

MAURO DE MEDEIROS OLIVEIRA

**ANÁLISE GENÔMICA DE BACTÉRIAS PROBIÓTICAS
BACTERIOCINOGENÉTICAS: IDENTIFICAÇÃO TAXONÔMICA E
CARACTERIZAÇÃO DE GENES RELACIONADOS À SÍNTESE DE
BACTERIOCINAS**

Relatório de Pós-doutorado realizado na
Universidade Estadual Paulista “Júlio de
Mesquita Filho” (UNESP), Faculdade de
Ciências Agrárias e Veterinárias (FCAV)

Supervisor(a): Alessandro de Mello
Varani

FAPESP – Nº 2023/04372-1

Jaboticabal

2025

RESUMO

Este estudo visa realizar a análise genômica de linhagens bacterianas probióticas produtoras de bacteriocinas, isoladas a partir de amostras da microbiota intestinal de frangos de corte (*Gallus gallus domesticus*) linhagem Cobb, suínos (*Sus scrofa domesticus*) mestiços das raças Landrace e Large white e peixes (tilápia-do-nilo, *Oreochromis niloticus*, e truta-arco-íris, *Oncorhynchus mykiss*). Para isso, serão realizados os seguintes objetivos específicos: (1) Realizar o sequenciamento genômico após o isolamento e a caracterização fenotípica de linhagens de Bacilos ácido lácticos (BALs) produtoras de bacteriocinas. (2) Realizar a genômica comparativa pela análise filogenética do 16S rRNA e do *sequence type* (MLST) obtidos *in silico*. (3) Realizar a identificação de genes de bacteriocinas através do uso de diferentes bancos de dados e ferramentas de bioinformática como AntiSMASH, BACTIBASE, BAGEL , BOA e NeuBI. Espera-se descobrir novas variantes de bacteriocinas e sua respectiva associação com determinadas linhagens, espécies e nichos ecológicos. Uma vez que as linhagens estejam genomicamente caracterizadas, o projeto visa contribuir para o desenvolvimento racional de suplementações probióticas destinadas a animais de produção.

Palavras chaves: Bacteriocinas; Genômica comparativa; Metagenômica; Probióticos; Análise de sequência de DNA

1. INTRODUÇÃO

A crescente resistência microbiana aos antibióticos convencionais representa uma ameaça crítica à saúde animal e pública, impulsionando a busca por alternativas sustentáveis na produção agropecuária. Nesse cenário, os probióticos, particularmente as bactérias ácido-láticas (BALs), emergem como estratégia promissora para prevenir e tratar infecções bacterianas em animais de consumo (SACHDEVA et al., 2025). Como destacado no projeto associado "Bioprospecção de bactérias probióticas bacteriocinogênicas: da otimização do cultivo à aplicação em sistemas de produção animal", essas BALs não apenas colonizam benéficamente o trato gastrointestinal de aves, suínos e peixes, mas também produzem biomoléculas bioativas essenciais (DENG et al., 2022). Entre estas, as bacteriocinas – peptídeos antimicrobianos de amplo espectro – desempenham papel crucial no controle de patógenos e na modulação da microbiota intestinal (SUGRUE et al., 2024; CHERUVARI et al., 2025).

Contudo, os efeitos probióticos são altamente específicos da linhagem bacteriana, e a produção otimizada de bacteriocinas está intrinsecamente ligada ao potencial genético de cada cepa (SUGRUE et al., 2024; CHERUVARI et al., 2025). Embora o projeto principal avance no isolamento, seleção e aplicação de BALs bacteriocinogênicas, uma caracterização aprofundada dos determinantes genéticos subjacentes à síntese, regulação e secreção dessas bacteriocinas permanece essencial para otimizar sua eficácia e segurança (LAHIRI et al., 2022; SACHDEVA et al., 2025). A identificação taxonômica precisa e o mapeamento dos genes de interesse (como bacteriocinogênicos, imunidade, transporte e regulação) são etapas fundamentais para: Validar a origem e segurança das cepas selecionadas; Compreender a diversidade e o potencial antimicrobiano do "mix" probiótico; Direcionar a engenharia genética para cepas superprodutoras; Garantir a estabilidade das bacteriocinas durante a encapsulação e aplicação.

Alinhado ao projeto de bioprospecção, este estudo propõe uma análise genômica integrada das BALs probióticas bacteriocinogênicas isoladas de aves, suínos e peixes. Utilizando abordagens de sequenciamento de nova geração (NGS) e ferramentas de bioinformática, para: Identificar taxonomicamente as cepas em nível de espécie e linhagem; Anotar e caracterizar agrupamento gênico associados à biossíntese de bacteriocinas; Correlacionar perfis genômicos com a capacidade produtiva e compatibilidade entre cepas; Detectar genes de resistência a antimicrobianos ou fatores de virulência.

Os resultados fornecem um alicerce molecular robusto para o desenvolvimento de probióticos veterinários de alta precisão, ampliando o impacto do projeto principal na redução de zoonoses e resistência microbiana, e consolidando a liderança brasileira no agronegócio sustentável.

2. OBJETIVOS

Caracterização genômica e rotulação de genes relacionados à síntese de bacteriocinas em bactérias probióticas.

Objetivos específicos

(1) Realizar o sequenciamento genômico após o isolamento e a caracterização fenotípica de linhagens de Bactérias ácido lácticas (BALs) produtoras de bacteriocinas.

(2) Realizar a genômica comparativa pela análise filogenética do 16S rRNA, tipagem de sequências multilocus (MLST), frequência de nucleotídeos e pangenoma.

(3) Realizar a identificação de genes de bacteriocinas através do uso de diferentes bancos de dados e ferramentas de bioinformática mesclando busca por similaridade (AntiSMASH e BAGEL) e aprendizado de máquina (Macrel e amPEP).

3. MATERIAIS E MÉTODOS

3.1. Isolamento de DNA e sequenciamento *Long reads*

O DNA genômico foi extraído de cada isolado bacteriano utilizando o PureLink™ Genomic DNA Mini Kit (Invitrogen™, Thermo Fisher Scientific, Wilmington, DE, EUA), seguindo os protocolos do fabricante. A quantidade e a pureza do DNA extraído foram avaliadas por espectrofotometria (NanoDrop One/One^c - Thermo Fisher Scientific, EUA) e fluorometria (Qubit 2.0™ - Thermo Fisher Scientific). As bibliotecas de DNA foram preparadas utilizando o protocolo PacBio HiFi e sequenciadas na plataforma PacBio HiFi REVIO™. Todos os procedimentos de preparação e sequenciamento das bibliotecas foram realizados no Arizona Genomics Institute.

3.2. Avaliação da qualidade das leituras brutas, montagem do genoma e identificação taxonômica

As leituras brutas foram montadas utilizando o HiFiAsm-meta (v0.3-r073) (FENG et al., 2022) e Flye (v2.9.6) (KOLMOGOROV et al., 2020). Os genomas microbianos circulares completos foram reorientados usando a ferramenta DNAApler (v0.7.0) (BOURAS et al., 2024). Os genomas montados foram analisados para identificação taxonômica e avaliação de qualidade usando a ferramenta GTDB-Tk (v2.3.2) (CHAUMEIL et al., 2022) e o Genome Taxonomy

Database (GTDB vR214) (PARKS et al., 2022), bem como CheckM (PARKS et al., 2015) e BUSCO (v5.6.1) (lactobacillales_odb10) (SIMÃO et al., 2015). Em paralelo, a identidade média de nucleotídeos (ANI) foi calculada usando as ferramentas OrthoANI (LEE et al., 2016) e ANIclustermap v1.4.0 (<https://github.com/moshi4/ANIclustermap>) para comparar os genomas de cada isolado contra os genomas de referência de cada espécie depositados no NCBI RefSeq. O processo de identificação e rotulação das sequências extracromossômicas (p. ex.: plasmídeos) foi realizado pelas ferramentas deeplasmid (ANDREOPOULOS et al., 2022) e PLASMe (TANG, et al., 2023)

3.3. Anotação estrutural e funcional do genoma

A anotação dos genomas foi realizada por meio das ferramentas Bakta (SCHWENGERS et al., 2021) NCBI Prokaryotic Genome Annotation Pipeline (PGAP - versão 2024-04-27) (TATUSOVA et al., 2016). Além disso, a ferramenta eggNOG Mapper v2.1.12 (CANTALAPIEDRA et al., 2021) nos auxiliará no mapeamento dos genes contidos nas sequências estruturalmente anotadas pelo PGAP em famílias gênicas conhecidas, fornecendo informações funcionais adicionais. A identificação de prófagos em genomas bacterianos foi realizada utilizando o PHASTEST (v3.0) (WISHART et al., 2021) e PhaBOX (SHANG et al., 2023). Representações circulares do genoma foram geradas utilizando o pipeline GenoVi (v0.4.3) (CUMSILLE et al., 2023). Para identificar genes de resistência a antibióticos (ARGs), sequências de nucleotídeos codificadores de proteínas de cada genoma foram comparadas com diferentes bancos de dados, incluindo NCBI AMRFinderPlus (FELDGARDEN et al., 2021), CARD (ALCOCK et al., 2023), ResFinder (BORTOLAIA et al., 2020), ARG-ANNOT (GUPTA et al., 2014), MEGARES (DOSTER et al., 2020) e PlasmidFinder (CARATTOLI et al., 2014), utilizando a ferramenta ABRICATE (v1.0.1) (<https://github.com/tseemann/abricate>).

3.4. Análise de Clusters de Genes Biossintéticos de Metabólitos Secundários

Para a análise de clusters gênicos envolvidos na biossíntese de metabólitos secundários dos isolados sequenciados, será utilizada a ferramenta antiSMASH v7.1 (BLIN et al., 2023) e BAGEL4 (VAN HEEL et al., 2018). Estas ferramentas empregam Modelos de Markov Ocultos (HMM) (EDDY, 1996) para detectar clusters de genes que codificam tais componentes.

Adicionalmente, para visualização detalhada de tais clusters, pretende-se gerar arquivos que detalham essas regiões com base na posição inferida durante a anotação gênica.

3.5. Identificação de novo de peptídeos sintetizados ribossomalmente e modificados pós-tradução

A identificação de peptídeos sintetizados ribossomalmente e modificados pós-tradução (RiPPS do inglês - Ribosomally synthesized and post-translationally modified peptides) foi realizada utilizando algoritmos de aprendizado de máquina que podem realizar a predição de RiPPS sem as limitações que as ferramentas baseadas em sequências de referências apresentam. Para esta tarefa, nós utilizamos as ferramentas MACREL (SANTOS-JUNIOR, et al., 2020) baseado em redes neurais e amPEP (LAWRENCE, et al, 2021) modelo baseado em florestas de decisão aleatórias (Random Forest).

