCCTGCTCCTCTACCTTGCGAAGTACAAGACTTCGTAACGAGGAGGACGTTCGACTACCCTGACCCTACGGCTTGCGGTCTAGTTTAAA

如图3所示，其中“TAATACGACTCACTATAGG”为合成时加入的T7启动子序列，作为体外转录实验中RNA聚合酶的结合位点，不是实际的内标序列；序列最末尾的“TTTAAA”是设计过程中加入的一个用于线性化质粒的平末端酶切位点，酶切后TTT即为内标序列的最后三个碱基；综上所述，该内标设计合成的序列总长度为1026 bp，实际内标序列长度为1004 bp。

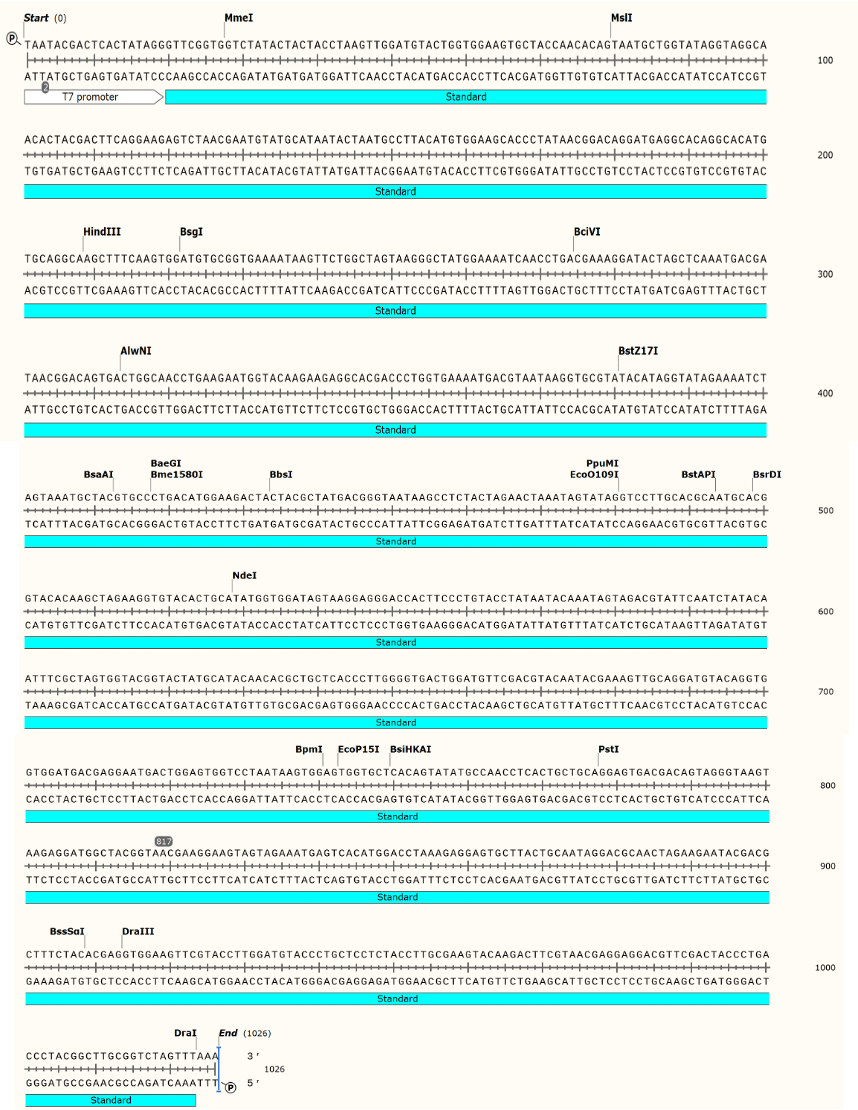


图3. 内标设计合成的结构示意图

## 附录2. 胶回收的具体实验步骤

Promega Wizard® SV Gel and PCR Clean-Up System：

1）电泳后，从琼脂糖凝胶上切下目的条带（完全线性化的质粒条带），并将凝胶放入1.5 mL离心管中；

2）每10 mg胶块添加10 μL Membrane Binding Solution，并在50–65°C下孵育（金属浴仪或烘箱皆可），直到凝胶片完全溶解，期间需每隔3-5分钟~~中~~颠倒混匀；

3）将SV Minicolumn插入收集管中（Collection Tube），并做好样品名称的标记；

4）将溶解的凝胶混合物冷却到室温，然后转移到Minicolumn管中，在室温下孵育1分钟；

5）以16,000 × g离心1分钟，丢弃收集管中的废液，然后将Minicolumn重新插入收集管中；

6）添加700 μL的Membrane Wash Solution（确保Membrane Wash Solution使用之前已经按照说明书的要求加入乙醇），以16,000 × g离心1分钟，丢弃收集管中的废液，然后将Minicolumn重新插入收集管中；

7）添加500μL的Membrane Wash Solution，以16,000 × g离心5分钟，丢弃收集管中的废液，然后将Minicolumn重新插入收集管中；

