

SGC PARA CLASSIFICAÇÃO TAXONÔMICA DE UMA CEPA DE *Bacillus paranthracis* ISOLADA DE UMA INDÚSTRIA FARMACÊUTICA, IDENTIFICADA POR MALDI-TOF MS COMO *Bacillus* spp.

WGS FOR TAXONOMIC CLASSIFICATION OF A *Bacillus paranthracis* STRAIN ISOLATED FROM A PHARMACEUTICAL FACILITY IDENTIFIED BY MALDI-TOF MS AS *Bacillus* spp.

Paula Araujo de Souza

<https://orcid.org/0000-0002-4381-8107>
paula.souza@bio.fiocruz.br

Juliana Nunes Ramos*

<https://orcid.org/0000-0002-7063-6720>
juliana.amos@bio.fiocruz.br

Ana Beatriz Lima dos Santos Tiburcio

<https://orcid.org/0009-0002-7066-0824>
ana.tiburcio@bio.fiocruz.br

Ana Luiza Sant'ana da Silva Guimaraes

<https://orcid.org/0009-0000-8565-5169>
alguimaraes@aluno.fiocruz.br

Talita Bernardo Valadão

<https://orcid.org/0000-0002-5877-8048>
talita.valadao@bio.fiocruz.br

Rebeca Vitória da Silva Lage de Miranda

<https://orcid.org/0000-0002-7131-6155>
rebeca.lage@bio.fiocruz.br

Andressa Sbano da Silva Ferreira

<https://orcid.org/0000-0003-0996-2292>
andressa.ferreira@bio.fiocruz.br

Luciana Veloso da Costa

<https://orcid.org/0000-0003-1873-4056>
luciana.costa@bio.fiocruz.br

Marcelo Luiz Lima Brandão

<https://orcid.org/0000-0003-1121-7312>
marcelo.brandao@bio.fiocruz.br

Obs: todos os autores da
Fundação Oswaldo Cruz, FIOCRUZ, RJ

***Autor por correspondência**

ARTIGO CIENTÍFICO
Submetido em: 20 06 2026
Aprovado em: 27 07 2026

RESUMO

A identificação microbiana acurada é necessária para o controle de contaminação durante a produção de medicamentos estéreis. Contudo, a identificação ao nível de espécie de *Bacillus* e gêneros relacionados permanece um desafio. Este estudo teve como objetivo caracterizar a cepa B0029.24, isolada de uma indústria de imunobiológicos no Rio de Janeiro, Brasil, previamente identificada no sistema MALDI Biotyper® como *Bacillus cereus/thuringiensis* (escore 1,7), e por sequenciamento do gene 16S rRNA como *Bacillus* sp. A caracterização genômica foi realizada por meio do sequenciamento do genoma completo (SGC) utilizando a plataforma Illumina NextSeq 2000. O SGC revelou um tamanho de genoma de 5.63 Mb, 93 contigs, N50=188620 pb, L50=8 e conteúdo GC de 35.14%. O *Type Strain Genome Server* classificou a cepa como *Bacillus paranthracis*, confirmado por ANIb (98,58%). No entanto, o valor de dDDH de 77,80% indica que o isolado representa uma provável nova subespécie. A completude e a contaminação foram estimadas em 99,07% e 0,39%, respectivamente, pelo CheckM2 utilizando a plataforma Galaxy. A cepa foi tipificada por *Multilocus Sequence Typing* (MLST) como ST-164 e por rMLST como rST69175, com *B. paranthracis* como o táxon previsto (100%). O genoma foi depositado no banco de dados do *National Center for Biotechnology Information* (NCBI) (nº de acesso JBVQSU000000000) e identificado pela ferramenta de taxonomia do NCBI como *B. paranthracis* (TaxID 2026186). Após a confirmação genômica, foi gerado o Main Spectrum Profile (MSP), contribuindo para a ampliação do banco de dados do MALDI Biotyper®. Conclui-se que a integração da genômica com a espectrometria de massas é uma estratégia eficaz para uma identificação microbiana rápida e acurada, permitindo decisões mais assertivas na estratégia de controle de contaminação.

Palavras-Chave: SGC. *Bacillus paranthracis*. MLST. Taxonomia. Indústria farmacêutica.

ABSTRACT

Accurate microbial identification is needed for contamination control strategy in sterile drug manufacturing. However, species-level identification of *Bacillus* and related genera, remains challenging. The study aimed to characterize the strain B0029.24, isolated from an immunobiological facility, previously identified by MALDI Biotyper® as *Bacillus cereus/thuringiensis* (score 1.7), and 16S rRNA sequencing yielded only *Bacillus* sp. In this study, comprehensive genomic characterization was achieved via whole genome sequencing (WGS) using Illumina NextSeq 2000. WGS of B0029.24 revealed a genome size of 5.63 Mb, 93 contigs, N50=188620 bp, L50=8, GC content of 35.14%. Type Strain Genome Server classified the strain as *Bacillus paranthracis*, confirmed by Average Nucleotide Identity (98.58%). However, digital DNA-DNA hybridization value of 77.80% indicates that the isolate represents a probable novel subspecies. Completeness and contamination were estimated at 99.07%, and 0.39% by CheckM2 using Galaxy Platform. The strain was typed by Multilocus Sequence Typing (MLST) as ST-164 and rMLST as rST69175, with *B. paranthracis* as the predicted taxon (100%). The genome was deposited in NCBI database (access n. JBVQSU000000000) and identified by NCBI Taxonomy tool as *B. paranthracis* (TaxID 2026186). After genomic confirmation, a Main Spectrum Profile (MSP) was created to enrich the MALDI Biotyper® database. This integration of genomics and mass spectrometry is crucial for precise microbial identification and effective contamination control decisions.

Keywords: WGS. *Bacillus paranthracis*. MLST. Taxonomy. Pharmaceutical facility.

1 INTRODUÇÃO

A identificação acurada de microrganismos é crucial para o controle de áreas assépticas na indústria farmacêutica (Brasil, 2022; EMA, 2022; Song *et al.*, 2024). A classificação taxonômica em nível de espécie contribui para a maior eficiência nas investigações de desvios relacionados a produtos ou processos, permitindo a implementação de medidas corretivas mais assertivas (Costa *et al.*, 2022; Song *et al.*, 2024; Ramos *et al.*, 2025). Esse rigor no controle microbiológico é indispensável para o cumprimento das Boas Práticas de Fabricação (BPF) e para assegurar a qualidade, eficácia e segurança do produto final (Kawai *et al.*, 2019; Brasil, 2022; EMA, 2022; Caldeira *et al.*, 2024).

