

CONSIDERAÇÕES SOBRE A ANÁLISE DE DADOS DE METAGENÔMICA OBTIDOS DE AMOSTRAS DE SOLO

Roberto Marafon Leandro,¹ Emerson Danzer,² Glacy Jaqueline da Silva,³
Eduardo Vinicius Kuhn⁴

RESUMO: Nas últimas décadas, diversos avanços tecnológicos e incentivos voltados à pesquisa científica vêm fazendo com que o Brasil obtenha destaque internacional no âmbito da produção agrícola. Apesar disso, verifica-se ainda significativa negligência em relação à importância das comunidades microbianas edáficas, seja por falta de informação ou pela subestimação de seu papel, o que resulta em práticas de manejo que não conseguem explorar plenamente a capacidade produtiva do solo. Para mudar esse cenário, uma abordagem promissora se dá através do uso de técnicas avançadas de biotecnologia, tal como o sequenciamento metagenômico de amostras de solo. Análises desses dados oriundos do sequenciamento podem revelar informações valiosas sobre os microrganismos presentes na amostra, permitindo assim uma gestão mais eficiente e sustentável dos solos agrícolas. Nesse contexto, tem-se por objetivo aqui reproduzir a *pipeline* para análise de dados de metagenômica descrita por Ordine et al. (2023), utilizando-se dos dados sequenciados disponibilizados publicamente. Os resultados obtidos mantiveram, em linhas gerais, o padrão mais comum de genes de resistência antimicrobiana (ARGs) e de fatores de virulência (VFs) encontrados, i.e., o *vanRO* e o *acpXL*, respectivamente. Contudo, os demais resultados diferiram, em termos de quantidades e variedades, daqueles apresentados em Ordine et al. (2023). Diferenças essas que podem ter ocorrido devido a erros na reprodução da *pipeline* e/ou por falta de informações acerca dos *softwares* (e.g., versões, parâmetros, banco de dados de referência) utilizados originalmente na *pipeline*; portanto, investigações futuras são necessárias para elucidar detalhes pertinentes à *pipeline* de análise implementada.

Palavras-chave: Bioinformática, metagenômica, *pipeline*, sequenciamento *shotgun*.

1 INTRODUÇÃO

O Brasil possui destaque internacional no setor agrícola, sendo o 2º maior exportador de grãos no ano de 2020 (ARAGÃO; CONTINI, 2023). No ano de 2022, todas as atividades relacionadas ao setor agrícola brasileiro somaram 1,836 trilhão de reais, correspondendo a 18% do PIB (CONFEDERAÇÃO DA AGRICULTURA E PECUÁRIA DO BRASIL (CNA), 2024). Todavia, esse cenário atual de prestígio no setor de produção de alimentos nem sempre foi uma realidade; afinal, nas décadas de 1960 e 1970 a produção agrícola era precarizada e incapaz de suprir as demandas nacionais, fazendo com que a maior parte da comida consumida fosse importada. Entretanto, uma série de políticas públicas e

¹  Graduando em Engenharia Eletrônica, Universidade Tecnológica Federal do Paraná (UTFPR), campus Toledo, Paraná, Brasil; ✉ rleandro@alunos.utfpr.edu.br.

²  Programa de Pós-graduação em Bioinformática, Universidade Federal de Minas Gerais (UFMG), campus Belo Horizonte, Minas Gerais, Brasil; ✉ emersonwdanzer@gmail.com.

³  Programa de Pós-graduação em Biotecnologia aplicada à Agricultura, Universidade Paranaense (UNIPAR), campus Umuarama/Toledo, Paraná, Brasil; ✉ glacyjaqueline@prof.unipar.br.

⁴  Departamento de Engenharia Eletrônica, Universidade Tecnológica Federal do Paraná (UTFPR), campus Toledo, Paraná, Brasil; ✉ kuhn@utfpr.edu.br.

