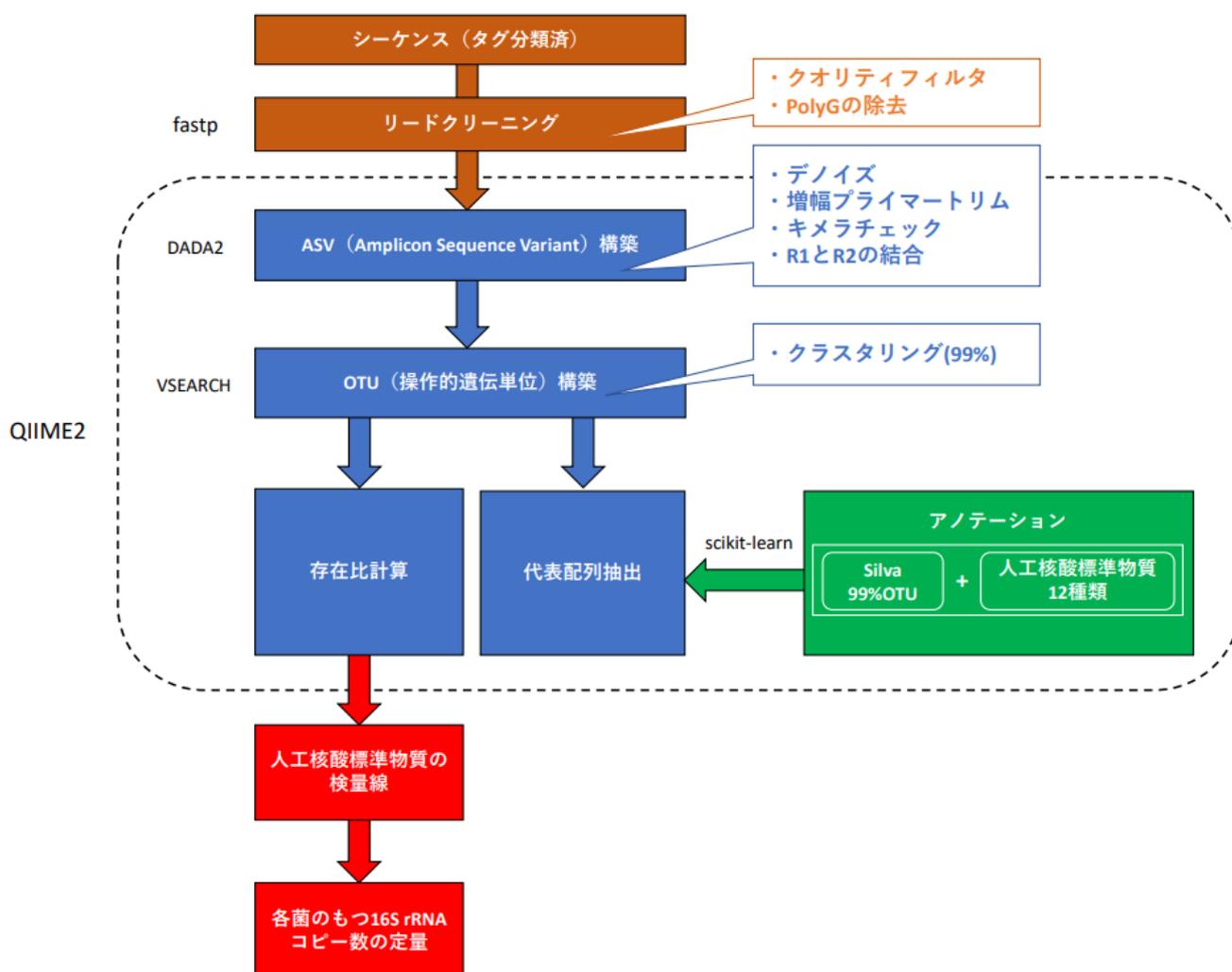


【実施例】 QIIME2を用いた定量的な16S rRNA細菌叢解析

16S DNA Quantitative Standard for Microbiome（製品コード NN0002）を用いた16S rRNAの定量解析は一般的な16S rRNA解析手法のいずれにおいても用いることが可能です。ここでは、QIIME2パイプラインを使用する実施例をご紹介します。細菌叢解析を行うためのパイプラインはmothurなど複数存在しますが、QIIME2は古くから使われていることに加え、Earth Microbiome Projectなど多くのプロジェクトで使用されています。そのため、ファイル書式に対応可能なこと、解析結果をwebブラウザで確認できることなどの使い勝手の良さから、QIIME2が広く使用されています。

本解析の主な流れは、まずシーケンスリードの前処理を行い、取得したリードからOTU（操作的分類単位: Operational Taxonomy Unit）を構築します。続いて、構築したOTUから代表配列を抽出を行い、OTUに対して細菌の系統分類を実施します。さらに系統分類された細菌と人工的な16S rRNA遺伝子の人工核酸標準物質の存在比を算出した後に、人工核酸標準物質における検量線に基づいて、細菌の16S rRNA遺伝子のコピー数を推定します。



使用するNGS試薬・キットとシーケンスデータについて

タカラバイオが提供する以下の試薬・キットを使用して、V3-V4 領域 (増幅領域長約 460 bp) を増幅し、シーケンスを行います。

- 16S (V3-V4) Metagenomic Library Construction Kit for NGS（製品コード R161A）
- 16S DNA Quantitative Standard for Microbiome（製品コード NN0002）

シーケンスの作業工程は、以下の通りです。

工程	条件
機種	NextSeq 2000
PhiX※	40 %
Read	250PE
Cycle number	251-8-8-251
Output format	Fastq