Bacillus e gêneros relacionados são frequentemente isolados de áreas limpas, contudo, a sua identificação ao nível de espécie ainda é um desafio (Costa *et al.*, 2022; Tiburcio *et al.*, 2026). Embora a espectrometria de massas por ionização e dessorção a laser assistida por matriz por tempo de voo (MALDI-TOF MS) e o sequenciamento do gene 16S rRNA sejam amplamente empregados e, em muitos casos, resolutivos, os mesmos podem não ser o suficiente para a diferenciação de espécies intimamente relacionadas (Thompson *et al.*, 2020; Ramos *et al.*, 2025). Nesses casos, tornam-se necessárias abordagens mais avançadas, como o sequenciamento do genoma completo (SGC) (Thompson *et al.*, 2020; Costa *et al.*, 2024; Li *et al.*, 2026).

Sendo assim, o objetivo deste trabalho teve como objetivo identificar ao nível de espécie a cepa B0029.24, isolada em uma indústria farmacêutica, através de análises de taxonomia genômica. A cepa foi previamente classificada por MALDI Biotyper® como *Bacillus cereus/thuringiensis* com baixos níveis de confiança (escore 1,7), e por sequenciamento do 16S rRNA como *Bacillus* sp.

2 METODOLOGIA

2.1 CEPA BACTERIANA

A cepa B0029.24 foi isolada em 2024, a partir de uma amostra de cultivo celular proveniente de uma indústria de imunobiológicos localizada no Rio de Janeiro e depositada na Coleção de Bactérias do Ambiente e Saúde/Fiocruz, sob o nº CBAS 1561 e Coleção de Bactérias Patogênicas/Fiocruz, sob o nº CBP P7845.

A identificação da cepa B0029.24 por MALDI-TOF MS foi realizada utilizando o equipamento MALDI Biotyper® (Bruker) seguindo as instruções do fabricante. Inicialmente, a cepa foi cultivada em ágar sangue (SBA) e incubada a 30-35°C por 24h. Em seguida, uma porção da colônia foi transferida a um poço da placa descartável para MALDI Biotyper®. Um microlitro de ácido fórmico 70% foi aplicado sobre o poço e, após secagem, foi aplicado 1 µL da matriz de ácido α -ciano-4-hidroxicinâmico (HCCA). Em um dos poços da placa, foi adicionado 1 µL do padrão de *Escherichia coli*, denominado BTS (*Bacterial Test Standard*) e, após secagem, 1 µL da matriz HCCA foi adicionada ao poço. Posteriormente, a placa foi inserida no equipamento, que foi calibrado, a partir do poço contendo o BTS, com a utilização do programa Flex Control v. 3.4, conforme instruções do fabricante. Os espectros de massa gerados foram analisados pelo programa MBT Compass HT v.5.1.100 e comparados com as bibliotecas disponíveis no equipamento (MBT Compass Library Revision G, 2023). A interpretação dos resultados foi realizada de acordo com o escore obtido: 0.00-1.69 - nenhuma identificação de micro-organismo possível; 1.70-1.99 - baixa confiança para identificação da espécie; e 2.00-3.00 - alta confiança para identificação da espécie.

2.2 SEQUENCIAMENTO, MONTAGEM E ANOTAÇÃO DO GENOMA

O DNA genômico bacteriano foi extraído utilizando o kit DNeasy Blood & Tissue Genomic DNA Isolation (Qiagen, Hilden, Alemanha). O sequenciamento de nova geração foi realizado pela Rede de Plataformas da Fiocruz (Instituto Oswaldo Cruz, Rio de Janeiro, Brasil) utilizando o sequenciador Illumina NextSeq 2000 (Illumina®, San Diego, CA, EUA). A preparação das bibliotecas para o sequenciamento foi realizada com o kit Nextera DNA Flex (Illumina®, San Diego, CA, EUA), e a corrida de sequenciamento utilizou os reagentes do Kit P2 XLEAP-SBST™ (Illumina®, San Diego, CA, EUA), com 300 ciclos (2 × 151 pares de bases, *paired-end*), de acordo com as recomendações do fabricante.

A qualidade das leituras brutas do sequenciamento foi avaliada inicialmente usando os programas FastQC v.0.12.1 (Andrews, 2010) e FASTP v.0.23.2 (Chen, 2025). Posteriormente, o programa Trimmomatic v.0.40 (Bolger; Lohse; Usadel, 2014) foi empregado para remover bases de baixa qualidade ($Q < 28$) de ambas as extremidades, sequências de adaptadores e leituras com tamanho inferior a 50 pb. Posteriormente, as leituras filtradas e trimadas foram montadas com o programa SPAdes v.4.2.0 (Prjibelski *et al.*, 2020). Para o refinamento da montagem, empregou-se o *software* Pilon, onde foram executadas correções de erros de montagem, inconsistências de bases e preenchimento de *gaps* (Walker *et al.*, 2014).

O CheckM2 v.1.2.4 (Parks *et al.*, 2015), integrado à Plataforma Galaxy v.25.0, foi utilizado para estimar a completude e a contaminação do genoma montado. A estimativa da presença de sequências contaminantes e quiméricas foi realizada com o programa GUNC v.1.0.5 (Orakov *et al.*, 2021), sendo considerados indicativos de contaminação ou quimerismo genômico valores de *Clade Separation Score* (CSS) superiores a 0,45.

O algoritmo Contest16S (Lee *et al.*, 2017) foi utilizado para detectar potenciais contaminações na sequência do gene 16S rRNA do genoma montado. O BUSCO v.6.0.0 (Simão *et al.*, 2015) foi usado para identificar genes ortólogos presentes, ausentes e fragmentados, e o QUAST v.5.3.0 (Mikheenko *et al.*, 2018) forneceu as estatísticas das métricas de montagem. Por fim, o genoma foi depositado no *GenBank/National Center for Biotechnology Information* (NCBI) sob o número de acesso JBVQSU000000000.

2.3 TAXONOMIA GENÔMICA E ANÁLISES FILOGENÔMICAS

A classificação taxonômica da cepa B0029.24 foi realizada utilizando rMLST (*Ribosomal Multilocus Sequence Typing*) (Jolley *et al.*, 2012), NCBI Taxonomy browser (Schoch *et al.*, 2020), com base no identificador taxonômico (TaxID) atribuído ao genoma, e o *Type Strain Genome Server* (TYGS) (Meier-Kolthoff *et al.*, 2022). Genomas de referência intimamente relacionados identificados pelo TYGS, foram utilizados para determinar os valores de Identidade Média de Nucleotídeos (ANI) usando o algoritmo Blast com *software* PyANI v.0.3.0-alpha (<https://github.com/widdowquinn/pyani>). A hibridização digital DNA-DNA (dDDH) foi calculada usando a Calculadora de Distância Genoma-a-Genoma (GGDC) com base na fórmula recomendada 2 (identidades/comprimento da HSP) (Meier-Kolthoff *et al.*, 2022).

