

# Metagenomics at your fingertips

Análise de Sequências Biológicas 2023-2024

Licenciatura em Bioinformática

Docente: Francisco Pina

Data: 27/06/2024

202200646 Joana Caetano

202200655 Lara Evangelista

202200239 Lucas Terlica

201901365 Rendrick Carreira

# Índice

<b>Índice de figuras .....</b>	<b>iii</b>
<b>1. Introdução .....</b>	<b>1</b>
1.1. Pipeline .....	2
<b>2. Objetivos .....</b>	<b>4</b>
<b>3. Métodos e Materiais .....</b>	<b>5</b>
3.1. Metadata.....	6
SRA para FASTQ .....	7
3.2. 16s Pipeline .....	7
<b>4. Resultados .....</b>	<b>8</b>
4.1. Heatmap .....	9
4.2. PCoA .....	11
4.3. Gráfico de abundância .....	12
<b>5. Discussão.....</b>	<b>13</b>
5.1. <i>Heatmaps</i> .....	13
5.2. PcoA .....	14
5.3. Gráficos de barras .....	16
<b>6. Contribuições.....</b>	<b>17</b>
<b>7. Bibliografia .....</b>	<b>18</b>
<b>8. Anexos.....</b>	<b>20</b>

## Índice de figuras

Figura 1.1 - Pipeline do Artigo.....	2
Figura 3.1 - <i>Workflow</i> para replicação de <i>outputs</i> .....	5
Figura 4.1 - <i>Heatmap</i> em várias amostras de DNA.....	10
Figura 4.2 - <i>Heatmap</i> em várias amostras de RNA.....	10
Figura 4.3 - PCoA.....	11
Figura 4.4 - Gráfico de barras em várias amostras de DNA e RNA.....	12
Figura 5.1 – Grafico NMDS.....	15

## 1. Introdução

Uma das principais fontes de emissão de gases de efeito estufa na agricultura é a produção de metano em solos de arrozais, como se pode ver no artigo (Rajendran et al., 2024). Tendo em conta que estes são ambientes ideais para a metanogênese devido às condições anaeróbicas criadas pela inundação dos campos, é importante termos em conta que, neste contexto, a dinâmica das comunidades microbianas metanogénicas, que são responsáveis pela produção de metano é fundamental para entender/analisar, e potencialmente mitigar essas emissões.

As comunidades microbianas metanogénicas são compostas por arqueias metanogénicas, nas quais desempenham o papel principal na decomposição da matéria orgânica em ambientes anóxicos (ambientes onde o oxigênio está ausente ou presente em concentrações extremamente baixas), que resultam na produção de metano, como se pode ver no artigo (Castillo et al., 2015)). A composição e a atividade destas comunidades são influenciadas por várias condições ambientais, incluindo a disponibilidade de água. Já o *stress* de dessecação, causado por períodos de seca seguidos de reidratação, pode alterar significativamente a estrutura e a função dessas comunidades microbianas, como se pode ver em (Wellsbury et al., 1993).

A resposta das comunidades metanogénicas ao *stress* de dessecação é um campo de estudo particularmente importante, especialmente em regiões que enfrentam variações sazonais de humidade. Estudos demonstraram que as comunidades microbianas em solos de sequeiro são mais resilientes ao *stress* de dessecação comparadas às de solos irrigados (Reim et al., 2017). Este fenómeno pode ser explicado pelas adaptações evolutivas que permitem às arqueias metanogénicas em solos de sequeiro sobreviverem e continuarem a produzir metano após a reidratação.

O estudo "Response of Methanogenic Microbial Communities to Desiccation Stress in Flooded and Rain-Fed Paddy Soil from Thailand" visou investigar como as comunidades microbianas metanogénicas respondem ao *stress* de dessecação em diferentes tipos de solos de arrozais.

A principal hipótese do estudo era que as comunidades microbianas em solos de sequeiro exibiriam maior resiliência ao *stress* de dessecação devido à sua adaptação a flutuações de humidade.

## 1.1. Pipeline

De acordo com o pipeline da figura 1, algumas das etapas são:

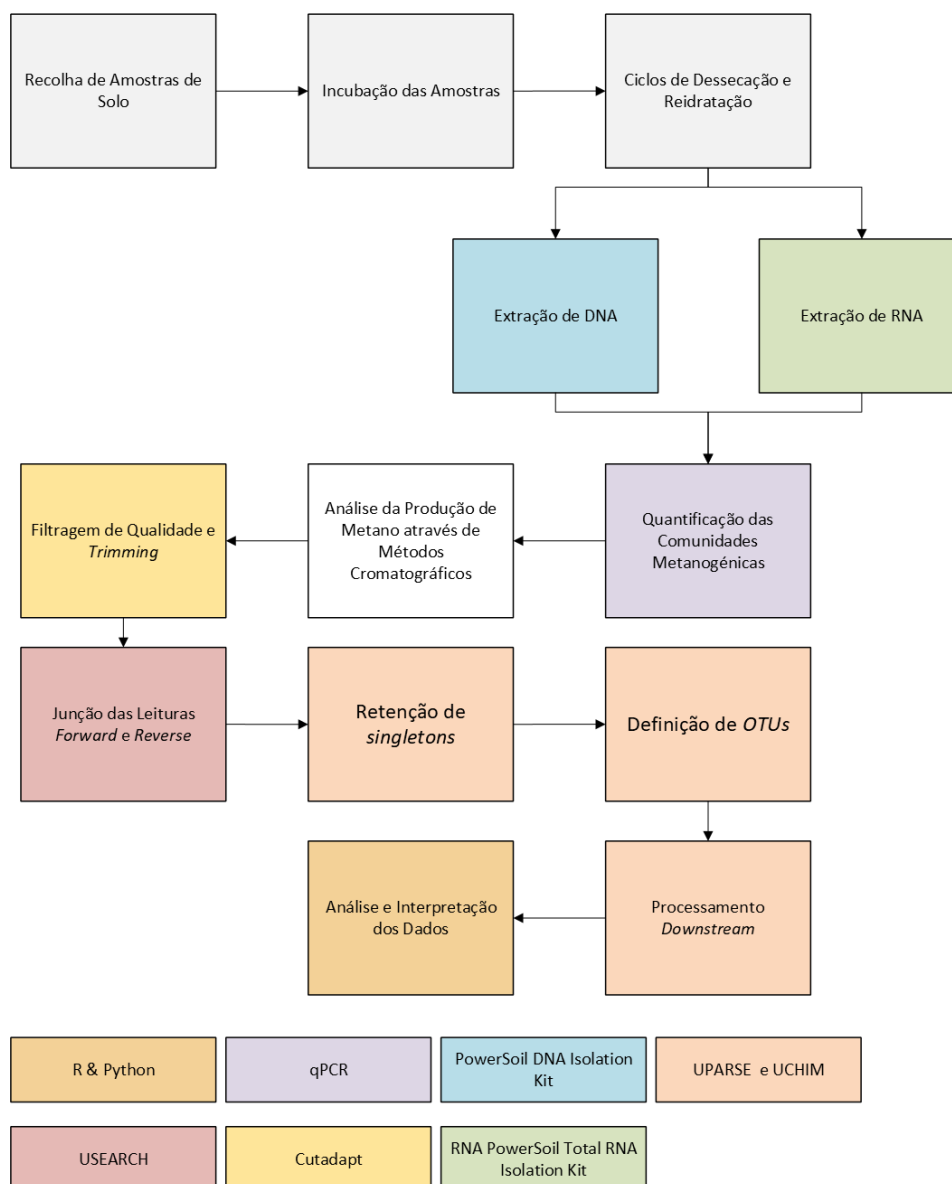


Figura 1.1 - Pipeline do Artigo

Tabela 1.1 - Etapas da pipeline

Recolha de amostras de solo	Amostras de solo foram recolhidas de arrozais inundados e de sequeiro na Tailândia.
Incubação	As amostras de solo foram incubadas sob condições controladas.
Ciclos de dessecação e reidratação	As amostras foram submetidas a ciclos de dessecação (seca) e reidratação para simular as variações naturais de humidade.
Extração de DNA e RNA	O DNA e o RNA foram extraídos das amostras para posterior análise.
Quantificação das comunidades metanogénicas por qpcr	A quantidade de arqueias metanogénicas foi medida usando a técnica de qPCR.
Análise da produção de metano através de métodos cromatográfico	A produção de metano foi quantificada utilizando métodos cromatográficos, como a cromatografia gasosa.

Além dos procedimentos experimentais que se encontram na Tabela 1.1, foi utilizado um *pipeline* bioinformático específico para analisar os dados de sequenciamento, que podemos também observar na Figura 1.1 no qual tem como etapas:

- Filtragem de qualidade e *trimming*: O primeiro passo foi a filtragem de qualidade e o *trimming* dos adaptadores *forward* e *reverse* das sequências usando a ferramenta cutadapt (Martin, 2011).
- Junção das leituras *forward* e *reverse*: Em seguida, as leituras *forward* e *reverse* foram unidas usando o comando *usearch fastq\_mergepairs* (Edgar, 2013).
- Processamento downstream: O processamento *downstream* foi realizado utilizando os pipelines UPARSE e UCHIME (Edgar et al., 2011)
- Descartamento de sequências curtas: Sequências menores que 250 bp foram descartadas.
- Retenção de *singletons*: A retenção de *singletons* refere-se à manutenção de sequências únicas (*singletons*) que aparecem apenas uma vez no conjunto de dados, ao invés de descartá-las.

- Definição de OTUs: As unidades taxonômicas operacionais (OTUs) foram definidas a um nível de identidade de sequência de 97%.

Estes procedimentos permitiram uma análise detalhada das mudanças nas comunidades microbianas e na produção de metano em resposta ao *stress* ambiental.

Por fim os resultados do estudo original indicaram que as comunidades metanogénicas em solos de sequeiro não apenas sobreviveram ao *stress* de dessecação, mas também revelaram um aumento na produção de metano após a reidratação, sugerindo uma robustez adaptativa a condições flutuantes de humidade.

Concluindo assim que resultados obtidos são consistentes com a literatura existente, que sugere que as comunidades microbianas em ambientes frequentemente sujeitos à seca desenvolvem mecanismos de resistência ao *stress*, como se pode ver nos artigos anteriormente referidos de Fetzer, Bak & Conrad (1993) e de Conrad (2007).

## 2. Objetivos

O objetivo do nosso estudo é reanalisar estes dados utilizando métodos bioinformáticos aprendidos em sala de aula, e comparar os resultados obtidos com os do artigo (Reim et al., 2017), ajudando a compreender melhor as comunidades microbianas presentes e os seus possíveis papéis no ecossistema.

Para isso serão realizadas várias tarefas, como:

- Obtenção dos ficheiros utilizadas no artigo original a partir do banco de dados especificado, utilizando os números de acesso fornecidos;
- Reanálise dos dados utilizando métodos bioinformáticos aprendidos em sala de aula, construção de gráficos de *heatmap* e PCoA e gráficos de abundância relativa;
- Comparação dos resultados obtidos com os do artigo original para verificar a compatibilidade;

- Garantir a reprodutibilidade das análises realizadas, disponibilizando todos os scripts e métodos utilizados em um repositório público, como GitHub.

### 3. Métodos e Materiais

Para a replicação dos resultados apresentados no artigo, tornou-se necessário o desenvolvimento de um *workflow*, que permitisse a criação de um *heatmap*, de uma análise de coordenadas principais (*Principal Coordinate Analysis*, PCoA), e de um gráfico de abundância relativa, como mostra a **Erro! A origem da referência não foi encontrada..**

