

利用CRISPR/Cas9系统减少基于16S rRNA基因扩增子测序中大量的宿主污染

发表日期：2020-06-03

第一作者：Luyang Song(宋露洋)

通讯作者：KabinXie(谢卡斌，kabinxie@mail.hzau.edu.cn)

单位：作物遗传改良国家重点实验室和湖北省植物病理学重点实验室

摘要Abstract

高通量细菌16S rRNA基因测序(16S-seq)是研究细菌群落结构的常用方法。然而，线粒体和质体中16S rRNA基因的污染阻碍了细菌16S-seq在植物微生物群体分析中的灵活性，特别是对于水稻等一些植物物种。到目前为止，在基于16S rRNA基因的扩增子测序中，能有效地减轻这种宿主污染而不会产生任何的偏好仍旧是一个挑战。

结果Results

我们开发了Cas-16S-seq方法来减轻在基于16S rRNA基因扩增子测序的微生物群体研究中的宿主污染。该方法利用Cas9核酸酶和特异性向导RNA（gRNA）在文库构建过程中剪切靶标16s rRNA，从而去除16S-seq中的宿主污染。以水稻为例，我们验证了Cas-16S-seq的可行性和有效性。我们也建立了一整套生物信息学流程来设计专门靶向水稻16s rRNA且不会脱靶到细菌的16S rRNA基因的gRNA. 我们也比较了Cas-16S-seq和常规的16S-seq的方法在人工混合16S rRNA基因群体、稻田土、水稻根和叶际样品中的效果。我们的结果表明，Cas-16S-seq显著降低了水稻16S rRNA基因序列在根系样品文库中的比例，宿主污染率从63.2%降低到2.9%，在叶片样品中，从99.4%到11.6%.因此，在分析植物关联微生物群体中Cas-16S-seq比常规16S-seq能检测到更多的细菌种类。更重要的是，在分析土壤样品时，Cas-16S-seq和常规16S-seq显示了几乎相同的细菌群落，这表明我们设计的具有宿主特异性gRNAs的Cas-16S-seq方法在水稻微生物群落分析中不会产生脱靶现象。

结论Conclusion

我们开发的Cas-16S-seq的方法可以有效地去除大量的宿主污染，并且不会对基于16S rRNA基因的扩增子测序产生任何的偏好，从而能以低成本和高灵活性实现更深入的细菌群体的分析研究。本研究为植物微生物组学研究提供了一个强有力的工具，也为更深入的优化微生物组学研究工具提供新思路

引言Background

微生物在动植物健康中发挥着重要作用。对植物微生物群的深入解将为调控植物生长、提高作物抗逆性、减少植物病害以及促进农业可持续发展提供新的策略。标记基因大量并行测序，如16S小亚基核糖体RNA基因(16S rRNA基因)，是一种独立的不依赖培养的微生物组研究方法。16S-seq采用通用引物扩增16S rRNA基因的一个或多个高变区(V1-V9)，从而推断出微生物群的组成和潜在功能。这些通用引物还靶向真核生物的线粒体和质体16S rRNA基因，因为这些基因与原核生物具有相同的进化来源。因此，在分析来自植物样本的微生物群落时，质体和线粒体序列占16S-seq数据的99%。宿主16S rRNA基因的污染大大限制了植物微生物组的研究。

PCR钳，最初设计用于在PCR期间抑制特定等位基因的扩增，从而能够显著减轻16S-seq中宿主16S rRNA基因的污染。PCR钳利用肽核酸(PNA)和锁定核酸(LNA)寡核苷酸来阻止PCR延伸或引物结合靶标DNA。少数宿主特异性PNA和LNA的寡核苷酸被设计用来特异性抑制质体和线粒体16S rRNA基因的扩增。然而，最近的一项研究发现，PNA寡核苷酸也阻止了某些细菌序列的扩增，从而在微生物群体分析中引入了明显的偏好。因此，迫切需要一种在16S-seq中高效且不引入任何偏好的特异性去除大量宿主16S rRNA基因的方法。

在本研究中，我们开发了一种Cas-16S-seq方法来减轻16S rRNA基因扩增子测序中的宿主污染。这种方法使用Cas9和特异性gRNA特异性切割宿主16S rRNA基因，从而在16S-seq扩增子文库制备过程中富集细菌序列(**图1**),以水稻为例,我们建立了一个生物信息学流程来设计特异性的gRNA,使其能够从数百万个原核16S rRNA基因中区分出水稻16S rRNA基因。我们利用人工混合的16S rRNA基因群体、稻田土壤、水稻根系和水稻叶片样品，将Cas-16S-seq与常规16S-seq方法进行比较，证明了Cas-16S-seq去除宿主16S rRNA基因的高效、特异，且不会产生任何的偏好性。

结果Results

Cas-16S-seq方法的工作流程

The workflow for Cas-16S-seq method

我们以水稻为例，利用16S-seq中的常用的四对通用引物包括515F-806R、27F-338R、799F-1193R和1114F-1392R，利用水稻DNA为模板进行PCR。我们发现水稻线粒体(515F-806R, 799F-1193R)和叶绿体(27F-338R, 515F-806R, 1114F-1392R)的16S rRNA基因都能被扩增出来(附加文件1: 图S2)。由于这些通用引物的退火位点位于16S rRNA基因最保守的区域，从而也能广泛的覆盖细菌序列。因此引物序列的任何改变都会影响其覆盖范围，并且也会产生显著的偏好性。为了解决这个问题，我们开发了Cas-16S-seq方法，该方法使用一种RNA介导的可编程核酸酶Cas9特异性去除16S-seq扩增子文库中宿主的污染。

