

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
Instituto de Ciências Biológicas  
Departamento de Genética, Ecologia e Evolução  
Programa de Pós-Graduação em Genética

César da Silva Santana Moura

**Caracterização genômica dos profagos e elementos parecidos com profagos nas  
bactérias lácticas do gênero *Weissella***

BELO HORIZONTE

2023

César da Silva Santana Moura

**Caracterização genômica dos profagos e elementos parecidos com profagos nas  
bactérias lácticas do gênero *Weissella***

Tese apresentada ao Programa de Pós-Graduação em Genética do Departamento de Genética, Ecologia e Evolução do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito para a obtenção do título de Doutor em Genética.

**Orientador: Dr. Álvaro Cantini Nunes**  
**Co-orientadora: Dr. Patrícia Costa Lima da Silva**

**Belo Horizonte**  
**2023**

043

Moura, César da Silva Santana.

Caracterização genômica dos profagos e elementos parecidos com profagos nas bactérias lácticas do gênero *Weissella* [manuscrito] / César da Silva Santana Moura. – 2023.

122 f. : il. ; 29,5 cm.

Orientador: Dr. Álvaro Cantini Nunes. Coorientadora: Dr. Patrícia Costa Lima da Silva.

Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Genética. 2. Probióticos. 3. *Weissella*. 4. Profagos. I. Nunes, Álvaro Cantini. II. Silva, Patrícia Costa Lima da. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 575



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
Instituto de Ciências Biológicas  
Programa de Pós-Graduação em Genética

**ATA DE DEFESA DE TESE**

<b>ATA DA DEFESA DE TESE</b>	<b>178/2023</b> <b>entrada</b>
<b>César da Silva Santana Moura</b>	<b>1º/2019</b> <b>CPF: 116.120.416-47</b>

Às quatorze horas do dia **27 de outubro de 2023**, reuniu-se, no Instituto de Ciências Biológicas da UFMG, a Comissão Examinadora de Tese, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: **“Caracterização genômica dos profagos e elementos parecidos com profagos nas bactérias lácticas do gênero Weissella”**, requisito para obtenção do grau de Doutor em **Genética**. Abrindo a sessão, o Presidente da Comissão, **Álvaro Cantini Nunes**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa do candidato. Logo após, a Comissão se reuniu, sem a presença do candidato e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

<b>Prof./Pesq.</b>	<b>Instituição</b>	<b>CPF</b>	<b>Indicação</b>
Álvaro Cantini Nunes	UFMG	678.159.606-34	Aprovado
Gabriel Magno de Freitas Almeida	Arctic University of Norway	050.471.476-73	Aprovado
Paula Luize Camargos Fonseca	UFMG	108.925.146-71	Aprovado
Savio Henrique de Cicco Sandes	Universidade Estadual de Campinas	014.941.576-16	Aprovado
João Locke Ferreira de Araújo	UFMG	099.811.766-80	Aprovado

Pelas indicações, o candidato foi considerado: **APROVADO**.

O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

**Belo Horizonte, 27 de outubro de 2023.**

Álvaro Cantini Nunes

Gabriel Magno de Freitas Almeida

Paula Luize Camargos Fonseca

Savio Henrique de Cicco Sandes

João Locke Ferreira de Araújo

Assinatura dos membros da banca examinadora:



Documento assinado eletronicamente por **Paula Luize Camargos Fonseca, Usuária Externa**, em 27/10/2023, às 14:02, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Alvaro Cantini Nunes, Professor do Magistério Superior**, em 27/10/2023, às 14:48, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **João Locke Ferreira de Araújo, Usuário Externo**, em 27/10/2023, às 17:34, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Gabriel Magno de Freitas Almeida, Usuário Externo**, em 28/10/2023, às 12:09, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Sávio Henrique de Cicco Sandes, Usuário Externo**, em 30/10/2023, às 19:26, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **2745718** e o código CRC **E27E0AB7**.



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
Instituto de Ciências Biológicas  
Programa de Pós-Graduação em Genética

## **FOLHA DE APROVAÇÃO**

**“Caracterização genômica dos profagos e elementos parecidos com profagos nas bactérias lácticas do gênero Weissella”**

**César da Silva Santana Moura**

Tese aprovada pela banca examinadora constituída pelos Professores:

Álvaro Cantini Nunes  
Universidade Federal de Minas Gerais

Gabriel Magno de Freitas Almeida  
Arctic University of Norway

Paula Luize Camargos Fonseca  
Universidade Federal de Minas Gerais

Savio Henrique de Cicco Sandes  
Universidade Estadual de Campinas

João Locke Ferreira de Araújo  
Universidade Federal de Minas Gerais

Belo Horizonte, 27 de outubro de 2023.

06/12/2023

SEI/UFMG - 2745722 - Folha



Documento assinado eletronicamente por **Alvaro Cantini Nunes, Professor do Magistério Superior**, em 27/10/2023, às 14:49, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **João Locke Ferreira de Araújo, Usuário Externo**, em 27/10/2023, às 17:34, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Gabriel Magno de Freitas Almeida, Usuário Externo**, em 28/10/2023, às 12:09, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Paula Luize Camargos Fonseca, Usuária Externa**, em 29/10/2023, às 22:36, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Sávio Henrique de Cicco Sandes, Usuário Externo**, em 30/10/2023, às 19:26, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **2745722** e o código CRC **456F362D**.

---

Referência: Processo nº 23072.265323/2023-01

SEI nº 2745722

## AGRADECIMENTOS

Aos meus amados pais, Nilce da Silva Santana Moura e Sergio César de Moura, quero expressar minha profunda gratidão por sua incrível generosidade e apoio incondicional. Suas qualidades de amor, paciência e incentivo foram como um farol a iluminar cada caminho que escolhi em minha vida. Sei que posso contar com vocês até o fim, e é inestimável o valor de ter pais tão dedicados e inspiradores.

Aos membros da minha família, em especial às minhas queridas irmãs Nadia e Naila, agradeço a paciência e compreensão demonstradas ao longo de toda a minha jornada. Sua presença e apoio tornaram esta caminhada ainda mais significativa e especial. A família é o alicerce que sustenta nossos sonhos, e sou profundamente grato por tudo o que fizeram para me apoiar durante meu doutorado.

Aos meus valiosos amigos de laboratório, em particular à Karen, Gabi, Ingrid, Livia, Gustavo, Mari e Sarah, agradeço do fundo do meu coração por estarem ao meu lado nesta jornada acadêmica. O apoio de vocês transcende as palavras e foi uma âncora fundamental para a conclusão deste trabalho. Sua amizade e apoio não apenas tornaram os momentos mais desafiadores suportáveis, mas também enriqueceram profundamente minha experiência acadêmica. Suas brilhantes ideias, a disposição para colaborar e o espírito de equipe que compartilham são um reflexo de sua dedicação excepcional à pesquisa e ao nosso laboratório. Ter amigos tão talentosos e inspiradores como vocês ao meu lado é uma dádiva e uma alegria. Obrigado por tornarem essa jornada acadêmica não apenas produtiva, mas também incrivelmente divertida.

Ao meu orientador, professor Álvaro Cantini Nunes, quero expressar minha gratidão pela oportunidade, orientação, ensinamentos, confiança e pelo apoio inestimável que ofereceu ao longo desta jornada. Sua dedicação incansável e orientação foram fundamentais para o sucesso desta caminhada acadêmica, e sou profundamente grato por sua contribuição para meu crescimento como pesquisador.

À minha coorientadora, Dra. Patrícia Costa Lima da Silva, agradeço do fundo do coração pela orientação, expertise e comprometimento que foram imprescindíveis para a realização deste trabalho. Nas discussões, ideias, companheirismo e paciência, sua contribuição foi inestimável, e sua presença foi como um farol de sabedoria que iluminou meu percurso acadêmico. Sua orientação ajudou a moldar não apenas este projeto, mas também o meu desenvolvimento como pesquisador.

Aos respeitadores professores cujas contribuições foram fundamentais para o desenvolvimento deste projeto, em especial a Dra. Elisabeth Neumann e o Dr. Jacques Robert Nicoli, quero expressar minha profunda gratidão pelo conhecimento e orientações valiosas que compartilharam. Suas contribuições foram essenciais para o sucesso deste trabalho e enriqueceram minha formação acadêmica.

À coordenação, professores e colegas do curso de Ciências Biológicas da UFMG e da Pós-Graduação em Genética da UFMG, quero manifestar meu profundo agradecimento por contribuírem para meu crescimento acadêmico e pessoal ao longo desses anos. Sua dedicação à educação e à pesquisa é inspiradora, e é uma honra fazer parte desta comunidade acadêmica.

Às agências de fomento que tornaram esta jornada possível, em especial à CAPES, que financiou minha bolsa de Mestrado e Doutorado, e à FAPEMIG e CNPq, que apoiaram os projetos de pesquisa fornecendo os recursos necessários para a realização deste trabalho, quero expressar minha sincera gratidão. Seu suporte financeiro desempenhou um papel fundamental na viabilização deste projeto e na minha formação acadêmica.

Aos respeitáveis membros da banca examinadora, agradeço por aceitarem o convite para avaliar minha tese e pela contribuição valiosa que ofereceram à minha pesquisa.

E a todas as pessoas que, de alguma forma, direta ou indiretamente, contribuíram para minha formação acadêmica, meu mais sincero obrigado.

Esta jornada não teria sido a mesma sem o apoio e colaboração de todos vocês. Obrigado por fazerem parte desta trajetória.

## RESUMO

Bactérias ácido lácticas (BAL) constituem um grupo de microrganismos amplamente distribuídos em alimentos e na microbiota de humanos e animais. Ao longo das décadas, essas bactérias desempenharam um papel fundamental na indústria de alimentos, servindo como culturas fermentadoras essenciais. Mais recentemente, elas emergiram como agentes probióticos, oferecendo benefícios à saúde humana. O gênero *Weissella*, pertencente às BAL, ainda não possui membros caracterizados como probióticos. No entanto, destaca-se como uma área de pesquisa intrigante, uma vez que algumas de suas linhagens têm demonstrado potencial para serem caracterizados como probióticos. E dentre esses candidatos, a caracterização de seus profagos, que desempenham um papel vital na transferência de genes e na modulação das comunidades microbianas, permanece notavelmente subexplorada. Neste estudo, adotamos uma abordagem abrangente que incluiu 313 genomas disponíveis no banco de dados do *National Center for Biotechnology Information* (NCBI) e o sequenciamento, anotação e análise funcional dos genomas de três novas linhagens de *Weissella*: *W. cibaria* 1BM, *W. confusa* GIR46L4 e *W. paramesenteroides* V3B-11. Nossos resultados revelaram uma característica intrigante das *Weissella*: a presença de sistemas de restrição e modificação de sequências, como RM/DRUANTIA/DISARM, que desempenham um papel vital na defesa contra fagos. Esses sistemas promovem a coevolução entre as bactérias e os fagos, facilitando a incorporação de profagos e genes virais em seus genomas e a modificação dos profagos já existentes. Esta dinâmica é crucial para a sobrevivência dessas bactérias em ambientes desafiadores e para sua capacidade de competir eficazmente com outras espécies bacterianas. Outro aspecto relevante abordado foi a análise do sistema CRISPR-Cas. Detectamos variações em sua ocorrência e eficácia entre as diferentes espécies de *Weissella*. O gênero, em geral, apresenta sistemas CRISPR-Cas menos desenvolvidos, com a exceção notável da espécie *W. soli*, que apresentou dois subtipos distintos desse sistema. No entanto, a presença de espaçadores em número limitado em algumas espécies sugere uma possível limitação na capacidade de combate a fagos. Além disso, identificamos a preservação de profagos específicos em diferentes espécies de *Weissella*, destacando a riqueza e complexidade da diversidade genética presente nesses microrganismos. Quando se trata de resistência a antibióticos, nossa investigação não encontrou genes com alta identidade que representassem um risco significativo de transferência genética horizontal. A maioria dos genes de virulência compartilhava semelhanças com genes virais estruturais, sem alta identidade. Também identificamos proteínas virais com potencial em aplicações biomédicas, ampliando as possíveis utilizações desses microrganismos. Em resumo, este estudo proporciona uma compreensão mais profunda dos profagos em bactérias do gênero *Weissella*, destacando seu papel na evolução bacteriana e apontando para uma relativa segurança em relação à disseminação de genes de importância clínica. Este estudo aborda uma lacuna na pesquisa relacionada ao gênero *Weissella*, seus profagos e suas implicações biológicas. Diante da crescente relevância das bactérias ácido lácticas (BAL) e dos fagos, bem como do interesse no potencial biotecnológico das *Weissella*, nosso objetivo foi contribuir para a classificação do gênero como status GRAS (*Generally Recognized As Safe*) ou status QPS (*Qualified Presumption of Safety*). Essa abordagem fornece insights importantes sobre as interações fago-bactéria, com implicações significativas para a segurança e aplicações biotecnológicas das *Weissella*.

**Palavras-chave:** *Weissella*, probióticos, profagos, evolução bacteriana

## ABSTRACT

Lactic acid bacteria (LAB) constitute a group of microorganisms widely distributed in food and the microbiota of humans and animals. Over the decades, these bacteria have played a fundamental role in the food industry, serving as essential fermenting cultures. More recently, they have emerged as probiotic agents, offering health benefits to humans. The genus *Weissella*, belonging to LAB, does not yet have members characterized as probiotics. Despite this, it stands out as an intriguing research area, as some of its strains have shown potential to be characterized as probiotics. Nevertheless, the characterization of their prophages, which play a vital role in gene transfer and the modulation of microbial communities, has remained notably underexplored. In this study, we adopted a comprehensive approach that included 313 genomes available at the National Center for Biotechnology Information (NCBI) and the sequencing, annotation, and functional analysis of the genomes of three new *Weissella* strains: *Weissella cibaria* 1BM, *Weissella confusa* GIR46L4, and *Weissella paramesenteroides* V3B-11. Our results revealed an intriguing feature of *Weissella*: restriction-modification systems, such as RM/DRUANTIA/DISARM, which play a vital role in defense against phages. These systems promote coevolution between bacteria and phages, facilitating the incorporation of prophages and viral genes into their genomes and modifying existing prophages. This dynamic is crucial for the survival of these bacteria in challenging environments and their ability to compete effectively with other bacterial species. Another relevant aspect addressed was the analysis of the CRISPR-Cas system. We detected variations in its occurrence and effectiveness among different *Weissella* species. The genus generally presents less developed CRISPR-Cas systems, with the notable exception of the *Weissella soli* species, which exhibited two distinct subtypes of this system. However, the limited number of spacers in some species suggests a potential limitation in their ability to combat phages. Additionally, we identified the preservation of specific prophages in different *Weissella* species, highlighting the richness and complexity of the genetic diversity present in these microorganisms. Concerning antibiotic resistance, our investigation did not find genes with high identity representing a significant risk of horizontal gene transfer. Most virulence genes shared similarities with structural viral genes but lacked high identity. We also identified viral proteins with potential biomedical applications, expanding the possible uses of these microorganisms. In summary, this study provides a deeper understanding of prophages in *Weissella* bacteria, highlighting their role in bacterial evolution and pointing to relative safety regarding the dissemination of clinically important genes. This study addresses a gap in research related to the *Weissella* genus, its prophages, and their biological implications. Given the increasing relevance of LAB and phages, as well as the interest in the biotechnological potential of *Weissella*, our goal was to contribute to the classification of the genus as GRAS (Generally Recognized As Safe) or QPS (Qualified Presumption of Safety) status. This approach offers important insights into phage-bacteria interactions, with significant implications for the safety and biotechnological applications of *Weissella*.