※ 他の多様性のあるLibraryと混合しシーケンスする場合のみPhiX不要

本解析では、1サンプルあたり15万リードペア（15万リード x2）以上を推奨します。

本解析のプロトコルについて

① シーケンスリードの前処理

- **fastpを使用したリードクリーニング**
シーケンサーから得られたリードは、アダプター配列、クオリティの低いリード、2色SBS特有のアーティファクト（polyGを含むリード）を除去する。

② OTU構築

- **DADA2を使用したデノイズ**
各サンプルから得られたタグ分類後のリードに対し、DADA2を使用したデノイズ、あわせてPCRプライマー配列の除去、ペアエンドリードのオーバーラップ、キメラチェック、100%一致する配列のマージを実施する。
- **VSEARCHを使用したクラスタリング**
DADA2を使用して得られた配列に基づいて、VSEARCHを使用して99%の一致率を閾値としてクラスタリングを行い、クラスターごとに最も長い配列を代表配列として抽出する。

③ 系統分類された菌叢存在比の算出

系統分類には、クラスタリングの結果から得られたOTU代表配列に基づいて実施する。まず、12種類の人工核酸標準物質を含むSilvaの99% OTUデータセットを使用して単純ベイズ分類器モデルを作成する。作成した分類器を使用してOTU代表配列の系統分類を行う。その結果と、各OTUのリード量から存在比を算出する。

④ 12種類の人工核酸標準物質の検量線に基づく菌叢の16S rRNA遺伝子のコピー数の推定

まず、リードが得られた、各サンプルの人工核酸標準物質について、系統分類レベル7（Taxonomic Level 7）のリード量をY軸、16S rRNA遺伝子のコピー数（理論値）をX軸にした時の検量線を作成する（Y値とX値はともにlog10で変換）。各サンプルで、検量線の範囲内のリード量を有する菌叢については、検量線に基づきコピー数を定量する。一方、検量線の範囲外の菌叢については、対象となる菌叢のリード量の総和を算出し、算出した総和が検量線の範囲内であればコピー数を定量するが、そうでない場合は定量不可とする。このような処理を、Taxonomic Level 1~7で実施する。