Cas-16S-seq工作流程如(**图1**)所示。Cas-16S-seq采用与常用的16S-seq方法相同的两步PCR法。第一轮PCR使用带有合适接头的通用引物来扩增16S rRNA基因的可变区域。第二轮PCR扩增第一轮PCR的产物，引物中含有Illumina测序引物(P5和P7)和标签序列(i5和i7)。在Cas-16S-seq中，引入Cas9和宿主16S rRNA特异性gRNA来酶解第一轮PCR产物。gRNA是针对宿主16S rRNA基因设计的，不会靶向细菌序列(**图1**)。第二轮PCR无法扩增出被剪切的宿主16S rRNA片段，因此不会出现在最终的测序文库中(**图1**)。与常用16s-seq方法相比 Cas-16S-seq在第一轮和第二轮PCR步骤之间只额外添加了一步(Cas9/gRNA处理)，因此，它可以很容易地整合到目前基于16S rRNA基因的扩增子测序工作流程中。

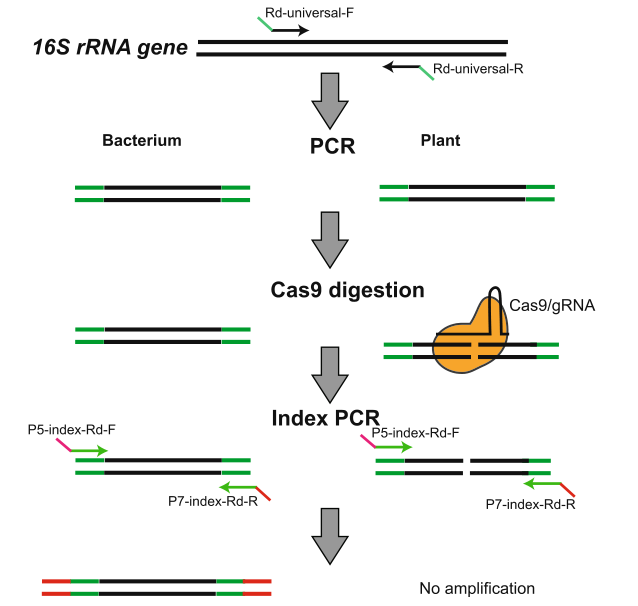


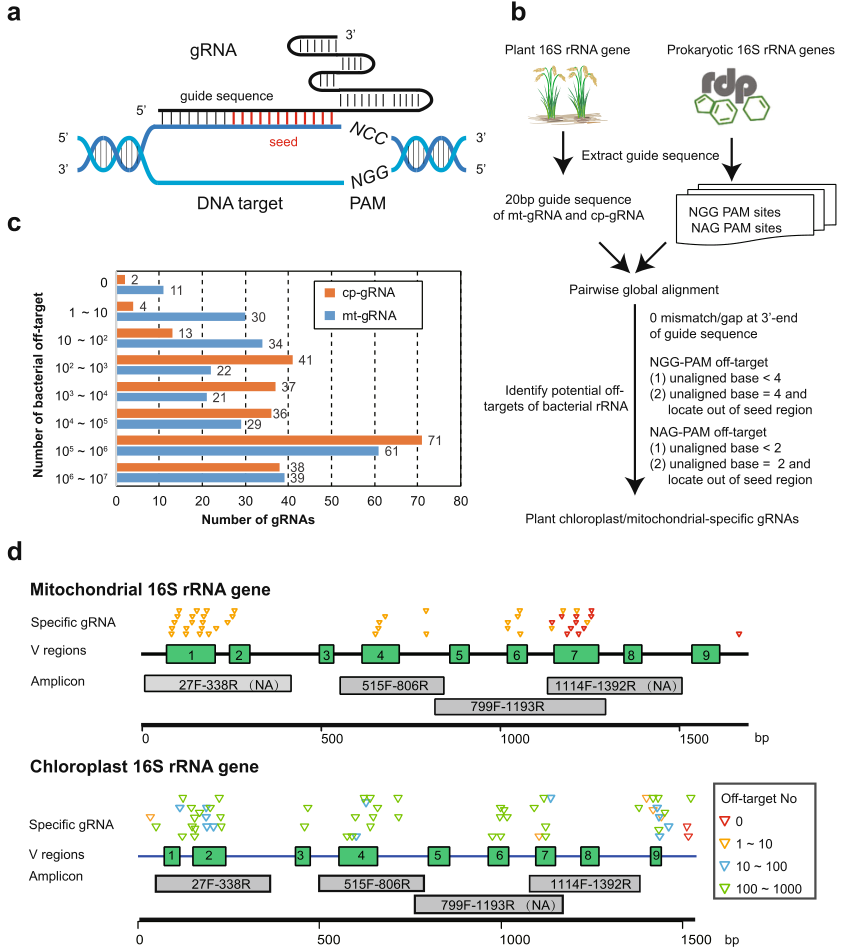
图1. Cas-16S-seq原理图. Cas-16S-seq是基于常规的16S-seq方法，使用两步PCR来扩增和标记16S rRNA基因。第一轮PCR扩增，使用带接头的通用引物(Rd-universal-F/R)扩增16S rRNA基因的可变区域。然后用Cas9和特异性gRNA剪切PCR产物中的宿主16S rRNA基因。第二轮的标记PCR过程中，使用含标签序列和Illumina测序仪接头序列的引物(P5-index-Rd-F和P7-index-Rd-R)扩增酶解产物，获得最终扩增子库。宿主的16S rRNA基因片段在第二轮PCR中不能进行扩增，从而在最后的16S-seq数据中被去除。

Fig. 1 Schematic diagram of Cas-16S-seq. Cas-16S-seq is based on the regular 16S-seq procedure using two-step PCR to amplify and index 16S rRNA genes. In the first PCR, universal primers with adaptors (Rd-universal-F/R) are used to amplify the variable regions of 16S rRNA genes. Then,Cas9 and specific gRNA is used to cleave the host 16S rRNA genes in the PCR product. In the second index PCR, primers containing index and Illumina sequencing adaptors (P5-index-Rd-F and P7-index-Rd-R) are used to amplify the digested products and obtain the final amplicon library.The host 16S rRNA gene fragments are not amplified in the second PCR and are therefore removed in the final 16S-seq data.