**Keywords:** *Weissella*, probiotics, prophages, bacterial evolution

## Lista de abreviações:

(p/v) – peso por volume

(v/v) – volume por volume

°C – Graus Celsius

%GC – Porcentagem conteúdo de guanina-citosina

%A – Porcentagem adenina

%T – Porcentagem timina

%G – Porcentagem guanina

%C – Porcentagem citosina

*att* – Sítio de ligação (*Attachment site*)

ARGS – Genes de resistência aos antibióticos (*Antibiotic resistance genes*)

BAL – Bactérias ácido lácticas

BLAST – Ferramenta de procura de alinhamentos básicos locais (*Basic Local Alignment Search Tool*)

BLASTn – Ferramenta de procura de alinhamentos básicos locais de sequências de nucleotídeos (*Basic Local Alignment Search Tool nucleotide*)

BLASTp – Ferramenta de procura de alinhamentos básicos locais de sequências de aminoácidos/proteínas (*Basic Local Alignment Search Tool protein*)

Cas – Proteína associada a CRISPR (*CRISPR associated protein*)

CARD – Banco de dados abrangente de resistência aos antibióticos (*The Comprehensive Antibiotic Resistance Database*)

CDS – Região de codificação (*CoDing Sequence*)

COGs – Aglomerado de grupos ortólogos (*Cluster of Orthologous Groups*)

CRISPR – Conjunto de repetições palindrômicas curtas regularmente interespaçadas (*Clustered Regularly Interspaced Short Palindromic Repeats*)

crRNA – CRISPR RNA

DNA – Ácido desoxirribonucleico (*Deoxyribonucleic acid*)

dsDNA – Dupla fita de DNA (*double stranded DNA*)

dsRNA – Dupla fita de RNA (*double stranded RNA*)

EPS – Exopolissacarídeos

GRAS – Geralmente reconhecido como seguro (*Generally Recognized As Safe*)

GTAs – Agentes de transferência gênica (*Gene transfer agents*)

ICNV – Comitê internacional de nomenclatura de vírus (*International Committee on Nomenclature of Viruses*)

ICTV – Comitê internacional de taxonomia de vírus (*International Committee Taxonomy Viruses*)

kb - Quilobase(*kilobase*), ou milhares de bases

LPS – Lipopolissacarídeos

Mbp – Milhões de pares de base (*Millions of base pairs*)

ML – Máxima verossimilhança (*Maximum Likelihood*)

NCBI – Centro nacional de informações biotecnológicas (*National Center for Biotechnology Information*)

ORFs – Fase de leitura aberta (*Open Reading Frames*)

pb - Pares de base

PICIs – Ilha cromossômica induzida por fago (*Phage-inducible chromosomal island*)

QPS – Presunção qualificada de segurança (*Qualified Presumption of Safety*)

rDNA – gene DNA ribossomal

RNA – Ácido ribonucleico (*Ribonucleic acid*)

RNase – Enzima envolvida na clivagem do RNA

ROS – Espécies reativas de oxigênio (*Reactive Oxygen Species*)

ssDNA – Fita simples de DNA (*single stranded DNA*)

ssRNA – Fita simples de RNA (*single stranded RNA*)

TCP – Pili co-regulado por toxina (*toxin-coregulated pili*)

TYGS – Servidor de genomas de linhagens tipo(*Type Strain Genome Server*)

tRNA – RNA transportador

tmRNA - RNA de transferência-mensageiro

VLEs – Elementos parecidos com vírus (*Virus-Like Elements*)

VFDB – Fatores de virulência de bactérias patogênicas (*Virulence Factor of Pathogenic Bacteria*)

## Lista de figuras e tabelas:

Figura 1. Classificação de fagos baseado na morfologia e tipo de material genético. Esquema da composição estrutural e genômica de diferentes fagos e suas respectivas imagens por microscopia eletrônica de varredura.

Figura 2: Espectro contínuo do estilo de vida dos fagos.

Figura 3: Representação do ciclo lítico e lisogênico em fagos.

Figura 4: Esquema da indução de profagos através do sistema SOS.

Figura 5: Diferentes papéis de fagos na patogenicidade e evolução bacteriana.

Figura 6: Mecanismos de defesa contra fagos em bactéria.

Figura 7: Fluxograma da metodologia para anotação dos genomas e análise.

Figura 8: Fluxograma da metodologia para detecção e análise dos profagos no gênero *Weissella*.

Tabela 1: Painel de características das diferentes bibliotecas produzidas.

Figura 9. Distribuição das CDS em categorias para cada montagem.

Figura 10. Distribuição funcional e contagem dos genes em COGs (*Cluster of Orthologous Groups*) para o genoma *Weissella cibaria* 1BM (VSF134).

Figura 11. Distribuição funcional e contagem dos genes em COGs (*Cluster of Orthologous Groups*) para o genoma *Weissella confusa* GIR46L4 (VSF140).

Figura 12. Distribuição funcional e contagem dos genes em COGs (*Cluster of Orthologous Groups*) para o genoma *Weissella paramesenteroides* V3B-11 (VSF142).

Figura 13: Distribuição da quantidade de genes encontrados entre os membros do gênero *Weissella* para cada tipo de sistema de defesa contra fagos.

Figura 14: Árvore filogenética dos membros do gênero *Weissella*.

Figura 15: Número de elementos parecidos com vírus (*Virus-like elements – VLEs*), VLEs íntegros e tamanho do genoma das espécies de *Weissella*.

Figura 16: Distribuição do comprimento e quantidade de elementos parecidos com vírus (*Virus-like elements – VLEs*) dentro do Gênero *Weissella*.

Figura 17: Gráfico de dispersão da correlação entre o número de elementos semelhantes a fagos (*Virus-like elements* – VLEs) e o tamanho do genoma em genomas de *Weissella*.

Figura 18: Pan-genoma dos profagos no gênero *Weissella*.

Figura 19: Core-genoma das sequências de profagos do gênero *Weissella* com a mesma morfologia predita.

Figura 20: Árvore filogenética de genomas completos de grupos similares de fagos encontrados em *Weissella*.

Figura 21: Presença de grupos de profagos similares em cada espécie de *Weissella*.

Figura 22: Mapa de calor ilustrando a presença de grupos similares de profagos em cada espécie e linhagem de *Weissella*.

Figura 23: Quantidade de genes e grupos de profagos associados à resistência e virulência presente nos fagos de *Weissella* analisados.

## SUMÁRIO

1. INTRODUÇÃO.....	18
1.1 Bacteriófagos.....	18
1.1.1. Ciclo lítico e ciclo lisogênico.....	22
1.1.2. Profagos.....	28
1.1.3. Ecologia dos bacteriófagos.....	30
1.2 Sistema de proteção contra fagos de bacterias.....	32
1.3 Trato gastrointestinal.....	36
1.4 Bactérias ácido lácticas e o gênero <i>Weissella</i> .....	38
2. JUSTIFICATIVA.....	43
3. OBJETIVOS.....	45
3.1 Objetivo geral.....	45
3.2 Objetivos específicos.....	45
4. METODOLOGIA.....	46
4.1 Montagem de genomas de <i>Weissella</i> potencialmente probióticas e análise dos mecanismos de defesa contra fagos presentes no gênero <i>Weissella</i> .....	46
4.1.1. Crescimento bacteriano.....	46
4.1.2. Sequenciamento genômico.....	46
4.1.3. Montagem genômica.....	49
4.1.4. Anotação genômica e análise gênica.....	50
4.2 Detecção e análise dos profagos no gênero <i>Weissella</i> .....	51
4.2.1. Detecção dos profagos.....	53
4.2.2. Anotação gênica dos profagos completos de <i>Weissella</i> .....	54
4.2.3. Pan- core- genoma.....	54
4.2.4. Agrupamento dos profagos similares e análises.....	55
4.2.5. Análise da presença de genes de virulência e resistência.....	56
4.2.6. Análise estatística.....	57
5. RESULTADOS E DISCUSSÃO.....	58
5.1 Sequenciamento.....	58
5.2 Anotação funcional.....	60
5.3 Sistema de defesa contra fagos.....	66
5.4 Detecção <i>in silico</i> e análise de distribuição de elementos semelhantes a profagos em espécies de <i>Weissella</i> .....	71

5.5	Características dos Profagos de <i>Weissella</i> .....	76
5.6	Montagem de fagos em grupos e filogenética .....	81
5.7	A contribuição da genética de profagos para a resistência bacteriana e virulência em <i>Weissella</i> .....	90
6.	CONCLUSÃO.....	96
7.	REFERÊNCIAS .....	98

# 1. INTRODUÇÃO

## 1.1 Bacteriófagos

Bacteriófagos (fagos) são vírus que infectam bactérias, e como todo vírus, eles são parasitas intracelulares obrigatórios, ou seja, necessitam do hospedeiro para poder se replicar. Descobertos por Fredrick Twort e Felix D'Herelle, respectivamente em 1915 e 1917 (TWORT, 1915; D'HERELLE, 1917), os fagos são considerados as entidades biológicas mais abundantes e diversas do planeta ( $10^{31}$  partículas de fagos), superando as bactérias em uma razão de 10 fagos para cada bactéria (WHITMAN *et al.*, 1998; WOMMACK *et al.* 2000; BRÜSSOW *et al.*, 2002; SUTTLE, 2007). Fagos são ubíquos sendo encontrados em grande quantidade e diversidade em todos os ambientes em que bactérias já foram descritas, desde o oceano (SUTTLE, 2005; NIGRO *et al.*, 2017; SUTTLE, 2007), o intestino humano (DUERKOP *et al.*, 2012; REYES *et al.*, 2013; MINOT *et al.* 2013; MANRIQUE *et al.*, 2016; BENLER *et al.* 2021), na produção industrial (KILIÇ *et al.*, 1996; JONES *et al.*, 2000; MARCÓ *et al.*, 2012; HALTER *et al.*, 2018; PUJATO *et al.*, 2019), no ambiente hospitalar (ROVIK *et al.*, 2000; BROWN-JAQUE *et al.*, 2016), em ambientes inóspitos, como fossas marítimas (CHEN *et al.*, 2020; SU *et al.*, 2023), e em diversos outros ambientes (APPELT *et al.*, 2014; AL-SHAYEB, 2020).

Fagos possuem papéis cruciais no controle populacional e evolução dos seus hospedeiros, participando desde a transferência horizontal de genes, gerando diversidade para seus hospedeiros e para si próprio, até o controle microbiológico através da eliminação de uma população em um ambiente. A relação fago/bactéria levou a várias descobertas dentro da biologia molecular e ao conhecimento genético como um todo, alguns exemplos são: as enzimas de restrições do tipo II ou o sistema CRISPR, ambos usados para edição genômica e clivagem; a T4 DNA ligase e a T7 DNA polimerase de fagos usadas para replicação, recombinação e reparo; os fagos lambda, M13, fd e P1, que, modificados, são amplamente usados como vetores (SALMOND *et al.*, 2015). Além da importância no desenvolvimento de técnicas moleculares, eles são estudados para diversas finalidades, como controle populacional de bactérias patogênicas (DI LALLO *et al.*, 2014; MALIK *et al.*, 2017; YEN *et al.*, 2017; DRILLING *et al.*, 2017; LAVERGNE *et al.*, 2018; IWANO *et al.*, 2018; CHAN *et al.*, 2018; FURFARO *et al.*, 2018), modulação de microbiota no trato gastrointestinal (KUTATELADZE *et al.*, 2008; MCCARVILLE *et al.*, 2016; LEE *et al.*, 2018),

das vias aéreas por nebulização (CAO *et al.*, 2015), em dispersar biofilme bacteriano (SIRINGAN *et al.*, 2011; SZAFRAŃSKI *et al.*, 2017), nos estudos sobre desenvolvimento de antivirais, veículo de vacinas, pesquisa de câncer e nanotecnologia (HUANG *et al.*, 2012; CLARK *et al.*, 2006; RONCA *et al.*, 2012; HYMAN *et al.*, 2012; HARADA *et al.*, 2018; LÜBKE *et al.*, 2020; LIU *et al.*, 2022; JIANG *et al.*, 2019; ZHOU *et al.*, 2022).

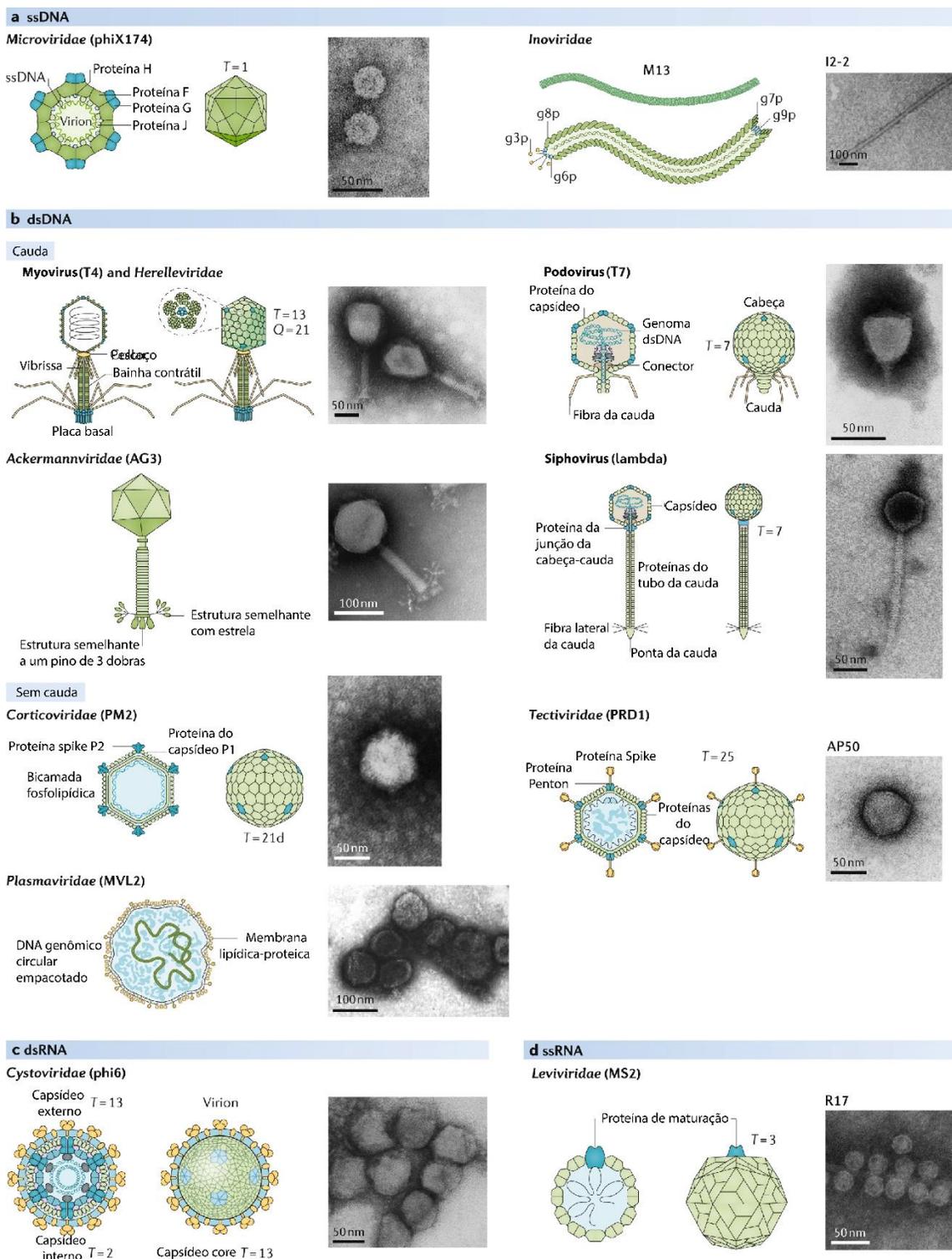
Historicamente, as primeiras classificações de fagos consideravam a sua morfologia (ACKERMANN, 2009). Em 1971, o Comitê Internacional de Nomenclatura Viral (ICNV) adotou o sistema de David Bradley de classificação, que classificava os fagos com cauda em 3 morfotipos, A (cauda contrátil), B (cauda longa e não contrátil) e C (cauda pequena e não contrátil), posteriormente enquadrados dentro das famílias, respectivamente, *Myoviridae*, *Siphoviridae* e *Podoviridae*. De 1981 até hoje, o Comitê Internacional de Taxonomia Viral (ICTV) é o responsável pela classificação e proposta de novas organizações da taxonomia dos vírus, o que inclui os fagos (TURNER *et al.*, 2021). Com o avanço das técnicas de sequenciamento e aumento dos genomas de fagos depositados nos bancos de dados públicos, tornou-se evidente a grande diversidade dessas três grandes famílias. Esse cenário levou à conclusão de que essas famílias não eram monofiléticas, resultando na necessidade de introdução de novas famílias e subdivisões (TURNER *et al.*, 2021).

Recentemente, houve uma mudança significativa na taxonomia de fagos, que aboliu as famílias tradicionais baseadas em morfologia, como *Myoviridae*, *Podoviridae* e *Siphoviridae*, juntamente com a eliminação da ordem *Caudovirales*. Essas classificações convencionais foram substituídas por uma taxonomia mais contemporânea que introduz a classe *Caudoviricetes*, uma categoria que engloba todos os vírus de cauda de DNA de fita dupla (dsDNA) que infectam hospedeiros bacterianos e arqueais. Essa reclassificação reflete uma mudança fundamental em nossa compreensão da diversidade e das relações entre fagos, enfatizando características genéticas e funcionais em vez de distinções morfológicas. (TURNER *et al.*, 2023).