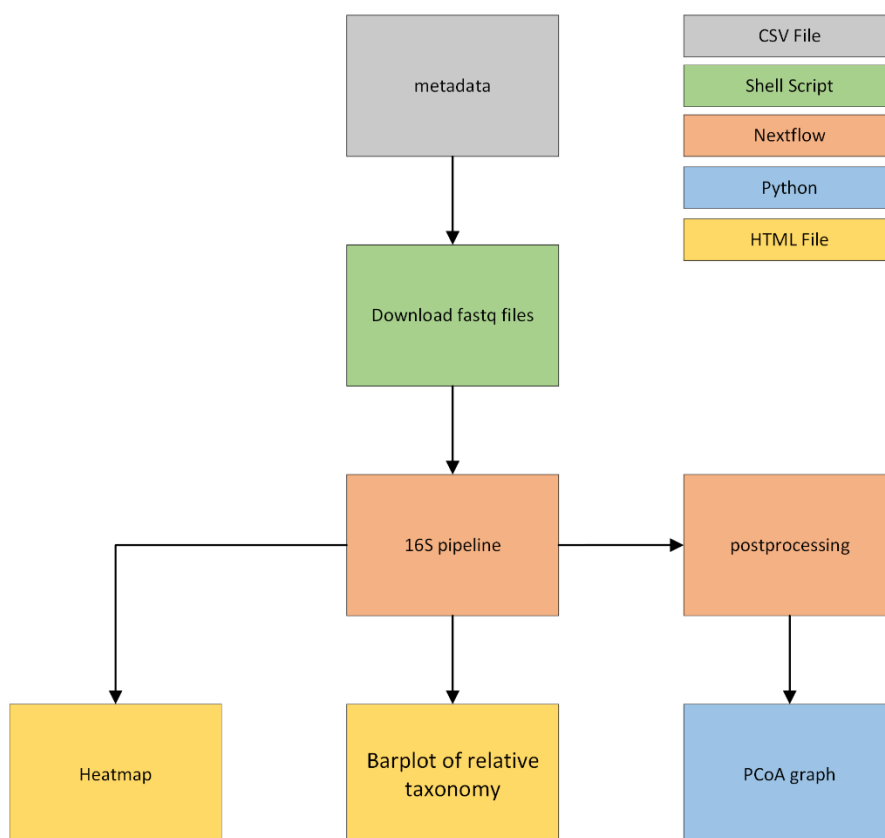


Figura 3.1 - *Workflow* para replicação de *outputs* com os respetivos *softwares/ferramentas* utilizados



## 1.2. Metadata

Antes de iniciar o *workflow*, foi necessário obter os metadados do artigo. Para tal, foi necessário ir à secção de *Bioproject* no *website* do NCBI (*National Center for Biotechnology Information*), pesquisando pelas referências PRJNA362531 e PRJNA362529. Através desta pesquisa, foi possível identificar os ficheiros SRA (*Sequence Read Archive*) associados a cada *bioproject*. Dado o elevado número de ficheiros, recorreu-se à ferramenta "*Send results to Run selector*" do NCBI, que permitiu o *download* de um ficheiro TXT (por cada referencia), que foi convertido para CSV, contendo todas as informações de cada ficheiro SRA (ver exemplo [aqui](#)), como:

- *Run*;
- *Assay Type*;
- *BioSample*;
- *Organism*;
- *Experiment*;
- Etc...

Contudo, no ficheiro de metadados, eram necessárias apenas duas colunas: "Run", que permite o *download* dos ficheiros, e "isolation\_source", que contém detalhes sobre cada ficheiro, como, por exemplo, "*DNA-dreid-irrigated-rep1*". Por conseguinte, todas as colunas, exceto as duas mencionadas, foram removidas manualmente juntamente com uma ligeira alteração no nome, como mostra

Tabela 3.1 - Metadados modificado do *bioproject* PRJNA362529

Run	isolation_source
SRR5192369	DNA-dreid-irrigated-rep1
SRR5192371	DNA-dreid-irrigated-rep2
SRR5192373	DNA-dreid-irrigated-rep3
SRR5192377	DNA-dreid-irrigated-rep4
SRR5192266	DNA-fresh-irrigated-rep1
SRR5192264	DNA-fresh-irrigated-rep2
SRR5192275	DNA-fresh-irrigated-rep3
SRR5192360	DNA-fresh-irrigated-rep4
....	.....

Os metadados alterados podem ser encontrados no repositório [Github](#).

## SRA para FASTQ

Com o ficheiro de metadados, foi utilizado um [script](#) em Shell(Blum, 2008), que permite fazer o *download* dos ficheiros FASTQ utilizando o SRA-Tools (v.3.11) (Sayers et al., 2022), e modificar o nome dos ficheiros para corresponder aos nomes presentes no ficheiro de metadados. Todos os ficheiros FASTQ (dos dois conjuntos de metadados) foram guardados na pasta "fastq", e todos eram *single-end*.

## 1.3. 16s Pipeline

Nesta etapa, foi utilizada uma *pipeline* desenvolvida por [hawaiidatascience](#) (*MetaFlow|mics: Nextflow Pipelines for Metagenomics — MetaFlow|mics 0.1 Documentation*, n.d.) para a análise de comunidades microbianas com sequências do gene 16S. Esta *pipeline* inclui etapas para filtragem e *denoising*, fusão de leituras, filtragem de contigues, agrupamento de OTUs, atribuição taxonómica e pós-processamento. Utiliza o Dada2 para desinfeção, o Mothur para agrupamento e filtragem de OTUs, além de diversas ferramentas para classificação taxonómica e cálculo de métricas de diversidade. Requer ficheiros FASTQ demultiplexados, e a base de dados SILVA, produzindo tabelas de abundância e taxonomia, sequências de OTUs e métricas de diversidade.

Além dos ficheiros necessários mencionados, é necessário também ter instalados o Java (*Learning Java - Patrick Niemeyer, Jonathan Knudsen - Google Books*, n.d.) (11.0.23), Docker (Boettiger, 2015)(26.1.4), Python (*Python*, 2006) (3.11.5) e Nextflow (Patel et al., 2024) (24.04.2.5914), sendo estas as versões utilizadas para correr a pipeline. Foram necessárias algumas alterações no ficheiro "nextflow.config" da pipeline, tais como:

- ***min\_overlap*** = 0: A sobreposição mínima entre leituras durante a fusão de sequências foi ajustada para 0;
- ***silva\_release*** = "seed": A versão da base de dados SILVA foi alterada para a versão "seed";

- ***clustering\_thresholds*** = "97": O limiar de agrupamento de OTUs foi ajustado para 97%, correspondendo a uma similaridade de 97%;
- ***beta\_diversity*** = "braycurtis-thetayc-sharedsobs-sharedchao": As métricas de diversidade beta foram configuradas para incluir Bray-Curtis, ThetaYC, SharedSobs e SharedChao;
- ***Single\_end*** = *true*: Esta configuração foi aplicada, visto que os ficheiros FASTQ utilizados eram *single-end*, e não *pair-end*.