设计特异性靶向水稻16S rRNA基因的gRNAs

In silico design of gRNAs specifically targeting rice 16S rRNA genes

Cas-16S-seq方法的可行性主要依赖于能够区分宿主和细菌16S rRNA基因的高特异性gRNAs. 近年来Cas9介导的基因组编辑的大量研究揭示了CRISPR/ Cas9的靶向规律，因此，可以根据这些规则预测给定gRNA的潜在脱靶。为了设计靶向水稻16S rRNA基因特异性gRNA,我们在CRISPR-PLANT的使用方法上进行了修改，使其能够在全基因组范围内识别出潜在脱靶到原核生物16S rRNA基因的gRNAs.具体特异性gRNAs的筛选方法如（**图2b**）所示。我们共计筛选出243和247个分别靶向水稻叶绿体和线粒体16S rRNA基因的可用gRNAs，并且根据其脱靶到RDP-rRNA的数量进行了排序如（**图2c**）所示。具体特异性的gRNAs在水稻线粒体和叶绿体16S rRNA基因不同扩增子区域的分布如（**图2d**） 所示。



**图2.**设计特异性靶向宿主16S rRNA基因的gRNAs使其在水稻中运用Cas-16S-seq.**a** Cas9和gRNA靶向DNA剪切的原理示意图。20bp的DNA靶标序列与gRNA向导序列互补配对。前间隔序列邻近基序(PAM, 5′-NGG-3′)是Cas9的结合是必不可少的一部分，它紧邻互补配对区域.红色标注的区域为

Cas9结合容错度较低的区域。**b** 评估gRNA特异性的生物信息学分析流程图。靶向水稻叶绿体和线粒体16S rRNA基因的gRNA向导序列(mt-gRNA和cp-gRNA) 根据Cas9/gRNA结合的要求被提取(**图.2a**)。然后，这些向导序列与RDP数据库中的原核16S rRNA基因进行比对。尽管Cas9对NAG-PAM识别能力较弱，本文也同时考虑了RDP rRNA中NGG PAM和NAG PAM的脱靶位点。每个gRNA的RDP-rRNA脱靶数量都是根据流程图中所示的标准进行确定的(**详见“方法”部分**)。最后，我们根据每个gRNA的细菌脱靶总数对gRNA特异性进行了排序(见附加文件2:表S1和S2)。**C** 根据脱靶到RDP-rRNA的数量，确定不同特异性等级的mt-gRNAs和cp-gRNAs数量。**d** 大多数特异性mt-gRNA(< 10个脱靶)和cp-gRNA(< 1000个脱靶)的分布。绿框表示16S rRNA基因的高度可变区(V1-V9)。灰色方框表示四个16S rRNA扩增子的区域。三角形表示gRNA的位置，并根据脱靶到RDP-rRNA的数量进行着色。NA,表示不能被扩增

Fig. 2 In silico design of host 16S rRNA gene-specific gRNAs for Cas-16S-seq in rice. a The schematics of targeted DNA cleavage using Cas9 and gRNA. The 20 bp DNA target is paired with the gRNA guide sequence. A protospacer-adjacent motif (PAM, 5′-NGG-3′), which is indispensable for Cas9 binding, immediately follows the paired region. Red indicates the seed region that is less tolerance to mismatches for Cas9 binding. b Theflowchart of the bioinformatics analysis procedure to evaluate gRNA specificities. The guide sequences of gRNA that target rice mitochondrial and chloroplast 16S rRNA gene (mt-gRNA and cp-gRNA) were extracted according to the requirements for Cas9/gRNA binding (Fig. 2a). Then,these guide sequences were aligned to prokaryotic 16S rRNA genes in the RDP database. Both NGG PAM and NAG PAM off-target sites in RDP-rRNA were considered here, although Cas9 weakly recognized NAG-PAM. The RDP-rRNA off-targets of each gRNA were identified using thecriteria shown in the flowchart (see details in the “Methods” section). Finally, the total number of bacterial off-targets for each gRNA was used to rank the gRNA specificities (see Additional file 2: Table S1 and S2). c Number of mt-gRNAs and cp-gRNAs with different specificity ranks according to the number of RDP-rRNA off-targets. d Distribution of most specific mt-gRNA (< 10 off-targets) and cp-gRNA (< 1000 off-targets). The green boxes indicate hypervariable regions (V1–V9) of 16S rRNA gene. Gray boxes indicate the region of four 16S rRNA amplicons. The triangles indicate the gRNA position and are colored according to the number of RDP-rRNA off-targets. NA, not amplified