Atualmente, o ICTV reconhece como vírus bacterianos 4 ordens, 47 famílias, 98 subfamílias, 1199 gêneros e 4079 espécies de vírus que infectam bactérias e arqueias (ANDRIAENSSENS *et al.*, 2020; TURNER *et al.*, 2023; ICTV report, 2023), contudo, essa lista deve aumentar nos próximos anos. O avanço nos estudos de fagos usando metagenomas e viomas são um dos principais

alavancadores dessas novas identificações (ZHONG *et al.*, 2015; QUAISER *et al.*, 2015; PAEZ *et al.*, 2016; ROUX *et al.*, 2016; KRISHNAMURTHY *et al.*, 2016; CALLANAN *et al.*, 2018; CREASY *et al.*, 2018; GONG *et al.*, 2018; SOMMERS *et al.*, 2019; WANG *et al.*, 2019; SHKOPOROV *et al.*, 2019; CALLANAN *et al.*, 2020; BAJAJ *et al.*, 2022; LAIO *et al.*, 2022; SALIH *et al.*, 2022; SCHACKART *et al.*, 2023; JI *et al.*, 2023 AKHTAR *et al.*, 2023). A abolição da ordem e famílias classificadas por morfologia resultou em um grande número de subfamílias e gêneros que muitas vezes se sobrepõem, dado a grande diversidade das sequências, e um enorme esforço da comunidade será necessário para melhor estabelecer e ranquear esses novos táxons (TURNER *et al.*, 2023).

Os genomas de fagos têm diversos tamanhos, variando desde 2,4 kb do fago L5 de *Leuconostoc* (*Leuconostoc* phage L5) até *Jumpophage* com mais de 735 kb (AL-SHAYEB *et al.*, 2020). O material genético dos fagos pode ser encontrado em fita simples ou dupla, tanto de DNA quanto de RNA, o que os torna extremamente versáteis, com uma grande diversidade genômica e com diversos papéis nas comunidades bacterianas. A maior parte dos fagos identificados, hoje, são dsDNA e possuem cauda, contudo, novos fagos dsDNA sem cauda e ssDNA já se mostraram dominantes em algumas comunidades, como no trabalho de Kauffman e colaboradores (2018) que demonstraram que fagos dsDNA sem cauda são mais abundantes que os fagos com cauda no oceano, e no trabalho de Roux e colaboradores (2019) que detectaram através de *machine learning* milhares de sequências de vírus ssDNA de genomas bacterianos e metagenomas, mostrando que a diversidade e o impacto desses fagos, no oceano, no solo e nos genomas bacterianos, era subestimada (BRUM *et al.*, 2013; KAUFFMAN *et al.*, 2018; LIM *et al.*, 2015; ROUX *et al.*, 2019; MOÏRA *et al.*, 2020). O material genético pode estar encapsulado de forma poliedral (*Microviridae*, *Corticoviridae*, *Tectiviridae*, *Leviviridae* e *Cystoviridae*), filamentoso (*Inoviridae*), pleomórfico (*Plasmaviridae*) ou conectado por uma cauda (*Caudoviricetes*) (ACKERMANN, 2009; MOÏRA *et al.*, 2020). (Figura 1)



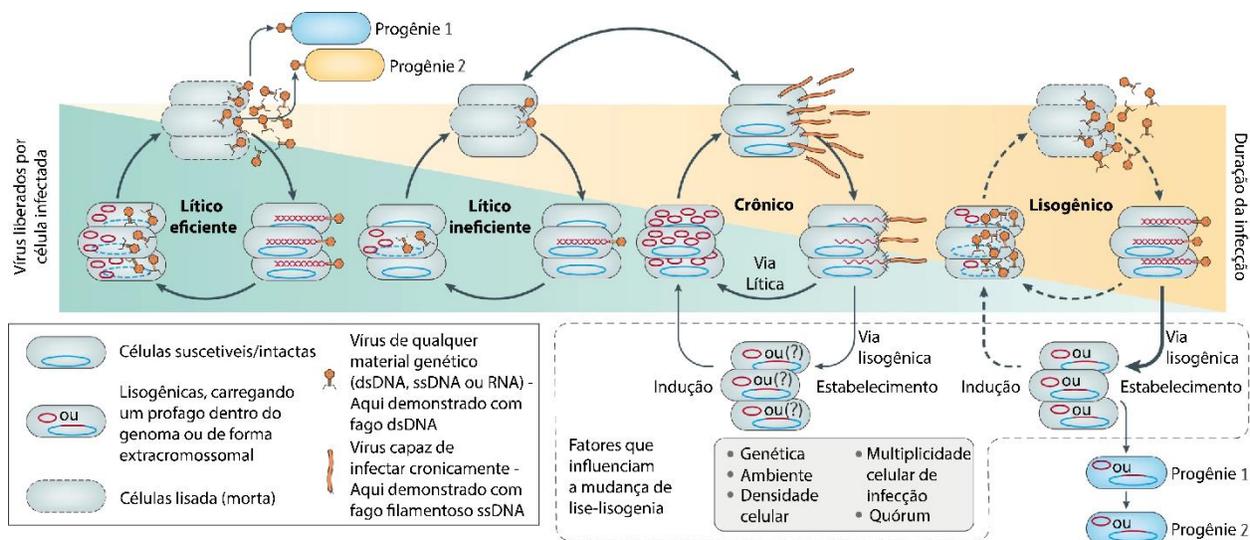
**Figura 1. Classificação de fagos baseado na morfologia e tipo de material genético. Esquema da composição estrutural e genômica de diferentes fagos e suas respectivas imagens por microscopia eletrônica de varredura. a:** Fagos que apresentam genoma ssDNA: *Microviridae* apresentam genoma ssDNA e capsídeo em formato icosaédrico; *Inoviridae* apresenta genoma ssDNA superenovelado empacotado em uma estrutura proteica filamentosa longa. **b:** Fagos que apresentam genoma dsDNA encapsulado por um capsídeo proteico: Com morfologia de miovírus são

fagos de cauda longa e contrátil; morfologia de podovírus possuem cauda curta não contrátil; *Ackermannviridae* tem a morfologia parecida com miovírus contudo possuem espinhos na ponta de suas caudas; morfologia de sifovírus possuem cauda longa não contrátil; *Corticoviridae* não possuem cauda, possuem genoma dsDNA circular e uma membrana lipídica interna ao capsídeo; *Tectiviridae* não possuem cauda, possuem genoma dsDNA linear e uma membrana lipídica interna ao capsídeo; Plasmaviridae possuem um genoma dsDNA circular envolto por uma membrana lipídica e não possui capsídeo. **c:** Fagos com material genético dsRNA são compostos por uma única família *Cystoviridae*, essa família apresenta dsRNA segmentado envolto por um capsídeo externo e duas camadas lipídicas internas. **d:** *Leviviridae* possuem material genético ssRNA que codifica somente quatro proteínas (*major capsid protein*, replicase, maturação e lise). O valor T representa o número de triangulação, definido como o número de proteínas por unidade assimétrica. Tendo T como faces equilaterais e Q para faces em um triângulo irregular (Modificado de MOÏRA, 2020).

Existem dois tipos de fagos; os fagos chamados virulentos que somente realizam o ciclo lítico; e os fagos temperados que podem caminhar para a lise do hospedeiro (ciclo lítico) ou podem ter seu genoma viral inserido no DNA do hospedeiro, denominado ciclo lisogênico. Os fagos que estão inseridos no DNA bacteriano recebem o nome de profago. Os profagos podem ser encontrados também em plasmídeos extracelulares lineares, como o fago P1 de *Escherichia coli* (*E. coli* phage P1) (ŁOBOCKA *et al.*, 2004), ou circulares como o fago N15 de *Escherichia coli* (*E. coli* phage N15) (RAVIN, 2011).

### 1.1.1. Ciclo lítico e ciclo lisogênico

Atualmente, o estilo de vida dos fagos é representado em um espectro contínuo, onde fagos apresentam, principalmente, dois polos: Com infecções eficientes, representado dentro do ciclo lítico em uma transmissão horizontal; e persistentes, representado dentro do ciclo lisogênico em uma transmissão vertical. Portanto, os profagos podem apresentar diversos estilos de vida entre os dois polos. (CORREA *et al.*, 2021). (Figura 2)

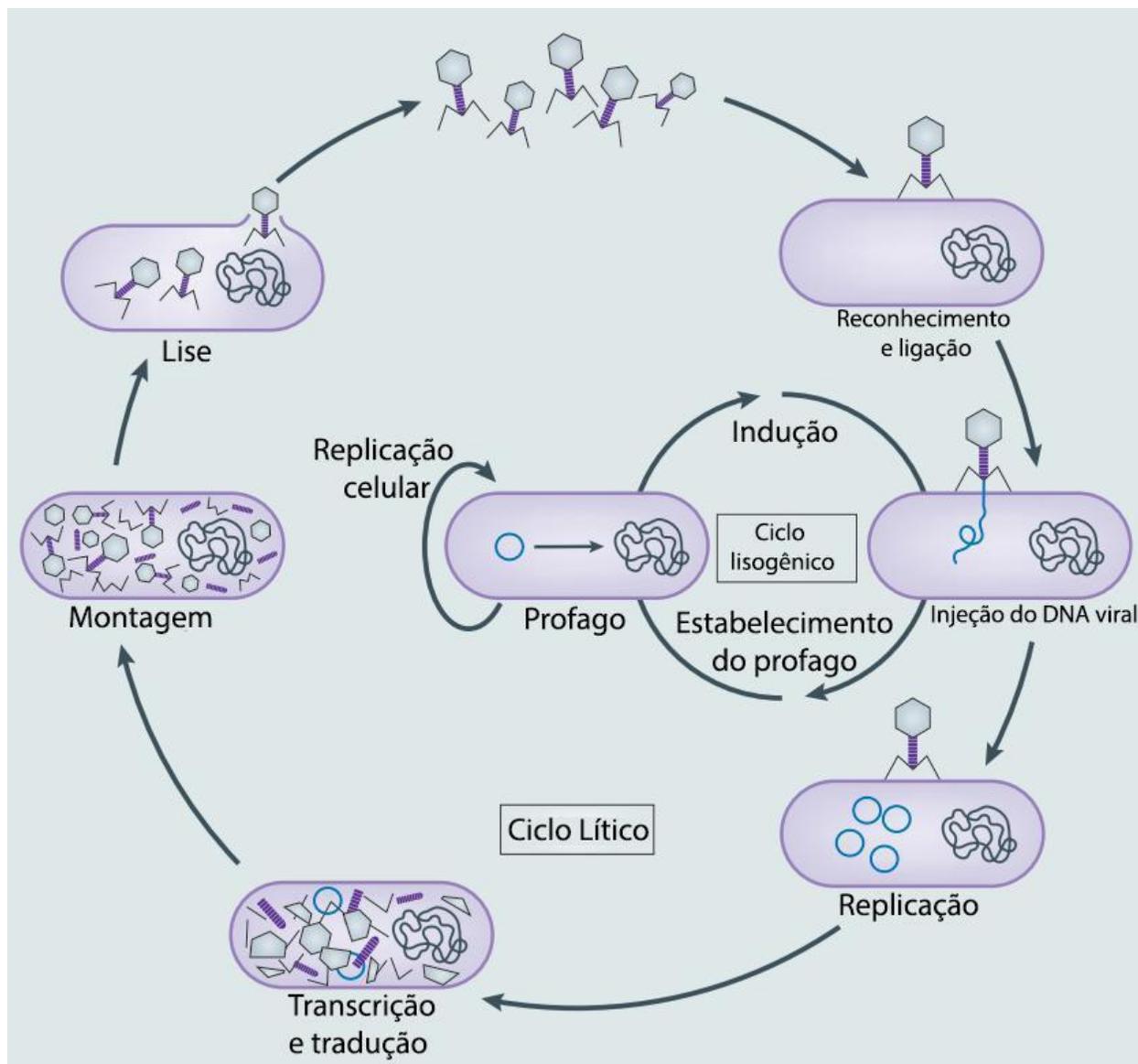


**Figura 2: Espectro contínuo do estilo de vida dos fagos.** À esquerda, são apresentadas as infecções eficientes dos fagos virulentos, enquanto à direita estão os fagos/profagos temperados persistentes. A figura ilustra o espectro contínuo em que os fagos podem se enquadrar, sendo influenciado pela quantidade de virions liberados (em verde) e a duração da infecção (em amarelo). Na parte inferior esquerda, encontra-se a legenda da figura, e na parte inferior direita, alguns possíveis fatores que influenciam a tomada de decisão são destacados. (Modificado de CORREA *et al.*, 2021)

Acredita-se que esse espectro seja influenciado por diversas condições que envolvem não apenas os fagos, mas também a população, seus hospedeiros e condições ambientais. Algumas dessas características ainda não foram totalmente caracterizadas. A fim de destacar as diferenças desses ciclos, iremos nos concentrar no que é tradicionalmente apresentado nos polos desse contínuo: um ciclo lítico e um ciclo lisogênico. No entanto, vale ressaltar que essas características são fluidas, o que significa que existe uma grande diversidade de estilos, muitas vezes sobrepostos.

O ciclo lítico se inicia na adsorção viral, onde as proteínas do vírion (partícula viral infecciosa) se ligam a receptores específicos na superfície da bactéria. Para os membros da classe *Caudoviricetes*, que representam a maior parte dos fagos descritos, a maquinaria para adsorção viral se encontra na cauda, variando em complexidade dentro da ordem, incluindo partes como fibras, espinhos, placa basal e outras estruturas proteicas (FOKINE *et al.*, 2014; CHATURONGAKUL *et al.*, 2014). Essas proteínas dos fagos podem reconhecer e se ligar a diversas estruturas da célula do hospedeiro. Dentre os alvos conhecidos estão proteínas do flagelo (SHIN *et al.*, 2012), porinas e proteínas osmorreguladoras (MARTI *et al.*, 2013), proteínas de bomba de efluxo (RICCI *et al.*,

2010), proteínas transportadoras (CASJENS *et al.*, 2005), LPS (PICKARD *et al.*, 2010; SHIN *et al.*, 2012; BAE *et al.*, 2013) e *pili* (BAE *et al.*, 2013; CHATURONGAKUL *et al.*, 2014). O reconhecimento de mais de um receptor pelo fago é possível, contudo, esse reconhecimento para a maioria dos fagos ocorre em receptor específico, gerando uma quantidade limitada de hospedeiros, ou seja, poucas espécies ou até mesmo somente algumas linhagens de uma única espécie (CHATURONGAKUL *et al.*, 2014; RAKONJAC *et al.*, 2011).



**Figura 3: Representação do ciclo lítico e lisogênico em fagos.** Para a infecção do hospedeiro, o primeiro passo é o reconhecimento e ligação a receptores específicos, seguido da adsorção e injeção do material genético. A progressão dependerá da estratégia empregada pelo fago; fagos virulentos continuarão pelo ciclo lítico enquanto fagos temperados terão de escolher entre ciclo lítico ou lisogênico. No ciclo lisogênico há o estabelecimento do profago

dentro do genoma bacteriano, o qual é replicado junto da replicação do hospedeiro. A indução do profago leva a excisão do profago do genoma e a volta ao ciclo lítico. No ciclo lítico, após a injeção do DNA viral no hospedeiro ocorre a replicação do material genético, seguido da transcrição e tradução das proteínas virais, montagem das partículas virais e a liberação da progênie por lise do hospedeiro (Modificado de SALMOND, 2015).

Nos fagos sem cauda, é possível ocorrer diversas formas de interações com o hospedeiro (MOÍRA *et al.*, 2020), desde o reconhecimento dos receptores do hospedeiro pelas proteínas de superfície do capsídeo, como no caso do fagos  $\Phi$ X174 (MCKENNA *et al.*, 1992); a modelos em que ocorre a transformação da membrana interna do fago em um tubo lipoproteico que permite furar a membrana do hospedeiro, como nos fagos que possuem membranas PRD1 e PM2 (PERALTA *et al.*, 2013); ou na ligação de uma membrana lipídica com a membrana do hospedeiro como nos fagos que não possuem capsídeo como MVL2 (KRUPOVIC *et al.*, 2018).

Após a ligação do fago ao hospedeiro, ocorre a injeção do material genético do fago no citoplasma, como demonstrado na figura 3. Usando a maquinaria celular do hospedeiro, o genoma do fago começa a transcrever proteínas virais que inibem a expressão das proteínas do hospedeiro, sequestrando a célula para priorizar a produção das proteínas virais. As proteínas virais são sintetizadas, o genoma é duplicado e ocorre a montagem da partícula viral, que se dá pela formação do capsídeo e da cauda, que ocorrem separadamente. Em seguida, ocorre o empacotamento do genoma viral dentro do capsídeo e acoplamento da cauda ao capsídeo, completando a estrutura do vírion (FORTIER *et al.*, 2013).

Ao final do ciclo lítico, ocorre a lise celular. Utilizando uma pequena proteína viral chamada holina, cria-se um poro na membrana citoplasmática da bactéria, permitindo que outras proteínas virais, como as endolisinas, possam ter acesso à camada de peptidoglicano, hidrolisando-a. Como consequência, rompem a bactéria e expõem o citoplasma para o exterior (WANG *et al.*, 2000). Após a lise celular, a progênie viral é liberada no ambiente podendo infectar novos hospedeiros susceptíveis e recomeçando o ciclo (FORTIER *et al.*, 2013) (Figura 3). Alguns fagos não realizam lise ao final do ciclo lítico, como por exemplo os fagos filamentosos que são liberados por brotamento atravessando a membrana (RAKONJAC *et al.*, 2011).