incentivos governamentais voltados para a pesquisa e ao desenvolvimento transformaram o cenário agrícola brasileiro, favorecendo a criação e o aperfeiçoamento de diversas técnicas de manejo; por exemplo, o plantio direto, o uso de agroquímicos e a avaliação de indicadores biológicos (**EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA (EMBRAPA), 2022; FAO, 2017; BUENO et al., 2021; FUENTES-LLANILLO et al., 2021**). Apesar dos diversos avanços no setor agrícola, ainda existe espaço para o aprimoramento e o desenvolvimento de novos métodos. Dessa forma, um fator usualmente negligenciado corresponde às comunidades microbianas que habitam o solo. Elas são importantes pois participam de diversos processos envolvendo a saúde do solo e das plantas, tais como de ciclagem de nutrientes, de resistência ao estresse e a patógenos e de regulação hormonal vegetal (**BANERJEE; VAN DER HEIJDEN, 2023; COMPANT et al., 2024**). Além disso, quando o solo é modificado através das práticas de manejo, o microbioma edáfico também sofre alterações, facilitando com que possíveis características se disseminem; por exemplo, a intensificação da resistência antimicrobiana (AMR) através da maior presença de genes de resistência a antibióticos (ARGs) e de fatores de virulência (VFs) (**ORDINE et al., 2023**).

Um método que pode ser empregado para o monitoramento do microbioma é o sequenciamento metagenômico do tipo *shotgun*, capaz de obter toda a informação genética de uma amostra extraída do ambiente sem a necessidade de cultivo ou isolamento de organismos. Quando comparada com outras técnicas, é possível contemplar informações mais completas acerca das comunidades analisadas, podendo-se identificar diversos genes como os ARGs e os VFs (**QUINCE et al., 2017; WARNECKE et al., 2007; BHARTI; GRIMM, 2019**). Entretanto, o processamento de dados advindos do sequenciamento metagenômico é uma tarefa desafiadora; afinal, o volume e a complexidade da informação gerada demandam recursos computacionais consideráveis, além do fato de existirem diversas ferramentas que podem ser utilizadas nas múltiplas etapas de processamento requisitadas. A combinação destes fatores, conjuntamente com a rápida e constante evolução tecnológica do setor, cria um cenário individual para cada estudo metagenômico, fazendo com que a *pipeline*¹ utilizada seja de difícil escolha. Evidencia-se que, até o momento, não existe uma *pipeline* padrão para a análise dos dados metagenômicos do solo (**BHARTI; GRIMM, 2019; LIU et al., 2021; RICHARDSON, L. et al., 2022**).

Neste contexto, usando os dados e a *pipeline* apresentada em **Ordine et al. (2023)**, tem-se por objetivo:

- reproduzir a *pipeline* de análise de dados metagenômicos;
- identificar genes de resistência a antibióticos (ARGs) e de fatores de virulência (VFs), componentes do resistoma;
- comparar os resultados obtidos com os resultados previamente encontrados.

¹ O termo se refere a um programa, a um *script* ou a um conjunto de programas (**LIU et al., 2021**).

2 METODOLOGIA

Para contextualizar o leitor, esta seção abordará diversas etapas metodológicas, sendo elas: 1) a explicação da estrutura típica de um trabalho metagenômico, 2) algumas das operações e dos *softwares* de processamento dos dados sequenciados e, por fim, 3) um breve detalhamento acerca do conjunto de dados e da *pipeline* presente em [Ordine et al. \(2023\)](#).

2.1 Etapas típicas da metagenômica

Geralmente são empregadas seis etapas. Em primeiro lugar, realiza-se a preparação da amostra, que consiste na obtenção do material que será estudado (o solo). O transporte e o armazenamento das amostras também deve ser realizado de forma atenciosa, evitando-se que o perfil do material genético sofra alterações. O máximo de informações (metadados) sobre os processos realizados e as características do local de coleta devem ser documentadas com o objetivo de servirem como dados complementares para as conclusões acerca dos futuros resultados obtidos ² ([QUINCE et al., 2017](#); [BHARTI; GRIMM, 2019](#)). Em segundo lugar, realiza-se a extração e a purificação do DNA das amostras. Diversos métodos podem ser utilizados, sejam eles mecânicos, químicos ou enzimáticos, sendo que cada procedimento carrega características próprias que alteram o material trabalhado de maneira específica. Diversos *kits* voltados para essa tarefa podem ser encontrados no mercado, como o *DNeasy PowerSoil Pro Kit* desenvolvido especificamente para amostras de solo e de lodo ([QUINCE et al., 2017](#); [QIAGEN, 2024a,b](#)). Em terceiro lugar, realiza-se a quantificação e o controle de qualidade do DNA extraído e purificado. Podem ser empregados os mais variados mecanismos nesta etapa, porém, destaca-se que a função principal é assegurar que o material genético apresente características adequadas de integridade e de pureza ([GREEN; SAMBROOK, 2012](#)). Em quarto lugar, realiza-se a construção da biblioteca, responsável por condicionar o DNA para os padrões da máquina de sequenciamento. Os fragmentos nucleicos são divididos em tamanhos específicos, normalmente entre 250 e 300 pares de bases e os adaptadores ³ são conectados a esses segmentos ([VAN DIJK; JASZCZYSZYN; THERMES, 2014](#); [ILLUMINA, 2024a](#)). A partir desse momento, o DNA poderá ser sequenciado, o que é realizado em quinto lugar pela etapa de sequenciamento. Ela ocorre em um equipamento que converte as informações nucleicas do mundo real para o ambiente eletrônico. Existem diversas gerações de sequenciamento e diversas máquinas disponíveis no mercado, porém, destaca-se o método apresentado pela empresa Illumina, uma plataforma relevante para o contexto de *Next Generation Sequencing* (NGS) abordado neste trabalho. Ela utiliza uma estrutura de sequenciamento por síntese (SBS); particularmente, os equipamentos relacionam valores de luminosidade com os nucleotídeos analisados, guardando as informações finais em um arquivo cuja extensão geralmente é