O mapa genômico circular comparativo entre B0029.24 e a cepa tipo *B. paranthracis* Mn5^T (nº acesso Genbank: GCA_001883995.1) foi construída utilizando a plataforma online Proksee (Grant *et al.*, 2023). O genoma da cepa B0029.24 foi previamente anotado por meio do *software* Prokka v.1.14.6 (Seemann, 2014), permitindo a exibição das coordenadas das regiões codificadoras (CDS) e demais elementos genéticos no mapa circular. Por fim, realizou-se a tipificação submetendo o genoma ao banco de dados PubMLST *B. cereus* (<https://pubmlst.org/organisms/bacillus-cereus>) para determinar o perfil alélico e o tipo de sequência (ST) correspondente. Atualmente, o esquema de *multi-locus sequence typing* (MLST) para o grupo *B. cereus* contempla 24 espécies, incluindo *B. paranthracis*.

2.4 AMPLIAÇÃO DA BASE DE DADOS NO MALDI Biotyper®

Para a ampliação da base de dados do sistema MALDI Biotyper® (Bruker Daltonics, Bremen, Alemanha), os procedimentos de aquisição e inserção do *Main Spectrum Profile* (MSP) na biblioteca foram realizados conforme a metodologia descrita por Costa *et al.* (2024).

A cepa B0029.24 foi semeada em SBA e incubada a 30-35 °C por 24 horas. Uma porção do crescimento foi transferido para um tubo tipo Eppendorf contendo 300 µL de água estéril e 900 µL de etanol absoluto, posteriormente homogeneizada em agitador tipo vórtex. A solução foi centrifugada a 13.000–15.000 rpm por 2 minutos e o sobrenadante foi descartado. Uma segunda centrifugação de 2 minutos foi realizada para eliminar o resíduo líquido. O tubo permaneceu destampado em temperatura ambiente por 3 a 5 minutos para secagem do *pellet*. Em seguida, após adição de 25 µL de ácido fórmico a 70% e 25 µL de acetonitrila, a suspensão foi homogeneizada por pipetagem e centrifugada novamente (13.000–15.000 rpm por 2 minutos). Aliquotas de 1 µL do sobrenadante foram aplicadas em oito poços da placa e, após secagem, foi adicionado 1 µL de matriz.

Cada poço contendo a cepa foi processado em triplicata por MALDI-TOF MS, por meio do programa Flex Control v. 3.4 no MALDI Biotyper®. A seleção dos espectros mais homogêneos e a eliminação dos espectros discrepantes (*outliers*) foi realizada no programa Flex Analysis v. 3.4. de acordo com as instruções do fabricante. A qualidade dos espectros foi avaliada na faixa de 6000 a 7000 Daltons (Da). Foram considerados espectros *outliers* aqueles que apresentaram diferença entre as massas acima de 500 ppm ou 2,5 Da. Os espectros discrepantes foram excluídos, de maneira que o MSP fosse criado com no mínimo 20 espectros.

Por meio do programa MBP Compass Explorer v. 4.1, o novo MSP foi comparado com todos os espectros existentes e, em seguida, foi inserido na biblioteca customizada. Finalmente, a nova versão desta biblioteca foi importada para o programa MBT Compass HT v. 5.1.100.

Por fim, foi realizada a verificação do MSP, onde a cepa foi novamente submetida à análise proteômica por MALDI-TOF MS, após incubação a 30-35 °C, por 24h em SBA, para que o MSP inserido pudesse ser desafiado, quanto a sua funcionalidade.

3 RESULTADOS E DISCUSSÃO

O SGC da cepa B0029.24 revelou um genoma de 5.63 Mb, com um conteúdo G+C de 35.14%, valor compatível com a faixa de 35-35.5%, reportada para a espécie no NCBI. A

avaliação de qualidade por meio do CheckM2 indicou completude de 99,70% e contaminação de 0,39%. Esses resultados classificam o genoma montado como de alta qualidade, considerando os critérios amplamente aceitos para genomas bacterianos, que estabelecem valores superiores a 95% de completude e inferiores a 5% de contaminação (Parks *et al.*, 2015). Adicionalmente, a análise realizada com o GUNC não detectou sequências quiméricas ou contaminantes, corroborando a elevada confiabilidade e integridade da montagem genômica obtida. A demais métricas de sequenciamento e anotação do genoma podem ser consultadas na **Tabela 1**.

Tabela 1. Métricas de qualidade da montagem do genoma completo da cepa B0029.24.

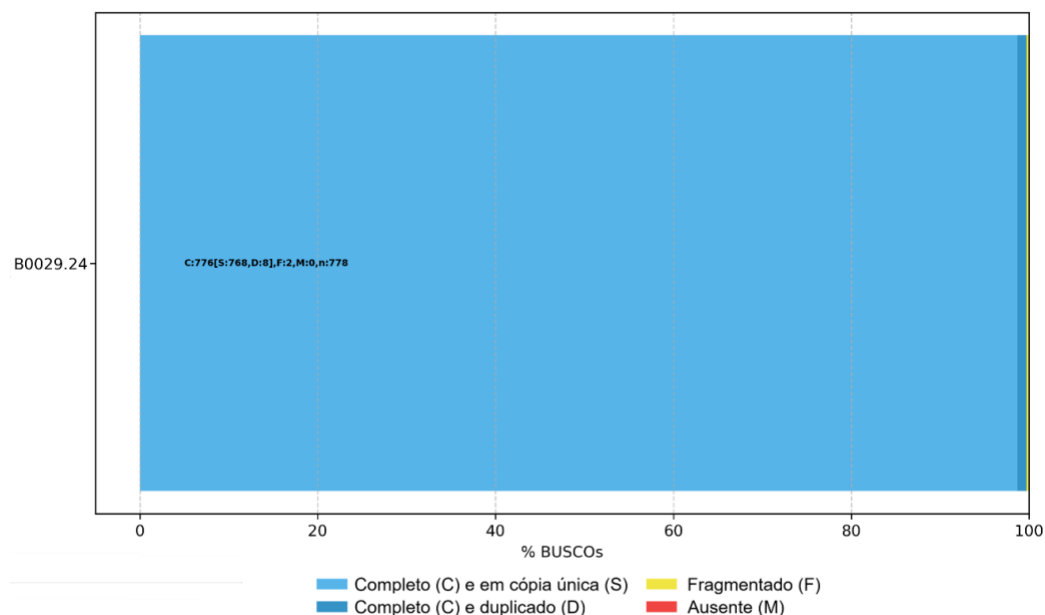
Métrica	Resultado
Tamanho do genoma (pb) ^a	5.63 Mb ^c
Profundidade	374x
N.º de contigs	93
N50	188620 pb
L50	8
Conteúdo G + C (%) ^b	35.14
Completude (%)	99.41
Contaminação (%)	0.29
Heterogeneidade	0

Legenda: ^apb, par de bases; ^bConteúdo de G+C, conteúdo de guanina e citosina; ^cMb, megabase.