No caso dos fagos temperados, além do ciclo lítico citado acima, eles também podem realizar o ciclo lisogênico. Quando o material genético do fago é injetado no citoplasma do hospedeiro, nos fagos temperados, ocorre uma decisão de qual ciclo seguir; continuar no ciclo lítico ou entrar no

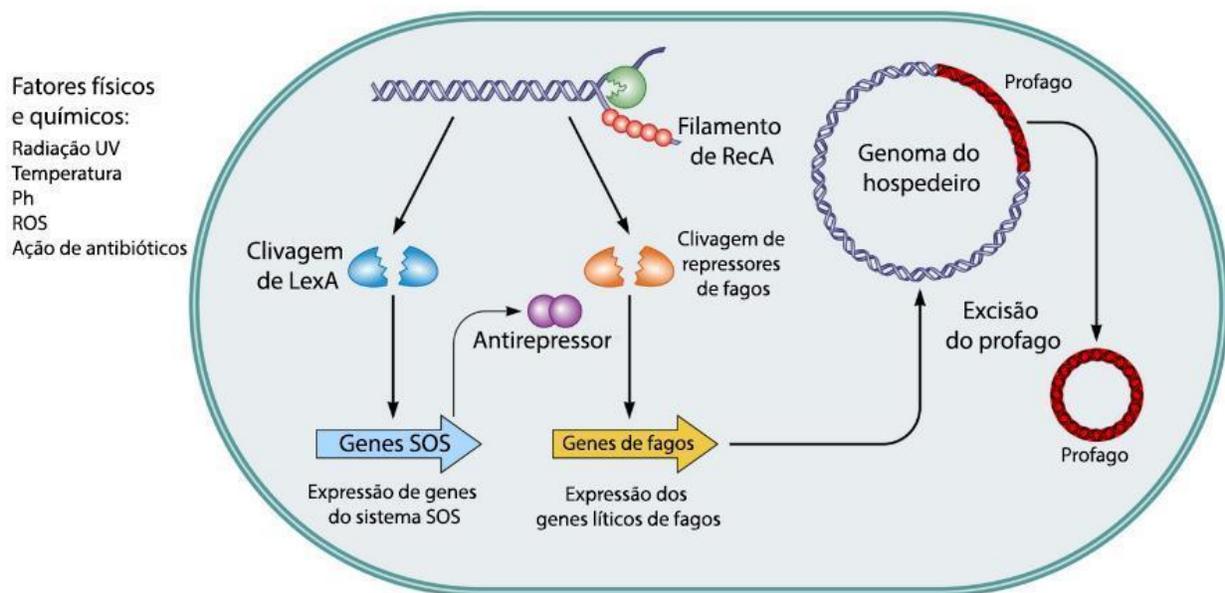
ciclo lisogênico (FORTIER *et al.*, 2013). Os fatores que levam a essa decisão são complexos e ainda não foram completamente definidos, contudo, alguns fatores já foram demonstrados influenciando a entrada no ciclo lisogênico, como a quantidade de partículas infectando o hospedeiro e a fisiologia do hospedeiro (KOURILSKY, 1973; KOURILSKY *et al.*, 1974; OPPENHEIM *et al.*, 2005; ZENG *et al.*, 2010). Com relação aos fagos do grupo spBeta que infectam *Bacillus* a decisão é guiada por um sistema chamado “*arbitrium*”, baseado na secreção de pequenos peptídeos (AimP) pela célula infectada, que permitem o fago perceber a densidade populacional. Quando a população de fagos em uma comunidade está alta, o meio apresenta uma grande quantidade de AimP, que é internalizado pelas bactérias. O AimP internalizado suprime um regulador negativo da lisogenia (AimX), induzindo o fago a começar o ciclo lisogênico. O contrário também acontece, quando a população de fagos está baixa, a concentração de AimP é insuficiente para reprimir o gene *AimX* levando a continuação do ciclo lítico (EREZ *et al.*, 2017).

Ao entrar no ciclo lisogênico, o fago produz uma enzima integrase sítio-específica que realiza a inserção do material genético do fago dentro do DNA bacteriano. Essa inserção é feita em regiões específicas da bactéria, chamadas *attB* (*attachment site B*), que são pequenas regiões de nucleotídeos que possuem algumas sequências idênticas a um sítio de ligação específico do fago, chamada *attP*. Após incorporado ao genoma bacteriano, o profago, nome dado ao fago inserido ao genoma bacteriano, apresenta a região *attL* e *attR* em cada uma das pontas, cada uma delas com uma parte da *attB* e *attP*. Essa incorporação, quando não revertida para o ciclo lítico, persiste verticalmente no genoma da bactéria, podendo causar até a perda de função de genes, como a inativação do gene da beta toxina de *S. aureus* pelo profago  $\Phi 13$  (COLEMAN *et al.*, 1991; SILVEIRA & ROHWER, 2016).

Após a incorporação, alguns dos genes do profago são silenciados para que a célula bacteriana mantenha seu funcionamento normal. Proteínas repressoras de expressão são produzidas pelo fago com a finalidade de suprimir genes relacionados com a montagem do vírion e a lise bacteriana, o que impede o genoma de continuar o ciclo lítico (DODD *et al.* 2005). O ciclo lítico pode ser retomado a partir da mudança ou presença de fatores físicos, como radiação UV, temperatura, pH; ou químicos, como espécies reativas de oxigênio (ROS); ação de antibióticos como a mitomicina C (MmC) e fluoroquinolonas (Fq). Esses fatores podem levar a danos no DNA,

ativando o sistema SOS da bactéria. Esse sistema leva a clivagem do repressor da expressão dos genes responsáveis pelo ciclo lítico, e por consequência, ao processo denominado indução (Figura 4) (NANDA *et al.*, 2015; LUNDE *et al.*, 2005; DRI *et al.*, 1994).

Com o ciclo lítico retomado, o profago é excisado do DNA do hospedeiro, terminando assim o ciclo lisogênico. Assim, serão retomados os processos essenciais para a produção de novas partículas virais, como a transcrição e tradução das proteínas virais, replicação do DNA viral oriundo da excisão, a montagem da partícula viral e subsequente lise (FORTIER *et al.*, 2013).



**Figura 4: Esquema da indução de profagos através do sistema SOS.** Fatores físicos e químicos podem levar à dano do genoma hospedeiro e a ativação do sistema de reparo SOS da bactéria, esse processo pode levar a indução do profago. RecA se ligam em seguimentos de ssDNA formando um segmento para o reparo do DNA danificado, durante esse processo essas proteínas levam a clivagem de LexA (repressor do sistema SOS) ou a clivagem de repressores CI-like de fagos. A clivagem de LexA, leva a expressão de outras proteínas de reparo como DNA polimerase, enzimas de reparo e inibidores da divisão celular. Dentre esses genes estão antirepressores das proteínas de repressão de fagos. A inativação desses repressores leva a ativação de genes relacionados com o ciclo lítico, levando a indução do profago, replicação, transcrição e tradução, montagem das partículas virais e lise celular. (Modificado de NANDA *et al.*, 2015)

Na indústria alimentícia, foram identificados profagos em culturas iniciadoras para a produção de bebidas fermentadas (LEDORMAND *et al.*, 2022), alimentos fermentados (KIM *et al.*, 2021), queijos, iogurtes e outros produtos lácteos (GUO *et al.*, 2016, LAVELLE *et al.*, 2018; OLIVEIRA *et al.*, 2018; ALEXEEVA *et al.*, 2021). Os estudos indicam que, apesar da presença desses profagos nas culturas bacterianas e de sua possibilidade de provocar eventos de extermínio por

meio de sua ativação, que levam à perda da produção, esses eventos são raros e difíceis de serem reproduzidos. Diversos destes profagos estão em uma relação de coevolução com a bactéria, podendo conferir super-imunidade contra fagos ou, até mesmo, genes que proporcionam vantagens às bactérias. Usualmente, esses profagos acabam sofrendo modificações que os tornam defeituosos e fixos no genoma bacteriano, contudo ainda proporcionam os benefícios esperados.

### 1.1.2. Profagos

Profagos é um termo cunhado por Lwoff em 1953, que em seu trabalho descobriu que os fagos faziam lisogenia como forma de replicação e propôs um modelo explicativo para o ciclo lisogênico para fagos (LWOFF, 1953; ROHWER *et al.*, 2002). Nos genomas bacterianos, tanto genes funcionais como genes inativos relacionados a bacteriófagos podem representar uma parte significativa do genoma, chegando a compor mais de 20% desse conjunto (CASJENS, 2003; TOUCHON *et al.*, 2016; MAGAZINER *et al.*, 2019; LIU *et al.*, 2020). A presença abundante dessas sequências de profagos desempenha um papel importante na promoção da diversidade genética bacteriana, podendo até mesmo resultar na incorporação de genes de resistência a antibióticos e genes que facilitam a adaptação a ambientes desafiadores, influenciando, assim, a aptidão das bactérias. Entretanto, vale ressaltar que nem todos os profagos e elementos semelhantes a fagos, também conhecidos como VLEs (*virus-like elements*), são funcionalmente ativos. Muitos permanecem inativos ou fragmentados nos genomas bacterianos, devido a fatores como sistemas de defesa bacteriana, decaimento mutacional ou a perda de genes essenciais ao longo das gerações bacterianas. Elementos semelhantes a fagos, como os VLEs, estão amplamente distribuídos nos genomas bacterianos. No entanto, vale destacar que profagos podem dar origem a regiões de tamanho variável nos genomas bacterianos, correspondendo a, em média, 5% do genoma de *Salmonella* e mais de 15% do genoma de *Escherichia coli* (TOUCHON *et al.*, 2009; BOBAY *et al.*, 2013). Foi revelado através do sequenciamento de genomas de bactérias patogênicas, que linhagens com mais profagos apresentam uma maior virulência e são mais transmissíveis, como por exemplo *Pseudomonas aeruginosa* que afeta pacientes com fibrose cística (MCCALLUM *et al.*, 2002; RICE *et al.*, 2009), *Staphylococcus aureus* e *E. coli* O157 em infecções hospitalares (BAE *et al.*, 2006; ASADULGHANI *et al.*, 2009; NEPAL, 2023), em *Salmonella*

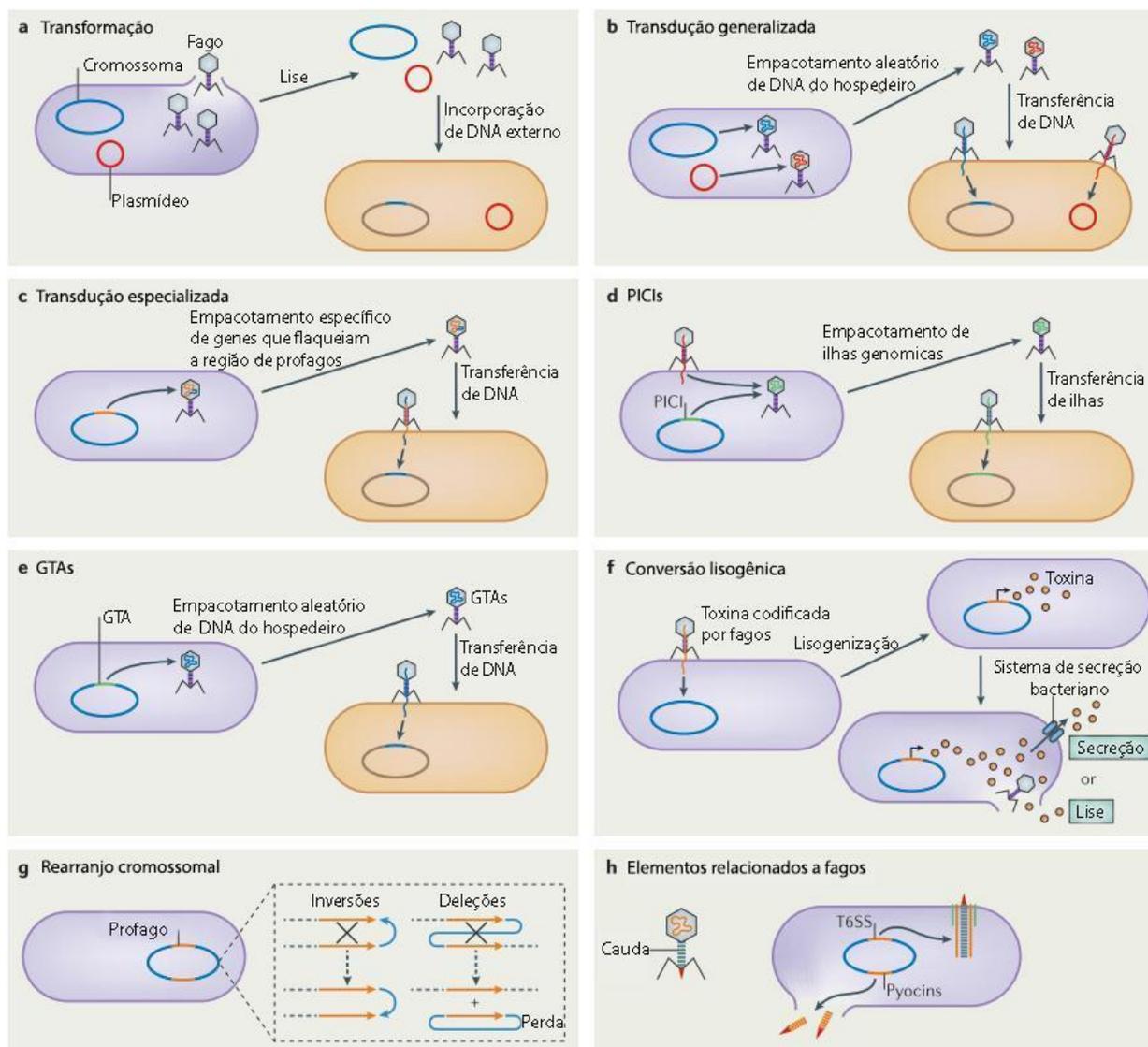
*enterica* sorovar Typhimurium em camundongos (FIGUEROA-BOSSI *et al.*, 1999), em *Pseudomonas aeruginosa* (SHMIDOV *et al.*, 2021; ALEDO *et al.*, 2023), em *Citrobacter rodentium* (MAGAZINER *et al.*, 2019), e em espécies de *Neisseria* (ORAZI *et al.*, 2022).

Por fazer parte do genoma das bactérias, os profagos podem sofrer modificações genéticas que impossibilitam a excisão, levando o profago a ficar preso no genoma bacteriano. Essas modificações podem ser induzidas por rearranjos genéticos, inserção de outros elementos móveis, mutações em sítios específicos. Quando um ou mais desses eventos ocorrem, os profagos perdem a capacidade de indução do ciclo lítico e passam a receber o nome de *cryptic prophage* (WANG *et al.*, 2010). Tanto os profagos intactos quanto os *crípticos* podem ter um importante efeito positivo no *fitness* de seus hospedeiros, que é vantajoso no caso dos fagos por terem uma relação simbiótica com os hospedeiros (WANG *et al.*, 2010).

A presença de um profago no genoma bacteriano pode levar à resistência da bactéria à infecção por outros fagos. Isso ocorre porque profagos possuem genes que inibem a injeção de DNA de outros fagos no citoplasma, ou modificam os receptores que os fagos usam, tornando a bactéria hospedeira imune à infecção de fagos, até mesmo fagos da própria espécie. Outros profagos podem ser reservatórios de genes de resistência, como por exemplo o *Rac prophage* que demonstra vantagem evolutiva na presença de carbenicilina (LIU *et al.*, 2015) ou no *Staphylococcus xylosus* phage  $\Phi$ JW4341 que possui gene de resistência a eritromicina (TAYLOR *et al.*, 2018). Outros profagos carregam toxinas, que muitas vezes são usadas pela bactéria hospedeira transformando-a em patogênica, como por exemplo a bactéria *Vibrio cholerae* que foi convertida de uma bactéria inofensiva para patógeno humano, após a aquisição do profago VPI $\Phi$ , possuidor da *toxin-coregulated pili* (TCP) (BONDY-DENOMY *et al.*, 2014) que serve como receptor para a infecção do fagos CTX $\Phi$  e carrega com si a toxina do cólera; outro exemplo é a neurotoxina de *Clostridium botulinum* que foi introduzida no genoma por um profago (FORTIER 2017). A lista de genes e funções novas é enorme, além de proteção contra fagos, exotoxinas e genes de resistência, às funções também abrangem resistência ao soro, a adesão da bactéria ao hospedeiro, fatores mitogênicos, invasão celular, remodelação/inativação da actina do hospedeiro, ativação/inativação de resposta pró inflamatória do hospedeiro e modulação da esporulação de bactéria (BOYD, 2012; BRÜSSOW *et al.*, 2004; FORTIER *et al.*, 2013).

### 1.1.3. Ecologia dos bacteriófagos

Os fagos são agentes importantes da evolução das bactérias, participando ativamente na transferência horizontal de genes. Alguns exemplos da atuação dos fagos na evolução das bactérias estão demonstrados na Figura 5.