² A anotação de metadados é essencial em todas as etapas do estudo metagenômico, devendo ser realizada desde o momento da coleta das amostras até às conclusões do estudo ([QUINCE et al., 2017](#)).

³ Os adaptadores são sequências nucleicas específicas que atuam como códigos de barra para a máquina de sequenciamento, para o começo e o fim do sequenciamento dos fragmentos ([VAN DIJK; JASZCZYSZYN; THERMES, 2014](#)).

no padrão .fastq (BHARTI; GRIMM, 2019; ILLUMINA, 2024b,c; HU et al., 2021). Por fim, resta a etapa de análises de bioinformática, indispensável para o estudo metagenômico e foco deste trabalho. É nesse ponto que se obtém realmente as informações acerca dos dados analisados. As ferramentas utilizadas variam de acordo com os objetivos do trabalho. Todavia, evidencia-se a possibilidade de identificação de ARGs e de VFs, componentes do resistoma do solo (BHARTI; GRIMM, 2019; LIU et al., 2021).

2.2 Algumas etapas e ferramentas de bioinformática

Os dados brutos gerados pela etapa sequenciamento oferecem informações demasiadamente complexas para se extraírem conclusões a respeito das comunidades microbianas que compõem as amostras. Assim, várias operações computacionais devem ser executadas para que algum resultado seja realmente obtido. Além disso, com constantes avanços nas técnicas de sequenciamento, expansão da capacidade computacional e a construção de novas ferramentas de bioinformática, a escolha do procedimento para as análises caracteriza uma etapa que exige planejamento e cuidado. Qualquer manipulação feita tanto nas etapas anteriores quanto nesta etapa afeta os resultados; então, evidencia-se a necessidade de uma documentação abrangente acerca das ações realizadas (QUINCE et al., 2017; BHARTI; GRIMM, 2019). Para este trabalho, quatro grandes operações são realizadas, em ordem, sobre os dados advindos do sequenciamento, sendo elas:

- 1) **Controle de qualidade:** é uma etapa fundamental para qualquer estudo metagenômico em razão de assegurar a qualidade e representatividade dos dados. Diversas operações podem ser realizadas, e.g. a remoção dos adaptadores, das leituras redundantes, dos trechos de qualidade indesejada e de genomas conhecidos que possam ter contaminado o experimento. Os mais variados programas realizam cada uma destas operações citadas; todavia, destaca-se o *fastp*, uma ferramenta *open-source* que unifica diversos processos de controle de qualidade de maneira conveniente e com simplicidade de utilização (CHEN, S., 2023; BHARTI; GRIMM, 2019; LANGMEAD; SALZBERG, 2012).
- 2) **Montagem:** os arquivos apresentam diversos fragmentos cujos tamanhos máximos são apenas algumas centenas de bases nucleicas, o que se demonstra insuficiente para a caracterização dos genes que se evidenciam nas amostras. O intuito da etapa de montagem, especificamente pelo método *de novo*, é conectar esses fragmentos de uma maneira que faça sentido⁴; assim, aumentando a informação genética disponível em cada segmento e viabilizando análises futuras (QUINCE et al., 2017). Um *software* que pode ser utilizado é o MEGAHIT, sendo *open-source* e, quando empregado em testes de *benchmarking*, apresenta usualmente resultados mais completos (LI et al., 2016; ZHANG et al., 2022).

⁴ O produto da etapa de montagem é conhecido como *contigs* e *scaffolds*, sendo este último gerado a partir dos *contigs* construídos (ISMAIL, 2023).