1. 使用する主なツール

本解析では以下のバージョンで動作確認を行い、正常に動作することを確認した。

Tool	Version
Conda	≥ 22.9.0
QIIME2	≥ 2023.7
Seqkit	≥ 0.10.0
fastp	≥ 0.23.4

※ OS環境：Linux x86_64

2. (Option) Linux x86_64向けのQIIME2解析環境の構築

既に構築済みの場合は本説明をスキップしてください。

※ Anaconda等を用いたconda仮想環境が予め構築されていることが必須です。

本解析で構築したQIIME2解析環境は以下の2つの手順で構築できます。

引用先:

- QIIME2 (version 2023.7) : <https://docs.qiime2.org/2023.7/install/native/>

(1) 本解析で構築したQIIME2 (version 2023.7) 解析環境をビルド。

```
# Linux x86_64向けのQIIME2 (version 2023.7) のYAML形式ファイルをダウンロード (URL内の空白は削除願います) 。
$ wget https://data.qiime2.org/distro/core/qiime2-2023.7-py38-linux-conda.yml
$ conda env create -n qiime2-2023.7 --file qiime2-2023.7-py38-linux-conda.yml
```

(2) 2つのRESCRIPT関連パッケージを逐次インストールして、QIIME2環境を再現。

```
# 仮想環境 ( qiime2-2023.7 ) を立ち上げて、以下2つのRESCRIPT関連パッケージを逐次インストール (URL内の空白は削除願います) 。
$ source activate qiime2-2023.7
$ pip install https://github.com/bokulich-lab/RESCRIPT/archive/refs/tags/2023.5.0.tar.gz
$ pip install https://github.com/bokulich-lab/q2-types-genomics/archive/refs/tags/2023.5.0.dev0.tar.gz
$ pip install ncbi-datasets-pylib

# qiimeの設定を読み直し、qiime環境が正常に機能することを確認。
$ qiime dev refresh-cache
QIIME is caching your current deployment for improved performance. This may take a few moments and should
only happen once per deployment.

# rescriptが正常にqiimeにインストールされていることを確認。
$ qiime rescript --version
QIIME 2 Plugin 'rescript' version 2023.5.0 (from package 'rescript' version 2023.5.0)
```

その他ツールのインストール

- QIIME2環境のactivateした状態 (`$ conda activate qiime2-2023.7` を実行) で以下のツールをインストール。

(1) Seqkitのインストール

```
$ conda install seqkit==0.10.0
```

(2) fastpのインストール

```
$ conda install -c bioconda fastp==0.23.4
```

*解析終了時は `$ conda deactivate qiime2-2023.7` を実行。

3. QIIME2を用いた定量的な16S rRNA細菌叢解析の準備

(a) (Option) 本解析に用いる、SILVA菌叢データベースと12種類の人工核酸標準物質から構成される参照配列の構築

本解析で構築した、SILVAデータベース由来の配列と12種類の人工核酸標準物質を含む参照配列、その分類系統ファイル、菌叢の存在比の算出に用いるファイルは[こちら](#)から入手可能です。

※ 該当する3つのファイルはQIIME2による解析仕様に合わせてqza形式ファイルに予め変換しております。

- 参照配列ファイル名: silva138.1_plus_std.qza

- ・ 分類系統ファイル名: silva138.1_plus_std_taxonomy.qza
- ・ 菌叢の存在比の算出用ファイル名: silva138.1_plus_std_classifier.qza

- (i) SILVAデータベースのダウンロードとアノテーション構築
- (ii) SILVA菌叢データベース登録された配列と12種類の人工核酸標準物質を統合（本解析向け）

引用先:

- ・ SILVA SSU database (release 138) : <https://www.arb-silva.de/documentation/release-138/>
- ・ 12種類の人工核酸標準物質 : Dieter M. Tourlousse, Satowa Yoshiike, Akiko Ohashi, Satoko Matsukura, Naohiro Noda and Yuji Sekiguchi, Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing, *Nucleic Acids Research*, 2017, Vol. 45, No. 4, e23