3.6. Equipamentos e infraestrutura

O laboratório "Varani's LAB: Genomics & Bioinformatics" apresenta infraestrutura computacional para realizar as diferentes atividades deste projeto tais como análise, processamento e armazenamento dos dados. Atualmente o laboratório conta com dois servidores nas seguintes configurações: Servidor emu1 - modelo Intel(R) Xeon(R), com 24 CPUs modelo E7540, velocidade 2.00GHz e memória RAM de 250G. Servidor xanthomonas - modelo Intel(R) Xeon(R), com 32 CPUs modelo E5-2630, velocidade de 2.40GHz e memória RAM de 377G. O acesso às duas máquinas é restrito a usuários cadastrados previamente. Além disso, a UNESP-FCAV oferece o suporte técnico para serviços relacionados ao gerenciamento de rede (Intranet e Internet), segurança digital e manutenção elétrica.

4. RESULTADOS E DISCUSSÃO

4.1. Genoma de Bactérias Ácidos Lácticas

4.1.1. Isolamento de DNA genômico e sequenciamento *Long reads*

Para obter os genomas das Bactérias Ácidos Lácticas, nós realizamos a seleção de 27 isolados representativos do projeto temático Bioprospecção de bactérias probióticas bacteriocinogênicas: da otimização do cultivo à aplicação em sistemas de produção animal (2018/25511-1) e que apresentavam maior potencial para produção de peptídeos antimicrobianos ou potencial

biotecnológico. A seleção dos isolados foi configurada com a subamostragem de 9 isolados oriundos de aves, 8 de suínos, 6 tilápia-do-nilo, 3 truta-arco-íris e 1 isolado de silagem de grãos. O DNA genômico foi extraído de cada isolado bacteriano utilizando o produto comercial "PureLink™ Genomic DNA Mini Kit" (Invitrogen™, Thermo Fisher Scientific, Wilmington, DE, EUA), seguindo os protocolos do fabricante.

4.1.2. Qualidade do DNA genômico

A qualidade do DNAg foi direcionada para atender a demanda da empresa de sequenciamento (Arizona Genomics Institute) consoante as determinações técnicas da plataforma de sequenciamento PacBio HiFi REVIO™. Para esta demanda, nós realizamos a quantificação do DNAg extraído nos equipamentos NanoDrop (para verificação de contaminantes; em geral proteínas) e Qubit (maior acurácia na determinação da concentração de ácidos desoxirribonucleicos). Os resultados obtidos por estas análises estão na **tabela A1 (Anexo)**. No geral, a quantificação e o volume de DNAg apresentaram adequados e concordantes com as determinações técnicas da plataforma e demandada pela empresa de sequenciamento.

4.1.3. Envio do DNA genômico

O valor financeiro disponibilizado pela agência de fomento para o sequenciamento era próximo de US\$7000 dólares. Para otimizar o uso do recurso financeiro e sequenciar o maior número possível de espécies isoladas, nós optamos pela montagem de conjuntos ("Pool") de DNAg contendo DNA de várias espécies bacterianas em diferentes gêneros. Cada conjunto foi considerado uma amostra e ao todo foram geradas 5 amostras para sequenciamento. Ademais, para formação dos conjuntos, nós consideramos os valores de concentração do DNA quantificados pelo fluorômetro Qubit™ de cada espécie. Na **tabela A2 (Anexo)** é apresentada a configuração dos conjuntos com as espécies isoladas para sequenciamento do DNAg. Por fim, cada conjunto de DNAg foi enviado para o sequenciamento em duplicata e com volume mínimo de 40 ng/uL. Após a consolidação do material genético em cada amostra, nós verificamos a integridade, qualidade e concentração do DNAg através da eletroforese em gel de agarose. A figura A1 (Anexo I) apresenta a fotografia do gel com a corrida do DNAg para cada amostra em duplicata. O envio de material ao exterior foi realizado pela empresa

BioTransporte (TRANSPORTE DE CARGA BIOLÓGICA EXPRESS LTDA com CNPJ:14.068.428/0001-80) seguindo as orientações técnicas e políticas vigentes no país para envio de material biológico ao exterior. A preparação das bibliotecas de DNAg e o sequenciamento foram realizados pelo Arizona Genomics Institute.

4.2. Construção dos Genomas Sequenciados

4.2.1. Qualidade geral do Sequenciamento

O Arizona Genomics Institute tem experiência em sequenciamento de DNAg para geração de sequências longas. Ademais, o mesmo já realizou essa tarefa com elevada qualidade em momentos anteriores para o grupo de pesquisa coordenado pelo professor doutor Alessandro de Mello Varani. O serviço oferecido pelo instituto para sequenciamento “Long reads” é a plataforma PacBio HiFi REVIO™. No geral, a plataforma apresenta alta precisão, e isso minimiza erros de leitura e aumenta a qualidade geral dos dados sequenciados. Além disso, a plataforma PacBio HiFi REVIO™ é eficaz na resolução de estruturas genômicas complexas, como regiões repetitivas e variantes estruturais, que são desafiadoras para outras plataformas de sequenciamento. Estas vantagens ocorrem principalmente pela geração de “reads” de DNA muito longa e com baixo viés de amplificação; fatores fundamentais para a obtenção de representações precisas e completas do genoma alvo. A **tabela A3 (Anexo)** apresenta o resumo médio da quantidade e qualidade do sequenciamento em cada amostra de DNAg. Um detalhe nesse resumo está no nível de precisão maior que Q40 (Qualidade média das “reads”). O valor observado permite a geração de genomas mais precisos e essa característica pode abrir espaço para aplicações mais refinadas, tais como edição genética, identificação de variantes de nucleotídeos raras ou mutações de resistência de baixo nível.

4.2.2. Montagem dos genomas Sequenciados

A montagem das amostras sequenciadas foi realizada por duas ferramentas: Flye (KOLMOGOROV, et al., 2019) e hifiasm_meta (FENG, et al., 2022). As duas apresentaram características próximas tais como: trabalhar com “Long reads”, correção de erros pré montagem e capacidade de lidar com regiões repetitivas do genoma. No entanto, a ferramenta Flye é uma ferramenta

de montagem genômica de sobreposição de grafo de De Bruijn (Montagem hierárquica) e foi projetada para genomas únicos. Por outro lado, a ferramenta hifiasm_meta foi direcionada para montagem de dados metagenômicos baseada em grafo de "string" e foi estruturada para montar genomas mistos em uma amostra. Inicialmente a montagem das amostras sequenciadas foi realizada no Arizona Genomics Institute com o uso da ferramenta Flye. Já a montagem hifiasm_meta foi executada no servidor xanthomonas localizado no "Varani's LAB:Genomics & Bioinformatics" na UNESP-FCAV Jaboticabal-SP. A seguir, a título de comparação, nós apresentamos os resultados da montagem realizada pela ferramenta Fyle. Com essa ferramenta foi possível gerar 14 sequências em cada um dos conjuntos sequenciados 1, 2 e 3. Nos primeiros conjuntos 1 e 2 às 14 sequências montadas estavam circularizadas e quatro destas eram maiores que 1,8 Mbases (Mb). Já para o conjunto 3, onze sequências estavam circularizadas e três ficaram lineares com 125, 800 e 900 Kbases (Kb) de comprimento. Entre as sequências circularizadas, três sequências apresentaram tamanho superior a 1,8 Mb. Na amostra 4 foram geradas 54 sequências e destas 22 ficaram circulares, com seis sequências apresentando tamanho superior a 1,8 Mb. Entre sequências lineares uma sequência ficou com tamanho próximo de 1,1 Mb, quatro com comprimento entre 100 e 400 Kb e nove sequências entre 46 e 90 Kb.

Esta situação é semelhante à amostra 3, na qual estas sequências são fragmentos oriundos de regiões cromossômicas que a ferramenta Flye não conseguiu resolver. No conjunto 5 foi gerado 24 sequências e desse total duas sequências ficaram lineares; uma com tamanho superior a 235 Kbases e a outra com 12 Kbases. Nas sequências circulares, oito ficaram acima de 1,8 Mb e seis entre 47 e 85 Kb. Outro ponto importante no processo de avaliação da montagem do genoma é o número de cobertura do mesmo; ou seja, o número de vezes que cada nucleótido foi sequenciado. Considerando os conjuntos 1, 2 e 3 os valores médios de cobertura foram acima de 300 vezes e os conjuntos 4 e 5 estes valores ficaram acima de 150 vezes. Estes números refletem a configuração de como as amostras foram configuradas, uma vez que nos três primeiros conjuntos receberam 4 isolados e os dois últimos 7 e 8 isolados. Para efeito mais direto da cobertura, nós podemos considerar apenas as sequências de cada conjunto com valores acima de 1,8 Mb. Este valor sugere que estas

sequências são oriundas da região cromossômica, pois este valor é o número recorrente para o tamanho da região cromossômica de BALs. Nessa condição, a cobertura correspondente para as sequências com até 1,8 Mb seria de 200 a 570 vezes; para faixa maior que 1,8 até 2,2 Mb esse valor estaria entre 100 a 310 vezes, na faixa que 2,2 até 2,5 Mb o valor ficaria entre 30 a 366 vezes e para as sequências maiores que 2,5 até 5,7 Mb a cobertura estaria entre 30 a 366 vezes.

Apesar da montagem apresentar sequências fragmentadas para região cromossômica em duas amostras sequenciadas, a montagem realizada pela ferramenta Flye pode ser classificada como apta e de boa qualidade. Em contrapartida a montagem anterior, nós realizamos uma nova montagem das amostras sequenciadas através da ferramenta hifiasm_meta. Com esta ferramenta foi possível gerar um número maior de sequências quando comparada com a montagem realizada pela ferramenta Flye. A montagem hifiasm_meta permitiu agrupar as sequências geradas em três subgrupos: sequências maiores que 1,5 Mb, sequências menores que 300 Kb circular e sequências menores que 300 Kb linear.

Nos conjuntos 1, 2 e 3 quatro sequências com tamanho maior que 1,8 Mb foram geradas e todas circulares. Porém, para os subgrupos de sequências menores que 300 Kb o número de sequências foi diferente entre os três conjuntos. Para o subgrupo circular foram montadas 8, 10 e 6 sequências para as amostras 1, 2 e 3. No subgrupo linear foram montadas 22, 34 e 28 sequências respectivamente. No conjunto 4 nove sequências com tamanho maior que 1,5 Mb foram montadas e destas apenas uma sequência ficou linear. Para o subgrupo de sequências menores que 300 Kb o número de sequências circulares foi de 18 sequências e 32 lineares. Por fim, na amostra 5, nós obtivemos oito sequências com tamanho maior que 1,8 Mb e todas circulares. Nos subgrupos de sequências menores que 300 Kb, os números de sequências ficaram em 15 circulares e 61 lineares. Estes resultados indicam que todos os isolados reunidos nas 5 amostras tiveram suas sequências cromossômicas resolvidas.