- 3) **Predição de gene:** visa caracterizar possíveis genes presentes nos *contigs* e nos *scaffolds* sem utilizar um banco de dados de referência. Para isso, diversas estratégias podem ser empregadas como a observação dos códons de início e de parada. Ademais, o DNA dos procariotos e dos eucariotos se apresenta em padrões distintos, exigindo que a ferramenta utilizada leve em consideração os tipos de organismos a serem observados. Ferramentas como o Prokka podem ser utilizadas nesta etapa, já que unifica outros *softwares* para realizar a predição de gene (HYATT et al., 2010; SEEMANN, 2014).
- 4) **Identificação de gene:** tem como objetivo relacionar sequências nucleotídicas com genes conhecidos e anotados, tendo como base um banco de dados de referência. Um *software* que pode ser utilizado é o ABRicate, programado especificamente com o objetivo de identificar genes relacionados com a resistência antimicrobiana (AMR) de determinado ambiente. Este programa carrega nativamente alguns bancos de dados de referência, como o CARD (*Comprehensive Antibiotic Resistance Database*) para a identificação de ARGs, e o VFDB (*Virulence Factor Database*), para os conteúdos relacionados aos VFs (HUNT et al., 2017; CHEN, L. et al., 2005; MCARTHUR et al., 2013; SEEMANN, TORSTEN, 2020).

Conforme descrito em Ordine et al. (2023), foram obtidas amostras de 8 diferentes sítios de coleta na região nordeste de São Paulo (SP), com cada localidade possuindo características próprias em relação à classificação geológica e às atividades humanas, conforme demonstrado na Tabela 1. Todos os detalhes acerca da coleta podem ser conferidos no trabalho original; porém, destaca-se que foi utilizado para a extração do DNA das amostras o *kit DNeasy PowerSoil* da QIAGEN, a verificação de qualidade foi realizada em conjunto com o *Nanodrop One* da Thermo Fisher e a plataforma de sequenciamento foi o NovaSeq 6000 da Illumina. Os dados do sequenciamento foram disponibilizados no formato *.fastq* no *Sequencing Read Archive* (SRA), os quais podem ser acessados através do *BioProject PRJNA900430*. Para o processamento dos dados foi utilizada uma *pipeline* que consiste em quatro programas a saber: 1) para a remoção dos adaptadores inseridos durante a etapa de sequenciamento bem como das leituras de qualidade indesejada, foi utilizado o *fastp* na versão 23.1; 2) para a montagem dos dados tratados pela etapa de controle de qualidade, empregou-se o *MegaHIT*; 3) a predição de gene foi realizada com o *Prokka*; e 4) a identificação de genes foi desempenhada pelo *ABRicate*, tendo-se como bases de dados para a pesquisa gênica o CARD, o VFDB, o ARG-ANNOT, o *PlasmidFinder* e o *ResFinder* (ORDINE et al., 2023). Poucos parâmetros de entrada foram descritos e a maioria das versões dos *softwares* não foi elucidada.

2.3 Pipeline implementada

Como no trabalho de Ordine et al. (2023) não foram apresentados os resultados de genes identificados correspondentes às bases de dados *ResFinder* e *ARG-ANNOT* (e o *PlasmidFinder* não encontrou

Tabela 1 – Principais informações relacionadas às amostras de solo consideradas [adaptado de (ORDINE et al., 2023, Table 1)].

Código e denominação da amostra	Classificação geológica	Síntese da atividade antrópica	Sítio e data de coleta
SRR22278226 (Agricultura)	Latossolo vermelho	Grande propriedade agrícola com cultivo de cana-de-açúcar e uso de agroquímicos.	-21,16643, -47,99004 (25/11/2019)
SRR22278232 (Aviário)	Neossolo quartzarênico	Pequena propriedade de agricultura familiar com criação de aves.	-21.95133, -47.89079 (01/09/2020)
SRR22278228 (Pasto)	Neossolo quartzarênico	Pequena propriedade familiar de agropecuária (criação de gado).	-21.921, -47.90373 (01/09/2020)
SRR22278227 (Pomar)	Neossolo quartzarênico	Pequena propriedade de agricultura familiar com pomar.	-21.92674, -47.87008 (01/09/2020)
SRR22278229 (Praça)	Latossolo vermelho	Área urbana com alta circulação de pessoas e pequenos animais.	-21.1134, -47.98762 (01/09/2020)
SRR22278231 (Gado)	Arenito Bauru	Grande propriedade agropecuária com alta circulação de pessoas e animais.	-21.48066, -48.54118 (01/04/2020)
SRR22278225 (APP)	Latossolo vermelho	Área de Proteção Permanente com acesso restrito à pesquisas.	-21.1662, -47.86036 (01/03/2020)
SRR22278230 (Gramado)	Latossolo vermelho	Área urbana com alta circulação de pessoas e pequenos animais.	-21.16511, -47.85944 (01/03/2020)