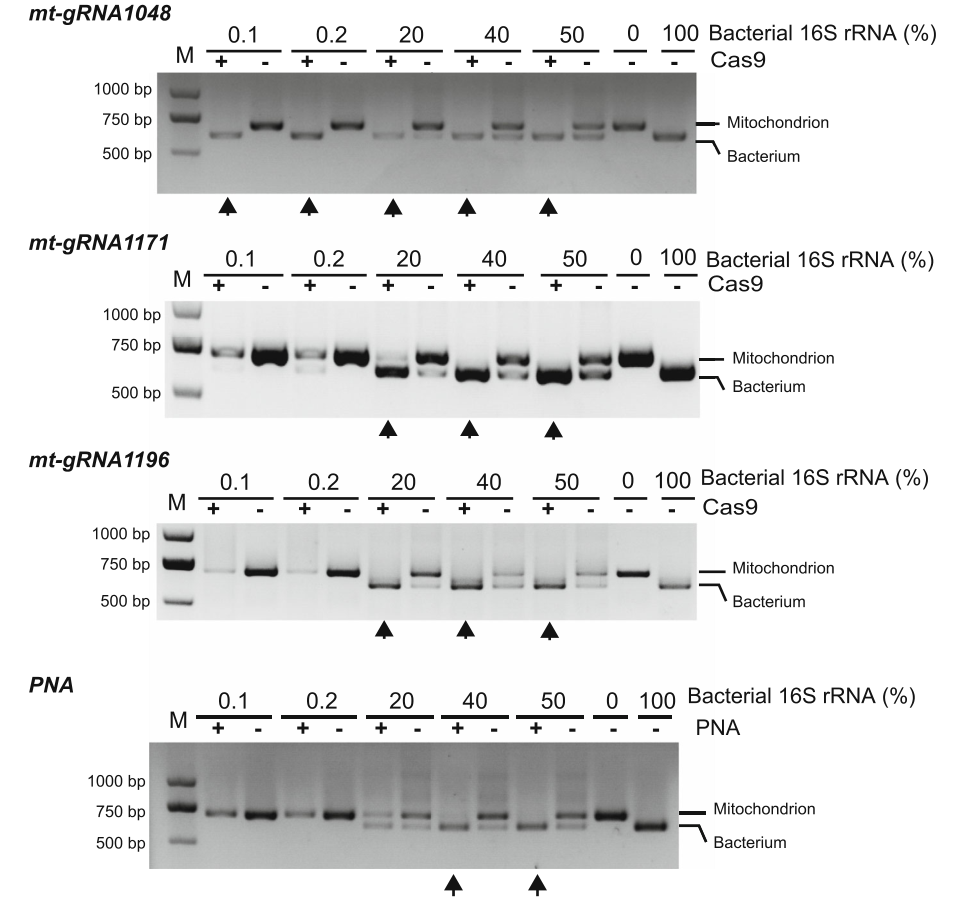
Cas9和mt-gRNA能有效消除人工构造样品中的宿主16S rRNA基因

Cas9 and mt-gRNA efficiently eliminate the host 16S rRNA

genes in mock samples

为了检测Cas9的剪切效率，我们利用纯化的水稻16S rRNA基因扩增子检测了12个mt-gRNAs和5个cp-gRNAs (附加文件1:图S3). 其中13个gRNAs在Cas9介导的水稻16S rRNA基因扩增子的剪切中显示出了90-100%的效率，4个gRNAs显示出了一般的剪切效率(65-84%)(附加文件1:图S3)。在5个具有100%剪切解效率的gRNAs中，我们选择mt-gRNA1048、mt-gRNA1171和mt-gRNA1196在Cas16S-seq中进行运用。这三种gRNAs都能靶向799F-1193R的扩增子片段，重要的是，该扩增子片段中不包含叶绿体16S rRNA基因。mt-gRNA1171和mt-gRNA1196不会脱靶到RDP-rRNA，而mt-gRNA1048根据我们的预测有10个RDP-rRNA的脱靶。(附加文件2:表S1).

然后，我们利用水稻线粒体和土壤细菌(NCBI accession number: AB658673) 799F-1193R扩增的DNA片段序列按不同比例混合制备的模拟群体，测试了Cas-16s-seq的有效性。由于细菌799F-1193R扩增子比水稻线粒体扩增子长度小84bp(附加文件1:图S1),因此PCR产物中线粒体16S rRNA的相对丰度可用凝胶电泳法测定。与常用的16S-seq方法相比，Cas-16S-seq的方法能大大减少扩增产物中的水稻16S rRNA的含量(**图3**)。我们还使用靶向水稻线粒体16S rRNA基因的PNA寡核苷酸检测了PCR钳法。并且将其与Cas-16S-seq的方法进行了比较(**图3**)。我们的结果表明，当细菌rRNA含量较低时，Cas-16S-seq比PNA PCR钳法更能有效消除水稻线粒体16S rRNA基因的污染(**图.3**)。



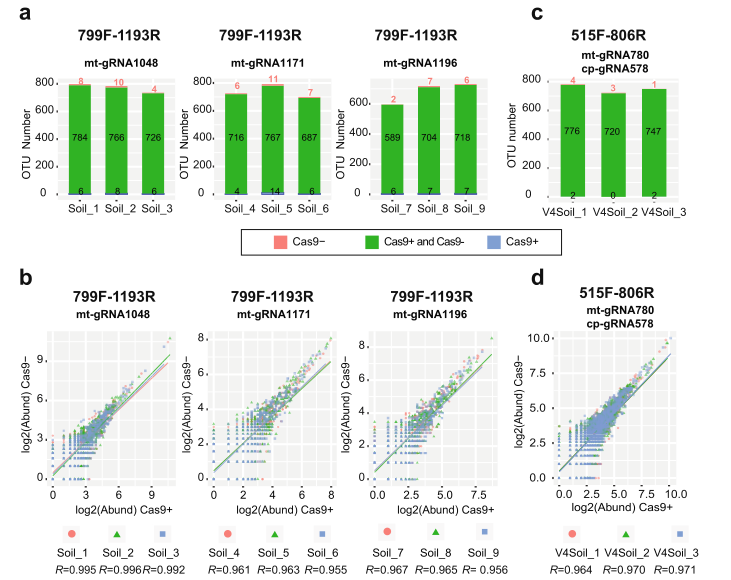
**图3.** Cas-16S-seq有效地去除了模拟群体中水稻线粒体16S rRNA基因片段。三种gRNAs (mt-gRNA1048、mtgRNA1171和mt-gRNA1196)被用来评估Cas-16S-seq的有效性。将含有0%、0.1%、0.2%、20%、40%、50%和100%细菌16S rRNA基因片段的模拟群体用Cas9/gRNA处理后进行PCR扩增。由于线粒体PCR产物比细菌产物大84 bp，因此可通过凝胶电泳来估算宿主16S rRNA基因的剪切效率。同样使用相同的模拟群体进行PNA PCR钳法实验。箭头表示能有效消除线粒体16S rRNA基因片段的处理组。M, DNA分子标记; bp, 碱基对