(b) Fastpを使用したリードフィルタリング

(c) Manifest file (manifest) の作成

(d) Fasting mapの作成

(a) (Option) 本解析に用いる、SILVA菌叢データベースと12種類の人工核酸標準物質から構成される参照配列の構築

- ・ 以下のコマンド操作により参照配列を構築する。ただし、下記に該当する3つのファイルをダウンロードした場合、以下の操作(i)と(ii)は不要。
 - 参照配列ファイル名: silva138.1_plus_std.qza
 - 分類系統ファイル名: silva138.1_plus_std_taxonomy.qza
 - 菌叢の存在比の算出用ファイル名: silva138.1_plus_std_classifier.qza

(i) SILVA菌叢データベース配列の抽出と前処理

```
$ mkdir -p reference
$ cd reference/

# SSUの99%クラスターとtaxonomyをダウンロード (SSURef_NR99で99% cluster)
$ qiime rescript get-silva-data \
--p-target 'SSURef_NR99' \
--p-include-species-labels \
--o-silva-sequences silva-138-ssu-nr99-seqs.qza \
--o-silva-taxonomy silva-138-ssu-nr99-tax.qza

# クオリティが低い配列 (8塩基以上のホモポリマー、5塩基以上のambiguous baseを含む配列) の除去 (産物ごとに配列異なるケースを考慮)
$ qiime rescript cull-seqs \
--i-sequences silva-138-ssu-nr99-seqs.qza \
--o-clean-sequences silva-138-ssu-nr99-seqs-cleaned.qza

# duplicate配列の除去 (データベースの冗長性をなくす)
$ qiime rescript dereplicate \
--i-sequences silva-138-ssu-nr99-seqs-cleaned.qza \
--i-taxa silva-138-ssu-nr99-tax.qza \
--p-mode 'uniq' \
--o-dereplicated-sequences silva-138-ssu-nr99-seqs-derep-uniq.qza \
--o-dereplicated-taxa silva-138-ssu-nr99-tax-derep-uniq.qza

# 菌叢の存在比の算出に用いる分類器の作成
$ qiime feature-classifier fit-classifier-naive-bayes \
--i-reference-reads silva-138-ssu-nr99-seqs-derep-uniq.qza \
--i-reference-taxonomy silva-138-ssu-nr99-tax-derep-uniq.qza \
--o-classifier silva_138.1_classifier.qza
```

(ii) SILVA菌叢データベース登録された配列と12種類の人工核酸標準物質を統合（本解析向け）

- 本解析における人工核酸標準物質 (fasta形式ファイル) とその分類系統 (tsv形式ファイル) は[こちら](#)から入手可能です。該当する2つのファイルをダウンロードした場合、下記2つのコマンド手順でSILVA+人工核酸標準物質12種の配列をQIIME2で使える形式に変換してください。

- 12種類の人工核酸標準物質ファイル名：16S_spikein_standard.fasta
- 12種類の分類系統ファイル名：16S_spikein_standard.tsv

SILVAデータベース由来の配列と人工核酸標準物質からなる参照配列（SILVA+12種類の人工核酸標準物質）の構築に向けた前処理

```
# SILVAデータベース由来の配列データをfastaファイル形式で取得
$ qiime tools export \
--input-path silva-138-ssu-nr99-seqs-derep-uniq.qza \
--output-path ./silva-138-ssu-nr99-seqs-derep-uniq

# SILVAデータベースからの系統分類をtsvファイル形式で抽出
$ qiime tools export \
--input-path silva-138-ssu-nr99-tax-derep-uniq.qza \
--output-path ./silva-138-ssu-nr99-tax-derep-uniq

# SILVAデータベース由来の配列と人工核酸標準物質12種の配列を統合
$ cat ./silva-138-ssu-nr99-seqs-derep-uniq/dna-sequences.fasta ./16S_spikein_standard.fasta >
silva138.1_plus_std.fasta
# SILVAデータベース由来の分類系統と12種の分類系統を統合
$ cat ./silva-138-ssu-nr99-tax-derep-uniq/taxonomy.tsv ./16S_spikein_standard.tsv >
silva138.1_plus_std.tsv
```