nenhum gene), a discussão dos resultados será restrita aos dados presentes no CARD e no VFDB ⁵. Dessa forma, para a reprodução da *pipeline*, utilizaram-se os mesmos programas que Ordine et al. (2023) com as seguintes características:

- 1) a versão do *fastp* foi a 0.23.1, tendo como entrada as *reads forward* e *reverse*, sendo as *flags* usadas o *-h* e o *-j* para a mudança de nomes dos arquivos estatísticos. Os arquivos gerados por esse *software* são ambas as *reads* filtradas no formato *.fastq*. Para o restante dos parâmetros foram utilizados os valores padrão.
- 2) o MEGAHIT empregado apresenta a versão v1.2.9, tendo como entrada os resultados filtrados pelo *fastp*. O arquivo de saída são os dados montados em um único arquivo de *contigs* no formato

⁵ Estes bancos de dados são utilizados para a identificação dos ARGs e dos VFs.

FASTA (uma extensão extremamente semelhante ao `fastq`, porém, com menos informações).

- 3) o Prokka foi utilizado na versão 1.14.6 com os seguintes parâmetros: `--kingdom Bacteria, --metagenome, --cpus 8, --norrna, --notrnae --outdir 'caminho/diretório/saída'`. O resultado de saída é um arquivo também do tipo FASTA referente aos genes preditos pelo programa.
- 4) Por fim, o ABRicate possui a versão 1.0.1, tendo sido utilizado os parâmetros `--threads 8` e `--db`. Como os resultados foram operados sobre duas bases de dados, o parâmetro `--db` foi utilizado uma vez na forma `--db card` e na outra como `--db vfdb`. Os resultados do ABRicate foram guardados usando o comando `>'arquivo.tsv'`. Antes do uso, as bases de dados foram atualizadas conforme ilustrado em Seemann, Torsten (2020), sendo as versões tanto do CARD quanto do VFDB referentes ao dia 04/08/2024. Os arquivos `.tsv` gerados são os resultados dos ARGs e VFs pela *pipeline* aplicada.

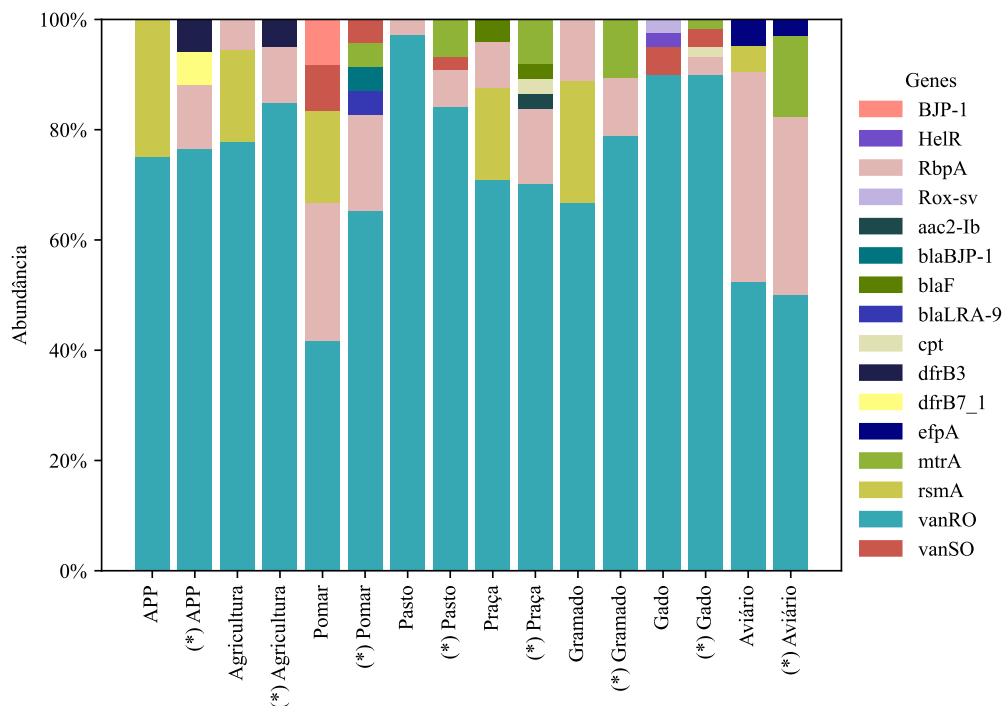
Destaca-se que, apesar dos programas gerarem diversos arquivos, foram evidenciados apenas os tipos de arquivos referentes às operações da *pipeline*. Além disso, como o trabalho de Ordine et al. (2023) descreveu poucos valores de parâmetros aplicados aos *softwares*, as *flags* não mencionadas foram mantidas nos valores padrão (inclusive o valor de *cut-off* utilizado no ABRicate, o único parâmetro explicitamente descrito).