8）重新以16,000 × g离心1分钟，以去除任何可能残留的乙醇；

9）小心地将Minicolumn转移到干净的1.5 mL离心管中；

10）向Minicolumn正中部加入20 μL Nuclease-free water。在室温下孵育1分钟，以16,000 × g离心1分钟；

11）丢弃Minicolumn，并将样品保存在-20 °C。

该实验方法参考：Promega公司 Promega Wizard® SV Gel and PCR Clean-Up System 试剂盒说明书。

## 附录3. VSS测定方法

实验材料：无菌超纯水，玻璃纤维滤盘，坩锅，烘箱，马弗炉，干燥器

实验步骤：

1）玻璃纤维滤盘的准备：组装过滤装置，打开真空泵，用20mL的无菌超纯水冲洗滤膜和滤杯，重复三次，确认滤膜和滤杯上的所有水都抽干后关闭真空泵，取下玻璃纤维滤盘，将其置于坩锅中，将坩锅和玻璃纤维滤盘一同放入马弗炉中550°C灼烧15min，灼烧完成适当冷却后取出置于干燥器内冷却至室温，称重。重复灼烧、干燥冷却和称重，直到获得恒重或重量变化小于4%或0.5mg（以较小值为准），记录玻璃纤维滤盘的自重，单位mg。将玻璃纤维滤盘储存在干燥器中备用。

***注意：***在组装过滤装置时，玻璃纤维滤盘带褶皱的一面应当朝上。

2）选择过滤器和样品：样品体积需满足能够产生2.5-200mg的干燥物，如果样品的过滤体积不能满足最低产量的要求，则将样品体积增加至1L；如果单个样品的过滤时间需要10min以上，则需要考虑增加过滤器的直径或减小样品体积，记录样品过滤体积，单位mL。

3）过滤：组装过滤器开始过滤，加入样品之前先向过滤系统中添加少量的无菌超纯水湿润过滤系统，确保待过滤样品充分混匀后正式开始过滤，样品过滤完成后，先不关闭真空泵。继续用10mL的无菌超纯水洗涤过滤器，并完全排水，再重复清洗过程两次，在清洗过滤完成后，继续抽吸3 min。

***注意：***对于固体含量特别高的样品可能需要额外的洗涤过程。

4）烘干称重：小心地从过滤器上取下玻璃纤维滤盘并转移到坩锅中，连同坩锅一起置于烘箱中在103-105°C下干燥至少1h，适当降温后取出置于干燥器内冷却至室温，称重。重复高温干燥、冷却和称重”直到获得恒重或重量变化小于4%或0.5mg（以较小值为准）。记录下此时质量，单位mg；表示烘干后玻璃纤维滤盘的自重和样品中总悬浮固体（Total Suspended Solids，TSS）的质量之和，即。

5）将滤盘重新置于坩锅中，再放入马弗炉中550°C灼烧15min\*，灼烧完成适当冷却后取出置于干燥器内冷却至室温，称重。重复灼烧、干燥冷却和称重直到获得恒重或重量变化小于先前称量的4%或0.5mg（以较小值为准）。记录下此时称重得到的质量，单位mg；表示马弗炉中550°C灼烧后玻璃纤维滤盘的自重和样品中灰分（Fixed Solids）的质量之和，即

注：\*通常200 mg残留物需要灼烧15-20min，但是一个以上的样品或样品中残留物质量大于200mg会加重马弗炉的负担，因此可能需要更长的灼烧时间。

***注意：***所有实验样品至少一式两份，重复测定的数据偏差应该小于平均值的5%。

6）计算

因为：

: 样品中的挥发性悬浮固体的质量，单位mg

: 过滤到玻璃纤维滤膜上的样品体积，单位mL

该测试方法参考：Association, A. P. H. (1998). [Standard methods for the examination of water and wastewater. 20th edition. American Public Health Association, American Water Works Association and Water Pollution Control Federation](https://www.standardmethods.org/)。

**致谢**

本文工作得到了科学技术部国家重点研发计划的资助（项目编号：2018YFE0110500）。内标物设计序列方案和合成方法参考Mary Ann Moran课题组发表的论文《Use of internal standards for quantitative metatranscriptome and metagenome analysis》；内标物的添加方法与生物信息学定量计算方法参考作者鞠峰博士发表的论文《Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes》。

**参考文献**

* + - 1. Ju, F., Beck, K., Yin, X., Maccagnan, A., McArdell, C.S., Singer, H.P., Johnson, D.R., and Zhang, T., Buergmann, H. (2018). [Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes](https://www.nature.com/articles/s41396-018-0277-8). *The ISME Journal* 13, 346–360
      2. Satinsky, B.M., Gifford, S.M., Crump, B.C., and Moran, M.A. (2013). [Use of internal standards for quantitative metatranscriptome and metagenome analysis](https://www.sciencedirect.com/science/article/pii/B9780124078635000125). *Methods in Enzymology* 531, 237-250