Com estas alterações, foi executado um [script](#) em Shell para iniciar a pipeline. Este script contém:

- A localização da pasta dos ficheiros FASTQ;
- A referência alinhada;
- A referência taxonómica;
- O número de CPUs utilizados.

Vale salientar que foi utilizado o *profile* Docker, para evitar a necessidade de instalar todos os pacotes necessários para correr o *script*, e que todos os parâmetros podem ser modificados. Após a conclusão da pipeline, é gerada uma pasta que contém todos os *outputs* desejados, incluindo um *heatmap* em formato HTML (*HyperText Markup Language*) (HTML & XHTML: The Definitive Guide: The Definitive, n.d.), e o ficheiro CSV "*beta-diversity.97.summary*", que proporciona um resumo das análises de diversidade beta realizadas. Utilizando um [script](#) em Python juntamente com o ficheiro CSV mencionado, foi possível criar um gráfico de PCoA.

## 4. Resultados

Neste capítulo, serão apresentados os resultados obtidos a partir do *workflow* desenvolvido. Contudo, é necessário salientar que, infelizmente, os resultados não são compatíveis com os do artigo de (Reim et al., 2017). Enquanto no referido artigo

é realizada uma separação entre bactérias e archaea, os resultados seguintes utilizam a base de dados SILVA 138\_1, que engloba ambos os grupos.

## 1.4. Heatmap

Os *heatmap* representam a diversidade e abundância de unidades taxonômicas operacionais (OTUs) de DNA associadas a diferentes condições de amostras ambientais. As amostras são identificadas verticalmente e incluem várias repetições (rep1, rep2, rep3) para condições de DNA fresco e reidratado ("*fresh*", "*rained*"), incubado e reidratado ("*incubated*", "*irrigated*"), e seco e reidratado ("*dried*", "*irrigated*").

Horizontalmente, as OTUs são categorizadas por espécies, com uma extensa variedade que inclui diversas linhagens de *Bacillus*, *Pseudomonas*, *Corynebacterium*, entre outras. A intensidade da cor no mapa de calor varia de azul-claro a azul-escuro, indicando respetivamente baixa a alta abundância relativa de OTUs nas amostras.

Observa-se na Figura 4.1 uma variação significativa na abundância de OTUs entre as amostras tratadas diferentemente, sugerindo que os métodos de tratamento (como secagem, irrigação e incubação) influenciam a composição bacteriana. Em particular, as amostras reidratadas após secagem, e aquelas frescas sujeitas à mostram padrões distintos de abundância em comparação com as amostras incubadas e irrigadas.

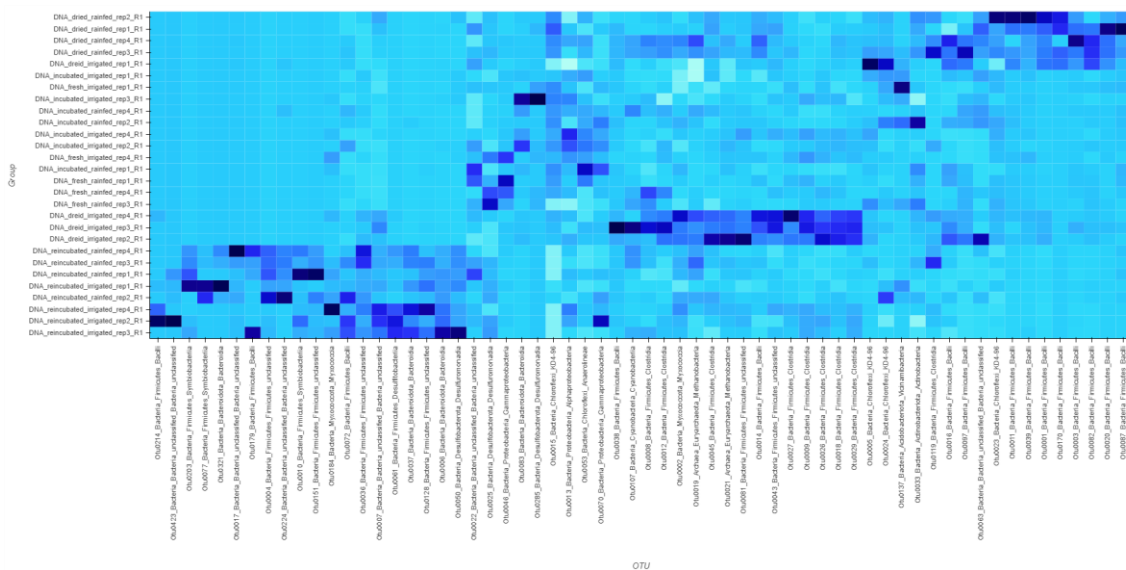


Figura 4.1 - *Heatmap* representando a abundância relativa de diferentes OTUs em várias amostras de DNA

Na Figura 4.2 representa o *heatmap* para as amostras de RNA, e é possível verificar apresenta resultados semelhantes a da figura anterior, e que apresenta um número maior de OTUs.