De acordo com o banco de dados do gênero *Bacillus* no *software* BUSCO, um total de 778 genes ortólogos foram avaliados. Dentre eles, 768 foram identificados como completos e de cópia única, oito como completos e duplicados, dois como fragmentados e nenhum ortólogo ausente, refletindo uma elevada completude do genoma (99,70%) (**Figura 1**).

Na avaliação da pureza genômica, não foi possível estimar a contaminação com base no gene 16S rRNA utilizando o algoritmo Contest16S, uma vez que apenas um fragmento de 1551 pb foi recuperado. A análise desse fragmento no banco de dados curado EzBioCloud revelou 100% de identidade com *B. paranthracis*, *Bacillus paramycoides* e *Bacillus nitratreducens*. Esses resultados evidenciam a limitada resolução taxonômica do gene 16S rRNA para a discriminação entre espécies filogeneticamente próximas pertencentes ao gênero *Bacillus*.

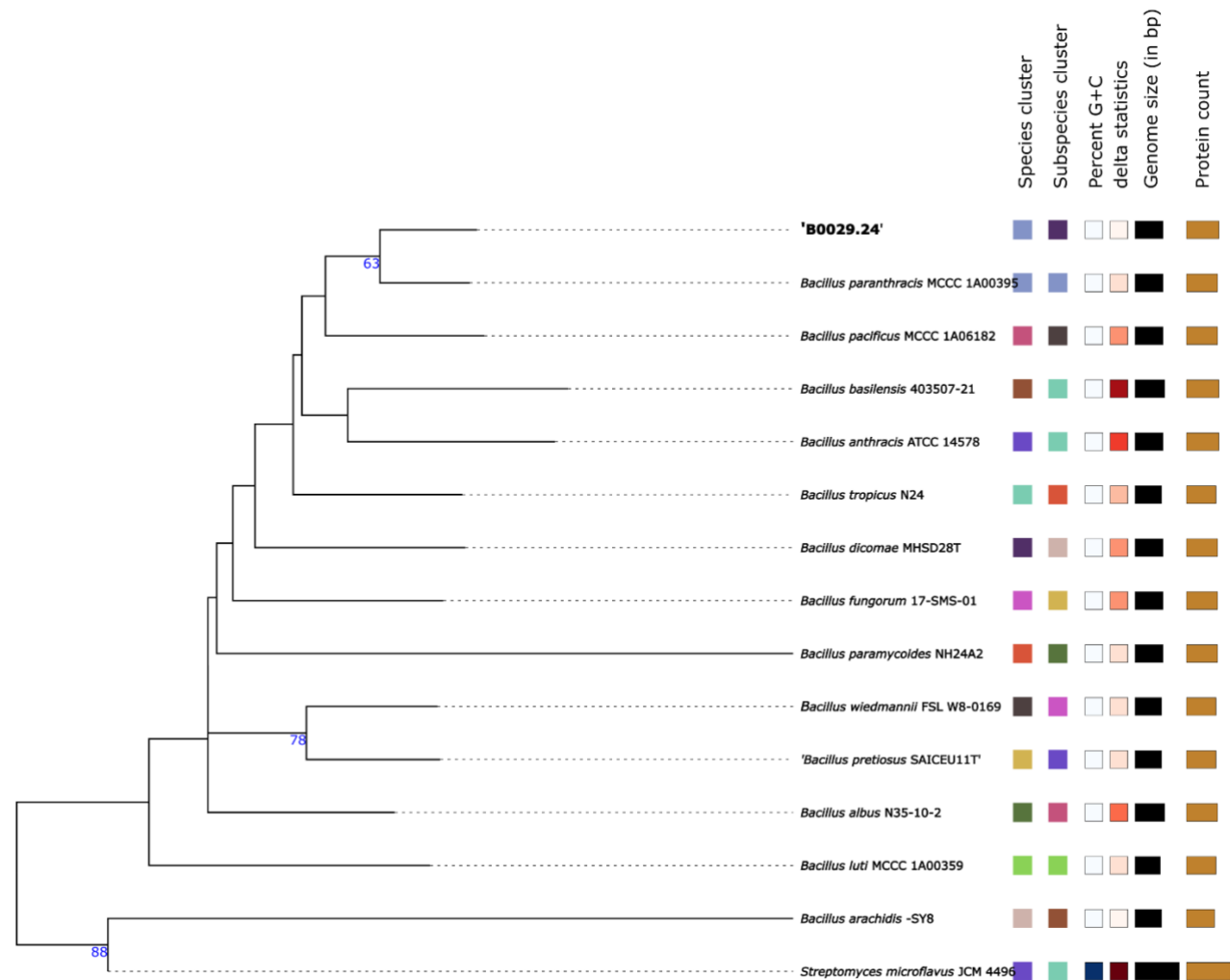
Figura 1. Avaliação da completude do genoma (genes ortólogos) da cepa B0029.24 utilizando o software BUSCO.



A cepa foi tipificada por MLST como ST-164. Segundo o banco de dados PubMLST, este ST possui distribuição global (Ásia, Europa e América do Sul), com isolamentos majoritariamente de fontes ambientais e produtos lácteos. Na literatura, o ST-164 é associado à produção da toxina cereulida e surtos alimentares (Biggel *et al.*, 2026; Vassileva *et al.*, 2007). Além disso, a cepa foi classificada no banco de dados rMLST como rST69175, com *B. paranthracis* como o táxon previsto com um valor de suporte de 100%. Os perfis rMLST são definidos com base nos alelos dos 53 genes ribossomais, como forma de integrar filogenia e a tipificação, sendo amplamente utilizado para classificação taxonômica microbiana (Bray *et al.*, 2022; Jolley *et al.*, 2012).

A análise filogenômica conduzida no TYGS (**Figura 2**) foi realizada comparando a cepa B0029.24 com as sequências genômicas das cepas tipo. O uso dessas cepas de referência é indispensável para garantir o rigor taxonômico e permitir a diferenciação de espécies intimamente relacionadas pertencentes ao grupo *Bacillus* (Freese *et al.*, 2026). A análise revelou que a cepa B0029.24 foi agrupada no mesmo clado que *B. paranthracis*, com suporte de ramo estatístico de 63%, indicando seu pertencimento a este táxon.

Figura 2. Árvore inferida com o FastME 2.1.6.1 a partir das distâncias GBDP calculadas com base em sequências genômicas.