Em relação a cobertura da montagem hifiasm_meta, nós observamos que as amostras 1, 2 e 3 apresentaram valores médios acima de 105 vezes e os conjuntos 4 e 5 valores acima de 90 vezes. Tendência semelhante a

montagem Flye. Se consideramos apenas as sequências isoladas de cada conjunto com valores acima de 1,8 Mb e agruparmos as sequências em grupos por tamanho, a cobertura correspondente seria de 63 a 115 vezes para as sequências de 1,8 Mb; 50 a 108 vezes para faixa maior que 1,8 até 2,2 Mb; 72 a 110 vezes para faixa maior que 2,2 até 2,5 Mb e 86 a 98 vezes para faixa maior que 2,5 até 5,7 Mb.

Com estes valores de cobertura, nós podemos inferir que a montagem hifiasm_meta apresentou amplitude menor que a observada nas sequências montadas pelo Flye. Ademais, a montagem hifiasm_meta apresentou um número maior de sequências montadas tanto circulares quanto lineares. Devido a estas particularidades nós escolhemos a montagem realizada pela ferramenta hifiasm_meta para continuidade das atividades de anotação e análise genética das amostras sequenciadas. Outro fator importante na montagem montagem hifiasm_meta foi a capacidade de montar a sequência de DNAg relacionada com o isolado *Limosilactobacillus vaginalis* que não havia sido realizada pela montagem Flye.

Outro aspecto observado na montagem hifiasm_meta foi a resolução da sequência cromossômica do isolado *Lactobacillus amylovorus* em sequência única com tamanho de 2 Mb. Este valor é próximo ao tamanho médio observado na literatura para *Lactobacillus amylovorus*; 2,1 Mb. Pois este isolado, quando consideramos a montagem realizada pela ferramenta Flye, foi apresentado com sequência cromossômica fragmentada em vários pedaços de tamanho variando entre 270, 410 Kb a 1,1 Mb.

4.2.3. Identificação e rotulação dos cromossomos

Para identificar e rotular as sequências cromossômicas, nós utilizamos a ferramenta GTDB-Tk (CHAUMEIL, et al., 2020) suportada pelos dados de referência do banco GTDB (Genome Taxonomy Database - release 214) (PARKS et al., 2021). A **tabela A4 (Anexo)** apresenta todas as espécies identificadas pela ferramenta GTDB-Tk. Os destaques em vermelho são referentes aos isolados identificados anteriormente pelo processo de MALDI-TOF (ANGELETTI, 2017) ou sequência parcial do gene 16S rRNA que tiveram alteração na sua identificação taxonômica após o sequenciamento do genoma.

A ferramenta GTDB-Tk oferece uma taxonomia consistente e abrangente para organismos procariontes com base em dados genômicos tais

como os genes RNA ribossômico 16S, proteínas conservadas e específicas para diferentes grupos taxonômicos. Ademais, apenas as sequências com comprimento superior a 1,5 Mb foram rotuladas como região cromossômica pela ferramenta GTDB-Tk.

4.2.4. Identificação e rotulação do Mobiloma

O Mobiloma é o conjunto de elementos genéticos móveis que podem ser encontrados nos genomas bacterianos, entre eles estão os transposons, plasmídeos, fagos, integrons e ilhas de patogenicidade. Nesse tópico as análises foram direcionadas para os seguintes elementos genéticos móveis: plasmídeos e fagos (bacteriófagos). Todas sequências não rotuladas pela ferramenta GTDB-Tk foram utilizadas no processo de identificação e rotulação do Mobiloma. O processo de identificação e rotulação dos plasmídeos foi realizado pelas ferramentas deeplasmid (ANDREOPOULOS et al., 2022) e PLASMe (TANG, et al., 2023); respectivamente.

A ferramenta deeplasmid é baseada em *machine learning* treinada com diferentes sequências de plasmídeos e cromossomos. Essa estrutura possibilita a ferramenta classificar a origem da sequência de DNA em duas classes: plasmídeo ou cromossomo. Em seguida, as sequências identificadas como plasmídeos são rotuladas pela ferramenta PLASMe. Essa ferramenta utiliza alinhamento para identificar plasmídeos já descritos na literatura (NCBI). Quando as sequências são divergentes da referência, a ferramenta utiliza o modelo de aprendizado de máquina *Transformer*. Esse modelo foi desenvolvido inicialmente para processamento de linguagem natural, mas recentemente foi adaptado para sequências genéticas com ótimos resultados.

A identificação e rotulação de fagos foi realizada pelas ferramentas PhaBox (SHANG, et al., 2023) e PHASTEST (WISHART, et al., 2023). A ferramenta PhaBox integra outras ferramentas tais como PhaMer (SHANG, et al., 2022), PhaGCN (SHANG, et al., 2021), CHERRY (SHANG, et al., 2022), e PhaTYP (SHANG, et al., 2023) para realizar a identificação, classificação taxonômica, estilo de vida e previsão de hospedeiro através da sequência de nucleotídeos pequena; em geral menores que 300 Kb. Já a ferramenta PHASTEST vai além do escopo da PhaBOX. A PHASTEST pode identificar, anotar e visualizar sequências de profagos em genomas bacterianos e plasmídeos. Os resultados das ferramentas deeplasmid e PLASMe foram

concordantes. Em resumo, para a amostra 1 foram identificados 8 plasmídeos circulares e 1 linear. No conjunto 2 foram identificados 9 plasmídeos circulares e 2 lineares e no conjunto 3 foram identificados 5 plasmídeos circulares e 3 lineares. Na amostra 4 nós identificamos 15 plasmídeos circulares e 4 lineares e para amostra cinco 14 plasmídeos circulares e 7 lineares. Já para os bacteriófagos foram identificados 4 com sequências lineares no conjunto 1. Já na amostra 2 foram identificados 1 fago com sequência circular e 6 lineares. Em seguida, no conjunto 3, nós identificamos 1 fago com sequência circular e 4 lineares. Na Amostra 4 foram identificados apenas 3 fagos lineares e para a amostra 5 foram identificados 2 fagos circulares e 5 lineares. É importante ressaltar que todos elementos genéticos foram identificados com pontuação de precisão acima de 99%.

4.3. Anotação das sequências do Genoma

A predição de genes e anotação das sequências de DNA foram realizadas por duas ferramentas PGAP (TATUSOVA, et al., 2016) e Bakta (SCHWENGERS, et al, 2021) para sequências de DNA cromossomal e/ou plasmidial. Todas as sequências foram re-orientadas para iniciar a sequência fasta com os genes de início de replicação dnaA para cromossomo, repA para plasmídeos e terL para fagos. As principais diferenças entre as ferramentas estão na predição de genes [PGAP usa o GeneMarkS-2+ (LOMSADZE, et al, 2018) e Bakta usa o Prodigal (HYATT, et al, 2010)] e no banco sequências de DNA de referência [RefSeq (LI, et al, 2021) para PGAP e UniProt (UNIPROT CONSORTIUM, 2019) para Bakta].

Após a predição dos genes, as sequências de aminoácidos das duas ferramentas foram utilizadas para anotação funcional através da ferramenta EggNOG-mapper (CANTALAPIEDRA, et al, 2021) com o banco de dados eggNOG database 6.0 (HUERTA-CEPAS, et al, 2019) como referência. Em paralelo a avaliação da completude dos genomas foi realizada pela ferramenta BUSCO (SIMÃO, et al, 2016) com identificação de linhagem automática considerando apenas sequências de procariotos.

4.3.1. Cromossomos

O número de genes preditos e anotação das sequências codificantes (CDSs) foram similares para as duas ferramentas; PGAP e Bakta. No entanto, a ferramenta PGAP apresenta algumas vantagens em relação a ferramenta Bakta tais como; apresentação dos arquivos de anotação padronizados para submissão ao banco de dados GenBak e o um preditor de genes GeneMarkS-2+. Esta ferramenta é baseada em modelos de aprendizado de máquina treinados com dados genômicos para melhorar a precisão da predição. Essa característica capacita o GeneMarkS-2+ identificar genes com estruturas complexas. Por isso, na **tabela A5 (Anexo)** o número de CDS identificada pela ferramenta PGAP é menor que o número observado na Bakta. Essa diferença pode estar associada com a fragmentação das CDSs realizadas pelo Prodigal na ferramenta Bakta. Basicamente, o tamanho das sequências cromossômicas e o número de CDSs foram proporcionais e este resultado é o esperado para cromossomos bacterianos. Além disso, a análise de completude realizada pela ferramenta BUSCO demonstrou que as sequências montadas e anotadas estão completas e com ótima qualidade. Uma vez que mais 99% das sequências cromossômicas apresentaram genes considerados essenciais e conservados ao longo da evolução para espécie identificada e anotada.

4.3.2. Plasmídeos

Para a anotação dos plasmídeos, nós utilizamos os dois subgrupos de sequências menores que 300 Kb gerados pela montagem hifiasm_meta: circular e linear. Estas sequências foram previamente identificadas e rotuladas pelas ferramentas deeplasmid e PLASMe para plasmídeo e as ferramentas PhaBOX e PHASTEST para bacteriófago. Considerando o subgrupo de sequências circular, as amostras 1 e 2 apresentaram números de sequências plasmidiais similares entre todas as espécies pertencentes nestes conjuntos. Em resumo, foram associados 1 plasmídeo para as espécies *Enterococcus faecium* e *Lactococcus lactis*; e dois para cada uma das espécies *Limosilactobacillus reuteri* e *Pediococcus pentosaceus*. Nestes dois conjuntos foi observado uma sequência plasmidial de 18 Kb parcial do plasmídeo pJIR4150 de *Clostridium perfringens* (NZ_LN835295). Ainda nestas amostras, nós observamos uma sequência de 13 Kb rotulada como plasmídeo que não apresentou similaridade com nenhuma sequência de DNA conhecida. Agora, quando consideramos o subgrupo linear, nós identificamos 1 sequência para o

plasmídeo pZL3 de *Lactiplantibacillus plantarum* para as duas amostras 1 e 2. Ademais, na amostra 2 nós verificamos que uma sequência linear foi duplamente rotulada como plasmídeo e bacteriófago. Essa sequência tem 49 Kb de tamanho e foi identificada como similar ao plasmídeo pF39_2 de *Enterococcus faecium* e assim como com o fago temperado de *Enterococcus* sp. FM11-1.

No conjunto 3, considerando apenas o subgrupo circular, nós encontramos um plasmídeo associado à espécie *Enterococcus faecalis* e dois para cada uma das espécies *Limosilactobacillus reuteri* e *Pediococcus pentosaceus*. No subgrupo linear foram observados 3 plasmídeos: o primeiro para pZL3 de *Lactiplantibacillus plantarum*; o segundo para pGTC3 de *Enterococcus faecalis* e terceiro para *Latilactobacillus curvatus*. Na amostra 4, considerando apenas o subgrupo circular, nós encontramos um plasmídeo associado às espécies *Lactobacillus amylovorus*, *Lactococcus lactis*, *Limosilactobacillus reuteri*, *Levilactobacillus brevis*, *Lactococcus* sp, *Pediococcus pentosaceus* e *Latilactobacillus curvatus*. Além destes, nós encontramos dois plasmídeos associados à espécie *Enterococcus faecalis* e uma sequência plasmidial de 18 Kb parcial do plasmídeo pJIR4150 de *Clostridium perfringens* (NZ_LN835295). Ainda para esta amostra, nós identificamos uma sequência de 13 Kb rotulada como plasmídeo que não apresentou similaridade com nenhuma sequência de DNA conhecida. Para subgrupo linear, nós observamos 3 plasmídeos: o primeiro para pLR581 de *Limosilactobacillus reuteri*; o segundo para *Latilactobacillus curvatus* e o último não apresentou similaridade com nenhuma sequência de DNA conhecida.