SILVA+12種類の人工核酸標準物質をQIIME2で使える形式に変換

```
# 12種類の配列についてfastaの取り込み
$ qiime tools import \
--type 'FeatureData[Sequence]' \
--input-path silva138.1_plus_std.fasta \
--output-path silva138.1_plus_std.qza

# アノテーション表を作成してからqzaに変換
$ qiime tools import \
--type 'FeatureData[Taxonomy]' \
--input-path silva138.1_plus_std.tsv \
--output-path silva138.1_plus_std_taxonomy.qza

# fastaとアノテーションをリンクさせる
$ qiime feature-classifier \
fit-classifier-naive-bayes \
--i-reference-reads silva138.1_plus_std.qza \
--i-reference-taxonomy silva138.1_plus_std_taxonomy.qza \
--o-classifier silva138.1_plus_std_classifier.qza
```

(b) fastpを使用したリードクリーニング

- 下記コマンド操作でリードクリーニングを実施。

```
$ fastp -w 4 \
-q 10 \
-u 40 \
-l 200 \
--trim_poly_g \
--i [R1].fastq.gz \
-I [R2].fastq.gz \
-o [filtered_R1].fastq.gz \
-O [filtered_R2].fastq.gz \
-h fastp.html \
-j fastp.json
```

(Option) seqkitを使用したダウンサンプリング

- 全サンプルを150000 readsに揃える場合、下記コマンドでダウンサンプリングを実施（-n optionでread数を指定）。

```
$ for i in *.fastq.gz
do
seqkit sample -s 100 -n 150000 ${i} > ds_150k_${i}
done
```

- seqkit sampleではread名に基づいて抽出を行うため、R1とR2に対してそれぞれ別個に実行しても同じreadが抽出されます。同じread setから別のsetをsamplingしたい場合は -s optionで乱数の値を変化させます（初期値は11）

(c) manifestファイルの作成

- 1行目にはtab区切りで「sample-id」「forward-absolute-filepath」「reverse-absolute-filepath」を入力。

```
$ echo -e "sample-id\tforward-absolute-filepath\treverse-absolute-filepath" >
${filepath}/template/manifest.tsv
```

- 2行目以降の1列目「sample-id」には検体名称を入力。
- 2行目以降の2列目「forward-absolute-filepath」および3列目「reverse-absolute-filepath」には、それぞれR1とR2のfastq.gzファイルの絶対パス（absolute-filepath）を入力。
- ファイル名は「manifest.tsv」とする。

(d) Fasting_mapファイルの作成

- 1行目がタブ区切りで「#SampleID」「BarcodeSequence」「LinkerPrimerSequence」の順で入力。

```
$ echo -e "#SampleID\tBarcodeSequence\tLinkerPrimerSequence\tDescription" >
${filepath}/template/Fasting_Map.txt
```

- 2行目以降の1列目の検体名称「#SampleID」について、manifestファイルの1列目の表記と一致することを確認。
- 2行目以降の2列目「BarcodeSequence」にはForwardの増幅primer、2行目以降の3列目「LinkerPrimerSequence」にはReverseの増幅primerの配列を入力する。
 - e.g. 341F-806Rの場合はそれぞれ「CCTACGGGNGGCWGCAG」「GGACTACHVGGGTWTCTAAT」
- ファイル名は「Fasting_map.txt」とする。