**Figura 5: Diferentes papéis de fagos na patogenicidade e evolução bacteriana.** Os fagos podem apresentar diferentes papéis na transferência horizontal gênica em seus hospedeiros, alguns exemplos são: **a.** Através da lise, os fagos expõem DNA para transformação de outras bactérias; **b.** Podem transferir fragmentos do cromossoma e plasmídeos para outras bactérias, denominado transdução generalizada. **c.** Fagos temperados podem carregar genes bacterianos durante a indução e transferir para outras bactérias, em um processo denominado transdução especializada; **d.** PICIs (Phage-inducible chromosomal islands) podem sequestrar estruturas virias durante uma

montagem de fagos, denominados *phage helpers*, e transferir ilhas inteiras a outras bactérias; e. GTAs são elementos parecidos com fagos, caracterizados por sua incapacidade de replicação autônoma, no entanto, têm a habilidade de montar partículas virais utilizando o DNA do hospedeiro; f. Fagos temperados podem carregar genes codificadores de proteínas específicas, como toxinas, que podem ser secretadas por bombas ou liberadas durante a lise. g. profagos têm a capacidade de desencadear eventos de recombinação, o que pode resultar na modificação genética tanto de seu próprio genoma quanto do hospedeiro. h. Elementos parecidos com fagos podem estar presentes em genomas bacterianos provendo vantagem evolutiva para as bactérias que as possuem, como por exemplo pyocinas que são bacteriocinas, e o T6SS, sistema de secreção do tipo 6, ambos com origem estipulada em cauda de profagos. (Modificado de SALMOND, *et al.* 2015).

Na figura 5, podemos observar a participação dos fagos na transformação (a), que através da lise celular expõe o DNA ao meio exterior podendo assim ser incorporado por outras bactérias naturalmente competentes (SALMOND *et al.*, 2015). Os itens b e c abordam as duas formas de transdução viral, sendo elas generalizada e especializada. A transdução generalizada ocorre durante a montagem do vírion, quando fragmentos de DNA do hospedeiro, tanto cromossomal quanto plasmidial, ou até mesmo plasmídeos inteiros, são incorporados no lugar do material genético viral. Posteriormente, esse novo material é injetado em outras bactérias (b); na transdução especializada (c), durante o processo de excisão do profago do genoma e montagem das novas partículas virais, esses fagos podem carregar consigo genes da bactéria hospedeira, podendo-os levar para outras bactérias após a lise (SALMOND *et al.*, 2015). Curiosamente, alguns *attB* se encontram em regiões de RNA transportador (tRNA), o que durante a transdução especializada, possibilita que alguns fagos possam excisar junto com genes de tRNA, podendo ser encontrado fagos carregando seus próprios genes de tRNA, como o *Lactobacillus johnsonii* phage Lj965 que possui 4 genes de tRNA (VENTURA *et al.*, 2003; FORTIER *et al.*, 2013).

O item d da figura 5 mostra as *phage-inducible chromosomal islands* (PICIs), que são uma família de elementos móveis, presente em bactérias gram-positivas (PENADÉS *et al.*, 2015) e gram-negativas (FILLOL-SALOM *et al.*, 2018), que exploram o sistema de empacotamento de fagos temperados (chamados de fagos *helper*). Durante a montagem do vírion essas PICIs substituem o genoma viral, e na lise bacteriana, essas partículas carregando essas PICIs, infectam outras bactérias, tornando-se um agente de transferência horizontal de ilhas inteiras de genes (SALMOND *et al.*, 2015). Já no item e, a ação dos elementos móveis *gene transfer agents* (GTAs) é demonstrada. GTAs são pedaços de DNA parecidos com fagos, pois possuem genes estruturais e de lise como os fagos, contudo, eles não são excisados do DNA bacteriano e empacotam

pedaços aleatórios do DNA do hospedeiro, gerando partículas similares a fagos que infectam e transferem pedaços de DNA de uma bactéria a outra (SALMOND *et al.*, 2015). Claramente GTAs derivam dos profagos, contudo, não podem ser considerados *fagos crípticos*, por não expressarem as proteínas virais estruturais e de lise, e não podem ser considerados fagos, por não replicarem seu próprio conteúdo gênico e se reproduzirem (LANG *et al.*, 2017).

Outros elementos vindos de fagos são as bacteriocinas *pyocins* e o sistema de secreção T6SS (Figura 5, item h), ambos se assemelham com as caudas de fagos, e são usados pelas bactérias. As *pyocins* são encontradas em *Pseudomonas aeruginosa* e são divididas em 3 tipos: Tipo S, R e F; elas diferem em morfologia e modo de ação; elas podem ser produzidas espontaneamente ou induzidas por dano no DNA (NEVES *et al.*, 2014). O sistema de secreção T6SS está distribuído por bactérias gram-negativas; em *P. aeruginosa* tem função de translação de proteínas e transporte de toxinas para outras bactérias por contato (CHEN *et al.*, 2015). Durante a fase lisogênica dos fagos temperados no DNA do hospedeiro, ocorre a incorporação não apenas dos genes virais, mas também de outros genes transportados por esses fagos. Esses genes podem ser expressos e secretados ou liberados durante a lise, conforme ilustrado na Figura 5, item f. Outro evento importante são os rearranjos cromossômicos, demonstrado no item g, onde profagos podem levar a inversões e deleções de regiões no cromossomo de seus hospedeiros (SALMOND *et al.*, 2015).

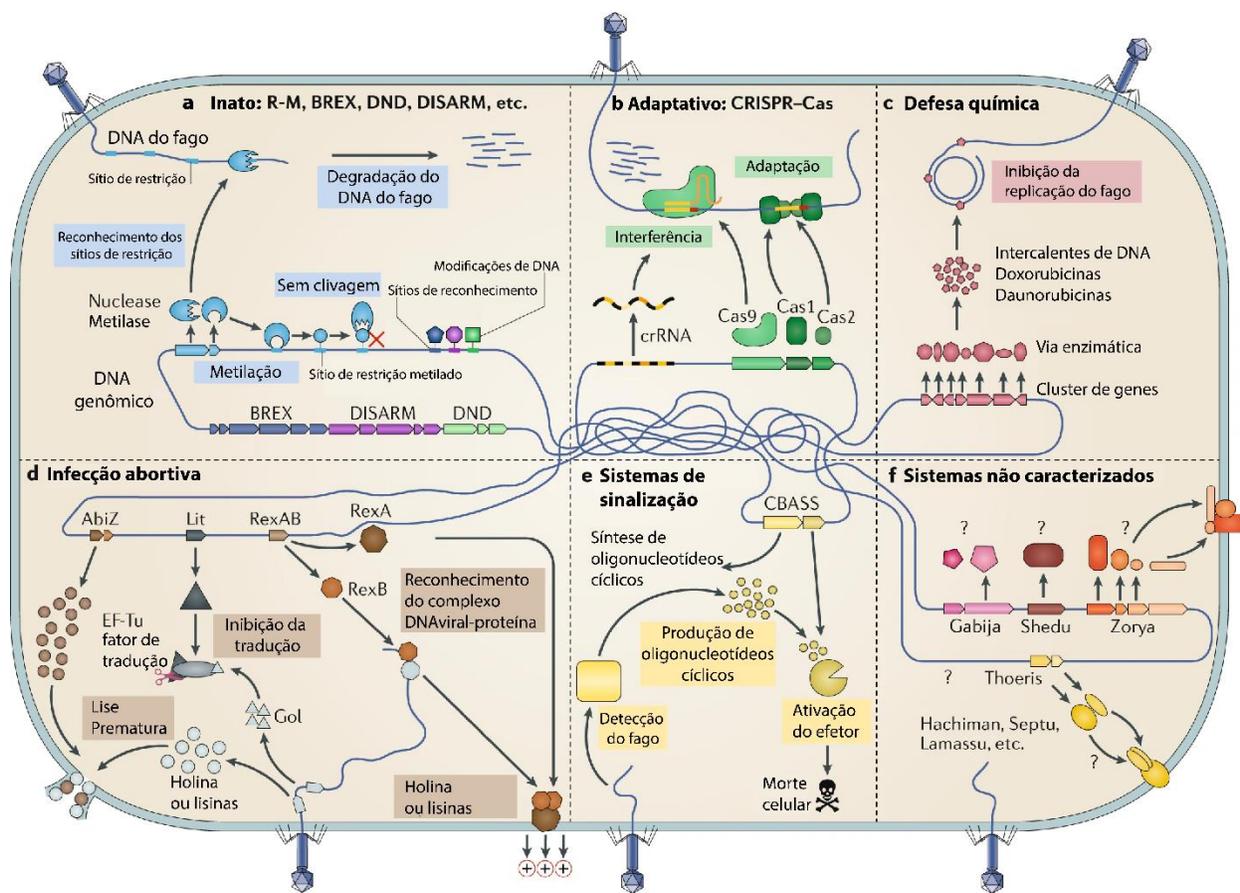
## **1.2 Sistema de proteção contra fagos de bacterias**

Bactérias e fagos estão envolvidos em uma coevolução constante, e sua relação pode ser compreendida pela hipótese da Rainha Vermelha. Proposta em 1973, essa hipótese tem como conceito básico a coevolução antagonista entre espécies ou populações. Ela sugere que, em uma interação entre duas espécies, ambas devem se adaptar, evoluir e proliferar continuamente para sobreviver enquanto competem com a espécie oposta nas mesmas condições, resultando em uma espécie de corrida armamentista, em ciclos de adaptação e contra-adaptação (LAGENRHANS, 2008). No contexto das bactérias e bacteriófagos, essa interação impulsiona a evolução de mecanismos de defesa antiviral nas bactérias e adaptações a esses sistemas nos

bacteriófagos. Devido à grande abundância e diversidade dos fagos, as bactérias desenvolveram diversos tipos de sistemas de defesa, incluindo sistemas inatos, adaptativos, químicos, abortivos, de sinalização, entre outros ainda não totalmente caracterizados. (BERNHEIM & SOREK *et al.*, 2020; DECAESTECKER & KING *et al.*, 2019; HAMPTON *et al.*, 2020)

Uma única espécie bacteriana pode apresentar diversos tipos de mecanismos de defesa ao mesmo tempo, apesar de que já foi demonstrando que esses mecanismos são frequentemente perdidos em genomas procariotos em curtos períodos da escala evolutiva, sugerindo que eles impondo que apesar de gerarem um fitness positivo para bactérias em defende-las contra fagos eles devem possuir um alto custo metabólico e outras implicações, que levam a um fitness negativo. (KOONIN *et al.*, 2017; VON HOUTE *et al.*, 2016)

A figura 6 ilustra de maneira gráfica os sistemas de defesa mais comuns em bactérias. O sistema de defesa inato é apresentado como um mecanismo de reconhecimento de ácidos nucleicos, sendo o primeiro componente viral a penetrar na célula durante uma infecção (Figura 6a). Geralmente, esse sistema utiliza a estratégia de restrição e modificação (R-M *Restriction-modification*). Os sistemas R-M clivam ou degradam o DNA a partir do reconhecimento de sítios específicos, por uma endonuclease, no genoma viral. Quando esses sítios específicos são encontrados no DNA bacteriano, geralmente passam por algum tipo de modificação, como metilação, que impede que o DNA bacteriano seja considerado um alvo para degradação. Diferentes tipos de modificações são empregados em vários sistemas para que a bactéria possa distinguir entre *self-DNA* e *non-self-DNA*. Entre esses sistemas estão R-M, BREX, DND, DISARM, e outros ainda não completamente caracterizados.



**Figura 6: Mecanismos de defesa contra fagos em bactéria.** Diferentes tipos de mecanismos de defesa contra fagos e suas características: **a. Sistema de defesa inata (R-M, BREX, DND, DISARM, etc):** Após a injeção do DNA viral, as endonucleases bacterianas reconhecem os sítios de restrição no DNA viral, clivando-os. Para evitar a clivagem do DNA bacteriano, esses sítios de restrição sofrem algum tipo de modificação. Na figura, diferentes modificações dos vários tipos de sistema R-M são exemplificadas por cores distintas; **b. Sistema de Defesa Adaptativa (CRISPR-Cas):** Este sistema compreende duas etapas distintas. Na etapa de adaptação, proteínas do complexo Cas identificam regiões específicas nos vírus e copiam parte dessas sequências nos espaçadores CRISPR no DNA bacteriano. Esses espaçadores geram RNA CRISPR (*crRNA*), que, quando processados, originam RNAs de interferência específicos para as regiões do vírus. Na etapa de interferência, os *crRNA* se ligam às regiões semelhantes no DNA viral, recrutando proteínas do complexo Cas que clivam o DNA de forma específica. **c. Sistema de defesa química:** Bactérias produzem moléculas que se ligam ao DNA viral e o inibem **d. Sistema de infecção abortiva:** São sistemas complexos, na figura temos como exemplo as proteínas *AbiZ* que levam a ativação das lisozimas e holinas dos fagos prematuramente, levando a lise prematura; ao expressar a proteína viral *Gol*, a proteína *Lit* bacteriana interrompe a tradução por meio da clivagem do fator de alongação *EF-Tu*; A proteína *RexA* reconhece um complexo de DNA-proteína, ativando *RexB* que é um canal de íons que despolariza a membrana, resultando na morte celular. **e. Sistemas de sinalização (CBASS):** O sistema CBASS detecta a presença do fago, gerando a produção de oligonucleotídeos cíclicos que ativam um efetor, culminando na morte celular. **f. Outros sistemas de defesa pouco caracterizados:** Outros sistemas de defesa com reconhecida ação anti-fagos, contudo, pouco caracterizados, como por exemplo *Gabija*, *Shedu*, *Zorya*, *Thoreris*, *Hachiman*, *Septu*, *Lamassu*, etc. (Modificado de BERNHEIM & SOREK, *et al.* 2020).

O sistema CRISPR-Cas é o principal sistema de defesa adaptativa das células bacterianas aos ataques de fagos (Figura 6b). Esse sistema possui proteínas que incorporam pequenos pedaços de DNA exógenos dentro de arranjos CRISPR (*clustered regularly interspaced short palindromic*

*repeats*). Essas regiões produzem os CRISPR RNAs (crRNA) e utilizando as proteínas Cas (*CRISPR associated proteins*) formam um complexo ribonucleoproteico, que ao reconhecer esse DNA exógeno faz a sua clivagem. Baseando-se nas proteínas Cas, organização do *loci* e na análise filogenética, o sistema CRISPR é dividido em classes, depois tipos e subtipos, seguido, de variantes (MAKAROVA *et al.*, 2020). Profagos podem inativar o sistema CRISPR-Cas, através de proteínas conhecidas como anti-CRISPR (SHIN *et al.*, 2017). Essas proteínas conseguem inibir diversas partes do sistema, como por exemplo interferir nas Cas (PAWLUK *et al.*, 2017) ou dificultar o reconhecimento do DNA viral (SHIN *et al.*, 2017).

Embora o sistema CRISPR-Cas forneça proteção contra DNA exógeno, essa vantagem genética pode ser acompanhada de desvantagens. Em um estudo envolvendo *P. aeruginosa*, observou-se que as linhagens com um sistema CRISPR-Cas ativo eram significativamente menores, cerca de 300 kb, em comparação com aquelas sem esse sistema, indicando uma menor diversidade genética dentro das linhagens com o sistema ativo (VAN BELKUM *et al.*, 2015). Curiosamente, alguns fagos contêm sistemas CRISPR-Cas completos em seus genomas, como exemplificado pelo fago *ICP1*, que infecta a bactéria *Vibrio cholerae*. Esse fago apresenta sistema CRISPR-Cas completo que desempenha um papel fundamental na proteção dos próprios fagos contra elementos semelhantes a PLE (*PICI-like elements*) presentes na bactéria hospedeira que tentam sequestrar as partículas de fago recém-montadas (SEED *et al.*, 2013).

Bactérias também possuem outros tipos de mecanismos como defesa química, infecção abortiva (Abi), sistema de sinalização e outros ainda não caracterizados. Defesa química já foi demonstrado em *Streptomyces sp.*, onde a bactéria produz moléculas que intercalam com o DNA viral, inibindo sua replicação (Figura 6c) (KRONHEIM *et al.*, 2018). O sistema Abi permite que a bactéria infectada consiga abortar infecção matando a si mesma ou impedir o seu metabolismo de funcionar, ambos eliminam o indivíduo, mas protegem a população e a comunidade local. Normalmente, esse sistema é ativado por algum componente do vírus que quando detectado leva a algum tipo de morte celular, como por exemplo depolarização de membrana, a ativação prematura da lise celular, a inibição do metabolismo celular (Figura 6d). Outro sistema muito parecido com os sistemas Abi são os sistemas de sinalização CBASS, nesse sistema, a partir da

sinalização de algum componente viral é produzido oligo nucleotídeos cíclicos que ativam um efector levando a morte celular (Figura 6e).