Fig. 3 Cas-16S-seq efficiently removed the rice mitochondrial 16S rRNA gene fractions in mock communities. Three gRNAs (mt-gRNA1048, mt-gRNA1171, and mt-gRNA1196) were used to evaluate the performance of Cas-16S-seq. Mock samples containing 0%, 0.1%, 0.2%, 20%, 40%, 50%,and 100% bacterial 16S rRNA gene fragments were treated with Cas9/gRNA and then amplified by PCR. The depletion efficiency of host 16S rRNA genes was estimated by gel electrophoresis since the mitochondrial PCR product is approximately 84 bp larger than that of bacteria. The PNA PCR clamping method was also performed using the same mock samples. Arrows indicate the treatments by which the mitochondrial 16S rRNA gene fraction was efficiently eliminated. M, DNA marker; bp, base pair

利用稻田土样品来评估Cas-16S-seq

Evaluation of Cas-16S-seq using paddy soil samples

由于我们想知道利用mt-gRNA1048、mt-gRNA1171和mt-gRNA1196的Cas-16S-seq方法在分析土壤细菌群落时是否引入了偏好。为此，我们使用Cas-16S-seq (Cas9+)和常规的16S-seq (Cas9-)制备扩增子库。将利用illumina双端250测序得到序列按照100%的相似度聚类成操作分类单元(OTUs)。将OTU表中的每个样品稀释抽样到16000条序列后，再进行多样性的分析(附加文件3:表S3)。Cas-16S-seq和16S-seq两种方法检测到的在所有样品中丰度大于> 0.02%的OTUs中，96.8–98.7%的OTUs在这两方法都能检测到(**图.4a**)。只有少数低丰度OTU仅被一种方法检测到，可能是由于PCR或测序过程中的人为操作造成的。并且这两种方法也没有发现细菌多样性的显著差异(Shannon和Simpson指数) (附加文件3:表S3). 重要的是，这些OTUs的丰度在Cas9+和Cas9−数据之间高度相关(皮尔森相关系数，*R* = 0.955-0.996，**图4b**)。并且我们也利用了DEseq2去检测了两种处理方法中OTU丰度的差异。在Cas-16S-seq数据中，只有一个OTU (OTU\_1)减少，即水稻线粒体16S rRNA基因(Wald检验，p < 0.05，附加文件3:表S4)。这并不奇怪，因为水稻土样品中有0.048-0.26%的水稻线粒体16S rRNA基因序列，这些序列在Cas-16S-seq中被去除。Cas-16S-seq和16S-seq在表征细菌结构的高度一致性表明Cas-16S-seq的方法对于土壤微生物群体的研究不会引入任何的偏好。



**图4.** 利用稻田土样品来评估Cas-16S-seq. 为了比较Cas-16S-seq (Cas9+)和常规的16S-seq (Cas9−)处理799F-1193R和515F-806R扩增子的区别，每个gRNA进行三次生物学复制。799F-1193R扩增子用mt-gRNA1048、mt-gRNA1171或mt-gRNA1196处理(**a,b**); 515F-806R扩增子被两个gRNAs (mt-gRNA780和cp-gRNA578处理, **c, d**)。OTU表被稀释抽样到16000条序列。**a, c** 检测到的OTUs(每个样品> 2e−4的平均相对丰度)在每个样品中Cas-16S-seq结果和常规的16S-seq结果高度一致。箱线图中两种方法都检测到的OTUs的数量(Cas9+和Cas9−，绿色标记)，只有Cas-16S-seq的检测到的(Cas9+,蓝色标记)，常规16S-seq检测到的(Cas9-,红色标记). **b, d** 在所有比较中，观察到的OTUs丰度在Cas-16S-seq (log2(Abund) Ca9+)和常规16S -seq (log2(Abund) Cas9−)数据之间相关性高度一致。OTU丰度进行log2转换。每个比较组见的相关系数(*R*)在底部显示。不同的生物学重复用不同颜色表示