全サンプルを対象としたリードクリーニング処理（手順(b)-(d)まで）

- QIIME2の入力としてmanifest.tsvとfasting_map.txtを出力

```
# FASTQファイルを置いている場所へ移動
$ cd /path/to/fastq
$ mkdir -p ./filtered_reads
$ echo -e "sample-id\tforward-absolute-filepath\treverse-absolute-filepath" >
./filtered_reads/manifest.tsv
$ echo -e "#SampleID\tBarcodeSequence\tLinkerPrimerSequence\tDescription" >
./filtered_reads/Fasting_Map.txt
```

- 全サンプルを対象にアダプタートリミング処理を実施

```
$ mkdir -p ./filtered_reads/log
$ for i in `ls *_R1_001.fastq.gz`; \
do \
if [[ $i != Undetermined* ]]; \
then \
```

```

fastp -g -w 4 -q 10 -u 40 -l 200 --trim_poly_g -i ${i} -I ${i%_R1_001.fastq.gz}_R2_001.fastq.gz -o
./filtered_reads/${i} -O ./filtered_reads/${i%_R1_001.fastq.gz}_R2_001.fastq.gz -h
./filtered_reads/log/${i%_R1_001.fastq.gz}_fastp.html -j
./filtered_reads/log/${i%_R1_001.fastq.gz}_fastp.json; \
echo -e
"${i%_R1_001.fastq.gz}\t${PWD}/filtered_reads/${i}\t${PWD}/filtered_reads/${i%_R1_001.fastq.gz}_R2_001.f
astq.gz" >> ./filtered_reads/manifest.tsv; \
echo -e "${i%_R1_001.fastq.gz}\tCCTACGGGNGGCWGCAG\tGGACTACHVGGGTWTCTAAT\t-" >>
./filtered_reads/Fasting_Map.txt; \
fi \
done

```

4. QIIME2を用いた定量的な16S rRNA細菌叢解析の実施

(a) 菌叢解析に必要な下記4つのファイルを確認

- manifestファイル : manifest.tsv
- Fasting_mapファイル : Fasting_Map.txt
- 参照配列ファイル : silva138.1_plus_std.qza
- 菌叢の存在比の算出に用いる菌叢分類用ファイル : silva138.1_plus_std_classifier.qza

※ 参照配列ファイルについては、SILVAデータベースに登録された菌種（ver138.1: filtering後441511種）と12種類の人工核酸標準物質を統合したものである。

※ 菌叢分類用ファイルについては、菌種情報と結び付けられたものである。

(b) QIIME2を使用した菌叢分類と、各菌叢の存在比の算出

- 下記コマンドを実施する。

```

# QIIME2の仮想環境を立ち上げる
$ source activate qiime2-2023.7

# 12種類の人工核酸標準物質を含む参照配列と菌叢分類用ファイルの場所を指定
$ refpath=[path to reference]

# 解析のset名を変数に与える (e.g. 日付YYMMDDなど)
$ setname=YYMMDD

# Fasting_mapファイルの名前を変数に与える
$ fmap=[path to Fasting_map]/Fasting_Map.txt

# Fasting_mapファイルをqzv形式に変換 ( QIIME2の中ではqza形式あるいはqzv形式でdataを扱う)
$ qiime metadata tabulate \
--m-input-file ${fmap} \
--o-visualization ${fmap%.txt}.qzv

# sequencingのfastq.gz dataをqza形式に変換
$ qiime tools import \
--input-path ${setname}_manifest.tsv \
--output-path ${setname}_sequence.qza \
--type 'SampleData[PairedEndSequencesWithQuality]' \
--input-format PairedEndFastqManifestPhred33V2

# denoiseの実施 ( 100%でのclustering)
# --p-trim-leftには 増幅primer長を入力 ( 検体により、あるいはFRで異なる場合には長い検体に統一 )
$ qiime dada2 denoise-paired \
--i-demultiplexed-seqs ${setname}_sequence.qza \
--p-trunc-len-f 0 \
--p-trunc-len-r 0 \