3 RESULTADOS E DISCUSSÕES

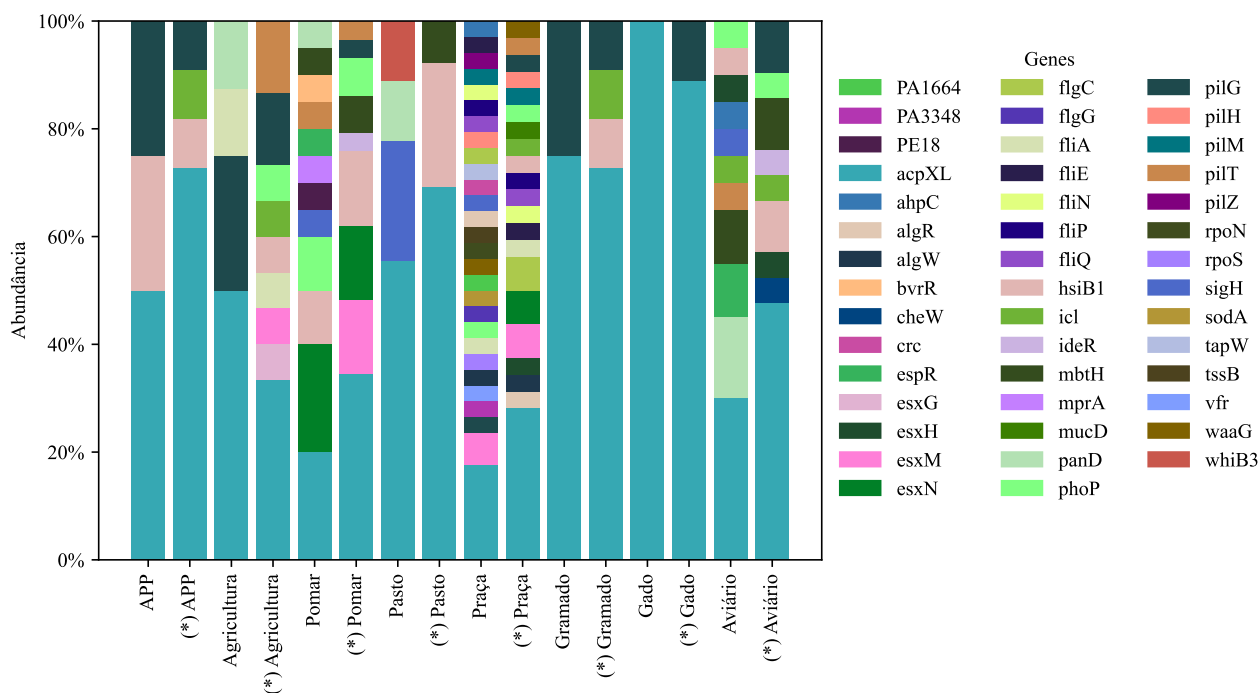
Observando a Figura 1, é possível concluir que os padrões de ARGs e de VFs encontrados na reprodução se distinguem consideravelmente do que foi identificado no trabalho de Ordine et al. (2023). A única tendência que se manteve foi a de que os genes `vanR0` e `acpXL` continuaram sendo encontrados em quantidades majoritárias. Os valores obtidos variaram completamente tanto em quantidade quanto em diversidade. Esses resultados consideravelmente diferentes podem ter diversas origens, a saber:

- **Erros de reprodução da *pipeline*:** A reprodução pode ter sido realizada com erros que não foram identificados durante as revisões dos códigos.
- **Versões diferentes:** Como as versões dos bancos de dados do ABRicate e de todos os outros programas além do Prokka não foram conferidas, os resultados podem ter apresentado variações simplesmente pelas atualizações dos *softwares*.
- **Parâmetros distintos:** Mudanças nos valores dos parâmetros podem alterar consideravelmente o resultado.
- **Erros no pós-processamento:** As funções construídas para operar sobre os dados processados podem apresentar erros não detectados.