Finalmente no conjunto 5, nós encontramos um plasmídeo associado às espécies *Enterococcus mundtii*, *Lactiplantibacillus plantarum* e *Staphylococcus aureus*. Além destes, nós encontramos dois plasmídeos associados às espécies *Pediococcus pentosaceus*, *Staphylococcus epidermidis*, *Limosilactobacillus reuteri* e *Ligilactobacillus salivarius*. Ademais, três sequências rotuladas como plasmídeo não apresentaram similaridade com nenhuma sequência de DNA conhecida. Para subgrupo linear nós identificamos 7 plasmídeos sendo que um apresentou similaridade com o plasmídeo pLR581 de *Limosilactobacillus reuteri*, três plasmídeos parciais para *Ligilactobacillus*

salivarius e três sequências com nenhuma similaridade com outras sequência de DNA conhecida.

Considerando todas as amostras, nós observamos que alguns plasmídeos apresentaram genes para produção de antimicrobianos, resistências a antibióticos e elementos transponíveis. Para confirmar estas observações, outras análises estão em curso para finalizar a anotação funcional destes plasmídeos.

4.3.3. Bacteriófagos

Devido às características do sequenciamento “Long reads” foi possível identificar bacteriófagos intactos na região cromossômica e/ou disponível no Mobiloma das cinco amostras. Devido a similaridade de espécies encontradas nas amostras 1 e 2, nós encontramos na sequência cromossômica destas amostras os mesmos bacteriófagos. Nestes conjuntos foi possível identificar 2 fagos nas espécies *Enterococcus lactis* (NC_025453 e NC_028671), *Lactococcus lactis* (NC_004066 e NC_031016) e *Pediococcus pentosaceus* (NC_004305 e NC_009812). Na espécie *Limosilactobacillus reuteri* destes conjuntos nos encontramos apenas 1 bacteriófago (NC_019486).

Em seguida, o conjunto 3 apresentou resultado semelhante aos conjuntos 1 e 2 para as espécies *Pediococcus pentosaceus* e *Limosilactobacillus reuteri*. No entanto, para este conjunto, as espécies *Lactococcus lactis* e *Enterococcus faecalis* apresentaram apenas 1 bacteriófago (NC_002667) e dois (NC_013648 e NC_013696) respectivamente. Para o conjunto 4 nos identificamos três bacteriófagos em *Pediococcus pentosaceus* (NC_009812, NC_004305 e NC_011801) e *Limosilactobacillus reuteri* (NC_019456, 019486 e NC_023719); dois nas espécies *Lactococcus lactis* (NC_004066, NC_031016) e *Enterococcus faecalis* (NC_013648 e 028671) e apenas um em *Limosilactobacillus vaginalis* (NC_019486).

No conjunto 5 foi observado a presença de 1 bacteriófago nas seguintes espécies: *Staphylococcus epidermidis* (NC_031241), *Paenibacillus* sp (NC_028805) e *Lactobacillus johnsonii* (NC_005354). Além destas, o conjunto 5 apresentou 2 bacteriófagos em outras três espécies; *Ligilactobacillus agilis* (NC_004305 e NC_029068), *Pediococcus pentosaceus* (NC_004305 e NC_009812) e *Enterococcus mundtii* (NC_047763 e NC_047763). No entanto, apenas a espécie *Limosilactobacillus reuteri* (NC_019486, NC_048625 e

NC_048790) apresentou três bacteriófagos no seu cromossomo. Todos os bacteriófagos relatados na sequência cromossômica estavam intactos de acordo com a ferramenta PHASTEST. Em relação ao Mobiloma, nós observamos um bacteriófago circular nos conjuntos 2 e 3; e dois bacteriófagos circular no conjunto 5. Estes bacteriófagos circulares apresentam estilo de vida temperado.

Os conjuntos 1 e 4 não apresentaram bacteriófagos circulares no Mobiloma. Por outro lado, todos os conjuntos apresentaram bacteriófagos com sequência linear. Os bacteriófagos lineares identificados apresentaram estilo de vida mais diverso variando entre virulento e temperado. No conjunto 1 foram observados 3 bacteriófagos temperados e um virulento; nas amostras 2 e 3 o número de bacteriófagos foi equilibrado em relação ao estilo de vida. Com amostra dois representada por 3 virulentos e 3 temperados e a amostra três por 2 temperados e 2 virulentos. Por fim, os conjuntos 4 e 5 apresentaram apenas bacteriófagos temperados; com três representantes para amostra 4 e cinco na amostra 5.

Considerando todos os bacteriófagos identificados nos conjuntos cromossômicos e Mobiloma foi possível observar a presença de genes associados a resistências a antibióticos e elementos transponíveis. Assim como em plasmídeos, estes resultados são preliminares e necessitam de análises complementares para concluir a anotação funcional do Mobiloma.

4.4. Análises genômicas específicas

4.4.1. Identificação de bacteriocinas ou RiPPS

A identificação de genes associados à produção de a bacteriocinas ou peptídeos sintetizados ribossomalmente e modificados pós-tradução (RiPPS do inglês - Ribosomally synthesized and post-translationally modified peptides) foi realizada por duas abordagens: a primeira identificação orientada por sequências conhecidas e a outro pelo processo de novo através de modelos baseados em algoritmos de aprendizado de máquina (Machine Learning Algorithm).

Para o primeiro caso nós usamos as ferramentas BAGEL4 (VAN HEEL, et al, 2018), antiSMASH (BLIN, et al, 2023), estas ferramentas trabalham de forma similar usando modelos baseados em profile HMM (profile HMM - do

inglês profile Hidden Markov Model) para identificação de RiPPS. De forma geral, o modelo profile HMM é um sistema de pontuação específico por posição e este sistema é adequado para identificar regiões conservadas e variáveis em sequências de peptídeos contra diferentes bancos de dados de referência. No entanto para sequências não descritas anteriormente, o modelo profile HMM pode apresentar resultados com grande variedade de falsos positivos (predição positiva para RiPPS mas a sequência não é um RiPPS) e falsos negativos (o contrário do falso positivo). Além disso, no modelo profile HMM outros atributos relacionados com os peptídeos não são considerados no processo de rotulação da sequência entre eles podemos destacar os atributos de hidrofobicidade relativa, estrutura secundária e exposição ao solvente. Por isso, modelos baseados em algoritmos de aprendizado de máquina podem realizar a predição de RiPPS sem as limitações que as ferramentas baseadas em sequências de referências apresentam; ou seja, podem identificar RiPPS de novo. Para realizar a predição de novo, nós utilizamos as ferramentas MACREL (SANTOS-JUNIOR, et al., 2020) baseado em redes neurais e amPEP (LAWRENCE, et al, 2021) modelo baseado em florestas de decisão aleatórias (Random Forest).

A **tabela A6 (Anexo)** apresenta os resultados de identificação e predição de RiPPs nas sequências cromossômicas. Os resultados apresentados pelas BAGEL4 e antiSMASH foram similares uma vez que as duas ferramentas utilizam abordagens semelhantes, assim como o conjunto de sequências de referência correlacionadas. No entanto, o processo realizado pela ferramenta anitSMASH é organizado no agrupamento de genes responsáveis pela biossíntese de metabólitos secundários, e essa característica permite a estruturação dos genes responsáveis pela produção, maturação, transporte e resistência RiPPs facilitando as análises posteriores.

Por fim, as predições de novo para peptídeos antimicrobianos (AMP do inglês - Antimicrobial peptides) realizadas pelas ferramentas MACREL e amPEP foram muito próximas, esse resultado é um indicador do potencial antimicrobianos que estas sequências podem apresentar, pois as ferramentas usam algoritmos de aprendizado de máquina diferentes e independentes. Ademais, a ferramenta MACREL classifica peptídeos em duas classes hemolíticos não hemolíticos e estas informações podem ser essenciais para o desenvolvimento de produtos biotecnológicos.

Inicialmente tínhamos apontado o uso das ferramentas BOA (<https://github.com/idoerg/BOA>) e NeuBI (HAMID, et al., 2019). As duas ferramentas são baseadas em Python, porém elas estão desatualizadas e a última atualização ocorreu há mais de 6 anos para NeuBI e 9 anos para BOA. Com isso, diferentes módulos estão apresentando conflitos de versão e os autores das ferramentas não apresentaram alternativas para execução das ferramentas que incluem o código, bibliotecas, dependências e variáveis de ambiente, garantindo que a ferramenta seja executada de maneira consistente, independentemente do ambiente em que é implantada.

Semelhantes às predições realizadas na região cromossômica, a predição de RiPPS nos plasmídeos foi concordante entre as duas análises: por referência e de novo. Com destaque para 2 plasmídeos circulares nos conjuntos 2 e 5 e outros dois plasmídeos lineares para os conjuntos 1 e 3 similares com o plasmídeo pZL3 de *Lactiplantibacillus plantarum*. Todos os plasmídeos apresentaram os genes para produção da bacteriocina Pediocina PA-1 assim como genes associados ao transporte, processamento e imunidade ao antimicrobiano. Aparentemente, estes plasmídeos são originários de *Pediococcus pentosaceus*, todos isolados de truta-arco-íris (*Oncorhynchus mykiss*). Recentemente, estudos in vivo com estes isolados estão em andamento.

Nós observamos dois plasmídeos com bacteriófagos incorporados. O primeiro foi encontrado no conjunto 2 com sequência linear. Essa sequência apresentou similaridade com o plasmídeo pF39_2 de *Enterococcus faecium*. O plasmídeo pF39_2 apresenta 8 genes de bacteriófagos na sua sequência de DNA e o seu tamanho é de 40 Kb e nenhum indicativo de genes com potencial antimicrobiano. Diferente do pF39_2, a sequência identificada na amostra 2 apresentou tamanho de 49 Kb, com 17/18 sequências de peptídeos antimicrobianos preditos pela análise de novo. Ademais, a sequência foi rotulada pela ferramenta deeplasmid como plasmídeo e pela ferramenta PhaBOX como bacteriófago; ambas com precisão elevada. Outro plasmídeo com bacteriófago incorporado foi identificado na amostra 5. O plasmídeo foi montado com 43 Kb e apresentou similaridade com o plasmídeo não nomeado de *Staphylococcus epidermidis* (NZ_CP016969). Esse plasmídeo apresenta tamanho de 23 Kb com três genes de bacteriófagos na sequência de DNA e

nenhum indicativo de genes com potencial antimicrobiano. Já no plasmídeo identificado na amostra 5 foi possível identificar 1 sequência de péptido com potencial antimicrobiano. Esse processo foi realizado pela análise de novo. Além disso, a sequência foi rotulada pela ferramenta deeplasmid como plasmídeo e pela ferramenta PhaBOX como bacteriófago; ambas com precisão elevada. Uma hipótese para estes resultados pode estar associada à capacidade que os bacteriófagos e plasmídeos têm de atuar na transferência horizontal de genes. Em especial durante infecções bacterianas, os bacteriófagos podem integrar seu material genético no genoma bacteriano hospedeiro e assim coexistir como elementos genéticos extracromossômicos, como no caso em discussão os plasmídeos. Uma outra alternativa para coexistência observada nos dois casos, é o compartilhamento entre bacteriófagos e plasmídeos do mecanismo de integração e mobilização da sequência de DNA. Este mecanismo pode resultar na presença conjunta de genes de bacteriófagos e plasmídeos na mesma sequência de DNA.

