Qiime2数据导入（针对于Earth Microbiome Project(EMP)标准格式的fastq数据类型）

1. 创建工作目录

mkdir qiime2-importing-tutorial

cd qiime2-importing-tutorial

2.创建单端、双端测序文件夹

mkdir emp-single-end-sequences

3．下载示例数据（序列样本标签）

wget \

-O "emp-single-end-sequences/barcodes.fastq.gz" \

<https://data.qiime2.org/2020.2/tutorials/moving-pictures/emp-single-end-sequences/barcodes.fastq.gz>

其中，“\”为换行，-O 参数即将下载文件存入emp-single-end-sequences目录并命名为barcodes.fastq.gz

（序列fastq文件,真正的序列）

wget \

-O "emp-single-end-sequences/sequences.fastq.gz" \

<https://data.qiime2.org/2020.2/tutorials/moving-pictures/emp-single-end-sequences/sequences.fastq.gz>

4.导入数据

qiime tools import \

--type EMPSingleEndSequences \

--input-path emp-single-end-sequences \

--output-path emp-single-end-sequences.qza

Qiime tools import 是qiime2导入命令，type指定语义类型，--input-path参数指定序列文件所在文件夹，--output-path参数指定输出文件格式为qza

qiime tools import --show-importable-types命令可用于查看语义类型，以下为语义类型：

DeblurStats

DistanceMatrix

EMPPairedEndSequences

EMPSingleEndSequences

ErrorCorrectionDetails

FeatureData[AlignedSequence]

FeatureData[Differential]

FeatureData[Importance]

FeatureData[PairedEndSequence]

FeatureData[Sequence]

FeatureData[Taxonomy]

FeatureTable[Balance]

FeatureTable[Composition]

FeatureTable[Frequency]

FeatureTable[PercentileNormalized]

FeatureTable[PresenceAbsence]

FeatureTable[RelativeFrequency]

Hierarchy

MultiplexedPairedEndBarcodeInSequence

MultiplexedSingleEndBarcodeInSequence

PCoAResults

Phylogeny[Rooted]

Phylogeny[Unrooted]

Placements

QualityFilterStats

RawSequences

SampleData[AlphaDiversity]

SampleData[BooleanSeries]

SampleData[ClassifierPredictions]

SampleData[DADA2Stats]

SampleData[FirstDifferences]

SampleData[JoinedSequencesWithQuality]

SampleData[PairedEndSequencesWithQuality]

SampleData[Probabilities]

SampleData[RegressorPredictions]

SampleData[SequencesWithQuality]

SampleData[Sequences]

SampleEstimator[Classifier]

SampleEstimator[Regressor]

SeppReferenceDatabase

TaxonomicClassifier

UchimeStats

那么，为什么要定义语义类型？有什么作用。

定义语义类型使我们能够确保传递给操作的数据对于将要执行的操作有意义。例如，一个biom.Table对象可能包含存在/不存在数据（即1表示在给定样本中至少一次观察到OTU，0表示在给定样本中至少一次未观察到OTU。样品）。但是，如果将其提供给计算定量分集度量的函数（其中将OTU丰度包括在计算中）（例如加权UniFrac），则该函数将成功完成，但结果将没有意义。

QIIME 2定义了语义类型，以帮助用户避免错误使用功能，并允许系统推断哪些QIIME 2 artifacts可以与哪些方法和可视化工具一起使用。

说白了，就是定义参数，省得白费功夫！

**通用语义类型（来自qiime2说明文档）**

除非另有说明，否则以下语义类型由[q2-types](https://github.com/qiime2/q2-types)插件定义并可以从其中导入。也可以在任何插件中定义语义类型，因此可用的语义类型不限于[q2-types中](https://github.com/qiime2/q2-types)定义的那些。即将添加有关如何完成此操作的说明。同时，您可以参考[q2-dummy-types](https://github.com/qiime2/q2-dummy-types)存储库以获取带注释的示例。

FeatureTable[Frequency]：一个功能表（例如，按OTU采样），其中每个值表示相应采样中OTU的频率，以原始计数表示。

FeatureTable[RelativeFrequency]：一个特征表（例如，按OTU采样），其中每个值表示相应样本中OTU的相对丰度，以使每个样本的值总计为1.0。

FeatureTable[PresenceAbsence]：一个特征表（例如，按OTU采样），其中每个值指示相应样本中是否存在OTU。

FeatureTable[Composition]：一个功能表（例如，按OTU采样），其中每个值表示相应样本中OTU的频率，并且所有频率均大于零。

Phylogeny[Rooted]：根系的系统树。

Phylogeny[Unrooted]：无根的系统发育树。

DistanceMatrix：距离矩阵。

PCoAResults：运行主坐标分析（PCoA）的结果。

SampleData[AlphaDiversity]：Alpha分集值，每个值与一个样本标识符关联。

SampleData[SequencesWithQuality]：具有质量得分的序列，其中每个序列集与一个样本标识符（即，解复用的序列）相关联。

SampleData[PairedEndSequencesWithQuality]：具有质量得分的成对末端序列（正向和反向），其中每对成对末端序列与一个样本标识符（即，解复用的成对末端序列）相关联。

FeatureData[Taxonomy]：与功能标识符关联的分类信息。

FeatureData[Sequence]：与特征标识符关联的单个未对齐序列（例如，代表性序列）。

FeatureData[AlignedSequence]：与特征标识符关联的单个比对序列，其中比对是与与所有其他特征标识符关联的比对序列有关的（即，当存在多个特征标识符时，表示多个序列比对）。

FeatureData[PairedEndSequence]：与特征标识符关联的成对末端序列（正向和反向）。

EMPSingleEndSequences：根据[Earth Microbiome Project测序协议](http://www.earthmicrobiome.org/protocols-and-standards/)生成的多路复用单端读取序列数据。有关[导入](https://docs.qiime2.org/2020.2/tutorials/importing/)此数据类型的详细信息，请参见[导入数据](https://docs.qiime2.org/2020.2/tutorials/importing/)。这种类型以前称为RawSequences。该RawSequences类型名称是过时的，但现在是一个别名EMPSingleEndSequences。您应该始终使用该名称EMPSingleEndSequences来代替RawSequences。目前，该类型是在[q2-demux](https://github.com/qiime2/q2-demux)插件中定义的，但将来可能会移至q2-types。

EMPPairedEndSequences：根据[地球微生物组计划测序协议](http://www.earthmicrobiome.org/protocols-and-standards/)生成的多对配对末端读取序列数据。有关[导入](https://docs.qiime2.org/2020.2/tutorials/importing/)此数据类型的详细信息，请参见[导入数据](https://docs.qiime2.org/2020.2/tutorials/importing/)。目前，该类型是在[q2-demux](https://github.com/qiime2/q2-demux)插件中定义的，但将来可能会移至q2-types。

TaxonomicClassifier：训练有素的分类器，可用于对序列进行分类分配。此类型在[q2-feature-classifier](https://github.com/qiime2/q2-feature-classifier)插件中定义。