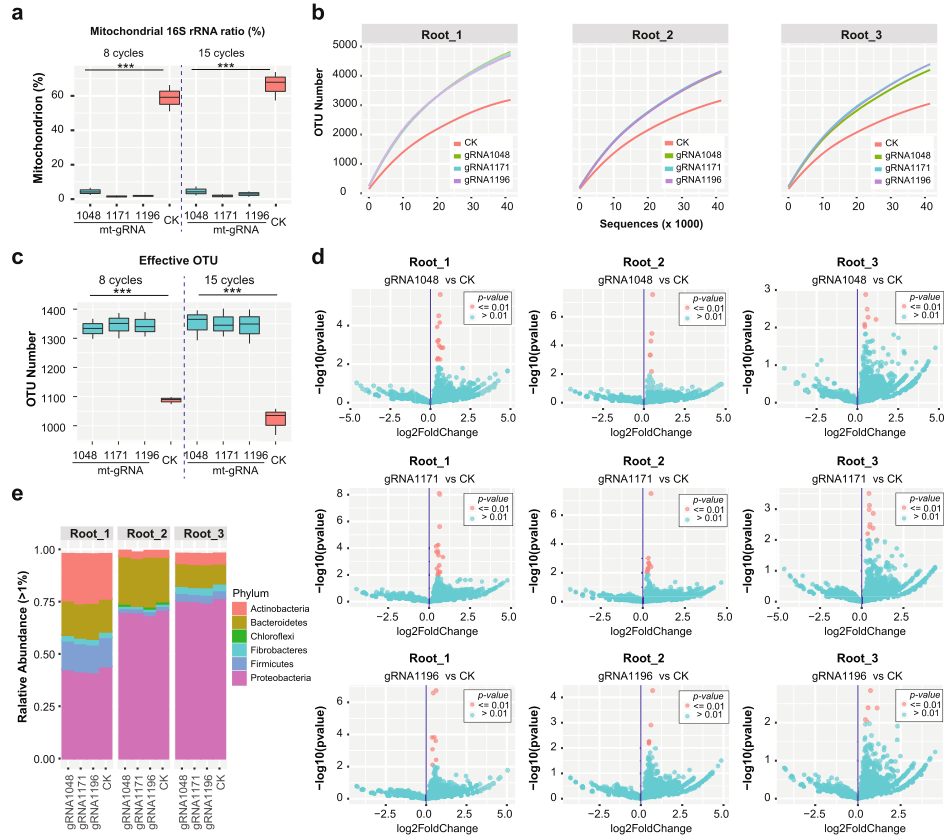
Fig. 4 Evaluation of Cas-16S-seq using paddy soil samples. Three biological replicates per gRNA were performed to compare the Cas-16S-seq(Cas9+) and regular 16S-seq (Cas9−) for 799F-1193R and 515F-806R amplicons. The 799F-1193R amplicons were digested using mt-gRNA1048,mt-gRNA1171, or mt-gRNA1196 (a, b); the 515F-806R amplicons were digested with two gRNAs (mt-gRNA780 and cp-gRNA578, c, d). The OTU tables were rarefied to 16,000 sequences. a, c The observed OTUs (average relative abundance per sample > 2e−4) were highly consistent between Cas-16S-seq and regular 16S-seq results for each sample. The boxplots show the number of OTUs detected by both methods (Cas9+ &Cas9−, colored in green), by Cas-16S-seq only (Cas9+, colored in blue) and by regular 16S-seq only (Cas9−, colored in red). b, d The abundances of observed OTUs are highly consistent between Cas-16S-seq (log2(Abund) Ca9+) and regular 16S-seq (log2(Abund) Cas9−) data in all comparisons. The OTU abundances were log2transformed. The correlation coefficients (R) for each comparison are shown at the bottom. The different biological replicates are indicated by color

Cas-16s-seq能有效地从根样品中去除宿主污染

Cas-16S-seq efficiently removes host contamination from

root samples

我们又探究了Cas9-16S-seq的方法去除水稻根系样品799F-1193R扩增子中大量宿主污染的性能，使用的水稻根系样品包含根围和根内两个部分。并且也对Cas-16S-seq和常规16S-seq的方法进行了比较。为排除PCR或测序过程中的人为操作造成的误差，使用不同的PCR循环数(8和15个循环)对每个处理进行了两次技术重复(Cas9+ vs Cas9−)。Cas-16S-seq扩增子中的水稻线粒体16S rRNA基因片段(2.9±1.5%)平均比16S-seq扩增子(63.2±8.4%)少21倍(**图5a**，附加文件4:表S5; ANOVA，F检验，*p* = 5e−16).并且三种gRNAs在去除线粒体16S rRNA基因片段的效率也有细微的差别(ANOVA, F 检验, *p*= 0.016)。综上所述，利用Cas-16S-seq方法能有效地去除了根系样品中大量的宿主污染。稀释曲线表明无论使用哪种mt-gRNA, Cas16S-seq都比16S-seq产生了更高的细菌丰富度**(图5b)**。且去除大量宿主污染后，Cas -16S-seq检测到的OTUs平均是16S-seq的1.3倍(ANOVA, F检验，*p* = 4.35 e−10，**图5c**)。因此，Cas-16S-seq中的细菌群落组成与16S-seq中存在显著差异(PERMANOVA, R2= 0 . 1 5 , p< 0.001 ; 附加文件1: 图S6)。为了确定与16S-seq相比，Cas -16S-seq中是否有细菌的16S rRNA基因被特异性去除，我们使用DESeq2分析了两种方法的差异OTUs。在Cas9/gRNA处理后，我们没有检测到相对丰度显著降低的细菌OTU(**图5d**和附加文件4:表S6)。说明我们没有在Cas-16S-seq中去除特定的细菌16S rRNA基因。Cas-16S-seq处理组中，包括*Burkholderiaceae*(10个OTUs)、*Fibrobacteraceae*(4个OTUs)、*Paludibacteraceae*(5个OTUs)、*Micromonosporaceae*(3个OTUs)的几个科等29个OTUs的丰度较16S-seq显著增加了1.2-2.2倍(**图5d**;附加文件4:表S 6, W al d 检验, *p* < 0.01)。我们还在门分类水平上比较了细菌群落组成，发现优势门的相对丰度在Cas-16S-seq和16S-seq数据上没有显著差异(**图5e**)。综上所述，Cas-16S-seq能有效的去除靶标宿主16S rRNA基因，并且富集了根样本中的细菌16S rRNA基因。