4.4.2. Genes associados à resistência a antimicrobianos

A identificação dos genes associados à resistência a antimicrobianos foi realizada pela ferramenta ABRICate usando diferentes bancos de dados para genes de resistência antimicrobianos NCBI AMRFinderPlus (FELDGARDEN, et al, 2019), CARD (JIA, et al, 2016), Resfinder (ZANKARI, et al, 2012), ARG-ANNOT (GUPTA, et al, 2014) e MEGARES (DOSTER, et al, 2020). Dos 27 isolados sequenciados, 14 apresentaram pelo menos 1 genes de resistência a antimicrobianos. Como por exemplo todos os isolados identificados como *Lactococcus lactis* apresentaram resistência bacteriana aos antibióticos lincosamidas devido a presença do gene *lmrD*; gene responsável pela codificação de uma bomba de efluxo. A principal função desta bomba é extrair lincosamidas do interior da célula bacteriana e bombear esses antibióticos para fora da célula. Já a resistência aos antibióticos da classe dos aminoglicosídeos foi mais presente no gênero *Enterococcus*. Três espécies, *Enterococcus lactis*, *Enterococcus hirae* e *Enterococcus mundtii*, apresentaram o gene para produção da enzima Aminoglycoside N-acetyltransferase AAC6'-PRIME. A principal função dessa enzima é modificar os aminoglicosídeos através da transferência de um grupo acetil da coenzima acetil-coenzima A (AcCoA), para

o grupo amino de aminoglicosídeos. A espécie *Enterococcus lactis* também apresentou o gene *msrC* responsável pela produção da proteína MsrC. Essa proteína atua como transportador de efluxo removendo ativamente os antibióticos do interior da célula bacteriana. As espécies *Lactobacillus amylovorus* e *Lactobacillus johnsonii* apresentaram o gene *erm* produtor de uma enzima metiltransferase. Essa enzima é responsável por metilar uma adenina específica no RNA ribossômico 23S, resultando em uma modificação que impede ou reduz a ligação dos antibióticos da classe dos macrolídeos ao ribossomo bacteriano. Essa enzima é responsável por conferir resistência a vários antibióticos da classe dos macrolídeos. O isolado *Paenibacillus* apresentou dois genes para resistência a antimicrobianos: um para aminoglicosídeos e outro para rifampicina. O maior destaque para esta análise está nos resultados das espécies *Enterococcus faecalis* e *Staphylococcus epidermidis*. Estas duas espécies apresentaram maior número de genes responsáveis pela resistência a antimicrobianos, sendo cinco genes *Enterococcus faecalis* e 12 genes em *Staphylococcus epidermidis*. A principal preocupação está associada ao tipo de resistência conferida por estes genes. No geral estes genes estão associados com a produção de proteínas transmembranares que atuando como sistemas de transporte ativo para extrair uma ampla variedade de compostos tóxicos ou terapêuticos do interior da célula bacteriana; conferindo a bactéria resistência a múltiplas drogas. Este tipo de resistência bacteriana é uma preocupação significativa em saúde pública, pois pode levar à falha do tratamento através de antibiótico e por conseguinte no aumento da disseminação de bactérias resistentes.

4.4.3. Genes associados à virulência

A identificação dos genes de virulência foi realizada pela ferramenta ABRICate (<https://github.com/tseemann/abricate>) usando banco de dados de fatores de virulência VFDB (CHEN, et al, 2016), mais os dados para fatores de virulência conhecidos em *Escherichia coli*; Ecoli_VF (INGLE, et al, 2016). Como esperado, apenas 4 isolados apresentaram genes de virulência e estes são reduzidos em duas espécies: *Enterococcus faecalis* e *Enterococcus lactis*. Os isolados da espécie *Enterococcus lactis* são originários de tilápia-do-nilo (*Oreochromis niloticus*) e a espécie apresentou dois genes de virulência: o gene *sgrA* que codifica uma proteína conhecida como SgrA (Cell Wall Anchored

Protein) responsável pela adesão a superfícies e interações com o ambiente hospedeiro; e o gene *acm* que codifica uma proteína conhecida como Acn (Collagen Adhesin). Essa adesina é uma molécula que permite que a bactéria se ligue a outras superfícies celulares ou moleculares tais como tecidos ricos em colágeno. Os isolados da espécie *Enterococcus faecalis* são originários de aves (*Gallus gallus*). A espécie em questão apresentou 28 genes de virulência. Em resumo esse grupo de genes estão distribuídos em mecanismos importantes de adesão e colonização da espécie tais como os genes *fss1* e *fss2* responsáveis pelas proteínas de superfície que se ligam ao fibrinogênio, facilitando a adesão da bactéria a superfícies biológicas; e os genes *ebpA*, *ebpB*, *ebpC*, *srtC* e *ace* produtores de proteínas envolvidas na formação de “pili de endocardite” e biofilme. Estes genes contribuem para a adesão e colonização da bactéria em diferentes superfícies. Em relação a patogenicidade e virulência o gene *prgB* está envolvido na produção de substâncias de agregação (surface protein Asc10) que pode ser responsável pela formação de biofilme e potencial patogenicidade da espécie. Além disso, duas enzimas proteolíticas também estão associadas com a virulência da bactéria, Estas enzimas são codificadas pelos genes *sprE* e *gelE* e podem estar envolvidos na degradação de proteínas no hospedeiro. Ademais, quatro genes (*cpsA*, *fsrA*, *fsrB* e *fsrC*) envolvidos na regulação da expressão gênica, modulação da virulência e a adaptação do patógeno ao ambiente estão presentes em *E. faecalis*. Por fim, onze (*cpsA*, *cpsB*, *cpsC*, *cpsD*, *cpsE*, *cpsF*, *cpsG*, *cpsH*, *cpsI*, *cpsJ* e *cpsK*) genes envolvidos na síntese da cápsula bacteriana, mecanismo responsável pela proteção da bactéria contra o sistema imunológico do hospedeiro. Esse mecanismo pode conferir resistência à fagocitose. Por fim, dois genes relacionados à disseminação da espécie, *ef0818* e *ef3023*, são responsáveis pela codificação de enzimas envolvidas na hidrólise de polissacarídeos, como a hialuronidase. E este mecanismo pode facilitar que a bactéria se espalhe e invada o tecido do hospedeiro pela bactéria. Acreditamos que estes resultados são fundamentais para tomada de decisão para uso e aplicação destes isolados em produtos biotecnológicos.

5. ORIENTAÇÕES

Desde Janeiro de 2024, o bolsista tem co-orientado o aluno de mestrado João Victor dos Anjos Almeida com projeto de pesquisa: "Decifrando o Potencial Prebiótico: Análise Genômica Comparativa da Microbiota Degradadora de Polissacarídeos em Animais de Produção". Este projeto tem atividades correlacionadas com o projeto Temático processo FAPESP 2018/25511-1. O estudante está matriculado no programa de pós-graduação em microbiologia agropecuária na UNESP-FCAV. Em janeiro de 2025, o projeto foi aprovado pela agência de fomento FAPESP com a concessão de bolsa de mestrado.

6. PRODUÇÃO

6.1. Artigos aceito

Outline

Highlights

Abstract

Keywords

Impact statement

Introduction

Results

Virulence factors and catalase test of LAB

Growth kinetics of LAB

Acidification kinetics of LAB

BLIS activity and minimum inhibitory concentrat...

Discussion

Materials and methods

CRediT authorship contribution statement

Ethics statement

Declaration of Competing Interest

Acknowledgments

Appendix A. Supplementary material

Data availability

References

Show full outline

Figures (6)





New Biotechnology
Volume 88, 25 September 2025, Pages 114-131



Phenotypic and genomic characterization of bacteriocin-producing lactic acid bacteria with probiotic and biotechnological potential for pathogen control in animal production

Pamela Oliveira de Souza de Azevedo ^{a b 1 2}, Mauro de Medeiros Oliveira ^{c 1 3}, Taís Mayumi Kuniyoshi ^{a 4}, Carlos Emilio Cabrera Matajira ^{a 5}, Elionio Galvão Frota ^{a 6}, Meriellen Dias ^{a 7}, Sebastián Armando Bermúdez-Puga ^{a 8}, Amanda Romana Santos Pessoa ^{a 9}, Anna Carolina Meireles Piazzentin ^{a 10}, Carlos Miguel Nóbrega Mendonça ^{a 11}, Wellison Amorim Pereira ^{a 12}, Attilio Converti ^{d 13}, José Manuel Domínguez ^{e 14} , Martin Gierus ^{b 15}, Alessandro M. Varani ^{c 16}, Ricardo Pinheiro de Souza Oliveira ^{a 17} 

Show more

Add to Mendeley

Share

Cite

<https://doi.org/10.1016/j.nbt.2025.04.008>

Get rights and content

Under a Creative Commons license

Open access

Highlights

- Three LAB strains show potential as probiotics to combat avian pathogens.
- LAB strains are tolerant to bile salts and low pH, ideal for poultry feed additives.

