Roteiro - Metabarcoding com QIIME 2

Autor: Juan Zegarra
Data: 11/2026

Objetivo

Este manual tem como objetivo guiar uma análise prática de dados de metabarcoding utilizando o QIIME 2. Ao final, o aluno deverá ser capaz de:

- Entender o fluxo básico de análise de dados de sequenciamento de amplicons
- Avaliar qualidade de dados
- Processar dados com DADA2
- Gerar e interpretar perfis taxonômicos
- Comparar amostras usando métricas de diversidade

Contexto Biológico

Metabarcoding é uma abordagem que utiliza sequenciamento de fragmentos específicos de DNA (amplicons) para identificar comunidades biológicas em uma amostra (ex: microbioma). Neste exercício, utilizamos dados de sequenciamento da região V4 do 16S rRNA para caracterizar comunidades de bactérias e arquéias.

São utilizadas amostras de solo do deserto do Atacama em diversas condições distintas de profundidade, pH, altitude, nível de humidade e temperatura, por exemplo.

► mSystems. 2017 May 30;2(3):e00195-16. doi: 10.1128/mSystems.00195-16

Significant Impacts of Increasing Aridity on the Arid Soil Microbiome

Julia W Neilson [a,✉], Katy Califf [b], Cesar Cardona [c], Audrey Copeland [a], Will van Treuren [d], Karen L Josephson [a], Rob Knight [e], Jack A Gilbert [f], Jay Quade [g], J Gregory Caporaso [b], Raina M Maier [a]

Editor: Haiyan Chu[h]

► Author information ► Article notes ► Copyright and License information

PMCID: PMC5451488 PMID: 28593197

Plataformas de software utilizadas

Micromamba
Micromamba é um gerenciador de ambientes computacionais que permite instalar e organizar programas e suas dependências de forma reprodutível, evitando conflitos entre diferentes ferramentas bioinformáticas.

QIIME2
QIIME 2 é uma plataforma de bioinformática voltada para análise de dados de microbioma e metabarcoding, permitindo processar sequências, controlar qualidade, inferir ASVs, atribuir taxonomia e gerar visualizações interpretáveis.

0. Preparando o ambiente de análise

Entre no servidor bioinfo com ssh

Entre também no seu diretório na web usando o link http://bionfo.icb.ufmg.br/bioufmg
(USER: bioufmg)
(SENHA: prof vai falar)

Ativar ambiente micromamba

```
micromamba activate barcode2
```

Preparando os arquivos necessários
De dentro da sua pasta pessoal, crie links para o fodler de dados e outros aquivos que vamos utilizar:

```
ln -s /home/treinamento/barcode/Data_paired_reads
ln -s /home/treinamento/barcode/sample-metadata.tsv
ln -s /home/treinamento/barcode/manifest_samples.tsv
ln -s /home/treinamento/barcode/classifier.qza
```

Inspecione o conteúdo da pasta Data
O que são esses arquivos?
Inspecione o arquivo metadata.tsv. Que informações são essas?
Falaremos sobre o arquivo manifest em breve.

---

1. Importação dos dados para o qiime

O QIIME 2 trabalha com arquivos em um formato próprio chamado "artifact" (.qza). Por isso, precisamos importar os arquivos FASTQ para esse formato.

Comando

```
qiime tools import \
--type 'SampleData[PairedEndSequencesWithQuality]' \
--input-path manifest_samples.tsv \
--output-path fastq-data.qza \
--input-format PairedEndFastqManifestPhred33V2
```

Perguntas

- O que é o arquivo manifest? Que informações ele contém? (Use o comando “less” para visualizar o arquivo)
- Por que precisamos importar os dados para o formato QIIME?

2. Avaliação da qualidade dos dados

Antes de processar os dados, é importante avaliar a qualidade das sequências. Essa etapa nos ajuda a decidir como tratar os dados depois.

Comando

```
qiime demux summarize \
--i-data fastq-data.qza \
--o-visualization fastq-data.qzv
```

```
qiime tools export --input-path fastq-data.qzv --output-path html_dir/sum1/
```

Perguntas

- Quantas reads contém cada amostra?
- Qual o tamanho delas?
- Como a qualidade muda ao longo da read?
- Há diferença de qualidade entre reads forward e reverse?
- Em que posição a qualidade começa a cair mais fortemente?
- Qual a qualidade (Phred score) das bases ruins? Essa qualidade é muito ruim?

3. DADA2 (Filtragem, Denoising, Merging)

DADA2 corrige erros de sequenciamento, gera ASVs (variantes de sequência exatas) e remove sequências quiméricas.

Comando

```
qiime dada2 denoise-paired \
  --i-demultiplexed-seqs fastq-data.qza \
  --p-trim-left-f 0 \
  --p-trim-left-r 0 \
  --p-trunc-len-f 150 \
  --p-trunc-len-r 150 \
  --p-max-ee-f 1.0 \
  --p-max-ee-r 1.0 \
  --p-trunc-q 10 \
  --o-table asv-table.qza \
  --o-representative-sequences asv-seqs.qza \
  --o-denoising-stats denoising-stats.qza
```

Trunc-len é a posição onde a sequência será cortada (truncada). Todas as bases após essas posição serão descartadas em todas as reads. Temos que escolher um valor para as reads forward (f) e reverso (r). Um bom valor de trunc-len:

- **Remove o máximo possível de regiões de baixa**

Visualização

```
qiime metadata tabulate \
--m-input-file denoising-stats.qza \
--o-visualization denoising-stats.qzv
```

```
qiime tools export --input-path denoising-stats.qzv --output-path html_dir/denoise
```

Perguntas

- Quantas reads foram removidas de cada amostra?
- Qual etapa remove mais dados?