As bactérias também empregam outros tipos de mecanismos de defesa, como defesa química, infecção abortiva (Abi), sistema de sinalização, e outros ainda não completamente caracterizados. A defesa química foi demonstrada em *Streptomyces* sp., onde a bactéria produz moléculas que se intercalam com o DNA viral, inibindo sua replicação (Figura 6c). O sistema Abi permite que a bactéria infectada aborte a infecção, causando sua própria morte ou impedindo o funcionamento do seu metabolismo. Ambos os métodos eliminam o indivíduo, mas protegem a população e a comunidade local. Normalmente, esse sistema é ativado por algum componente do vírus que, quando detectado, leva a algum tipo de morte celular, como a depolarização da membrana, ativação prematura da lise celular ou inibição do metabolismo celular (Figura 6d). Existem ainda alguns sistemas que demonstraram participar na defesa contra fagos, mas que ainda não foram completamente caracterizados (Figura 6f). Um exemplo é o sistema GABIJA que possui semelhanças com o sistema R-M, pois possui uma endonuclease que permanece suprimida sob condições normais. No entanto, é ativada quando os níveis de NTP e dNTP são esgotados devido aos processos de replicação e transcrição de fagos invasores, ativando novamente a endonuclease e clivando tanto o DNA viral quanto o DNA bacteriano, levando à morte do indivíduo, assemelhando-se a um sistema Abi. Contudo, alguns pontos desse sistema ainda não foram completamente explicados.

### 1.3 Trato gastrointestinal

O trato gastrointestinal, com uma área de superfície superior a  $200\text{m}^2$ , é uma das áreas mais densamente populosas de microrganismos do corpo humano (HOOPER *et al.*, 132). Alterações na microbiota intestinal, ou disbiose intestinal, já se mostrou contribuir para alterações no metabolismo, câncer, doenças gastrointestinais e autoimunes (DAVE *et al.*, 2012). O intestino humano abriga por volta de  $10^{15}$  partículas de fagos, e pesquisas sobre a microbiota humana indicam a existência de uma ampla variedade de fagos ainda não catalogados em nenhum banco de dados (DALMASSO *et al.*, 2014; BARR *et al.*, 2013; NAYFACH *et al.*, 2021).

Mudanças da composição de fagos e indução de profagos, podem influenciar na ecologia da microbiota intestinal, seja por predação, seja por introdução de novos genes. Isso mostra que a comunidade de fagos consegue ditar a comunidade bacteriana, ao mesmo tempo que a comunidade bacteriana dita a comunidade de fagos, em uma relação de presa-predador, criando um modelo de coevolução influenciado pela hipótese evolutiva da Rainha Vermelha (MAGILL *et al.*, 2017).

Em meio aos diversos modelos conceituais que delineiam a dinâmica entre fagos e bactérias no contexto intestinal, destaca-se o "*Piggyback-the-Winner*" (PtW). Este modelo aborda a transição de ciclo de vida de fagos temperados, enfatizando seu papel no controle da predação bacteriana. Essa regulação se manifesta especialmente durante períodos de elevada abundância microbiana e taxas de crescimento, onde os fagos adotam a forma de profagos integrados ao genoma bacteriano. Tal integração confere imunidade ao hospedeiro contra superinfecções e facilita a transferência vertical de genes, apesar do ônus energético adicional para a bactéria hospedeira (SILVEIRA & ROHWER, 2016).

A produção de muco no intestino desempenha um papel crucial ao fornecer uma camada protetora às células, reduzindo o atrito e dessecação, e promovendo sua preservação contra a ação de ácidos e enzimas digestivas. No entanto, essas camadas mucosas estabelecem um habitat propício para a colonização por bactérias e fagos. A quantidade significativa de fagos presente no muco, em comparação com o ambiente circundante, em média de 4,4 vezes maior, confere uma defesa adicional ao hospedeiro produtor do muco contra bactérias (BARR *et al.*, 2013). Este fenômeno é respaldado pelo modelo de aderência de bacteriófagos ao muco (BAM - *Bacteriophage Adherence to Mucus*), que postula a capacidade de interação entre os domínios das proteínas estruturais dos fagos e os domínios parecidos com imunoglobulinas presentes nas glicoproteínas mucinas, componentes predominantes do muco. Assim, a elevada concentração de fagos no muco, conforme sugerido pelo modelo BAM, possui o potencial de mitigar a patogênese bacteriana e conferir uma imunidade mucosal ao intestino (BARR *et al.*, 2013).

Curiosamente, o processo inflamatório do hospedeiro, também pode ser um fator de mudança externa, levando ao aumento da transferência gênica entre linhagens bacterianas no intestino através dos fagos. Como exemplo, estudos realizados com a bactéria *Salmonella Typhimurium*

(*S. Tm*), duas linhagens foram introduzidas no trato gastrointestinal; a linhagem *S. Tm* SL1344 era possuidora do profago SopEΦ que codifica genes de virulência, aumentando a invasão celular do hospedeiro e exacerbando a enteropatia; e a linhagem *S. Tm* ATCC144028S, a qual não possuía esse profago. A ativação da resposta inflamatória do hospedeiro frente a invasão da linhagem *S. Tm* SL1344, ativa o sistema SOS dessa bactéria, resultando na indução do profago SopEΦ. O fago SopEΦ, então, infecta a linhagem *S. Tm* ATCC144028S, a qual não possuía esse profago, tornando-a patogênica. Quando comparado com hospedeiros vacinados protegidos da invasão por IgA, portanto, não ativando a resposta inflamatória, foi observado uma diminuição de  $10^5$  na transferência desse profago para a linhagem *S. Tm* ATCC144028S, mostrando que o hospedeiro também modula a interação bactéria-fagos e a transferência de genes na microbiota intestinal (PABST *et al.*, 2020; DIARD *et al.*, 2017).

#### **1.4 Bactérias ácido lácticas e o gênero *Weissella***

Bactérias ácido lácticas (BAL) pertencem a um grupo não monofilético de bactérias gram-positivas com a capacidade de fermentar açúcares produzindo ácido lático como produto final, com morfologia de bacilos (*bacilli*) ou cocos (*cocci*), são anaeróbicas ou microaerófilas e não formam esporos. As BAL podem ser divididas em dois grupos, as homofermentativas e as heterofermentativas. As homofermentativas produzem ácido lático como principal produto da sua fermentação, enquanto as heterofermentativas podem produzir diversos outros produtos da fermentação, como dióxido de carbono, álcoois e ácido acético. No ano de 2020, procedeu-se a uma reclassificação das bactérias ácido lácticas (BAL), resultando em sua atual inclusão em uma única família, denominada *Lactobacillaceae*. Nesse contexto, identificam-se 31 gêneros distintos, a saber: *Lactobacillus*, *Paralactobacillus*, *Pediococcus*, *Weissella*, *Fructobacillus*, *Convivina*, *Oenococcus*, *Leuconostoc*, juntamente com a adição de 23 novos gêneros previamente categorizados como espécies pertencentes ao gênero *Lactobacillus* (ZHENG *et al.*, 2020). BAL são encontradas em quase todos os ambientes, com destaque para os alimentos, mas também sendo encontradas na cavidade oral, genital e intestinal de humanos e animais (FLOROU-PENERI *et al.*, 2013).

BAL vêm sendo usadas durante décadas na indústria alimentícia, não só pela habilidade de fermentação rápida e acidificação de comidas e bebidas, mas também, pela produção de moléculas como álcoois, exopolissacarídeos (EPS), compostos aromáticos e pela capacidade de manter a qualidade, reprodutividade e segurança dos processos. Por conta dessas características, elas podem ter diferentes componentes nutricionais, sensoriais e com propriedades organolépticas e reológicas distintas (LEROY *et al.*, 2004), participando como cultura iniciadora na produção de diversos produtos, como leites fermentados, kefir, iogurtes, queijos. Participa também no processamento e fermentação de bebidas alcoólicas, como vinhos, *kombucha*, cervejas; em carnes, como em linguiças e carnes curadas; e vegetais, como *picles*, *kimchi*, café, cacau (BRANDT *et al.*, 2014; LEROY *et al.*, 2004; TAMANG *et al.*, 2016). Muito além das aplicações das BAL em alimentos, tem participações na produção industrial de bioquímicos, cosméticos e farmacêuticos. Recentemente, elas têm ganhado atenção por seu caráter probiótico (KAIDI *et al.*, 2020).

Probióticos são microrganismos vivos, os quais quando administrados em quantidades adequadas conferem melhora na saúde do hospedeiro, podem ser bactérias, leveduras e bolores, tendo as BAL e *Bifidobacterium* como as mais comumente usadas (FAO/WHO, 2012; HILL *et al.*, 2014; ROOBAB *et al.*, 2020). Os critérios para classificação de uma bactéria como probiótica são: a habilidade de sobreviver a passagem no trato gastrointestinal; isolamento de um animal sadio; não ser patógeno ou estar associado a doenças em seu hospedeiro; ser geneticamente estável, portanto, não carregar genes de resistência em elementos genéticos móveis (LEBEER *et al.*, 2008; COLLINS *et al.*, 2009). Os benefícios providos pelos probióticos são extensos, alguns merecem menção como: ação antimicrobiana e de modulação da disbiose intestinal, ação anti-inflamatória ou pró-inflamatória, ação antialérgica, atividade de enzimas digestivas, ação antidepressiva (ROOBAB *et al.*, 2020, KAIDI *et al.*, 2020). Pelo fato de BAL serem um dos grupos de microrganismos mais profundamente estudado da literatura, muitos de seus membros possuem o *status* GRAS (*Generally Recognized As Safe*) pelo FDA (*Food and Drug Administration, USA*) na regra final efetivada em outubro de 2016 (FDA, 2016); ou o *status* QPS (*Qualified Presumption of Safety*) pela Autoridade Europeia para a Segurança Alimentar (EFSA - *European Food Safety Authority*) (EFSA, 2005). Contudo, não existe um padrão para determinar a segurança dessas

bactérias, mas alguns pilares para escolha são definidos, sendo eles: o nível taxonômico a que pertence, levando em consideração os seus semelhantes; a familiaridade, o quanto se sabe, o quanto já foi estudado; a patogenicidade, a presença de genes de virulência ou toxigênia, bem como elementos genéticos móveis; e o uso final, onde será usado o organismo.

As bactérias do gênero *Weissella*, pertencem ao filo *Bacillota*, classe *Bacilli*, ordem *Lactobacillales* e família *Lactobacillaceae*. O gênero *Weissella* surgiu em 1993 quando Collins e colaboradores (1993) sugeriram uma nova classificação de diversas BAL baseado no sequenciamento do 16S rRNA e na fenotipagem. Nessa reclassificação, o gênero *Weissella* recebeu membros classificados anteriormente como *Leuconostoc* e *Lactobacillus*. Recentemente, Zheng e colaboradores (2020) sugeriram mudar a descrição da família *Lactobacillaceae*, fazendo incluir todos os gêneros anteriormente em *Leuconostocaceae*, unindo-os em uma única família. Outra mudança foi o nome do filo, de *Firmicutes* para *Bacillota* (OREN & GARRITY, 2021; OREN & GARRITY, 2022; OREN *et al.*, 2023). Todas as *Weissella* pertencem ao grupo BAL, são aeróbicas facultativas e heterofermentativas obrigatórias, produzindo ácido lático e ácido acético como principais produtos da sua fermentação.

Na natureza, bactérias do gênero *Weissella* já foram encontradas em diversos tipos de ambientes, em alimentos fermentados, vegetais, carnes, fezes, em vertebrados e invertebrados (FUSCO *et al.*, 2015; BELDA *et al.* 2011). Apesar da presença de *Weissella* em vertebrados, somente *W. confusa* e *W. ceti* foram descritas em eventos de infecções patogênicas oportunistas em humanos, eventos esses raros, ocorrendo devido a resistência natural à vancomicina dessas bactérias, acometendo principalmente imunossuprimidos ou já debilitados de outras conformidades (FUSCO *et al.*, 2015). *Weissella ceti* é o agente etiológico da weissellose, uma doença hemorrágica da truta-arco-íris (*Oncorhynchus mykiss*), capaz de infectar diferentes órgãos do peixe, com alta transmissão pela água e presente em diversos continentes, com surtos recentes na China, Brasil, Estados Unidos da América e Japão (FUSCO *et al.*, 2015; HOAI *et al.*, 2018). *Weissella* apresenta características que supõem seu emprego como probiótico: com auxílio de suas bacteriocinas que possuem ação antimicrobiana, inibição de formação de biofilme e modulação da resposta imune de hospedeiros (FUSCO *et al.*, 2015; SPIEGELHAUER *et al.*, 2020). Considerando as características apresentadas, as *Weissella* propostas como probiótico precisam

ser cuidadosamente estudadas e caracterizadas, pois ainda nenhuma possui o status de segura para consumo humano (TEIXEIRA *et al.*, 2021).

Por serem ubíquos, a ocorrência de contaminação por fagos é algo comum, e problemas relacionados com essa contaminação já foram relatados em diversas indústrias como a de alimentos, farmacêutica e química. As vezes tratados como “amigos”, devido à sua especificidade os fagos são agente ideais no controle de bactérias indesejadas, às vezes como “inimigos”, eliminando a cultura iniciadora mudando qualidade do produto, atrapalhando a acidez, a textura, aroma dos produtos, entre outros (PUJATO *et al.*, 2018; GARNEAU *et al.*, 2011). Mosimann e Ritter (1946) foram os primeiros a descreverem *Leuconostoc* afetados por fagos, onde fagos causavam alteração do aroma da manteiga. Estudos relacionados com o uso de bactérias de importância biotecnológica, em culturas iniciadoras ou em probióticos, têm levantado questionamento sobre o quanto os fagos interferem em seus usos na indústria e como eles podem influenciar na seleção e evolução dessas bactérias (MEADEN *et al.*, 2013; KOT *et al.*, 2014). Fagos de *W. cibaria*, isolados na fermentação de pepinos, mostraram poder infectar mais de uma espécie de *Weissella* e em alguns casos até outros gêneros (LU *et al.* 2012, KLEPPEN *et al.* 2012b). Há avanço nos estudos de isolamento e caracterização de novos fagos em *Weissella*, principalmente fagos ambientais de importância biotecnológica: como fago  $\Phi 22$  e  $\Phi YS61$  de *Weissella cibaria* na fermentação de *Nham* (linguiça Tailandês) (PRINGSULAKA *et al.*, 2011; KLEPPEN *et al.*, 2012b); os fagos de *Weissella cibaria*,  $\Phi 3.8.43^1$ ,  $\Phi 7.2.50^1$ ,  $\Phi 3.8.18^1$ ,  $\Phi WC51^2$ ,  $\Phi WC52^2$ ,  $\Phi WC005^2$ ,  $\Phi WC130^2$ ,  $\Phi WC248^2$  em fermentação de *Kimchi* (LU *et al.*, 2012<sup>1</sup>; KONG *et al.*, 2020<sup>2</sup>; LEE *et al.*, 2021<sup>2</sup>); o fago WCP30 de *Weissella cibaria* em laticínios (LEE *et al.*, 2017); em *Weissella ceti*, bactéria patogênica de truta-arco-íris, já foi caracterizado o fago lítico PWc (HOAI *et al.*, 2018). Contudo, somente três genomas estão depositados no GenBank e todos pertencendo à classe *Caudoviricetes*: *Weissella cibaria* YS61 phage  $\phi YS61$  com morfologia podovírus); *Weissella* phage PWc e *Weissella* phage WCP30 com morfologia sifovírus.

A linhagem WpK4 de *Weissella paramesenteroides* vem sendo estudada e apontada como possível probiótico. Isolada da nasofaringe de um suíno recém-desmamado, a linhagem demonstrou características necessárias para ser classificada como um probiótico; *in vitro*: exibiu resistência ao suco gástrico, tolerância aos sais biliares, alta hidrofobicidade da parede celular,

produção de H<sub>2</sub>O<sub>2</sub>, síntese de EPS, sensibilidade aos principais antimicrobianos usados no tratamento de humanos e animais, capacidade de adesão a parede da mucosa intestinal; e *in vivo* (camundongos) em modelo de salmonelose (infecção por *Salmonella Typhimurium*): atenuou a perda de peso, preservou a integridade e arquitetura intestinal e promoveu a sobrevivência (ALVIM *et al.*, 2016). Em modelo de giardíase (infecção por *Giardia lamblia*), a linhagem WpK4 demonstrou uma redução de aproximadamente 70% da carga parasitária nos animais que consumiram o probiótico e uma redução do dano causado pelo patógeno às vilosidades intestinais, diminuindo os sintomas da doença (FONSECA *et al.*, 2019). Sandes e colaboradores (2020) demonstraram através da imunomodulação do hospedeiro a participação no eixo do trato gastrointestinal e o sistema nervoso central (*gut-brain axis*) da WpK4, reduzindo os sinais de depressão e ansiedade em camundongos. Em modelo de colite, amébrica e química, animais tratados com WpK4 tiveram uma redução na permeabilidade intestinal, redução de necrose, e recuperação do epitélio intestinal (SANDES *et al.*, 2020; PRADO *et al.*, 2020).