Pamela Oliveira de Souza de Azevedo, Mauro de Medeiros Oliveira, Taís Mayumi Kuniyoshi, Carlos Emilio Cabrera Matajira, Elionio Galvão Frota,

Meriellen Dias, Sebastián Armando Bermúdez-Puga, Amanda Romana Santos Pessoa, Anna Carolina Meireles Piazzentin, Carlos Miguel Nóbrega Mendonça, Wellison Amorim Pereira, Attilio Converti, José Manuel Domínguez, Martin Gierus, Alessandro M. Varani, Ricardo Pinheiro de Souza Oliveira,

Phenotypic and genomic characterization of bacteriocin-producing lactic acid bacteria with probiotic and biotechnological potential for pathogen control in animal production, *New Biotechnology*, Volume 88, 2025, Pages 114-131, ISSN 1871-6784, <https://doi.org/10.1016/j.nbt.2025.04.008>. (<https://www.sciencedirect.com/science/article/pii/S1871678425000469>)

*Abstract: The emergence of antibiotic-resistant pathogens has raised significant concerns in the poultry industry, driving the search for alternatives to antibiotics as growth promoters in animal feed. Probiotics, particularly those belonging to the Lactic Acid Bacteria (LAB) group, represent a promising solution by mitigating the risk of infectious disease. However, a uniform concentration of probiotic LAB is not suitable for feed additives due to varying growth kinetics. Additionally, the genomic and physiological profiles of the LAB strains involved must be thoroughly evaluated. In this study, we provide an analytical framework to comprehensively assess LAB as potential antibiotic alternatives in poultry farming. Three LAB strains – *Pediococcus pentosaceus* (isolated from corn silage), *Ligilactobacillus salivarius* (from the poultry gut) and *Lactococcus lactis* (from the gut of rainbow trout) – were sequenced and characterized, with a focus on evaluating their probiotic potential and safety at the genomic level. The analyses included co-culturing LAB with pathogens, assessing viable cells, and determining the minimum inhibitory concentration of bacteriocin-like inhibitory substances (BLIS). In addition to demonstrating effective antimicrobial activity against avian pathogens (*Salmonella* spp., *Clostridium* spp. and *Campylobacter coli*), the results revealed notable probiotic traits in all three LAB strains, such as tolerance to bile salts and acidic environment and adhesion to intestinal cells. In conclusion, our analytical framework and results highlighted the potential of the tested LAB strains as biotechnological tools for developing zootechnical additives. These probiotics show promise as viable alternatives to antibiotics for enhancing poultry health and productivity.*

Keywords: Lactic acid bacteria; Probiotics; Bacteriocins; Antibiotic resistance; Feed additive

6.2. Artigos submetidos ou em processo de submissão

Comparative Genomics Reveals Functional Complementarity among Probiotic Candidate Strains from Poultry and Swine Gut Microbiota

João Victor dos Anjos Almeida, Mauro de Medeiros Oliveira, Taís Mayumi Kuniyoshi, Fernando Moisés Mamani Sanca, Carlos Miguel Nóbrega Mendonça, Ricardo Pinheiro de Souza Oliveira, Alessandro de Mello Varani*

Abstract

Probiotics are live microorganisms that provide health benefits to the host by improving digestion, enhancing nutrient absorption, and modulating the immune system. Among them, lactic acid bacteria are known for producing vitamins and short-chain fatty acids, both essential for intestinal health. In this study, we performed high-fidelity (PacBio HiFi) whole-genome sequencing and comprehensive comparative genomic analysis of five Lactobacillales strains (Enterococcus lactis, Enterococcus mundtii, Ligilactobacillus agilis, Limosilactobacillus reuteri, Limosilactobacillus vaginalis) isolated from the intestinal microbiota of chickens and pigs. The assembled genomes ranged from 1.8 to 2.8 million base pairs, with more than 98 percent completeness and less than 1.31 percent contamination. Taxonomic classification, presence of antimicrobial resistance genes, bacteriocin biosynthetic potential, carbohydrate-active enzyme repertoires and vitamin biosynthesis pathways, and capacity to degrade plant polysaccharides were investigated. Functional characterization identified 65 families of carbohydrate-active enzymes, with E. mundtii presenting the greatest diversity (43 families) and absolute number (100 terms) of enzymes. Metabolic reconstruction revealed functional specialization among strains, with xylooligosaccharide degradation exclusive to E. mundtii and pectin utilization limited to E. lactis. Genes related to the biosynthesis of B-complex vitamins, including riboflavin, folate, and menaquinone, showed heterogeneous and complementary distribution among strains. These results highlight the potential for metabolic complementarity and cross-feeding, where metabolites produced by one strain serve as precursors for biosynthetic pathways in others. This study provides a genomic basis for developing multi-strain probiotics to improve intestinal health and feed efficiency in animal production.

Keywords Animal welfare · Microbial biotechnology · Gut microbiota · Prebiotics · Metabolic supplementation

Submetido para Probiotics and Antimicrobial Proteins 28/07/2025.

Genome-Driven Analysis Reveals the Biotechnological Potential of a Novel Paenibacillus sp. Isolated from Crude Oil

João Victor dos Anjos Almeida, Carlos Miguel Nóbrega Mendonça, Leandro Marcio Moreira, Ricardo Pinheiro de Souza Oliveira, Alessandro de Mello Varani and Mauro de Medeiros Oliveira*

Abstract

Microbial biotechnology plays a pivotal role in confronting environmental challenges and promoting sustainability. Here, we report the complete genome sequence of Paenibacillus sp. strain 210, previously isolated from Brazilian crude oil and noted for its levan metabolism and biosurfactant production. The 5.7 Mb assembled genome harbors four prophage regions and 13 antimicrobial biosynthetic gene clusters, including those encoding fusaricidin, paenicidin A, paenilan, paeninodin, and tridecaptin. Phylogenomic analysis combined with average nucleotide identity (ANI) measurements indicates that this strain does not cluster with any recognized Paenibacillus species, supporting its designation as a potential novel species. Notably, the identification of 259 carbohydrate-active enzyme (CAZyme) genes points to a robust capacity for degrading complex polysaccharides (e.g., cellulose, xylan, and pectin), positioning this bacterium as a promising candidate for biofuel production. Furthermore, the presence of complete metabolic pathways for B-vitamin biosynthesis underscores its potential utility in soil bioremediation through enhanced nutrient availability. In silico structural analyses of selected hydrolytic enzymes (GH1, GH5, GH11) reveal homology to functionally validated and crystallized proteins, highlighting their practical relevance. Collectively, these findings underscore the genetic versatility of Paenibacillus sp. strain 210 and its potential for diverse biotechnological applications, including ecosystem restoration, biofuel production, plant growth promotion, and biocontrol.

Keywords: bacteriocins, bioenergy, bioremediation, industrial microbiology, metabolic potential, Paenibacillus sp.

Submetido para MicrobiologyOpen em 28/03/2025

Enterococcus lactis LBM-M3D: Characterization and Antimicrobial Activity of the Potential Probiotic Strain Against Swine Pathogens and a Novel Low Molecular Weight Bacteriocin

Fernando Moisés Mamani Sanca, Mauro de Medeiros Oliveira, Alessandro de Mello Varani and Ricardo Pinheiro de Souza Oliveira*

ABSTRACT

Swine production is vital for global food security, yet its sustainability is undermined by the overuse of antibiotics, driving the emergence of multidrug-resistant pathogens. Probiotics and bacteriocins offer promising, natural alternatives to enhance animal health and reduce antibiotic dependence. In this study, Enterococcus lactis LBM-M3D—isolated from the porcine gastrointestinal tract—was evaluated for antimicrobial activity and probiotic potential. The strain effectively inhibited key swine pathogens, including Streptococcus suis, Staphylococcus hyicus, and Enterococcus faecalis, while showing tolerance to bile salts and acidic pH, and lacking major virulence factors. In silico analyses confirmed the absence of clinically relevant, acquired antimicrobial resistance genes, aligning with regulatory guidelines and supporting the strain's

safety as a probiotic candidate. Furthermore, we observed strong antimicrobial activity against *Listeria* pathogens. This characteristic may be associated with the four bacteriocin-like sequences (enterolysin A, mundticin ATO6, enterocin SE-K4 and enterocin P) identified in the genome of the LBM-M3D strain. These findings position our isolated as a promising probiotic for swine production, fostering sustainable livestock practices and mitigating antimicrobial resistance within the One Health framework.

Keywords: Bacteriocin, lactic acid bacteria, *Enterococcus lactis*, pig, peptides.

Submetido para Probiotics and Antimicrobial Proteins 16/05/2025.

A Novel *Lactococcus lactis* Isolate Demonstrates Genomic Safety and Modulates Intestinal Morphology in Rainbow Trout (*Oncorhynchus mykiss*)

Wellison Amorim Pereira, Sebastián Ignacio Ávila Puentes, Osvaldo Merino Painen, Iván Valdebenito, Adrian Jesus Hernandez Arias, Mauro de Medeiros Oliveira, João Victor dos Anjos Almeida, Alessandro de Mello Varani, Herman Patrício Dantagnan, Jorge G. Farías, Elías Figueroa Villalobos, Ricardo Pinheiro Souza Oliveira*

Abstract

*In rainbow trout aquaculture, probiotics offer sustainable alternatives to antibiotics, though strain-specific safety and efficacy require validation. This study evaluated a novel *Lactococcus lactis* isolate (LBMB2) through genomic safety assessment and an 8-week in vivo trial with 240 juvenile rainbow trout (13.58 ± 0.6 g). Fish were fed diets supplemented with 1×10^9 CFU/g LBMB2 or an unsupplemented control (triplicate tanks/treatment). Complete genome sequencing confirmed the taxonomic identity (*L. lactis*) and revealed the absence of virulence genes and antibiotic resistance markers. Furthermore, sequencing demonstrated the presence of biosynthetic clusters for bacteriocins, genes associated with exopolysaccharide (EPS) production, and conserved biosynthetic pathways for vitamins B2, B9, and K2. Dietary LBMB2 supplementation significantly enhanced survival ($P < 0.05$) and reduced hepatosomatic index ($P < 0.05$), indicating improved stress resistance and liver health. Intestinal morphometry showed increased goblet cell density ($P < 0.05$) alongside reduced villus dimensions and muscular layer thickness. No significant differences occurred in growth performance, feed conversion, or hematological parameters. These findings, supported by robust genomic safety data, demonstrate the potential of LBMB2 as a safe probiotic for rainbow trout that may contribute to intestinal barrier and stress resilience.*

Keywords: sustainable aquaculture; probiotic; fish nutrition; *Oncorhynchus mykiss*

Functional and genomic insights into *Pediococcus pentosaceus* as probiotic candidates with antibacterial properties against fish pathogens.

Elionio Galvão Frota; Taís Mayumi Kuniyoshi; Luara Lucena Cassiano; Taciana Freire de Oliveira; Amanda Romana Santos Pessoa; Fernando Moises Mamani Sanca; João Victor dos Anjos Almeida; Mauro de Medeiros Oliveira; Pamela Oliveira de Souza de Azevedo; Meriellen Dias; Nathalia Vieira Porphirio Veríssimo; Alessandro M. Varani; Leonardo Tachibana; Ricardo Pinheiro de Souza Oliveira*

Abstract

*This study aimed to determine the functional and genomic characteristics of *Pediococcus* spp. isolated from fish of the genus *Cichlasoma* spp., to establish their probiotic and biotechnological potential. The isolated strains BE2, BE6, BE8 and BE9 were considered safe for use with absence of virulence factors and antibiotic resistance profile compatible with applications in biotechnological products. The strains presented excellent probiotic properties such as high tolerance to stress in gastrointestinal tract environments (losses <2 log), good adhesion capacity in intestinal epithelial cells (~30%) and production of B vitamins (B2, B9 and B12). It was evidenced that the strains ferment their substrate in synergy and are capable of selectively fermenting prebiotics such as mannan-oligosaccharide (MOS), fructo-oligosaccharides (FOS) and inulin (INU) expanding their applications in synbiotic formulations. Furthermore, it was demonstrated that the strains can produce antimicrobial compounds during fermentation, which showed broad antimicrobial activity against important bioindicator pathogenic bacteria in aquaculture (*Streptococcus agalactiae* 1b, *Aeromonas hydrophila*, *Francisella orientalis* and *Lactococcus petauri*). Genomic analysis revealed genes related to the observed functions, supporting the phenotypic data. Therefore, the isolated strains present promising functional and genetic characteristics for use as probiotics and symbiotic consortia.*

Keywords: Lactic acid bacteria, Prebiotic, Mannan, vitamin B.