```

```

--p-trim-left-f 17 \
--p-trim-left-r 18 \
--output-dir ${setname}_dada2 \
--p-n-threads 32

$ cd ${setname}_dada2

# Silvaの菌種に基づき99%OTUを構築
$ qiime vsearch cluster-features-open-reference \
--p-threads 8 \
--i-table table.qza \
--i-sequences representative_sequences.qza \
--i-reference-sequences ${refpath}/silva138.1_plus_std.qza \
--p-perc-identity 0.99 \
--o-clustered-table ${setname}_table-or-99.qza \
--o-clustered-sequences ${setname}_clustered_seqs.qza \
--o-new-reference-sequences ${setname}_new-ref-seqs-or-99.qza

# OTU代表配列のmetricを取得 ( qzaで出力された内容をqzv形式で取得 )
$ qiime feature-table tabulate-seqs \
--i-data ${setname}_clustered_seqs.qza \
--o-visualization ${setname}_clustered_seqs.qzv

# qzv fileの内容をhtmlに出力
$ qiime tools export \
--input-path ${setname}_clustered_seqs.qzv \
--output-path ${setname}_representative_sequences

# それぞれのOTU代表配列に菌種情報を付与
$ qiime feature-classifier classify-sklearn \
--i-classifier ${refpath}/silva138.1_plus_std_classifier.qza \
--i-reads ${setname}_clustered_seqs.qza \
--o-classification ${setname}_taxonomy.qza

# 付与された菌種の情報をqzvでも出力
$ qiime metadata tabulate \
--m-input-file ${setname}_taxonomy.qza \
--o-visualization ${setname}_taxonomy.qzv

# qzvの内容をhtmlに出力
$ qiime tools export \
--input-path ${setname}_taxonomy.qzv \
--output-path .

# それぞれの菌種のabundanceをbarchartに出力
$ qiime taxa barplot \
--i-table ${setname}_table-or-99.qza \
--i-taxonomy ${setname}_taxonomy.qza \
--m-metadata-file ../${fmap} \
--output-dir ${setname}_taxa-barplot

# barchartをhtmlに出力
$ cd ${setname}_taxa-barplot
$ qiime tools export \
--input-path visualization.qzv \
--output-path .

```

(c) 人工核酸標準物質の検量線に基づく菌叢の16S rRNA遺伝子のコピー数の推定

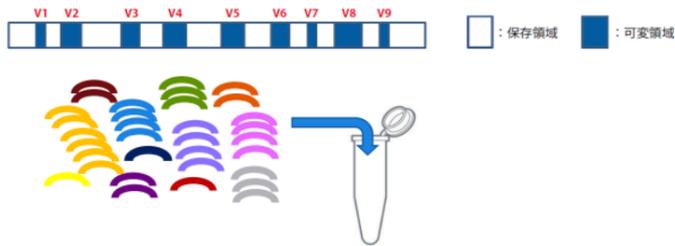
- 12種類の人工核酸標準物質の16S rRNAコピー数（理論値）とその存在比の表について（表1）