Legenda: MCCC - Microbial Type Culture Collection and Gene Bank; ATCC – American Type Culture Collection.

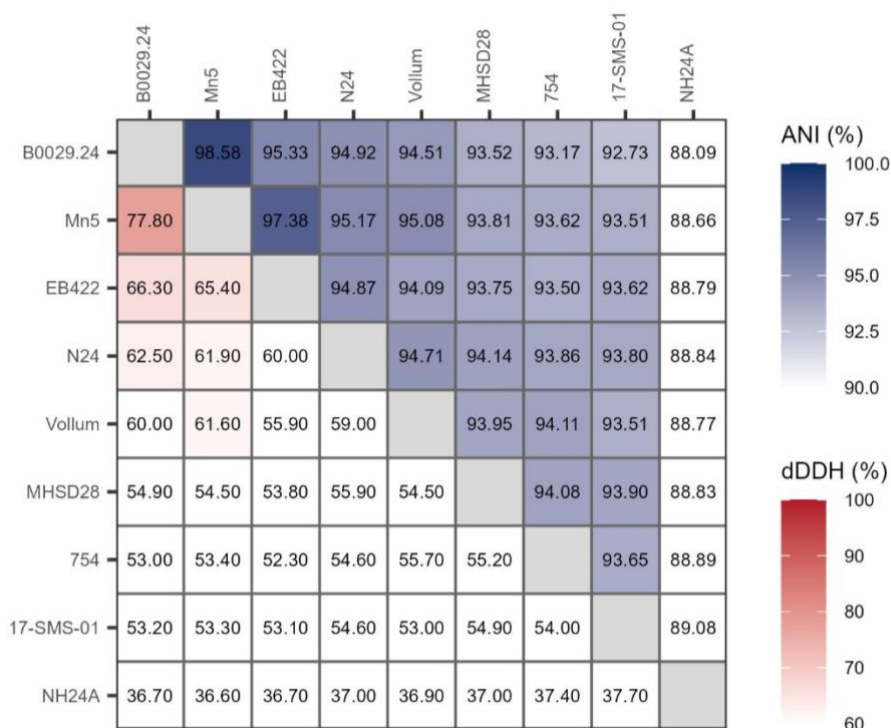
Essa classificação foi corroborada pelos valores de ANIb (98,58%) e dDDH (77,80%), ambos superiores aos limiares de corte estabelecidos para a delimitação de espécies bacterianas ($\geq 95\text{--}96\%$ para ANI e $\geq 70\%$ para dDDH) (Hu *et al.*, 2022) (**Figura 3**). Contudo, o valor obtido para a dDDH permaneceu abaixo do limiar de 79% descrito na literatura para diferenciação em nível de subespécie (Tian *et al.*, 2024), sugerindo a classificação da cepa B0029.24 como uma possível nova subespécie (**Tabela 2**). Essa hipótese é reforçada pela análise filogenômica conduzida no TYGS, na qual a cepa B0029.24 foi alocada em um ramo diferente da cepa tipo de *B. paranthracis* (**Figura 2**). Adicionalmente, a ferramenta de taxonomia do NCBI também classificou a cepa B0029.24 como *B. paranthracis* (TaxID 2026186).

Tabela 2. Valores de ANIb e dDDH para a cepa B0029.24 em comparação com as cepas tipo mais próximas.

Cepas tipo	N.º de acesso GenBank	ANIb ^a	dDDH ^b
<i>B. paranthracis</i> Mn5	GCA_001883995.1	98,58	77,80
<i>B. pacificus</i> EB422	GCA_001884025.1	95,33	66,30
<i>B. tropicus</i> N24	GCA_001884035.1	94,92	62,50
<i>B. anthracis</i> Vollum	GCA_022221345.1	94,51	60,00
<i>B. dicomae</i> MHSD28	GCA_006494425.1	93,52	54,90
<i>B. basilensis</i> 754	GCA_050630465.1	93,14	53,00
<i>B. fungorum</i> 17-SMS-01	GCA_002746455.1	92,73	53,20
<i>B. paramycoides</i> NH24A	GCA_001884235.1	88,09	36,70

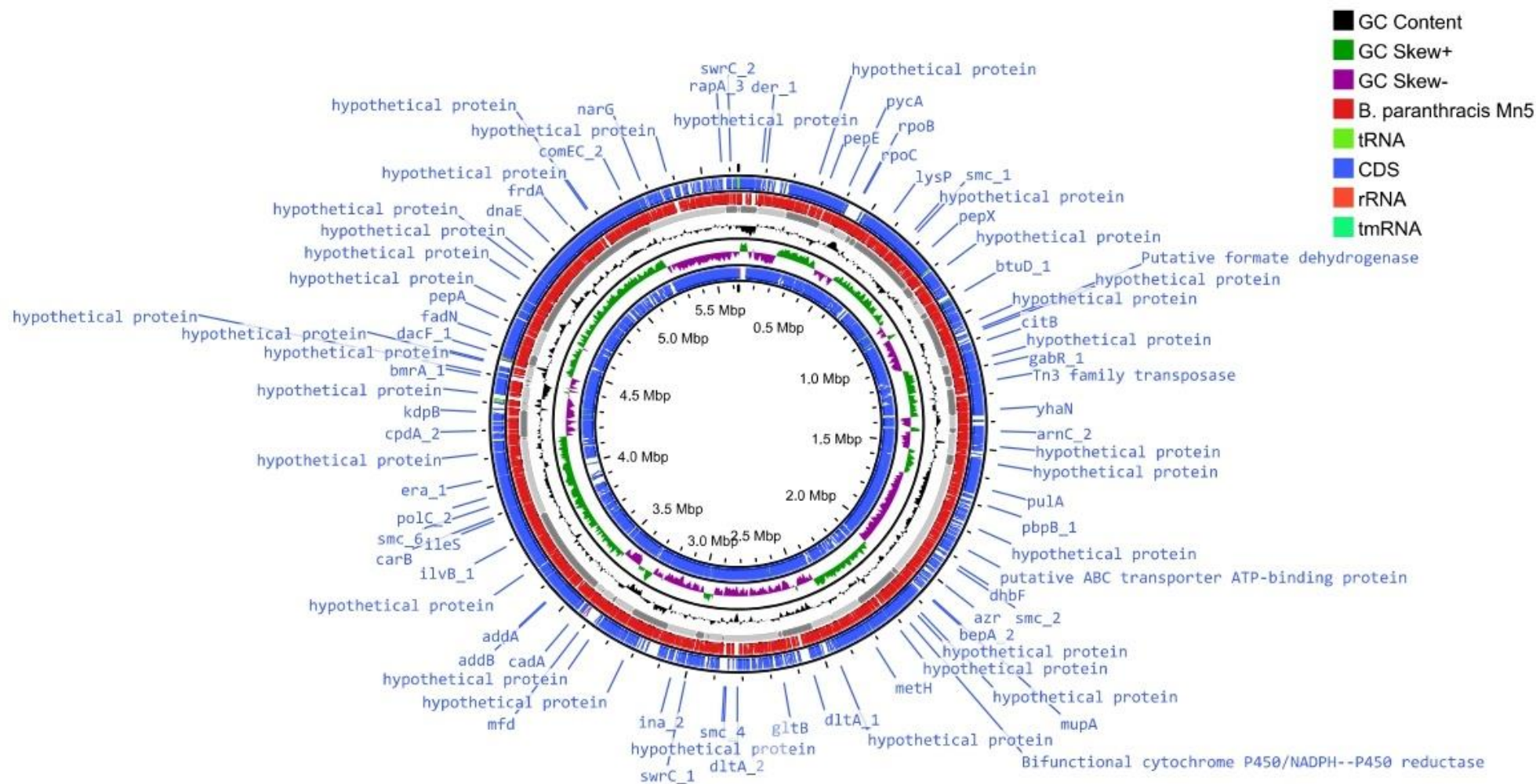
Legenda: ^aIdentidade Média de Nucleotídeos usando o algoritmo Blastn; ^bHibridização digital DNA-DNA.