---

4. Tabela de Features

A **feature table** do QIIME 2 é uma tabela que registra quantas vezes cada sequência/ASV foi observada em cada amostra. Em termos práticos, ela é a matriz principal de abundância usada nas etapas seguintes da análise, como diversidade alfa/beta, comparação entre grupos e visualização da composição das comunidades.

Veja a que foi gerada pela sua análise:

Comando

```
qiime feature-table summarize \
--i-table asv-table.qza \
--m-metadata-file sample-metadata.tsv \
--o-summary asv-table.qzv \
--o-sample-frequencies sample-frequencies.qza \
--o-feature-frequencies asv-frequencies.qza
```

```
qiime tools export --input-path asv-table.qzv --output-path html_dir/asv_table
```

Investigue o arquivo **asv_table**

Perguntas

- Quantas ASVs existem?
- Algumas amostras têm poucas reads?
- Isso pode afetar a análise?

5. Sequências representativas

As **rep-seqs** do QIIME 2 são as sequências representativas de cada feature/ASV presente na análise. Enquanto a *feature table* diz **quantas vezes** cada ASV aparece em cada amostra, o arquivo de *rep-seqs* contém a **sequência de DNA** correspondente a cada ASV, sendo usado para etapas como atribuição taxonômica, alinhamento e construção de árvores filogenéticas.

Comando

```
qiime feature-table tabulate-seqs \
--i-data asv-seqs.qza \
--o-visualization asv-seqs.qzv
```

```
qiime tools export --input-path asv-seqs.qzv --output-path html_dir/asv_seqs
```

Investigue o arquivo **asv-seqs.html**

Perguntas

- Qual sequência é mais abundante?
- O que faz uma sequencia ser mais abundante na amostra?
- Isso significa que é biologicamente mais importante?

6. Filtragem

As opções de filtragem no QIIME 2 permitem remover amostras ou sequências com baixa qualidade, baixa abundância ou critérios específicos definidos pelo usuário, tornando a análise mais confiável e adequada ao objetivo do estudo.

Vamos filtrar com base em quantas amostras cada ASV foi identificada.

Comando

Para filtrar a feature-table:

```
qiime feature-table filter-features \
--i-table asv-table.qza \
--p-min-samples 2 \
--o-filtered-table asv-table-ms2.qza
```

Para filtrar as sequências representativas:

```
qiime feature-table filter-seqs \
--i-data asv-seqs.qza \
--i-table asv-table-ms2.qza \
--o-filtered-data asv-seqs-ms2.qza
```

7. Classificação taxonômica

O **classificador taxonômico** atribui nomes taxonômicos às ASVs comparando suas sequências representativas com um banco de referência. No QIIME 2, isso é frequentemente feito com o comando classify-sklearn, que usa um classificador **Naive Bayes** treinado com sequências do banco **SILVA**, uma base de dados curada e amplamente usada para genes ribossomais, como o 16S rRNA.

Comando

```
qiime feature-classifier classify-sklearn \
--i-classifier classifier.qza \
--i-reads asv-seqs-ms2.qza \
--o-classification taxonomy.qza
```

Perguntas

- Como funciona a classificação taxonômica?
- O que significa baixa confiança?

---

8. Barplot taxonômico

O **barplot taxonômico** do QIIME 2 é uma visualização interativa que mostra a composição taxonômica das amostras em diferentes níveis, como filo, classe, ordem, família ou gênero. Ele permite comparar visualmente quais grupos microbianos são mais abundantes em cada amostra ou grupo experimental, a partir da combinação entre a *feature table* e a taxonomia atribuída às ASVs.

Comando

```
qiime taxa barplot \
--i-table asv-table-ms2.qza \
--i-taxonomy taxonomy.qza \
--m-metadata-file sample-metadata.tsv \
--o-visualization taxa-bar-plots.qzv
```

Visualize o gráfico.

```
qiime tools export --input-path taxa-bar-plots.qzv --output-path html_dir/taxa
```

Perguntas

- Coloque o nível taxonômico em 2 “Sort Samples By” em “average-soil-relative-humidity”.
- Agora mude o nível taxonômico para 7 e “Sort Samples By” em “average-soil-relative-humidity”. Ficou mais fácil de ver a diferença?

---

9. Diversidade beta

Vocês lembram da diversidade beta? e da alfa? Vamos calcular algumas métricas comuns de diversidade de amostras. A ideia é compararmos os cenários de raiz e de solo com base na diversidade deles. Como temos três amostras de cada cenário, eles vão funcionar como réplicas

```
qiime diversity core-metrics \
--i-table asv-table-ms2.qza \
--m-metadata-file sample-metadata.tsv \
--p-sampling-depth 100 \
--p-n-jobs 2 \
--output-dir diversity
```

Visualize

```
qiime tools export --input-path diversity/bray_curtis_emperor.qzv
--output-path html_dir/diversity
```

Perguntas

- Coloram por “average-soil-temperature” (coloquem a coloração em viridis e configurem para escala contínua).
- Agora mudem para “average-soil-relative-humidity”.