## 2. JUSTIFICATIVA

O acervo de genomas de fagos depositados no banco de dados do *National Center for Biotechnology Information* (NCBI) tem crescido a cada ano, mas o número atual ainda é notavelmente limitado. Essa limitação contradiz a reconhecida diversidade genética dos fagos, destacando-se que metade dos fagos com sequências depositadas têm hospedeiros restritos a apenas sete gêneros bacterianos de interesse clínico e biotecnológico (*Mycobacterium*, *Streptococcus*, *Escherichia*, *Pseudomonas*, *Gordonia*, *Lactococcus* e *Salmonella*) (DION *et al.*, 2020; SOUSA *et al.*, 2021).

A relevância dos fagos na evolução bacteriana, na modulação do comportamento bacteriano e na disseminação de genes de virulência destaca a necessidade premente de compreender a dinâmica entre profagos e bactérias. Essa compreensão é particularmente crucial em bactérias com aplicações biotecnológicas, exigindo uma investigação mais aprofundada sobre o impacto dos profagos em sua fisiologia.

Dentro do perfil atribuído a microrganismos probióticos, destaca-se a notável ênfase na estabilidade e segurança genética. Os profagos, identificados em linhagens iniciadoras na indústria (ZABURLIN *et al.*, 2017) e em contextos probióticos (MERCANTI *et al.*, 2016), emergem como agentes promotores de evolução para as bactérias e fagos associados. Uma pesquisa conduzida por Wei *et al.* (2020) revelou que o profago presente na linhagem comercial de *Lactobacillus plantarum* NCU116 não origina partículas infecciosas em cenários de estresse celular ou em condições induzidas, corroborando assim a estabilidade genética e a segurança inerentes à utilização dessa linhagem na esfera industrial.

A análise do genoma de *Acinetobacter baumannii*, disponível no *GenBank* (NCBI), revelou uma considerável diversidade de profagos, com genes impactando na virulência e no fitness bacteriano (COSTA *et al.*, 2018; LÓPEZ-LEAL *et al.*, 2020; LOH, *et al.*, 2020). Cripto-profagos de *Lactobacillus*, encontrados no gene *msrA* e transportando genes do sistema toxina-antitoxina bacteriano, indicam a variedade de funções dos profagos (DENOUE *et al.*, 2008). Trabalhos como o de Javan *et al.* (2019) destacam a influência dos profagos no sucesso evolutivo de linhagens de *Streptococcus*, enquanto Bruggermann *et al.* (2013) discutem o potencial terapêutico dos

profagos PRFR1 e PRFR2 contra cepas de *P. acnes* associadas a doenças. Todos esses estudos abordam diferentes formas pelas quais profagos influenciam bactérias de interesse biotecnológico.

O conhecimento sobre o gênero *Weissella*, seus profagos e sua caracterização biológica é limitado (PRINGSULAKA *et al.*, 2011; KLEPPEN *et al.*, 2012b; LEE *et al.*, 2017; PANTHEE *et al.*, 2019; SHARMA *et al.*, 2023; WAN *et al.*, 2023). Considerando a ubiquidade de fagos e bactérias lácticas ácido (BAL), juntamente com as diversas formas de transferência gênica, e o crescente interesse na biotecnologia relacionada ao gênero *Weissella* (TEIXEIRA *et al.*, 2021; ALVIM *et al.*, 2016; SANDES *et al.*, 2020; PRADO *et al.*, 2020; FONSECA *et al.*, 2019), emerge a necessidade de explorar e compreender melhor essas interações.

Portanto, este estudo propõe contribuir para a classificação do gênero *Weissella* como GRAS (Generally Recognized As Safe) ou QPS (Qualified Presumption of Safety) (ABRIOUEL *et al.*, 2015; FESSARD & REMIZE, 2017). Para atingir esse propósito, a pesquisa visa identificar sistemas de proteção contra fagos nos genomas do gênero *Weissella*, investigar a presença de profagos nesses genomas e caracterizar a diversidade genética desses elementos. Utilizando genomas já depositados e genomas de três linhagens de *Weissella* isoladas por nosso grupo com características probióticas, esta abordagem abrangente permitirá uma análise detalhada das interações entre fagos e as bactérias do gênero *Weissella*, proporcionando implicações importantes para a segurança e as aplicações biotecnológicas dessas bactérias.

### 3. OBJETIVOS

#### 3.1 Objetivo geral

Analisar a prevalência dos sistemas de proteção contra fagos em genomas do gênero *Weissella* e identificar e caracterizar profagos presentes dentro destes genomas.

#### 3.2 Objetivos específicos

- Realizar a anotação e análise funcional dos genomas de três linhagens de *Weissella*, as quais foram previamente identificadas por apresentarem características probióticas distintas.
- Utilizar o banco de dados PADS Arsenal para identificar e caracterizar genes específicos associados a mecanismos de defesa contra fagos presentes nos genomas do gênero *Weissella*.
- Identificar os profagos presentes dentro de genomas do gênero *Weissella* depositados no *National Center for Biotechnology Information* (NCBI) e em linhagens probióticas isoladas por nosso grupo de pesquisa.
- Verificar a correlação entre o tamanho do genoma bacteriano e a quantidade de profagos presentes no gênero *Weissella*.
- Anotar, usando diferentes banco de dados, e comparar, através da sintenia e similaridade das sequências, a composição genômica dos profagos.
- Analisar a presença de genes de virulência e resistência a antimicrobianos nos profagos utilizando banco de dados curados.

## 4. METODOLOGIA

### 4.1 Montagem de genomas de *Weissella* potencialmente probióticas e análise dos mecanismos de defesa contra fagos presentes no gênero *Weissella*

#### 4.1.1. Crescimento bacteriano

A montagem intitulada VSF134 nesse trabalho, tem origem na linhagem 1BM, isolada da mucosa nasal de um suíno, do trabalho de Alvim (2011); a montagem intitulada VSF140, tem origem na linhagem GIR46L4, isolada do leite de uma vaca da raça girolando, do trabalho de Steinberg (2016); a montagem intitulada VSF142 nesse trabalho, tem origem na linhagem V3B-11, isolada da cavidade oral de uma novilha bovina, do trabalho de Sandes (2013). Os isolados das montagens VSF140 e VSF142 foram classificadas em seus respectivos trabalhos como pertencendo à espécie *Weissella paramesenteroides*, já o isolado da montagem VSF134 pertencendo a espécie *Weissella cibaria*, essas classificações se deram, a partir do sequenciamento e análise da região V3-V4 do 16S rRNA. As três linhagens, mantidas congeladas em glicerol a -80°C, foram reativadas em inóculos de 2% (v/v) em 5 mL de caldo De Man, Rogosa e Sharpe (MRS) em aerobiose por 18h a 37°C, seguida de uma repicagem em 5 mL de caldo MRS nas mesmas condições.

#### 4.1.2. Sequenciamento genômico

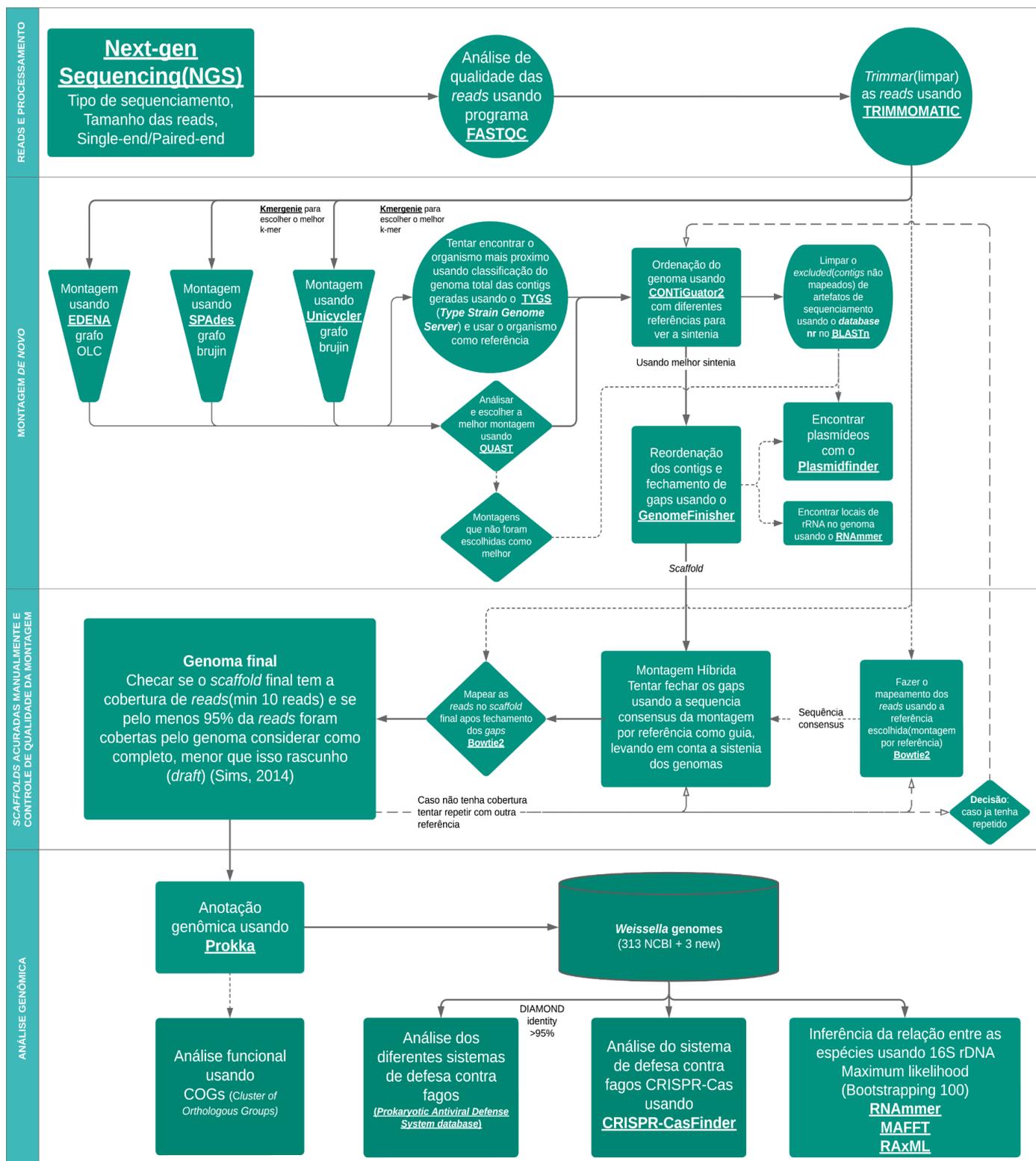
O DNA genômico foi extraído utilizando uma adaptação do protocolo de Doyle & Doyle, (1987). Resumidamente, 5 mL da cultura bacteriana crescida overnight foi centrifugada a 12.000 rpm por 2 min para remoção do sobrenadante. O precipitado celular resultante foi ressuscitado em tampão de digestão (constituído de 20% de sacarose; 0,01M Na<sub>2</sub>HOP<sub>4</sub>; liozima 0,1g/mL; pH 7) e incubado a 37°C por 1 hora. Após incubação, a lise celular foi acentuada pela adição de proteinase K (20mg/mL), tampão Tris-EDTA (10mM de Tris-HCl; 1mM de EDTA) e SDS 10%, seguida por uma reincubação a 56°C por 1 hora. Em seguida, o lisado celular foi submetido a tratamento com RNaseA – essa etapa aumenta a qualidade do sequenciamento – em uma nova incubação a 37°C por 1 h. Foram adicionados CTAB (brometo de hexadeciltrimetilamônio) em uma concentração de 2% e NaCl 0.7M, seguido por uma nova incubação de 10 min a 65°C. O DNA foi purificado por meio de uma partição de fases de clorofórmio/álcool isoamílico, e em seguida, precipitado com

isopropanol 100% a  $-20^{\circ}\text{C}$  por 20 min. O pellet de DNA foi obtido por centrifugação, seco ao ar e finalmente ressuspendido com água ultrapura.

A avaliação do DNA quanto à concentração e qualidade foi feita por eletroforese em gel de agarose 1% e *Qubit® Fluorometer Quantitation* (LIFE TECHNOLOGIES, 2014). Após verificado a qualidade e quantidade, o material genético foi enviado para sequenciamento pelo método *shotgun* na plataforma Illumina HiSeq 2500, utilizando bibliotecas genômicas Nextera (*paired-end*) de 150 pb. Foram gerados dois arquivos para cada biblioteca, cada arquivo com um par das *reads* (*paired-ends reads*).

O fluxograma da montagem pode ser observado na figura 7. A qualidade do sequenciamento de cada um dos arquivos foi verificada usando o programa FastQC (ANDREWS *et al.*, 2010). O *Phred quality score* é uma pontuação que mede a qualidade da identificação do nucleotídeo gerado pelo sequenciamento, sendo esse um score logarítmico inversamente ligado a probabilidade de adivinhar a base errada, logo um valor menor implica a uma maior chance de a base errada ser adicionada e um valor maior implica a menor chance de uma base errada ser adicionada. Os critérios de avaliação foram:

- (1) A qualidade por base das *reads*: Levando em consideração os valores das pontuações *Phred* gerados para cada posição em todas as *reads*.
- (2) Avaliação do score gerado: Considerando a distribuição das *reads*, suas pontuações *Phred*, e observando se a maior parte das *reads* possuem um score maior que 35.
- (3) A variação do conteúdo %GC: Levando em conta que %A=%T e %G=%C e sua distribuição em cada posição.
- (4) Distribuição normal da média do conteúdo %GC das *reads*, comparando com a distribuição teórica delas.
- (5) Porcentagem de N (base não identificada) nas *reads* por posição.
- (6) A distribuição do tamanho das *reads*: Levando em consideração que o sequenciamento Hiseq produz *reads* de tamanho 150pb.



**Figura 7: Fluxograma da metodologia para anotação dos genomas e análises dividida em 4 partes:** Sequenciamento das amostras e processamento das reads; montagem de novo; controle de qualidade e análise genômica.

### 4.1.3. Montagem genômica

A montagem das sequências foi feita *de novo* guiada pela montagem de genomas de referência (Figura 7) (SCHNEEBERGER *et al.*, 2011; MARIANO, 2015; LISCHER, *et al.*, 2017). Após o controle de qualidade, as *reads* foram submetidas a montagem *de novo* utilizando os programas:

- (1) Edena: Um montador que o seu algoritmo é baseado em grafos *Over-Layout-Consensus (OLC)* (HERNANDEZ *et al.*, 2008);
- (2) SPAdes: Um montador que seu algoritmo é baseado em grafos de *De Buijn* com diferentes tamanhos de k-mer (BANKEVICH *et al.*, 2012);
- (3) Unicycler: Uma pipeline de montagem que utiliza uma versão otimizada do SPAdes, logo, seu algoritmo é baseado em grafos de *De Buijn* com a comparação e predição do melhor k-mer para a montagem (WICK *et al.*, 2017).

O servidor TYGS (*Type (strain) genome server*) (<https://tygs.dsmz.de/>) foi usado para identificar a possível espécie e linhagem da montagem (MEIER-KOLTHOFF *et al.*, 2019). Esse servidor, procura e extrai a sequência do gene *16S rDNA* dos *contigs* e alinha essa sequência a um banco de dados de referência, determinando assim as possíveis espécies e linhagens. A referência encontrada pelo TYGS foi usada no mapeamento dos *reads* usando o programa Bowtie2, o consenso da sequência gerou a montagem com genoma de referência.

Posteriormente, os resultados das montagens *de novo* foram avaliados usando o programa QUAST (MIKHEENKO *et al.*, 2018) e utilizando como referência a espécie e linhagem de *Weissella* obtida no servidor TYGS. Dentro dos critérios de avaliação da qualidade da montagem foram considerados:

- (1) A quantidade de *contigs* gerados;
- (2) O tamanho do maior e menor *contigs*;
- (3) O tamanho total do genoma e sua comparação com o genoma de referência;
- (4) O valor N50 e L50, respectivamente, o tamanho do *contigs* que se encontra na metade o genoma total e quantidade de *contigs* para se chegar à metade do genoma total;
- (5) Conteúdo %GC e sua comparação com o genoma de referência.

A montagem de melhor qualidade foi selecionada e, a partir do genoma completo de diferentes linhagens da espécie referência, teve seus *contigs* mapeados e ordenados, em cada um desses genomas, pelo programa *CONTIGuator2* (GALARDINI *et al.*, 2011), seguida da avaliação da melhor sintenia e cobertura desses genomas. Após a ordenação, o *scaffold* gerado foi submetido ao fechamento dos gaps em duas etapas.