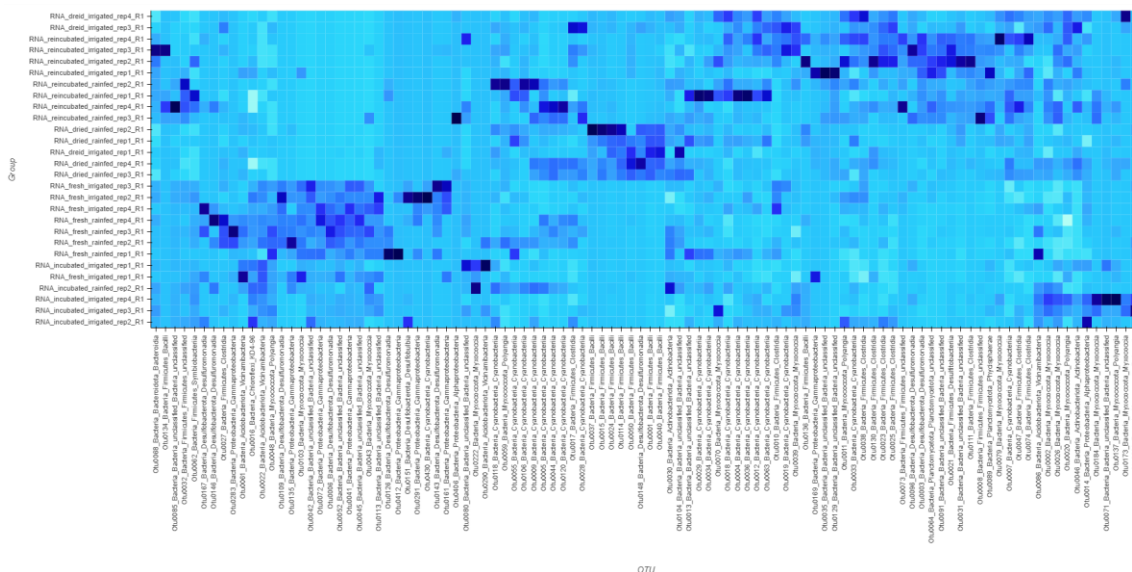


Figura 4.2 - *Heatmap* representando a abundância relativa de diferentes OTUs em várias amostras de RNA

Ambos os heatmaps tinha predominantemente OTUs do grupo *Bacteria*, contudo havia também presentes numa escala menor (2) OTUs do grupo *Archaea*.

## 1.5. PCoA

A análise do PCoA sugere a existência de uma estruturação marcante nas amostras, que parece ser influenciada tanto pelas condições de irrigação quanto pelo tipo de material genético examinado (DNA ou RNA).

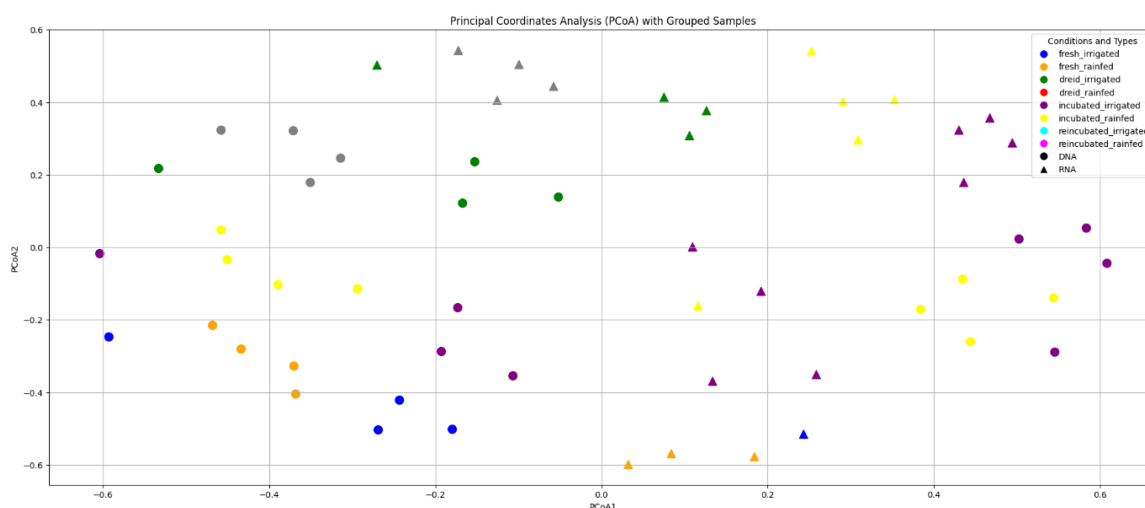


Figura 4.3 - PCoA

Na Figura 4.3 é possível verificar que as amostras de DNA (representadas por círculos) e RNA (representadas por triângulos) tendem a formar grupos distintos ao longo do eixo PCoA1, o que pode indicar diferenças significativas na composição ou na expressão genética entre os dois tipos de material.

As cores das amostras indicam diferentes tratamentos e estados, e é possível verificar que esses grupos parecem mostrar uma certa agrupação baseada nesses fatores

Por fim as amostras “*fresh*”, seja “*irrigated*” ou de “*rainfed*”, tendem a agrupar-se mais perto do centro do gráfico. Enquanto que as amostras “*dreid*” e “*incubated*”

mostram maior dispersão no espaço do PCoA, especialmente as *incubated*, que se estendem mais ao longo do eixo PCoA2.

## 1.6. Gráfico de abundância

Na figura Figura 4.4 apresentada é um gráfico de barras empilhadas. É possível observa-se uma significativa variação na composição microbiana entre os diferentes tratamentos, e condições de irrigação.