**图5.** Cas -16S-seq有效且特异性地减轻了水稻根微生物群体分析中的线粒体16S rRNA基因的的污染。每个mt-gRNA处理和对照分别进行3个生物重复(Root\_1-Root\_3)和2个技术重复(8个和15个PCR循环)。OTU表被稀释抽样到每个样本41500条序列。**a** Cas-16S-seq与常规16S-seq (CK)测序结果中水稻线粒体16S rRNA基因含量比较。用3个mt-gRNAs和Cas9处理后，宿主16S rRNA基因含量显著降低。\*\*\*ANOVA, *p* = 5 e−16. **b** Cas-16S-seq与常规16S-seq稀疏曲线分析。稀疏曲线分析使用一个技术重复(8个循环) 的数据结果进行在这里展示。**c** 与常规的16S-seq相比，Cas -16S-seq能检测到更多的OTUs， OTUs为每个样品 > 1的平均计数。\*\*\*ANOVA，*p* = 4.35e−10(另见附加文件4:表S6)。**d** Cas-16S-seq与常规16S-seq (CK)差异OTUs分析。在所有比较中(与常规的16S-seq相比，Cas9/gRNA处理后OTU的相对丰度均未显著降低)。相比之下，Cas -16S-seq数据中有29个OTUs显著增加(红点表示; Wald检验，*p* < 0.01;参见附加文件4:表S6)。e 在门分类水平上的细菌群体组成分析。图中显示了相对丰度为>1%的菌门

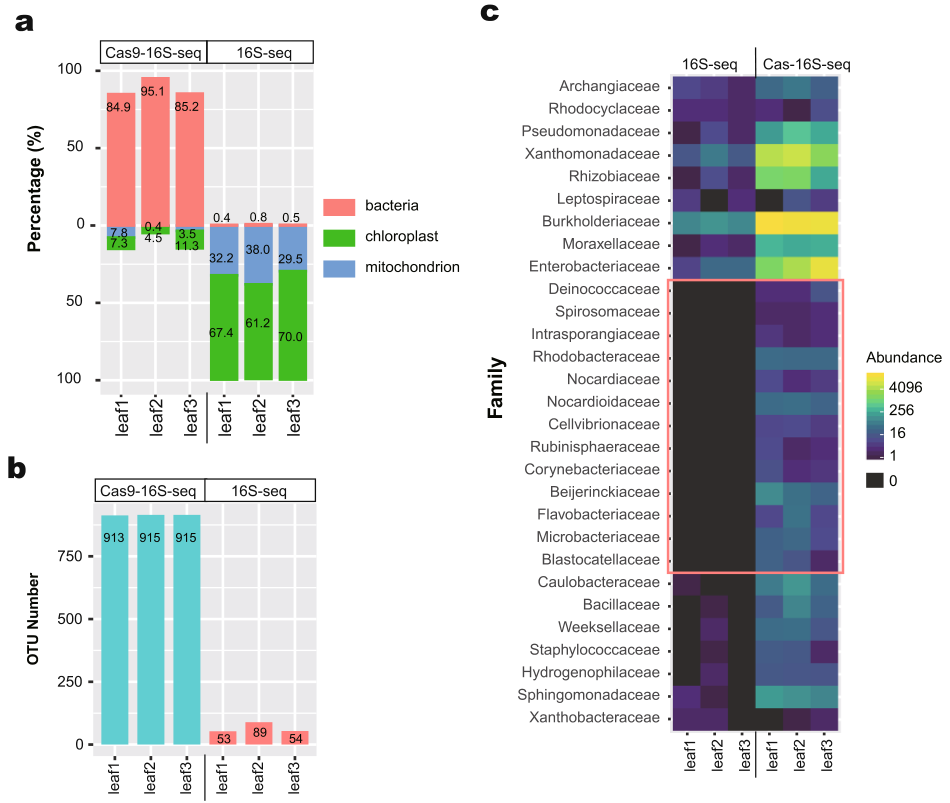
Fig. 5 Cas-16S-seq efficiently and specifically mitigated mitochondrial 16S rRNA gene fractions in rice root microbiota profiling. Three biological replicates (Root\_1–Root\_3) and two technique replicates (8 and 15 index PCR cycles) were performed for each mt-gRNA treatment and control.The OTU table was rarefied to 41,500 reads per sample. a Comparison of rice mitochondrial 16S rRNA gene contents in Cas-16S-seq and regular 16S-seq (CK) sequencing results. The host 16S rRNA gene was drastically reduced after Cas9 treatment with 3 mt-gRNAs. \*\*\*ANOVA, p = 5 e−16. b Rarefaction curve of Cas-16S-seq and regular 16S-seq data. The rarefaction analysis result using the data from one technique repeat (8 cycles) is shown here. c Cas-16S-seq observed more OTUs with average count per sample > 1 than regular 16S-seq. \*\*\*ANOVA, p = 4.35e−10 (see also Additional file 4: Table S6). d Analysis of differential bacterial OTUs between Cas-16S-seq and regular 16S-seq (CK). No OTU was significantly reduced in relative abundance after Cas9/gRNA treatments in all comparisons (Cas-16S-seq vs regular 16S-seq). In contrast, 29 OTUs were significantly increased in the Cas-16S-seq data (indicated by red dots; Wald test, p < 0.01; see also Additional file 4: Table S6). e Composition of bacterial communities at the phylum level. Bacterial phyla with relative abundances > 1% were shown here