Advances in Engineering for Sustainable Bacteriocin Production: A Circular Economy Approach for Food Preservation

Taciana Freire de Oliveira, Larissa Pereira Brumano, Mauro de Medeiros Oliveira Elionio Galvão Frota, Alessandro de Mello Varani, Júlio César dos Santos, Ricardo Pinheiro de Souza Oliveira*

Abstract

The growing emphasis on environmental sustainability has driven a paradigm shift that integrates economic, environmental, and societal dimensions. This transformation aligns with bioeconomy principles, minimizing waste and optimizing resource use to promote sustainable and economically viable systems. In the food industry, these principles are reshaping production to meet consumer demands for health- and environmentally-conscious practices. Bacteriocins, antimicrobial peptides produced by certain microorganisms, have emerged as effective natural

biopreservatives and alternatives to chemical additives, particularly in ready-to-eat products requiring extended shelf life and enhanced food safety. However, challenges persist in optimizing fermentation and downstream processes to achieve industrial scale-up and long-term sustainability. Recent engineering strategies focus on using low-cost, food-grade substrates derived from agro-industrial by-products, thereby promoting waste valorization and aligning with the principles of a circular economy. Process engineering innovations – such as advanced bioreactor designs, enhanced fermentation control strategies, and techno-economic analyses have improved production yields, cost-effectiveness, and environmental performance. This review critically examines recent advances in bacteriocin research, focusing on strategies for process optimization, scalable production, and the valorization of agricultural by-products. By integrating innovations in fermentation technology, techno-economic analysis, and life cycle assessment, this work bridges laboratory findings with industrial applications, supporting the development of sustainable biopreservation methods.

Keywords: Bacteriocin, Biopreservatives, Biomanufacturing process, Sustainability, Circular economy.

6.3. Artigos com parceiros



Volume 13
2024

Article Contents

Abstract
Data Description
Context
Methods
Conclusions
Additional Files
Abbreviations
Acknowledgements
Author Contributions
Funding
Data Availability
Competing Interests
References
Author notes
Supplementary data

[< Previous](#) [Next >](#)

JOURNAL ARTICLE

Genomic decoding of *Theobroma grandiflorum* (cupuassu) at chromosomal scale: evolutionary insights for horticultural innovation

Rafael Moysés Alves , Vinicius A C de Abreu , Rafaely Pantoja Oliveira , João Victor dos Anjos Almeida , Mauro de Medeiros de Oliveira , Saura R Silva , Alexandre R Paschoal , Sintia S de Almeida , Pedro A F de Souza , Jesus A Ferro ... [Show more](#)

[Author Notes](#)

GigaScience, Volume 13, 2024, giae027, <https://doi.org/10.1093/gigascience/giae027>

Published: 05 June 2024 **Article history** ▼

 PDF  Split View  Annotate  Cite  Permissions  Share ▼

Abstract

Background

Theobroma grandiflorum (Malvaceae), known as cupuassu, is a tree indigenous to the Amazon basin, valued for its large fruits and seed pulp, contributing notably to the Amazonian bioeconomy. The seed pulp is utilized in desserts and beverages, and its seed butter is used in cosmetics. Here, we present the sequenced telomere-to-telomere genome of cupuassu, disclosing its genomic structure, evolutionary features, and phylogenetic relationships within the Malvaceae family.

Findings

The cupuassu genome spans 423 Mb, encodes 31,381 genes distributed in 10 chromosomes, and exhibits approximately 65% gene synteny with the *Theobroma cacao* genome, reflecting a conserved evolutionary history, albeit punctuated with unique genomic variations. The main changes are pronounced by bursts of long-terminal repeat retrotransposons at postspecies divergence, retrocopied and singleton genes, and gene families displaying distinctive patterns of expansion and contraction. Furthermore, positively selected genes are evident, particularly among retained and dispersed tandem and proximal duplicated genes associated with general fruit and seed traits and defense mechanisms, supporting the

Rafael Moysés Alves, Vinicius A C de Abreu, Rafaely Pantoja Oliveira, João Victor dos Anjos Almeida, Mauro de Medeiros de Oliveira, Saura R Silva, Alexandre R Paschoal, Sintia S de Almeida, Pedro A F de Souza, Jesus A Ferro, Vitor F O Miranda, Antonio Figueira, Douglas S Domingues, Alessandro M Varani, Genomic decoding of *Theobroma grandiflorum* (cupuassu) at chromosomal scale: evolutionary insights for horticultural innovation, GigaScience, Volume 13, 2024, giae027, <https://doi.org/10.1093/gigascience/giae027>

Abstract

Background: *Theobroma grandiflorum* (Malvaceae), known as cupuassu, is a tree indigenous to the Amazon basin, valued for its large fruits and seed pulp, contributing notably to the Amazonian bioeconomy. The seed pulp is utilized in desserts and beverages, and its seed butter is used in cosmetics. Here, we present the sequenced telomere-to-telomere genome of cupuassu, disclosing its genomic structure, evolutionary features, and phylogenetic relationships within the Malvaceae family.

Findings: The cupuassu genome spans 423 Mb, encodes 31,381 genes distributed in 10 chromosomes, and exhibits approximately 65% gene synteny with the *Theobroma cacao* genome, reflecting a conserved evolutionary history, albeit punctuated with unique genomic variations. The main changes are pronounced by bursts of long-terminal repeat retrotransposons at postspecies divergence, retrocopied and singleton genes, and gene families displaying distinctive patterns of expansion and contraction. Furthermore, positively selected genes are evident, particularly among retained and dispersed tandem and proximal duplicated genes associated with general fruit and seed traits and defense mechanisms, supporting the hypothesis of potential episodes of subfunctionalization and neofunctionalization following duplication, as well as impact from distinct domestication process. These genomic variations may underpin the differences observed in fruit and seed morphology, ripening, and disease resistance between cupuassu and the other Malvaceae species.

Conclusions: The cupuassu genome offers a foundational resource for both breeding improvement and conservation biology, yielding insights into the evolution and diversity within the genus *Theobroma*.

ORIGINAL RESEARCH article
Front. Plant Sci., 29 April 2025
Sec. Plant Abiotic Stress
Volume 16 - 2025 | <https://doi.org/10.3389/fpls.2025.1567201>

This article is part of the Research Topic
Role of Plant Cell Wall in Biotic and Abiotic Stress Resistance
[View all 8 articles >](#)

Download article

1,237
Total views
328
Downloads

View article impact >

1
View altmetric score >

Share on

Edited by
Jong Hee Im

Reviewed by
MALAY KUMAR ADAK
Benedetta Mattei

Table of contents

Abstract
1 Introduction
2 Materials and methods
3 Results
4 Discussion

Cell wall modulation by drought and elevated CO₂ in sugarcane leaves

Alexandre Junio Borges Araujo¹ Amanda Pereira de Souza¹ Débora Pagliuso
Mauro de Medeiros Oliveira Bruno Viana Navarro Adriana Grandis Marcos Silveira Buckeridge^{*}

Laboratório de Fisiologia Ecológica de Plantas, Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil

Climate change poses significant challenges to global agriculture, with elevated atmospheric CO₂ (eCO₂) concentrations and increased frequency of droughts affecting crop productivity. Understanding how economically important crops like sugarcane respond to these combined stresses is essential for developing resilient cultivars. This study explores the effects of eCO₂ and drought stress on sugarcane growth and cell wall composition. Sugarcane plants were cultivated under CO₂ treatments (390 ppm and 780 ppm) and subjected to drought stress. Leaf biomass, cell wall composition, and global transcriptome sequencing were analyzed. eCO₂ (780 ppm) significantly increased leaf biomass by 64%, attributed to enhanced photosynthesis and water-use efficiency. Conversely, drought reduced leaf biomass by 45%, highlighting sugarcane's sensitivity to water scarcity. When both conditions were combined, eCO₂ mitigated drought's negative impact, maintaining biomass at levels comparable to ambient conditions. Despite notable changes in biomass, cell wall biomass was only slightly affected. Under drought, a 14% reduction in cell wall biomass was observed alongside compositional changes, including reduced arabinosylation in glucuronoarabinoxylan (GAX). This alteration, supported by decreased xylan arabinosyl transferase (XAT) expression and reduced arabinose content, suggests stronger associations between GAX and cellulose, potentially enhancing drought tolerance by modifying cell wall rigidity and flexibility. Under eCO₂, cell wall composition was altered, with reductions in glucose and uronic acid in specific fractions, indicating decreased mixed-linkage glucan (MLG) and pectin. These changes likely increased cell wall flexibility, supporting rapid growth. Combined eCO₂ and drought treatments amplified specific modifications, such as enhanced fucosylation of xyloglucan (XG) and potential MLG expansion, both linked to stress adaptation. Overall, the findings underscore the critical role of cell wall plasticity in sugarcane's response to abiotic stress. While eCO₂ boosted growth and partially alleviated drought effects, structural changes in cell wall composition under these conditions

Araujo AJB, de Souza AP, Pagliuso D, de Medeiros Oliveira M, Navarro BV, Grandis A and Buckeridge MS (2025) Cell wall modulation by drought and elevated CO₂ in sugarcane leaves. *Front. Plant Sci.* 16:1567201. doi: 10.3389/fpls.2025.1567201

Abstract

Climate change poses significant challenges to global agriculture, with elevated atmospheric CO₂ (eCO₂) concentrations and increased frequency of droughts affecting crop productivity. Understanding how economically important crops like sugarcane respond to these combined stresses is essential for developing resilient cultivars. This study explores the effects of eCO₂ and drought stress on sugarcane growth and cell wall composition. Sugarcane plants were cultivated under CO₂ treatments (390 ppm and 780 ppm) and subjected to drought stress. Leaf biomass, cell wall composition, and global transcriptome sequencing were analyzed. eCO₂ (780 ppm) significantly increased leaf biomass by 64%, attributed to enhanced photosynthesis and water-use efficiency. Conversely, drought reduced leaf biomass by 45%, highlighting sugarcane's sensitivity to water scarcity. When both conditions were combined, eCO₂ mitigated drought's negative impact, maintaining biomass at levels comparable to ambient conditions. Despite notable changes in biomass, cell wall biomass was only slightly affected. Under drought, a 14% reduction in cell wall biomass was observed alongside compositional changes, including reduced arabinosylation in glucuronoarabinoxylan (GAX). This alteration, supported by decreased xylan arabinosyl transferase (XAT) expression and reduced arabinose content, suggests stronger associations between GAX and cellulose, potentially enhancing drought tolerance by modifying cell wall rigidity and flexibility. Under eCO₂, cell wall composition was altered, with reductions in glucose and uronic acid in specific fractions, indicating decreased mixed-linkage glucan (MLG) and pectin. These changes likely increased cell wall flexibility, supporting rapid growth. Combined eCO₂ and drought treatments amplified specific modifications, such as enhanced fucosylation of xyloglucan (XG) and potential MLG expansion, both linked to stress adaptation. Overall, the findings underscore the critical role of cell wall plasticity in sugarcane's response to abiotic stress. While eCO₂ boosted growth and partially alleviated drought effects, structural changes in cell wall composition under these conditions

supporting rapid growth. Combined eCO₂ and drought treatments amplified specific modifications, such as enhanced fucosylation of xyloglucan (XG) and potential MLG expansion, both linked to stress adaptation. Overall, the findings underscore the critical role of cell wall plasticity in sugarcane's response to abiotic stress. While eCO₂ boosted growth and partially alleviated drought effects, structural changes in cell wall composition under these conditions further contribute to stress resilience, emphasizing the adaptive mechanisms of sugarcane to environmental challenges. This is the first report in which eCO₂, and drought are combined to evaluate the response of sugarcane to the impact of climate changes.