※ 原液16S DNA Quantitative Standard（ 1.3×10^8 copies/ μ l）の組成を表しています。

identifier	GenBank accession number	taxa description	16S rRNA(copies/ μ l)	16S rRNA copies(%)
Std_01	LC140931	d_Standard;p__Proteobacteria;c__Gammaproteobacteria; o__Enterobacteriales;f__Enterobacteriaceae;g__Escherichia;s__coli5001	6.31×10^7	47.39
Std_02	LC140932	d_Standard;p__Proteobacteria;c__Gammaproteobacteria; o__Enterobacteriales;f__Enterobacteriaceae;g__Escherichia;s__coli5002	3.32×10^7	24.93
Std_03	LC140933	d_Standard;p__Proteobacteria;c__Gammaproteobacteria; o__Enterobacteriales;f__Enterobacteriaceae;g__Escherichia;s__coli5003	1.75×10^7	13.14
Std_04	LC140934	d_Standard;p__Proteobacteria;c__Gammaproteobacteria; o__Enterobacteriales;f__Enterobacteriaceae;g__Escherichia;s__coli5004	9.20×10^6	6.91
Std_05	LC140935	d_Standard;p__Proteobacteria;c__Gammaproteobacteria; o__Enterobacteriales;f__Enterobacteriaceae;g__Escherichia;s__coli5005	4.84×10^6	3.63
Std_06	LC140936	d_Standard;p__Proteobacteria;c__Gammaproteobacteria; o__Enterobacteriales;f__Enterobacteriaceae;g__Escherichia;s__coli5501	2.55×10^6	1.91
Std_07	LC140937	d_Standard;p__Proteobacteria;c__Gammaproteobacteria; o__Enterobacteriales;f__Enterobacteriaceae;g__Escherichia;s__coli5502	1.34×10^6	1.01
Std_08	LC140938	d_Standard;p__Proteobacteria;c__Gammaproteobacteria; o__Enterobacteriales;f__Enterobacteriaceae;g__Escherichia;s__coli6001	7.06×10^5	0.53
Std_09	LC140939	d_Standard;p__Bacteroidetes;c__Bacteroidia; o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__vulgatus5501	3.72×10^5	0.28
Std_10	LC140940	d_Standard;p__Chloroflexi;c__Chloroflexi; o__Chloroflexales;f__Chloroflexaceae;g__Chloroflexus;s__aurantiacus5501	1.96×10^5	0.15
Std_11	LC140941	d_Standard;p__Gemmatimonadetes;c__Gemmatimonadetes; o__Gemmatimonadales;f__Gemmatimonadaceae;g__Gemmatimonas;s__aurantiaca5501	1.03×10^5	0.08
Std_12	LC140942	d_Standard;p__Spirochaetes;c__Spirochaetes;o__Spirochaetales; f__Spirochaetaceae;g__Treponema;s__bryantii5501	5.42×10^4	0.04

• 人工核酸標準物質の検量線の作成と、菌叢の16S rRNA遺伝子コピー数の推定

- (1) 各サンプルについて、Qiime2の出力結果Taxonomic Level 7のファイル（[prefix]_taxa-barplotフォルダ下のファイル名level-7.csv）を開き、その中から12種類の人工核酸標準物質（列名: d_Standard〜）のリード量に該当するカラムを抽出。
- (2) エクセル表計算ツールなどを用いて、人工核酸標準物質のリード量と16S rRNA遺伝子のコピー数をシートに貼り付け、人工核酸標準物質のリード量をY軸、その16S rRNA遺伝子のコピー数（理論値）をX軸として、検量線（Y値とX値をともにlog₁₀で変換）を作成。ただし、リードが得られた人工核酸標準物質を対象とする。
- (3) 検量線については以下の2つの項目を確認。
 - 回帰直線の決定係数 $R^2 > 0.90$ であること。
 - 各16S rRNA配列の16S rRNA copy数とリード量のpearson相関係数 ≥ 0.96 となること。
- (4) 作成した検量線に、興味のある菌のリード量を当てはめ、該当する菌の16S rRNA遺伝子のコピー数を推定。
- (5) 上記 (1) - (4) の処理を、系統分類レベル1~7（Taxonomic Level 1~7）で実施。



- 人工核酸標準物質の特徴
- ・ 12種類の配列既知の核酸が異なる存在比で混合された標準物質
 - ・ 可変領域 (V1~V9) は自然界に存在しない人工DNA配列
 - ・ V1-V2、V3-V4といったどの可変領域についても定量解析可能
 - ・ 解析サンプルの菌叢解析結果に影響を与えない
 - ・ あらゆる生体サンプルにスパイクイン可能な普遍的な標準品