Figura 3. Heatmap dos valores de ANIb (em azul) e dDDH (em vermelho) entre a cepa B0029.24 e as cepas tipo mais próximas.



O mapa genômico circular observado na **Figura 4** revelou um alto grau de conservação e similaridade entre os genomas da cepa B0029.24 e a cepa tipo de *B. paranthracis* Mn5. A imagem expõe a anotação gênica da cepa B0029.24 nos anéis externos, com CDS representadas em azul na fita *forward* e cinza na fita *reverse*. O anel interno demonstra o alinhamento, com uso da ferramenta *Basic Local Alignment Search Tool* (BLAST), dos dois genomas, indicando uma cobertura quase contínua entre as sequências, evidenciando mais uma vez a classificação taxonômica da cepa como pertencente a espécie *B. paranthracis*.

Figura 4. Mapa genômico circular da cepa B0029.24 e sua análise comparativa com o genoma da cepa tipo de *B. paranthracis* Mn5 gerado usando a ferramenta Proksee.



Legenda: GC- relação guanina e citosina; tRNA – ácido ribonucleico transportador; CDS - coordenadas das regiões codificadoras; rRNA - ácido ribonucleico ribossomal; tmRNA - ácido ribonucleico de transferência-mensageiro.

Após a confirmação taxonômica, criou-se o MSP com a finalidade de ampliar o banco de dados do MALDI Biotyper®. Posteriormente, realizou-se a verificação do MSP recém-inserido à biblioteca por meio de uma nova análise a partir de um cultivo fresco da linhagem B0029.24. O resultado foi satisfatório, uma vez que o espectro da cepa B0029.24, que havia sido inicialmente identificado como *B. cereus/thuringiensis* (**Figura 5A**), na análise após expansão, foi corretamente identificado como *B. paranthracis* como primeira possibilidade de espécie, apresentando um escore de alta confiabilidade (2,25) (**Figura 5B**).

Figura 5. Relatórios gerados pelo MALDI-TOF MS para cepa B0029.24. (A) Relatório anterior a expansão da base de dados (B) Relatório posterior a expansão da base de dados.

A				B			
Rank	Matched pattern	log(score) (Conf.)	NCBI	Rank	Matched pattern	log(score) (Conf.)	NCBI
1	Bacillus cereus_thuringiensis_PG_III BFR_BA_01162 BFR	1.7 (±)	86661	1	Bacillus paranthracis B0029.24	2.25 (±±±)	145076783
2	Bacillus cereus_thuringiensis_PG_III 191127_10 QLAB	1.57 (±)	86661	2	Bacillus cereus BC10	2.1 (±±)	145076783
3	Bacillus pseudomycolides_PG_I BFR_BA_00736 BFR	1.51 (±)	64104	3	Bacillus cereus_thuringiensis_PG_III 191127_09 QLAB	2.07 (±±±)	86661
4	Bacillus cytotoxicus_PG_VII 2833 CVUA	1.5 (±)	86661	4	Bacillus cereus_thuringiensis_PG_III BFR_BA_01090 BFR	2.03 (±±±)	86661
5	Bacillus cereus_thuringiensis_PG_III 191127_09 QLAB	1.5 (±)	86661	5	Bacillus cereus_thuringiensis_PG_III BFR_BA_01162 BFR	2.05 (±±±)	86661
6	Bacillus cereus_thuringiensis_PG_III 191127_02 QLAB	1.49 (±)	86661	6	Bacillus cereus_thuringiensis_PG_III BFR_BA_01015 BFR	2.05 (±±±)	86661
7	Bacillus pseudomycolides_PG_I BFR_BA_00699 BFR	1.48 (±)	64104	7	Bacillus cereus_thuringiensis_PG_III 18259419_2 MVD	2.03 (±±±)	86661
8	Bacillus cereus_thuringiensis_PG_III 15147149_1 MVD	1.47 (±)	86661	8	Bacillus cereus_thuringiensis_PG_III 191127_06 QLAB	2.02 (±±±)	86661
9	Bacillus cereus_thuringiensis_PG_III 161005557_1 HLG	1.46 (±)	86661	9	Bacillus cereus_thuringiensis_PG_III BFR_BA_01152 BFR	2.01 (±±±)	86661
10	Bacillus cytotoxicus_PG_VII DSM 22905T DSM_2	1.44 (±)	86661	10	Bacillus cereus_thuringiensis_PG_III 191127_02 QLAB	2.01 (±±±)	86661

Cabe ressaltar como uma limitação do estudo, que a etapa posterior à incorporação do novo espectro à biblioteca constituiu uma verificação preliminar e não uma validação. Como o espectro adicionado foi desafiado apenas contra a própria cepa de origem, os elevados escores obtidos já eram esperados. Para a realização da validação, é necessária a execução de testes adicionais desafiando o novo MSP contra outras linhagens de *B. paranthracis* e espécies correlatas.

Contudo, como observado na **Figura 5B**, outras espécies do gênero *Bacillus* também apresentaram alta correspondência com a cepa B0029.24, com valores de escore > 2,00, demonstrando a elevada similaridade proteômica compartilhada por esses microrganismos, inclusive a cepa de *Bacillus cereus* BC10, cujo MSP foi inserido na base de dados do equipamento pelo nosso grupo de pesquisa. Tal resultado ressalta a necessidade de saturação da base de dados do MALDI Biotyper® com MSPs de cepas *in house* previamente caracterizadas. Essa estratégia visa aumentar a acurácia da identificação, estabelecendo uma biblioteca personalizada e mais fidedigna à microbiota da indústria em questão (Costa *et al.*, 2022).