- **Erros no trabalho original:** O próprio trabalho original pode ter apresentado acidentalmente informações errôneas acerca do que foi executado.



(a) Genes de resistência a antibióticos (ARGs)



(b) Genes de fatores de virulência (VFs)

Figura 1 – O símbolo ‘(*)’ apresentado no eixo horizontal corresponde aos valores obtidos originalmente por [Ordine et al. \(2023\)](#). Esta figura tem como objetivo ilustrar as diferenças entre os resultados, esperando que o leitor compare um par de colunas por vez.

4 CONSIDERAÇÕES FINAIS

Os resultados aqui obtidos demonstram a falha na reprodução exata da *pipeline* desenvolvida em Ordine et al. (2023); porém, apesar das diferenças em termos de abundância e de diversidade, ratifica-se que genes de resistência a antibióticos e de fatores de virulência foram detectados nas amostras de solo avaliadas. Sugestões de trabalhos futuros, visando a continuidade da pesquisa, envolvem a investigação das possíveis causas das diferenças nos resultados, bem como o desenvolvimento de *pipelines* para análise taxonômica e funcional das comunidades microbianas.

AGRADECIMENTOS

Este trabalho foi parcialmente financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e pela UTFPR através do Edital “TCC-Inovação”.

REFERÊNCIAS

- ARAGÃO, Adalberto; CONTINI, Elisio. **O AGRO NO BRASIL E NO MUNDO: UMA SÍNTESE DO PERÍODO DE 2000 A 2020**. [S.l.: s.n.], 2023. <https://www.embrapa.br/documents/10180/62618376/O+AGRO+NO+BRASIL+E+NO+MUNDO.pdf>.
- BANERJEE, Samiran; VAN DER HEIJDEN, Marcel GA. Soil microbiomes and one health. **Nature Reviews Microbiology**, Nature Publishing Group UK London, v. 21, n. 1, p. 6–20, 2023.
- BHARTI, Richa; GRIMM, Dominik G. Current challenges and best-practice protocols for microbiome analysis. **Briefings in Bioinformatics**, v. 22, n. 1, p. 178–193, dez. 2019. DOI: [10.1093/bib/bbz155](https://doi.org/10.1093/bib/bbz155).
- BUENO, A. D. F. et al. Challenges for adoption of integrated pest management (IPM): the soybean example. **Neotropical Entomology**, v. 50, p. 5–20, 2021. DOI: [10.1007/s13744-020-00788-w](https://doi.org/10.1007/s13744-020-00788-w).
- CHEN, Lihong et al. VFDB: a reference database for bacterial virulence factors. **Nucleic acids research**, Oxford University Press, v. 33, suppl_1, p. d325–d328, 2005.
- CHEN, Shifu. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. **Imeta**, Wiley Online Library, v. 2, n. 2, e107, 2023.
- COMPANT, Stéphane et al. Harnessing the plant microbiome for sustainable crop production. **Nature Reviews Microbiology**, v. 22, n. 8, p. 567–584, ago. 2024. DOI: [10.1038/s41579-024-01079-1](https://doi.org/10.1038/s41579-024-01079-1).
- CONFEDERAÇÃO DA AGRICULTURA E PECUÁRIA DO BRASIL (CNA). **Panorama do Agro**. [S.l.: s.n.], 2024. <https://www.cnabrazil.org.br/cna/panorama-do-agro>.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA (EMBRAPA). **TRAJETÓRIA DA AGRICULTURA BRASILEIRA**. [S.l.: s.n.], 2022. <https://www.embrapa.br/visao/trajetoria-da-agricultura-brasileira>.