使用Cas-16S-seq同时去除叶绿体和线粒体16S rRNA基因

Simultaneous removal of chloroplast and mitochondrion

16S rRNA genes using multiplexed Cas-16S-seq

为了进一步证实我们的Cas-16S-seq的高适用性，我们还测试了使用Cas-16S-seq同时去除线粒体和叶绿体16S rRNA基因的效率和特异性。选取mt-gRNA780和cp-gRNA578去靶向土壤和叶际中水稻515F-806R扩增子。在土壤样品中，我们在Cas-16S-seq和常规16S-seq中检测到几乎相同的细菌OTUs，并且其丰度相似(**图4 c, d**)。说明同时使用mt-gRNA780和cp-gRNA578的Cas-16S -seq与16S-seq的结果几乎相同且没有去除土壤中细菌序列(未产生偏好)。相比之下，Cas-16S-seq处理的叶际样品中，宿主16S rRNA基因的比例从99.4±0.2%下降到11.6±5.9%(**图6a**). 这大大提高了在叶际样品中检测细菌OTUs的能力。16S-seq从叶际样品中的288条细菌序列中平均只得到65个OTUs(平均丰度> 1)，而Cas-16S-seq从44,216条细菌序列中平均得到914个OTUs(**图6b**，附加文件5:表S7和S8, Student’s t检验，*p* = 1. 8 e−4). 我们也在科分类水平对细菌组成进行了分析，发现Cas-16S-seq检出29个科分类水平的细菌，而16S-seq仅检出16个科分类水平的细菌(**图6c**)。与我们之前对799F-1193R扩增子的分析一样，我们在土壤和叶际样品中也没有检测到相对丰度明显低于16S-seq的细菌OTUs，说明在Cas-16S-seq中，cp-gRNA578和mt-gRNA780没有脱靶到细菌序列。总之，Cas-16S-seq可以在一次反应中高效、专一地同时去除线粒体和叶绿体的16S rRNA基因，而不会脱靶到细菌序列，从而极大地提高了植物微生物群体样品中细菌16S宏基因组学检测的灵敏性。



**图6.** 利用Cas-16S-seq同时去除叶际样品中的宿主线粒体和叶绿体16S rRNA基因。两种同时靶向515F-806R扩增子(V4区)的gRNAs (mt-gRNA780和cp-gRNA578)用于Cas-16S-seq中分析三个叶际样品(Leaf 1-3).OTU表被稀释抽样到50000条reads(参见附加文件5:表S7和S8).**a** 叶际样品中细菌、线粒体和叶绿体16S rRNA基因的占有比。**b** 每个叶际样品中平均丰度>1的OTU数量的比较(另见附加文件5:表S7).

c 在科分类水平细菌群落结构的比较(Cas-16S-seq与常规16S-seq)。热图显示了每个叶际样品中细菌科的丰度。红框标记了在常规16S-seq文库中检测不到的科分类水平的细菌

讨论Discussion

在本研究中，我们开发了一种Cas-16S-seq的方法，可以减轻扩增子文库中宿主细胞器对16S rRNA基因的污染。我们也建立了一整套生物信息学流程来设计靶向宿主16S rRNA基因的gRNAs，并且这些gRNAs在原核16S rRNA基因上几乎不会造成脱靶。为了验证这种方法的可行性和有效性，我们使用了包括mt-gRNA1048、mtgRNA1171、mt-gRNA1196、mt-gRNA780和cp-gRNA578在内的5种高度特异的gRNAs，靶向水稻799F-1193R和515F-806R扩增子进行Cas-16S-seq实验。利用模拟群体、稻田土、水稻根以及水稻叶际样品和常规16S-seq数据进行综合比较，我们结果表明，Cas-16S-seq显著降低了宿主的16S rRNA的比例，并且不会引入任何的偏好。因此，我们预期Cas-16S-seq将成为植物微生物学研究中一个有用工具。

Cas-16S-seq的主要优点如下：第一点，Cas-16S-seq能高效的去除靶标。我们的结果表明，使用任何一种gRNA的Cas-16S-seq分别可以将根系样本中的寄主污染率从63.2降低到2.9%，叶际样本中的寄主污染率从99.4降低到11.6%(**图5a、6a**). 当细菌DNA含量低于20%时，Cas-16S-seq也比PNA PCR钳法表现出更好的性能(**图3**).第二点，Cas9和gRNA用于Cas-16S-seq的额外费用可以忽略不计。根据我们的经验，Cas9和gRNA处理一个样品的成本仅仅需要8元人民币(约合1.15美元)，而Cas-16S-seq额外酶解步骤也只需要花费不到6小时。由于Cas9酶切去除了大量的宿主16S rRNA基因片段，因此在一次测序中可以合并更多的标记的Cas-16S-seq扩增子。因此，总的来说和常规的16S-seq相比使用Cas-16S-seq大大降低了每个样品的测序成本。第三，Cas-16S-seq很容易集成到目前基于16S rRNA基因的扩增子测序工作流程中。因为只需在标准的16S-seq文库构建中额外添加一个Cas9/gRNA酶解的步骤。

宿主特异性gRNA对Cas-16S-seq至关重要，我们以前期CRISPR-Plant所使用的gRNA设计工具为基础，通过比较植物和细菌的16S rRNA基因序列，设计出了分别靶向水稻线粒体和叶绿体的不同高特异性的mt-gRNAs和cp-gRNAs，并且也建立了一整套的生物信息学流程以供在其他植物物种中参考使用。由于，宿主特异性gRNAs的设计是基于已知的16S rRNA基因，因此Cas-16S-seq也需要考虑对未鉴定细菌的脱靶风险。因此，建议像我们在本研究中所做的那样，验证所选gRNAs的特异性。在未来，我们也可以根据样本类型使用更具体的参考数据集来进一步细化Cas-16S-seq的gRNA设计。

结论Conclusions

我们开发了Cas-16S-seq的方法，使用CRISPR/Cas9去除基于16S rRNA基因的扩增子测序中大量的宿主污染。此方法具有高效稳健、成本低廉且不会对基于16S rRNA基因扩增子测序结果产生任何偏好等优势。本研究也为植物微生物组学研究提供了一个强有力的工具，也为更深入的优化微生物组学研究工具提供新思路。

具体的材料与方法以及补充材料详见文章：https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-020-00859-0