Na comparação entre os relatórios obtidos antes (**Figura 5A**) e após (**Figura 5B**) a expansão da base de dados do equipamento, observou-se que os MSPs originalmente presentes na base de dados do fabricante (*Bacillus_cereus_thuringiensis_PG_III*) não apresentaram alta similaridade com a cepa B0029.24 na primeira análise, evidenciada por valores de escore <2,00. Esse resultado pode estar relacionado às diferenças metodológicas aplicadas nos protocolos de preparo das amostras.

Na primeira análise, foi utilizado o método direto com a adição prévia de ácido fórmico a 70%, que possivelmente não foi suficiente para promover a extração eficiente de proteínas de uma linhagem Gram-positiva, cuja parede celular apresenta maior resistência à lise (**Figura 5A**). Em contrapartida, na análise de verificação (**Figura 5B**), aplicou-se o protocolo de superextração, baseado na utilização sequencial de etanol, ácido fórmico e acetonitrila. Esse método é indicado pelo fabricante como mais eficiente para a obtenção de espectros de alta qualidade para análise em MALDI-TOF MS, o que pode explicar os valores de escore >2,00 obtidos para o MSPs originalmente presentes na base de dados do equipamento.

Portanto, o presente estudo demonstra que a integração de técnicas genômicas com a espectrometria de massas contribui significativamente para a identificação microbiana mais precisa. Enquanto o SCG fornece alta resolução taxonômica para diferenciação de espécies intimamente relacionadas, o MALDI-TOF MS fornece a rapidez e o alto rendimento exigidos pela rotina industrial (Thompson *et al.*, 2020; Costa *et al.*, 2022; Song *et al.*, 2024; Li *et al.*, 2026). Com o uso integrado das técnicas, a precisão taxonômica deixa de ser apenas um dado acadêmico e passa a subsidiar decisões eficazes no controle de contaminações.

Por fim, apesar de *B. paranthracis* ser um microrganismo comumente isolado de solo, água e alimentos, no presente estudo, o isolado foi recuperado a partir de amostras de cultivo celular em indústria farmacêutica. Embora a detecção de *B. paranthracis* em ambientes industriais ainda não tenha sido reportado anteriormente na literatura, espécies do gênero *Bacillus* são frequentemente isoladas nesse setor (Costa *et al.*, 2022, 2024; Caldeira *et al.*, 2024). A ocorrência dessas bactérias na indústria pode ser justificada pela sua capacidade de formar esporos e biofilme, evidenciando a relevância crítica do monitoramento rigoroso e da identificação precisa de *B. paranthracis* no controle microbiológico.

4 CONSIDERAÇÕES FINAIS

Conclui-se que as métricas genômicas de ANIb (98,58%) e dDDH (77,80%), bem como a classificação taxonômica por TYGS, confirmam a identificação da cepa B0029.24 como *B.*

paranthracis. Os dados também apontam a possibilidade de classificação da cepa como uma nova subespécie. Os resultados obtidos demonstram ainda que a integração das técnicas SGC e MALDI-TOF MS representa uma estratégia indispensável para a discriminação precisa de espécies estreitamente relacionadas, sendo essencial para o direcionamento de ações resolutivas no controle de contaminações.

AGRADECIMENTOS

Os autores agradecem ao suporte da Fiocruz em forma de Bolsa referente ao Programa de Pós-Doutorado Junior (Inscrição número 641-20251230), à FAPERJ (E-26/210.563/2025, E-26/204.245/2025 e E-26/200.546/2025), ao suporte do CNPq em forma Bolsa de Produtividade em Pesquisa (Processo: 304877/2024-7) e à Plataforma Genômica, Sequenciamento de Nova Geração - RPT01J - Rede de Plataformas Tecnológicas FIOCRUZ.

REFERÊNCIAS

- ANDREWS, S. **FastQC: a quality control tool for high throughput sequence data**. Babraham Bioinformatics, 2010. Disponível em: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Acesso em: 2 mai. 2026.
- BIGGEL, M. *et al.* Foodborne intoxication most likely linked to *Bacillus cereus* in a school canteen hot ready-to-eat meal. **Journal of Food Protection**, p. 100721, 2026. DOI: 10.1016/j.jfp.2026.100721.
- BOLGER, A.M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114-2120, 2014. DOI: <https://doi.org/10.1093/bioinformatics/btu170>.
- BRASIL. Ministério da Saúde. Agência Nacional de Vigilância Sanitária. Resolução – RDC nº 658, de 30 de março de 2022. Dispõe sobre as Diretrizes Gerais de Boas Práticas de Fabricação de Medicamentos. **Diário Oficial [da] República Federativa do Brasil**, Brasília, 31 mar. 2022. Edição 62, Seção 1, p. 320.
- BRAY, J.E. *et al.* Ribosomal MLST nucleotide identity (rMLST-NI), a rapid bacterial species identification method: application to *Klebsiella* and *Raoultella* genomic species validation. **Microbial genomics**, v. 8, n. 9, p. 000849, 2022. DOI: <https://doi.org/10.1099/mgen.0.000849>.
- CALDEIRA, N.G.S. *et al.* Characterization by MALDI-TOF MS and 16S rRNA gene sequencing of aerobic endospore-forming bacteria isolated from pharmaceutical facility in Rio de Janeiro, Brazil. **Microorganisms**, v. 12, n. 4, p. 724, 2024. DOI: <https://doi.org/10.3390/microorganisms12040724>.

CHEN, S. fastp 1.0: An ultra-fast all-round tool for FASTQ data quality control and preprocessing. **Imeta**, v. 4, n. 5, p. e70078, 2025. DOI: <https://doi.org/10.1002/imt2.70078>.

COSTA, L.V. *et al.* MALDI-TOF MS database expansion for identification of *Bacillus* and related genera isolated from a pharmaceutical facility. **Journal of Microbiological Methods**, v. 203, p. 106625, 2022. DOI: <https://doi.org/10.1016/j.mimet.2022.106625>.

COSTA, L. V. *et al.* *Bacillus lumedeiriae* sp. nov., a Gram-Positive, Spore-Forming Rod Isolated from a Pharmaceutical Facility Production Environment and Added to the MALDI Biotyper® Database. **Microorganisms**, v. 12, n. 12, p. 2507, 2024. DOI: <https://doi.org/10.3390/microorganisms12122507>.

EUROPEAN MEDICINES AGENCY (EMA). The Rules Governing Medicinal Products in the European Union. Volume 4: European Union Guidelines for Good Manufacturing Practice for Medicinal Products for Human and Veterinary Use. **Annex 1: Manufacture of Sterile Medicinal Products**. Bruxelas, 2022.