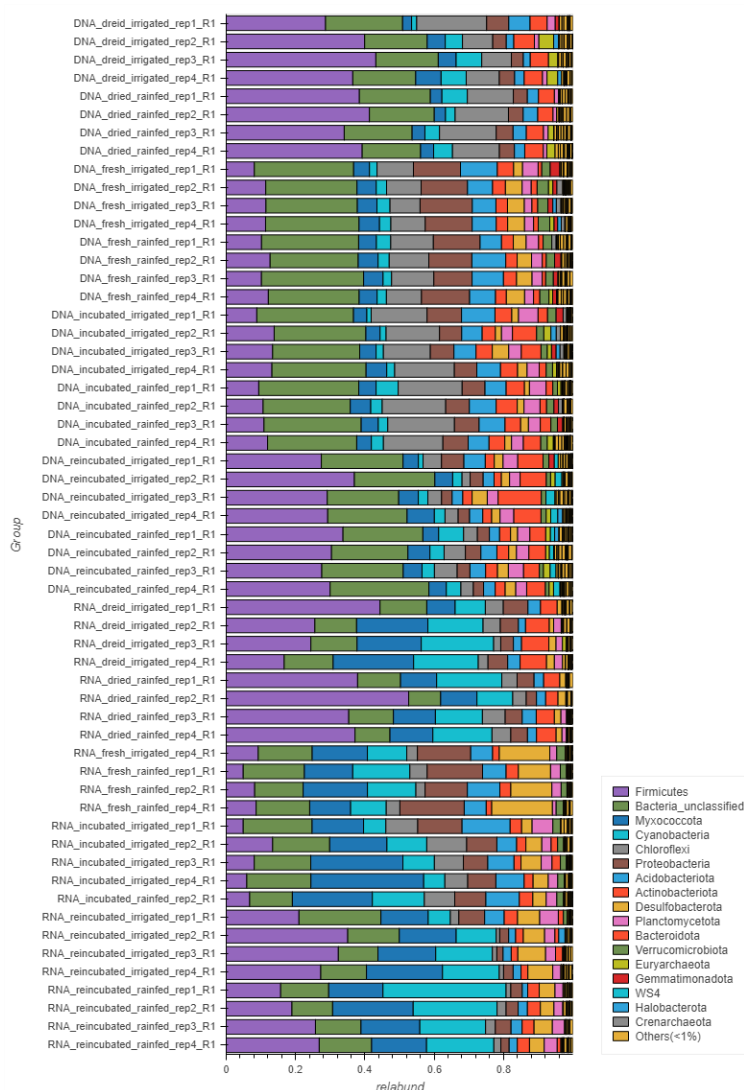


Figura 4.4 - Gráfico de barras empilhadas de abundância relativa de diferentes grupos taxonómicos em várias amostras de DNA e RNA.

Em muitas das amostras, os *Firmicutes* (indicados em roxo) parecem ser o grupo dominante, especialmente em amostras tratadas com condições "dried" e "reincubated".

Comparando amostras de DNA e RNA, nota-se uma diferença na composição microbiana. Por exemplo, em amostras de RNA, a abundância de *Proteobacteria* (azul-claro) parece ser maior em comparação com as amostras de DNA. Também existe uma presença notável de grupos taxonómicos menores, indicados por várias cores diferentes, apesar de menos abundantes individualmente, contribuem para a diversidade total do microbioma em cada amostra.

Por fim a categoria "Others(<1%)", que esta com maior presença no RNA "fresh" indica a existência de uma rica diversidade microbiana que não é dominada por nenhum grupo específico.

## 5. Discussão

### 1.7. Heatmaps

Nas comparar as figuras 5 e 7 do artigo (Reim et al., 2017) com a Figura 4.1 e Figura 4.2, observou-se que, apesar de algumas diferenças na apresentação, os resultados obtidos são essencialmente os mesmos. A principal distinção reside na forma como os dados estão visualizados nos gráficos e na combinação das comunidades microbianas num único gráfico, invés de estarem separados em *Archaea* e *Bacteria* como no artigo.

No artigo(Reim et al., 2017), as amostras de *Archaea* e *Bacteria* estão separadas em gráficos diferentes, permitindo-nos entender as dinâmicas específicas de cada grupo microbiano. Nos *heatmaps* replicados, ambas as comunidades estão integradas numa única visualização, permitindo ainda uma análise comparativa geral. Adicionalmente, as figuras do artigo apresentam uma árvore filogenética e uma divisão de grupos (fornecendo uma compreensão mais profunda das relações



evolutivas e das afinidades entre os diferentes OTUs) que não estão presentes nos *heatmaps* replicados.

No que toca aos resultados/visualizações obtidas ambos os conjuntos de gráficos mostram que as condições de incubação e secagem têm um impacto significativo na composição microbiana. Também é possível concluir que o impacto das condições a que foram expostos varia ligeiramente entre amostras de DNA e RNA. Esta visualização é obtida tanto nos *heatmaps* originais como nos replicados.

Apesar das diferenças na apresentação, e no método de análise os resultados são consistentes entre os dois conjuntos de gráficos. Ambos mostram que as condições experimentais e o tipo de composição das amostras influenciam significativamente a composição microbiana.

## **1.8. PcoA**

Ao comparar a Figura 5.1 (que corresponde a Figura 4 do artigo) com a replicação do mesmo Figura 4.3, observou-se que, apesar de algumas diferenças na apresentação, os resultados obtidos são essencialmente os mesmos. A principal distinção reside na forma como os dados estão visualizados nos gráficos e no método de análise utilizado.

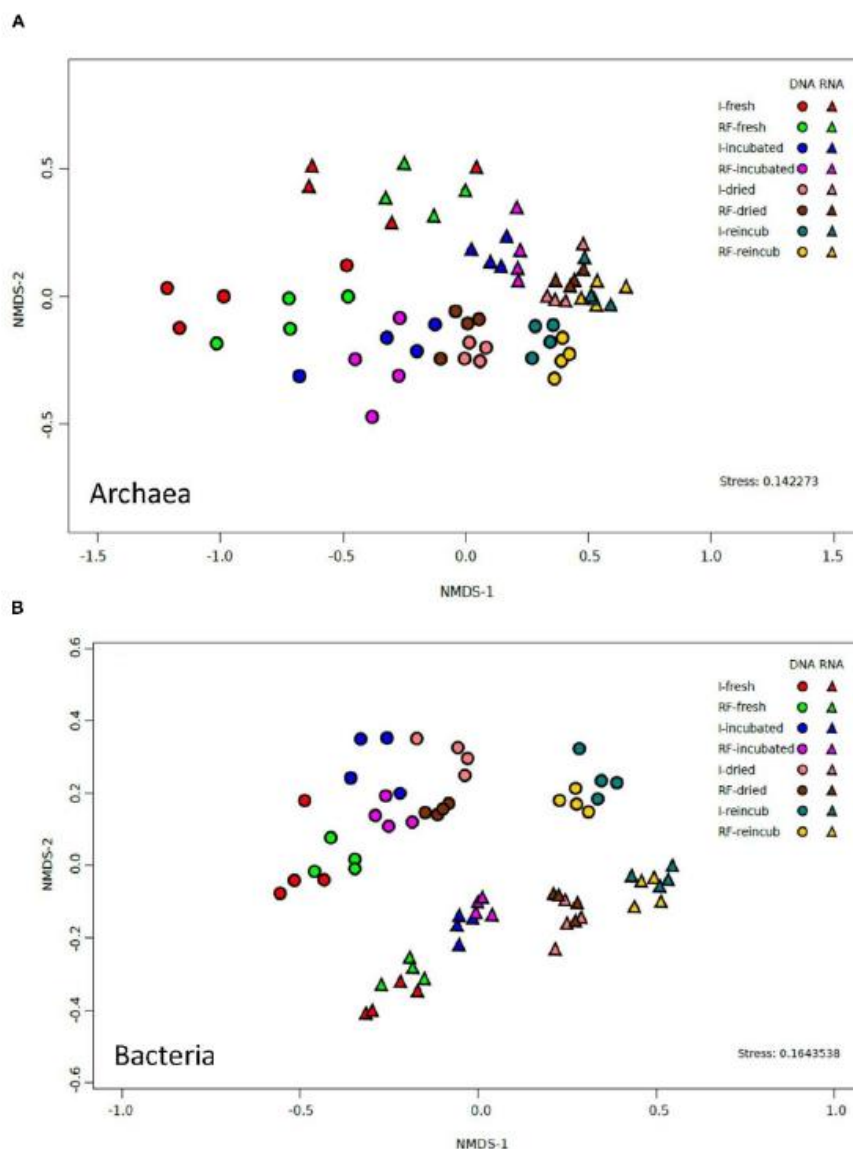


Figura 5.1 – Gráfico NMDS retirado de (Reim et al., 2017)