- FAO. **Soil Management and Conservation for Small Farms: Strategies and Methods of Introduction, Technologies and Equipment**. Rome: Food e Agriculture Organization of the United Nations, 2017.
- FUENTES-LLANILLO, R. et al. Expansion of no-tillage practice in conservation agriculture in Brazil. **Soil and Tillage Research**, v. 208, p. 104877, 2021. DOI: [10.1016/j.still.2020.104877](https://doi.org/10.1016/j.still.2020.104877).
- GREEN, Michael R; SAMBROOK, Joseph. **Molecular Cloning: A Laboratory Manual**. 4th edition. [S.l.]: Cold Spring Harbor Laboratory Press, 2012.
- HU, Taishan et al. Next-generation sequencing technologies: An overview. **Human Immunology**, v. 82, n. 11, p. 801–811, 2021. Next Generation Sequencing and its Application to Medical Laboratory Immunology. ISSN 0198-8859. DOI: <https://doi.org/10.1016/j.humimm.2021.02.012>. Disponível em: [↗](#).
- HUNT, Martin et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. **Microbial genomics**, Microbiology Society, v. 3, n. 10, e000131, 2017.
- HYATT, Doug et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. **BMC bioinformatics**, Springer, v. 11, p. 1–11, 2010.
- ILLUMINA. **Dream big**. [S.l.: s.n.], 2024. <https://www.illumina.com>.
- ILLUMINA. **Sequence file formats for a variety of data analysis options**. [S.l.: s.n.], 2024. <https://www.illumina.com/informatics/sequencing-data-analysis/sequence-file-formats.html>.
- ILLUMINA. **Sequencing platforms**. [S.l.: s.n.], 2024. <https://www.illumina.com/systems/sequencing-platforms.html>.
- ISMAIL, Hamid D. **Bioinformatics: A Practical Guide to Next Generation Sequencing Data Analysis**. [S.l.]: Chapman e Hall/CRC, 2023.
- LANGMEAD, Ben; SALZBERG, Steven L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, Nature Publishing Group, v. 9, n. 4, p. 357–359, 2012.
- LI, Dinghua et al. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. **Methods**, Elsevier, v. 102, p. 3–11, 2016.
- LIU, Yong-Xin et al. A practical guide to amplicon and metagenomic analysis of microbiome data. **Protein & cell**, Oxford University Press, v. 12, n. 5, p. 315–330, 2021.
- MCARTHUR, Andrew G. et al. The Comprehensive Antibiotic Resistance Database. **Antimicrobial Agents and Chemotherapy**, v. 57, n. 7, p. 3348–3357, 2013. DOI: [10.1128/aac.00419-13](https://doi.org/10.1128/aac.00419-13). eprint: <https://journals.asm.org/doi/pdf/10.1128/aac.00419-13>. Disponível em: [↗](#).
- ORDINE, João Vitor Wagner et al. Metagenomic insights for antimicrobial resistance surveillance in soils with different land uses in Brazil. **Antibiotics**, MDPI, v. 12, n. 2, p. 334, 2023.
- QIAGEN. **DNA**. [S.l.: s.n.], 2024. <https://www.qiagen.com/us/product-categories/discovery-and-translational-research/dna-rna-purification/dna-purification>.

QIAGEN. **DNeasy PowerSoil Pro Kits**. [S.l.: s.n.], 2024.

<https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/microbial-dna/dneasy-powersoil-pro-kit>.

QUINCE, Christopher et al. Shotgun metagenomics, from sampling to analysis. **Nature biotechnology**, Nature Publishing Group US New York, v. 35, n. 9, p. 833–844, 2017.

RICHARDSON, Lorna et al. MGnify: the microbiome sequence data analysis resource in 2023. **Nucleic Acids Research**, v. 51, n. D1, p. d753–d759, dez. 2022. DOI: [10.1093/nar/gkac1080](https://doi.org/10.1093/nar/gkac1080).

SEEMANN, Torsten. Prokka: rapid prokaryotic genome annotation. **Bioinformatics**, Oxford University Press, v. 30, n. 14, p. 2068–2069, 2014.

SEEMANN, TORSTEN. **ABRicate**. [S.l.: s.n.], 2020.
<https://github.com/tseemann/abricate>.

VAN DIJK, Erwin L.; JASZCZYSZYN, Yan; THERMES, Claude. Library preparation methods for next-generation sequencing: Tone down the bias. **Experimental Cell Research**, v. 322, n. 1, p. 12–20, 2014. ISSN 0014-4827. DOI: <https://doi.org/10.1016/j.yexcr.2014.01.008>.

WARNECKE, Falk et al. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. **Nature**, Nature Publishing Group UK London, v. 450, n. 7169, p. 560–565, 2007.

ZHANG, Zhenmiao et al. Benchmarking de novo assembly methods on metagenomic sequencing data. **bioRxiv**, Cold Spring Harbor Laboratory, p. 2022–05, 2022.