7. DISCIPLINAS MINISTRADAS COM A RESPECTIVA CARGA HORÁRIA E/OU A REALIZAÇÃO DE ATIVIDADES DE EXTENSÃO.

7.1. Atividade de extensão: Curso livre “Introdução ao LaTeX”.

Oferecimento de minicurso presencial na segunda edição da "Escola de Verão em Microbiologia Agropecuária". Evento organizado por docentes do Conselho do Programa de Pós-Graduação em Microbiologia Agropecuária (PPG Micro), da FCAV – Unesp, com o apoio de outros docentes deste programa de pós-graduação assim como de discentes do programa.

Minicurso 11 – “Introdução ao LaTeX”.

Carga horária total: 12 h.

Número de vagas: 20.

Responsável: Prof. Dr. Alessandro de Mello Varani.

Colaboradores: João Victor dos Anjos Almeida, Mauro de Medeiros Oliveira e Vitor Trinca.

Local: Sala de aulas do prédio do IPBEN, Unesp, Campus de Jaboticabal.

7.2. Participação como professor colaborador on-line em disciplina do Programa de Pós-Graduação em Microbiologia e Imunologia do Instituto de Microbiologia Paulo de Góes da UFRJ

Professor convidado online para ministrar aulas teóricas e práticas na disciplina “**Introdução à análise de genomas bacterianos**” (IMM-818) no Programa de Pós-Graduação em Microbiologia e Imunologia do Instituto de Microbiologia Paulo de Góes da Universidade Federal do Rio de Janeiro, durante o período de 26 a 30 de maio de 2025.

Nome da disciplina: Introdução à análise de genomas bacterianos

Código: IMM-818 Carga horária (h): 30

Coordenador(es): Marinella Silva Laport

Professor(es) envolvido(s): Bruno Francesco Rodrigues de Oliveira (PPGMPA-UFF e PPGBM-IEAPM/UFF) e Mauro de Medeiros Oliveira (UNESP/FCAV)

8. COMENTÁRIOS DO SUPERVISOR

9. REFERÊNCIAS

- 1 - SACHDEVA, Angel et al. Exploring probiotics as a sustainable alternative to antimicrobial growth promoters: mechanisms and benefits in animal health. *Frontiers in Sustainable Food Systems*, v. 8, p. 1523678, 2025.
- 2 - DENG, Zhaoxi et al. The probiotic properties of lactic acid bacteria and their applications in animal husbandry. *Current microbiology*, v. 79, n. 1, p. 22, 2022.
- 3 - CHERUVARI, Athira; KAMMARA, Rajagopal. Bacteriocins future perspectives: Substitutes to antibiotics. *Food Control*, v. 168, p. 110834, 2025.
- 4 - SUGRUE, Ivan; ROSS, R. Paul; HILL, Colin. Bacteriocin diversity, function, discovery and application as antimicrobials. *Nature Reviews Microbiology*, v. 22, n. 9, p. 556-571, 2024.
- 5 - LAHIRI, Dibyajit et al. Bacteriocin: A natural approach for food safety and food security. *Frontiers in Bioengineering and Biotechnology*, v. 10, p. 1005918, 2022.
- 6 - FENG, Xiaowen et al. Metagenome assembly of high-fidelity long reads with hifiasm-meta. *Nature methods*, v. 19, n. 6, p. 671-674, 2022.
- 7 - KOLMOGOROV, Mikhail et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nature methods*, v. 17, n. 11, p. 1103-1110, 2020.
- 8 - BOURAS, George et al. Dnaapler: a tool to reorient circular microbial genomes. *Journal of Open Source Software*, v. 9, n. 93, p. 5968, 2024.
- 9 - CHAUMEIL, Pierre-Alain et al. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics*, v. 38, n. 23, p. 5315-5316, 2022.
- 10 - PARKS, Donovan 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 research*, v. 50, n. D1, p. D785-D794, 2022.

11 - PARKS, Donovan H. et al. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research*, v. 25, n. 7, p. 1043-1055, 2015.

12 - SIMÃO, Felipe A. et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, v. 31, n. 19, p. 3210-3212, 2015.

13 - LEE, Imchang et al. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *International journal of systematic and evolutionary microbiology*, v. 66, n. 2, p. 1100-1103, 2016.

14 - ANDREOPOULOS, William B. et al. DeepPlasmid: deep learning accurately separates plasmids from bacterial chromosomes. *Nucleic acids research*, v. 50, n. 3, p. e17-e17, 2022.

15 - TANG, Xubo et al. PLASMe: a tool to identify PLASMid contigs from short-read assemblies using transformer. *Nucleic Acids Research*, v. 51, n. 15, p. e83-e83, 2023.

16 - SCHWENGERS, Oliver et al. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microbial genomics*, v. 7, n. 11, p. 000685, 2021.

17 - TATUSOVA, Tatiana et al. NCBI prokaryotic genome annotation pipeline. *Nucleic acids research*, v. 44, n. 14, p. 6614-6624, 2016.

18 - CANTALAPIEDRA, Carlos P. et al. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular biology and evolution*, v. 38, n. 12, p. 5825-5829, 2021.

19 - WISHART, David S. et al. PHASTEST: faster than PHASTER, better than PHAST. *Nucleic Acids Research*, v. 51, n. W1, p. W443-W450, 2023.

20 - SHANG, Jiayu et al. PhaBOX: a web server for identifying and characterizing phage contigs in metagenomic data. *Bioinformatics Advances*, v. 3, n. 1, p. vbad101, 2023.

21 - CUMSILLE, Andrés et al. GenoVi, an open-source automated circular genome visualizer for bacteria and archaea. *PLoS Computational Biology*, v. 19, n. 4, p. e1010998, 2023.

- 22 - FELDGARDEN, Michael et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Scientific reports*, v. 11, n. 1, p. 12728, 2021.
- 23 - ALCOCK, Brian P. et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic acids research*, v. 51, n. D1, p. D690-D699, 2023.
- 24 - BORTOLAIA, Valeria et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *Journal of Antimicrobial Chemotherapy*, v. 75, n. 12, p. 3491-3500, 2020.
- 25 - GUPTA, Sushim Kumar et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrobial agents and chemotherapy*, v. 58, n. 1, p. 212-220, 2014.
- 26 - DOSTER, Enrique et al. MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic acids research*, v. 48, n. D1, p. D561-D569, 2020.
- 27 - CARATTOLI, Alessandra et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrobial agents and chemotherapy*, v. 58, n. 7, p. 3895-3903, 2014.
- 28 - BLIN, Kai et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic acids research*, v. 51, n. W1, p. W46-W50, 2023.
- 29 - VAN HEEL, Auke J. et al. BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic acids research*, v. 46, n. W1, p. W278-W281, 2018.
- 30 - EDDY, Sean R. Hidden markov models. *Current opinion in structural biology*, v. 6, n. 3, p. 361-365, 1996.
- 31 - SANTOS-JUNIOR, Célio Dias et al. Macrel: antimicrobial peptide screening in genomes and metagenomes. *PeerJ*, v. 8, p. e10555, 2020.
- 32 - LAWRENCE, Travis J. et al. amPEPpy 1.0: a portable and accurate antimicrobial peptide prediction tool. *Bioinformatics*, v. 37, n. 14, p. 2058-2060, 2021.

- 33 - ANGELETTI, Silvia. Matrix assisted laser desorption time of flight mass spectrometry (MALDI-TOF MS) in clinical microbiology. *Journal of microbiological methods*, v. 138, p. 20-29, 2017.
- 34 - SHANG, Jiayu et al. Accurate identification of bacteriophages from metagenomic data using Transformer. *Briefings in Bioinformatics*, v. 23, n. 4, p. bbac258, 2022.
- 35 - SHANG, J.; JIANG, J.; SUN, Y. Bacteriophage classification for assembled contigs using graph convolutional network. *Bioinformatics* 37: i25–i33. 2021.
- 36 - SHANG, Jiayu; SUN, Yanni. CHERRY: a Computational methOd for accuratE pRediction of virus–pRokarYotic interactions using a graph encoder–decoder model. *Briefings in Bioinformatics*, v. 23, n. 5, p. bbac182, 2022.
- 37 - SHANG, Jiayu; TANG, Xubo; SUN, Yanni. PhaTYP: predicting the lifestyle for bacteriophages using BERT. *Briefings in Bioinformatics*, v. 24, n. 1, p. bbac487, 2023.
- 38 - LOMSADZE, Alexandre et al. Modeling leaderless transcription and atypical genes results in more accurate gene prediction in prokaryotes. *Genome research*, v. 28, n. 7, p. 1079-1089, 2018.
- 39 - HYATT, Doug et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics*, v. 11, p. 1-11, 2010.
- 40 - HUERTA-CEPAS, Jaime et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic acids research*, v. 47, n. D1, p. D309-D314, 2019.
- 41 - LI, Wenjun et al. RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. *Nucleic acids research*, v. 49, n. D1, p. D1020-D1028, 2021.
- 42 - HAMID, Md-Nafiz; FRIEDBERG, Iddo. Identifying antimicrobial peptides using word embedding with deep recurrent neural networks. *Bioinformatics*, v. 35, n. 12, p. 2009-2016, 2019.
- 43 - CHEN, Lihong et al. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic acids research*, v. 44, n. D1, p. D694-D697, 2016.
- 44 - INGLE, Danielle J. et al. In silico serotyping of E. coli from short read data identifies limited novel O-loci but extensive diversity of O: H serotype

combinations within and between pathogenic lineages. Microbial genomics, v. 2, n. 7, p. e000064, 2016.

10. ANEXOS