No gráfico da Figura 5.1 utilizou-se NMDS (*Non-metric Multidimensional Scaling*), que é uma técnica não-métrica de ordenação, particularmente útil para dados de biodiversidade, onde as distâncias não obedecem necessariamente a uma escala linear, enquanto na Figura 4.3 utilizou-se PCoA (*Principal Coordinates Analysis*), que é uma técnica métrica de ordenação, frequentemente utilizada para visualizar a similaridade ou dissimilaridade entre amostras. Apesar das diferenças na apresentação e no método de análise, os resultados são consistentes entre os dois

estudos. Ambos os gráficos mostram que as condições de incubação e secagem têm um impacto significativo na composição microbiana, tanto de *Archaea* quanto de *Bacteria*.

## 1.9. Gráficos de barras

Ao analisar as Figuras 3 e 6 do artigo (Reim et al., 2017) em comparação com a sua replicação Figura 4.4, percebeu-se que, embora haja algumas discrepâncias na apresentação, os resultados fundamentais são congruentes. A principal divergência reside na abordagem de visualização dos dados nos gráficos.

Os gráficos do artigo foram separados em quatro gráficos de barras diferentes. Os *Archaea* e *Bacteria* são apresentadas em gráficos diferentes e, dentro dessa divisão, dividiram-se novamente por DNA e RNA, permitindo uma análise mais detalhada e específica das dinâmicas dentro de cada grupo microbiano. Em contrapartida, o gráfico replicado, contém toda a informação dos gráficos originais num só gráfico. Apesar desta diferença, o gráfico replicado suporta as principais conclusões do artigo original.

Os resultados essenciais são consistentes entre o gráfico replicado e os gráficos originais, ambos os gráficos mostram que as condições de incubação, secagem e irrigação têm um impacto significativo na composição microbiana e as diferenças na resposta das comunidades microbianas entre amostras de DNA e RNA. Isto é, o tipo de amostra e as condições em que se encontra influencia a composição microbiana.

Podemos concluir que, embora haja diferenças na apresentação dos dados, os resultados obtidos são compatíveis e reforçam as conclusões originais sobre o impacto das condições experimentais na composição das comunidades microbianas.

## 6. Contribuições

Cada membro do grupo teve um tarefas específica consoante a sua área de especialidade:

- Joana: Ficou responsável pela introdução, a criação do pipeline do artigo, os objetivos e na conclusão. Durante o desenvolvimento destas tarefas, foi possível melhorar a forma de interpretar artigos científicos, bem como aprimoraras minhas habilidades de escrita e análise crítica.
- Lara: Pelo desenvolvimento dos gráficos, e do script para melhorar o gráfico de barras.
- Lucas: Ficou responsável pela discussão, parte da introdução e pipeline do artigo. Sente-se que melhorou as suas habilidades analíticas através da discussão e crítica dos resultados habilidades técnicas no design e gestão do pipeline.
- Rendrick: Ficou responsável pelo desenvolvimento do tópico dos materiais e métodos (*Workflow*, github e *Shell Scripts*), assim como na apresentação dos resultados. Com estas tarefas foi possível aprimorar a capacidade de apresentar de forma critica os resultados obtidos.

## 7. Bibliografia

- Blum, R. (2008). *Linux command line and shell scripting bible*. 809.  
[https://books.google.com/books/about/Linux\\_Command\\_Line\\_and\\_Shell\\_Scripting\\_B.html?id=dKxCcDn24glC](https://books.google.com/books/about/Linux_Command_Line_and_Shell_Scripting_B.html?id=dKxCcDn24glC)
- Boettiger, C. (2015). An introduction to Docker for reproducible research. *ACM SIGOPS Operating Systems Review*, 49(1), 71–79.  
<https://doi.org/10.1145/2723872.2723882>
- Castillo, N. A., Valdez, A. L., & Fariña, J. I. (2015). Microbial production of scleroglucan and downstream processing. *Frontiers in Microbiology*, 6(OCT), 145963.  
<https://doi.org/10.3389/FMICB.2015.01106/BIBTEX>
- Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* 2013 10:10, 10(10), 996–998.  
<https://doi.org/10.1038/nmeth.2604>
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27(16), 2194–2200. <https://doi.org/10.1093/BIOINFORMATICS/BTR381>
- HTML & XHTML: The Definitive Guide: The Definitive Guide - Chuck Musciano, Bill Kennedy - Google Books*. (n.d.). Retrieved June 27, 2024, from [https://books.google.pt/books?hl=en&lr=&id=O5Vpww0wTYUC&oi=fnd&pg=PR7&dq=html&ots=Y0t\\_wqBLXM&sig=oVA67KW6o-98qFYEVqcapeTy9LE&redir\\_esc=y#v=onepage&q=html&f=false](https://books.google.pt/books?hl=en&lr=&id=O5Vpww0wTYUC&oi=fnd&pg=PR7&dq=html&ots=Y0t_wqBLXM&sig=oVA67KW6o-98qFYEVqcapeTy9LE&redir_esc=y#v=onepage&q=html&f=false)
- Learning Java - Patrick Niemeyer, Jonathan Knudsen - Google Books*. (n.d.). Retrieved June 27, 2024, from [https://books.google.pt/books?hl=en&lr=&id=T9abAgAAQBAJ&oi=fnd&pg=PR5&dq=java&ots=vclkiHR7oG&sig=4oT1icEWF11HxKfpShzYvsVjlZs&redir\\_esc=y#v=onepage&q&f=false](https://books.google.pt/books?hl=en&lr=&id=T9abAgAAQBAJ&oi=fnd&pg=PR5&dq=java&ots=vclkiHR7oG&sig=4oT1icEWF11HxKfpShzYvsVjlZs&redir_esc=y#v=onepage&q&f=false)
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, 17(1), 10–12.  
<https://doi.org/10.14806/EJ.17.1.200>