FREESE, Heike M. *et al.* TYGS and LPSN in 2025: a Global Core Biodata Resource for genome-based classification and nomenclature of prokaryotes within DSMZ Digital Diversity. **Nucleic Acids Research**, v. 54, n. D1, p. D884-D891, 2026. DOI: <https://doi.org/10.1093/nar/gkaf1110>.

GRANT, J.R. *et al.* Proksee: in-depth characterization and visualization of bacterial genomes. **Nucleic acids research**, v. 51, n. W1, p. W484-W492, 2023. DOI: <https://doi.org/10.1093/nar/gkad326>.

HU, S. *et al.* New insights into the threshold values of multi-locus sequence analysis, average nucleotide identity and digital DNA–DNA hybridization in delineating *Streptomyces* species. **Frontiers in Microbiology**, v. 13, p. 910277, 2022. DOI: <https://doi.org/10.3389/fmicb.2022.910277>.

JOLLEY, K.A. *et al.* Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. **Microbiology**, v. 158, n. 4, p. 1005-1015, 2012. DOI: <https://doi.org/10.1099/mic.0.055459-0>.

KAWAI, M. *et al.* Culture independent approach reveals domination of human-oriented microbes in a pharmaceutical manufacturing facility. **European Journal of Pharmaceutical Sciences**, v. 137, p. 104973, 2019. DOI: <https://doi.org/10.1016/j.ejps.2019.104973>.

LEE, I. *et al.* ContEst16S: an algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences. **International journal of systematic and evolutionary microbiology**, v. 67, n. 6, p. 2053-2057, 2017. DOI: <https://doi.org/10.1099/ijsem.0.001872>.

LI, Q. *et al.* Leveraging whole-genome sequencing for microbial contamination tracking and risk assessment in pharmaceutical manufacturing. **Frontiers in Microbiology**, v. 17, p. 1807989, 2026. DOI: <https://doi.org/10.3389/fmicb.2026.1807989>.

MEIER-KOLTHOFF, J. P. *et al.* TYGS and LPSN: A database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. **Nucleic Acids Res.**, v. 50, n. D1, p. D801-D807, 2022. DOI: <https://doi.org/10.1093/nar/gkab902>.

MIKHEENKO, A. *et al.* Versatile genome assembly evaluation with QUAST-LG. **Bioinformatics**, v. 34, n. 13, p. i142-i150, 2018. DOI: <https://doi.org/10.1093/bioinformatics/bty266>.

ORAKOV, A. *et al.* GUNC: detection of chimerism and contamination in prokaryotic genomes. **Genome biology**, v. 22, n. 1, p. 178, 2021. DOI: <https://doi.org/10.1101/2020.12.16.422776>

PARKS, D.H. *et al.* CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. **Genome research**, v. 25, n. 7, p. 1043, 2015. DOI: <https://doi.org/10.1101/gr.186072.114>.

PARKS, D.H. *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. **Nucleic Acids Res.**, v. 50, n. D1, p. D785-D794, 2022. DOI: <https://doi.org/10.1093/nar/gkab776>.

PRJIBELSKI, A. *et al.* Using SPAdes de novo assembler. **Current protocols in bioinformatics**, v. 70, n. 1, p. e102, 2020. DOI: <https://doi.org/10.1002/cpbi.102>.

RAMOS, J.N. *et al.* Challenges in the identification of environmental bacterial isolates from a pharmaceutical industry facility by 16S rRNA gene sequences. **DNA**, v. 5, n. 3, p. 33, 2025. DOI: <https://doi.org/10.3390/dna5030033>.

SCHOCH, C.L. *et al.* NCBI Taxonomy: a comprehensive update on curation, resources and tools. **Database**, v. 2020, p. baaa062, 2020. DOI: <https://doi.org/10.1093/database/baaa062>.

SEEMANN, Torsten. Prokka: rapid prokaryotic genome annotation. **Bioinformatics**, v. 30, n. 14, p. 2068-2069, 2014. DOI: <https://doi.org/10.1093/bioinformatics/btu153>.

SIMÃO, F. A. *et al.* BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. **Bioinformatics**, v. 31, n. 19, p. 3210-3212, 2015. DOI: <https://doi.org/10.1093/bioinformatics/btv351>.

SONG, M. *et al.* A comprehensive technology strategy for microbial identification and contamination investigation in the sterile drug manufacturing facility—A case study. **Frontiers in Microbiology**, v. 15, p. 1327175, 2024. DOI: <https://doi.org/10.3389/fmicb.2024.1327175>.

THOMPSON, C.C. *et al.* Microbial genomic taxonomy. **Trends in the systematics of bacteria and fungi**, p. 168-178, 2020. DOI: <https://doi.org/10.1079/9781789244984.016>.

TIAN, X. *et al.* Genome characterisation and comparative analysis of *Schaalia dentiphila* sp. nov. and its subspecies, *S. dentiphila* subsp. *denticola* subsp. nov., from the human oral cavity. **BMC microbiology**, v. 24, n. 1, p. 185, 2024. DOI: <https://doi.org/10.1186/s12866-024-03346-w>.

SOUZA, P.A. de; RAMOS, J.N.; TIBURCIO, A.B.L. dos S.; GUIMARAES, AA.L.S. da S.; VALADÃO, T.B.; MIRANDA, R.V. da S.L. de; FERREIRA, A.S. da S.; COSTA, L.V. da; BRANDÃO, M.L.L.
SGC para classificação taxonômica de uma cepa de *Bacillus paranthracis* isolada de uma indústria farmacêutica, erroneamente identificada por MALDI-TOF MS como *Bacillus cereus/thuringiensis*
Revista Científica do UBM, v. 28, n. 55, 2 sem. 2026, p. 440-457, e-ISSN 2764-5185

TIBURCIO, A. B. L. DOS S.; MIRANDA, R. V. DA S. L. DE; COSTA, L. V. DA.
Identificação de espécies do grupo *Bacillus cereus* isoladas em indústrias farmacêuticas: uma revisão da literatura. **Revista Científica do UBM**, v. 28, n. 54, p. 124-142, 2026. DOI: <https://doi.org/10.52397/rcubm.v28i54.2592>.

VASSILEVA, M. *et al.* A new phylogenetic cluster of cereulide-producing *Bacillus cereus* strains. **Journal of Clinical Microbiology**, v. 45, n. 4, p. 1274-1277, 2007. DOI: <https://doi.org/10.1128/jcm.02224-06>.

WALKER, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. **PloS one**, v. 9, n. 11, p. e112963, 2014. DOI: <https://doi.org/10.1371/journal.pone.0112963>.