- MetaFlow|mics: Nextflow pipelines for metagenomics — MetaFlow|mics 0.1 documentation.* (n.d.). Retrieved June 27, 2024, from <https://metagenomics-pipelines.readthedocs.io/en/latest/#>
- Patel, Y., Zhu, C., Yamaguchi, T. N., Bugh, Y. Z., Tian, M., Holmes, A., Fitz-Gibbon, S. T., & Boutros, P. C. (2024). NFTest: automated testing of Nextflow pipelines. *Bioinformatics*, 40(2). <https://doi.org/10.1093/BIOINFORMATICS/BTAE081>
- Python.* (2006).
- Rajendran, S., Park, H., Kim, J., Park, S. J., Shin, D., Lee, J. H., Song, Y. H., Paek, N. C., & Kim, C. M. (2024). Methane Emission from Rice Fields: Necessity for Molecular Approach for Mitigation. *Rice Science*, 31(2), 159–178. <https://doi.org/10.1016/J.RSCI.2023.10.003>
- Reim, A., Hernández, M., Klose, M., Chidthaisong, A., Yuttitham, M., & Conrad, R. (2017). Response of methanogenic microbial communities to desiccation stress in flooded and rain-fed paddy soil from Thailand. *Frontiers in Microbiology*, 8(MAY), 260002. <https://doi.org/10.3389/FMICB.2017.00785/BIBTEX>
- Sayers, E. W., O'Sullivan, C., & Karsch-Mizrachi, I. (2022). Using GenBank and SRA. *Methods in Molecular Biology*, 2443, 1–25. [https://doi.org/10.1007/978-1-0716-2067-0\\_1](https://doi.org/10.1007/978-1-0716-2067-0_1)
- Wellsbury, P., Herbert, R. A., & John Parkes, R. (1993). Incorporation of [methyl-3H]thymidine by obligate and facultative anaerobic bacteria when grown under defined culture conditions. *FEMS Microbiology Ecology*, 12(2), 87–95. <https://doi.org/10.1111/J.1574-6941.1993.TB00020.X>

## **8. Anexos**

1 - Materiais e Métodos do artigo (Reim et al., 2017)

Materiais e Métodos Usados no artigo “Response of Methanogenic Microbial Communities to Desiccation Stress in Flooded and Rain-Fed Paddy Soil from Thailand” de Conrad (2017) são:

**“Sampling”:**

- Solos recolhidos de arrozais irrigados e de sequeiro na província de Ang Thong, Tailândia.
- Solos transportados em estado fresco, de seguida congelados e armazenados a -18°C até o início dos experimentos.

**Condições de Incubação:**

- Preparação de suspensões anóxicas de solo fresco: 10 g de solo fresco colocados em tubos de pressão de vidro de 26 ml.
- Incubação a 25°C por 1 mês.
- Secagem ao ar por 2 semanas a 35°C.
- Nova inundação simulada: solo seco (6 g) misturado com 6 ml de água estéril anóxica, incubado em tubos de vidro de 26 ml a 25°C por mais 7 semanas.
- Controlo de metagénese acetoclástica com fluoreto de metila (3%).

**Medição de Gases:**

- CH<sub>4</sub> e CO<sub>2</sub> foram analisados por cromatografia gasosa (GC).
- $\delta^{13}\text{C}$  foi analisado por espectrometria de massa por razão isotópica de combustão GC (GC-C-IRMS).
- $\delta^{13}\text{C}$  da matéria orgânica foi analisado pelo Centro de Pesquisa e Análise de Isótopos Estáveis (KOSI) da Universidade de Göttingen usando um analisador elementar acoplado a um IRMS.



- A fração (fH<sub>2</sub>) da produção de CH<sub>4</sub> por metagénese hidrogenotrófica foi calculada por balanço de massa como descrito anteriormente (Conrad et al., 2010) usando a fórmula:

### **Extração de DNA e RNA e qPCR:**

- De seguida foi extraído ácido nucleico usando uma versão modificada do protocolo de Bürgmann et al. (2001).
- Purificação de ácido nucleico por extração com fenol/clorofórmio.
- Tratamento de subset de ácido nucleico total com DNase, verificação da digestão completa do DNA por PCR do rRNA 16S bacteriano e transcrição reversa para cDNA.
- Determinação da abundância de rRNA 16S bacteriano, rRNA 16S arqueano e genes mcrA metanogénicos por qPCR.

### **Preparo da Biblioteca Illumina e Sequenciamento:**

- Amplificação da região V4 do gene rRNA 16S com primers específicos.
- Condições de PCR: desnaturação inicial a 94°C por 5 min, seguida por 28 ciclos de 94°C por 30 s, 50°C por 30 s, 68°C por 30 s, e uma extensão final a 68°C por 10 min.
- Purificação e quantificação dos amplicons.
- Sequenciamento utilizando o protocolo MiSeq 2 bp × 300 bp.

### **Análises Bioinformáticas:**

- Filtro de qualidade e corte dos adaptadores das sequências utilizando cutadapt.
- União das leituras utilizando usearch fastq\_mergepairs.
- Definição de OTUs a um nível de identidade de sequência de 97% utilizando UPARSE e UCHIME.

### **Análises Estatísticas e Classificação de OTUs:**

- Normalização dos dados a partir de transformações de Hellinger utilizando o pacote vegan no R.
- Diversidade beta analisada utilizando a ordenação NMDS das distâncias de Hellinger.
- Análise de componentes principais (PCA) dos dados transformados por Hellinger.
- Agrupamento hierárquico da matriz de distância usando o método "ward".

### **Análise de Taxonomia:**

Alinhamento de sequências representativas de cada OTU contra o banco de dados SILVA utilizando o classificador bayesiano ingênuo.