Na primeira etapa, foi utilizado a ferramenta *fgap* do programa *GenomeFinisher*, essa ferramenta utiliza o grafo *De Buijn* para, a partir do *scaffold*, fechar os gaps gerados com os *contigs* não mapeados e os *contigs* gerados pelos outros montadores que não foram selecionados, como parâmetros foram utilizados uma identidade mínima de 99%, com extremidade de *contigs* de 500pb e um máximo de *e-value* de  $1e-7$ .

Na segunda etapa foi feito um fechamento guiado por referência. As regiões de gaps tiveram a sequência dos seus 500 pb flaqueadores alinhadas contra a montagem com o genoma de referência usando a ferramenta *Basic Local Alignment Search Tool* (BLAST), levando em conta a sintenia do genoma, e a partir do consenso das *reads* mapeado com a referência, foram fechados gaps de regiões repetitivas.

#### **4.1.4. Anotação genômica e análise gênica**

Durante a primeira e segunda etapa, todas as alterações na sequência e cobertura de *reads* dessas sequências foram verificadas manualmente utilizando o programa *Ugene* (OKONECHNIKOV *et al.*, 2012). A profundidade do sequenciamento foi verificada utilizando a ferramenta *SamTools* (LI *et al.*, 2019) e levando em consideração os critérios apresentados por Sims e colaboradores (2014). Anotação automática do genoma foi feita usando *Prokka v.1.14.6* (SEEMANN, 2014) nos parâmetros *default*, e pelo servidor *RAST* (<https://rast.nmpdr.org/rast.cgi>) (AZIZ *et al.*, 2008), nos parâmetros *default* com o auxílio do *SEED viewer* (OVERBEEK *et al.*, 2005) para indexação de função das regiões de codificação (CDS) aos seus possíveis ortólogos usando *COGs* (*Cluster of Orthologous Groups*), quando presentes.

Usando a sequência de aminoácidos geradas, uma busca por genes de virulência e resistência a antibióticos foi conduzida utilizando o programa *DIAMOND* (BUCHFINK *et al.*, 2021), com os parâmetros padrões para *Blastp*, e os bancos de dados curados o *CARD* (*The Comprehensive*

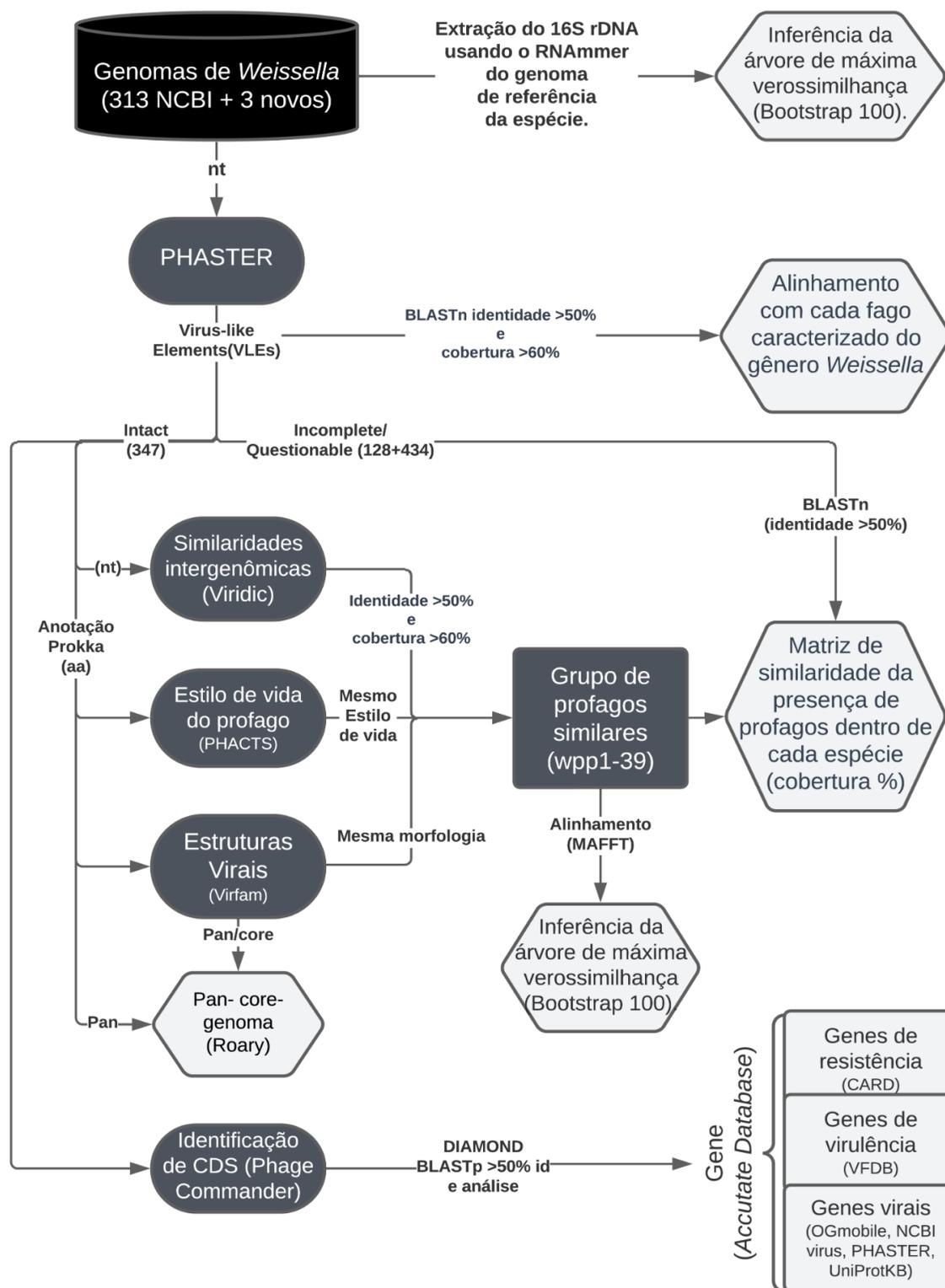
*Antibiotic Resistance Database*) (ALCOCK *et al.*, 2023) e o VFDB (*Virulence Factor of Pathogenic Bacteria*) (LIU *et al.*, 2019), respectivamente, para valores superiores a 70% de identidade e *e-value* menores que  $1e-10$ .

Para observar a presença de genes ligados à defesa das bactérias frente a fagos, todos as sequências de genomas do gênero *Weissella* depositadas no NCBI (*National Center for Biotechnology Information*) (Março, 2023) foram obtidas, totalizando 316 genomas, incluindo as três apresentadas nesse trabalho, e, novamente, o DIAMOND foi empregado usando como banco de dados curados ARSENAL (*Prokaryotic Antiviral Defense System*) (ZHANG *et al.*, 2020); e para aprofundar na avaliação de possíveis presenças de sistemas CRISPR-Cas, foi utilizado a ferramenta CRISPR-CasFinder (COUVIN *et al.*, 2018).

Para melhor entender a relação filogenética do gênero, uma árvore filogenética dos genomas de referência de cada espécie foi gerada a partir do gene *16S rRNA* usando TYGS (Type Strain Genome Server) (MEIER-KOLTHOFF *et al.*, 2022), e a relação foi verificada usando o próprio database do TYGS, que usa referências mais antigas para cada espécie. Para um melhor controle dos resultados, posteriormente, usamos o RNAmmer (LAGESEN *et al.*, 2007) para extrair as sequências do gene *16S rRNA*, alinhamos usando o programa MAFFT (NAKAMURA *et al.*, 2018), seguido da inferência da árvore filogenética usando *maximum likelihood* (ML) com bootstrap de 100 com o programa RAxML (STAMATAKIS *et al.*, 2014) e visualização da árvore usando iTOL (*Interactive Tree of Life*) web page (LETUNIC & BORK, 2021).

#### **4.2 Detecção e análise dos profagos no gênero *Weissella***

O fluxograma de trabalho para a detecção e posterior análises dos profagos está demonstrado na figura 8.



**Figura 8:** Fluxograma da metodologia para detecção e análise dos profagos no gênero *Weissella*. Fluxograma da metodologia empregada.

#### 4.2.1. Detecção dos profagos

No total, foram utilizados 316 genomas pertencentes ao gênero *Weissella* para esta análise. Destes, 313 foram obtidos a partir de montagens disponíveis no *National Center for Biotechnology Information* (NCBI) até março de 2021, compreendendo 40 montagens de genomas completos e 276 montagens em nível de *scaffold*. Adicionalmente, três genomas de *Weissella* potencialmente probióticas foram sequenciados especificamente para este estudo.

A detecção de profagos nos genomas de *Weissella* foi realizada empregando o programa PHASTER (*PHAge Search Tool Enhanced Release*), uma ferramenta amplamente reconhecida e amplamente utilizada para a anotação de profagos em genomas bacterianos (ARNDT *et al.*, 2016). O algoritmo subjacente do PHASTER emprega diversas técnicas, incluindo a predição de genes com o auxílio do programa GLIMMER, a identificação de tRNA e tmRNA utilizando o ARAGORN e o tRNAscan-SE, e a subsequente comparação das regiões preditas com um banco de dados por meio do programa BLAST, utilizando um limite de valor de *e* (*e-value*) superior a  $10^{-4}$ . O referido banco de dados é subdividido em três partes essenciais: um banco de dados de sequências bacterianas, um banco de dados de profagos/vírus e um banco de dados de fragmentos de DNA. Essas bases de dados foram elaboradas e são mantidas atualizadas por meio da inclusão de sequências provenientes de fagos, profagos e bactérias, originárias de genomas arquivados no NCBI, bem como pelo banco de dados de profagos desenvolvido por Srividhya e colaboradores (2006).

Os genes previstos são então agregados, e regiões candidatas a conterem profagos são propostas com base na avaliação da densidade local desses genes, utilizando o algoritmo DBSCAN. O PHASTER atribui uma pontuação a essas regiões, levando em consideração diversos critérios, como a presença de genes com similaridade a genes de outros fagos, o número total de genes codificadores de proteínas (CDS), o tamanho da região, bem como as funções e características dos genes que compõem a região. Com base nesses critérios, as regiões são classificadas em três categorias distintas: "intactos", "questionáveis" e "incompletos" (ARNDT *et al.*, 2016).

#### 4.2.2. Anotação gênica dos profagos completos de *Weissella*

A anotação gênica foi conduzida de maneira individualizada, utilizando a ferramenta Prokka v.1.14.6 (SEEMANN, 2014), com os parâmetros padrão e um banco de dados específico para genes virais. O Prokka faz uso do programa Prodigal (HYATT *et al.*, 2010) para identificar as *Open Reading Frames* (ORFs) nos genomas analisados.

Para a predição da possível morfologia viral, empregou-se a ferramenta Virfam (LOPES *et al.*, 2014). Essa ferramenta foi originalmente desenvolvida para identificar famílias virais com base em características morfológicas, mas, como mencionado anteriormente, as famílias preditas pela ferramenta foram consideradas obsoletas. No entanto, a ferramenta ainda possui um banco de dados de perfis de *Hidden Markov Models* (HMM) para genes estruturais virais, o que a torna uma escolha valiosa para a predição da morfologia dos vírus. A fim de classificar o estilo de vida dos profagos, foi utilizada a ferramenta PHACTS (MCNAIR *et al.*, 2012), que dispõe de um banco de dados contendo informações sobre genes de fagos lisogênicos e fagos líticos, incluindo genes relacionados à integração, excisão, lisogenia, genes estruturais e replicação. A ferramenta compara essas informações com os genes previamente anotados pelo Prokka e fornece uma probabilidade estatística do estilo de vida provável do profago.

#### 4.2.3. Pan- core- genoma

Para compreender de forma mais abrangente a diversidade dos profagos no gênero *Weissella*, foi realizada uma análise do pan-genoma e core-genoma desses profagos por meio da ferramenta ROARY (PAGE *et al.*, 2015). Dada a notável diversidade genética observada em fagos (TURNER *et al.*, 2021; BRUM *et al.*, 2013), aplicamos configurações que não separam os clusters em parálogos e exigiram que os genes do core estivessem presentes em pelo menos 50% dos profagos intactos. Além disso, utilizamos uma porcentagem de identidade para a análise de BLASTp que variou de 40% a 90%, conforme previamente demonstrado (MARQUES *et al.*, 2021). Posteriormente, a mesma estratégia foi empregada para profagos intactos que compartilhavam a mesma morfologia.

#### 4.2.4. Agrupamento dos profagos similares e análises

Todos os profagos considerados intactos foram submetidos a alinhamentos utilizando a ferramenta BLAST contra os três fagos líticos previamente caracterizados para o gênero *Weissella*: *Weissella cibaria* YS61 phage phiYS61, *Weissella* phage PWc e *Weissella* phage WCP30.

Um segundo alinhamento foi realizado utilizando o programa MAFFT para comparar os genomas completos das sequências nucleicas, e foi utilizado o argumento "--adjustdirection" para compreender sequências invertidas. Em seguida, uma matriz que engloba similaridade, cobertura e tamanho foi criada por meio do programa VIRIDIC (*Virus Intergenomic Distance Calculator*) (MORARU *et al.* 2020).

Com base nos resultados obtidos por essas ferramentas, os profagos foram agrupados quando apresentavam uma similaridade superior a 50%, uma cobertura superior a 60%, além de compartilharem o mesmo estilo de vida e morfologia (STRÖMSTEN *et al.*, 2003; VILLION *et al.*, 2009; COSTA *et al.*, 2018). Para representar cada grupo, foi selecionado o membro que apresentava a maior cobertura com os outros membros do grupo e o maior tamanho, uma vez que um genoma maior proporciona uma maior variedade de proteínas acessórias dentro do grupo. Com essas sequências selecionadas, foi conduzido um novo alinhamento usando MAFFT, utilizando esses genomas de referência do grupo. O programa jModeltest2 (DARRIBA *et al.*, 2012) foi utilizado para determinar o melhor modelo de substituição de bases, que posteriormente foi aplicado ao programa RAxML (STAMATAKIS *et al.*, 2014) para inferir a árvore filogenética dessas sequências utilizando o método de máxima verossimilhança (*maximum likelihood, ML*), com a realização de 100 replicatas de bootstrap. Por fim, a árvore filogenética resultante foi visualizada na página da web iTOL (*Interactive Tree of Life*) (<https://itol.embl.de/>) (LETUNIC & BORK, 2021).

Com o objetivo de superar um desafio significativo associado à utilização de montagens de genoma em nível de rascunho (*draft*), onde os profagos intactos podem estar fragmentados em vários contigs, resultando na classificação equivocada desses profagos pelo PHASTER como múltiplos profagos incompletos ou questionáveis, realizamos alinhamentos rigorosos usando a ferramenta BLASTn dessas sequências fragmentadas com cada representante dos profagos intactos. Estabelecemos um critério de classificação que considerou as sequências como

componentes desses profagos referência se apresentassem uma similaridade superior a 50%, e atribuímos um valor de porcentagem de cobertura relativa ao profago correspondente. Dessa forma, geramos uma matriz de cobertura para esses profagos classificados como questionáveis e incompletos pelo PHASTER em comparação com os profagos intactos.

#### 4.2.5. Análise da presença de genes de virulência e resistência

Em uma nova abordagem, todas as ORFs (*Open Reading Frames*) dos profagos considerados intactos dentro do gênero *Weissella* foram previamente identificadas utilizando o pipeline PhageCommander (LAZEROFF *et al.*, 2020). Esse pipeline emprega nove programas diferentes de predição de ORFs, cada um com algoritmos e abordagens distintas: Glimmer; GeneMark; GeneMark.hmm; GeneMarkS; GeneMark; GeneMarkS2; Prodigal; MetaGene; e ARAGORN para detecção de tRNA. Para minimizar a ocorrência de ORFs falsos positivos e falsos negativos, foram selecionadas apenas as ORFs que obtiveram um consenso de predição de quatro ou mais anotadores (LAZEROFF *et al.*, 2020; SALISBURY *et al.*, 2019).

Após a identificação das ORFs, realizamos a reanotação de seus genes utilizando o programa DIAMOND (BUCHFINK *et al.*, 2021), empregando os parâmetros padrão para a busca BLASTp com modo de sensibilidade (exigindo uma identidade superior a 30%). Para identificar genes relacionados à resistência e virulência, utilizamos os *database* CARD (*The Comprehensive Antibiotic Resistance Database*) e VFDB (*Virulence Factor of Pathogenic Bacteria*) (ALCOCK *et al.*, 2023; LIU *et al.*, 2019). Para genes virais, recorreremos às seguintes *databases*: mobileOG-db, um banco de dados preciso de elementos móveis que abrange todos os genomas de fagos conhecidos, bem como plasmídeos, profagos e outros elementos móveis; nr(*non-redundant*) NCBI *virus*, que engloba todos os genomas virais já depositados no NCBI e suas respectivas anotações; PHASTER, que contém sequências de profagos e fagos previamente anotados com precisão; e UniProtKB/Swiss-Prot, um banco de dados manualmente anotado e não redundante (BROWN *et al.*, 2022; BRISTER *et al.*, 2015; ARNDT *et al.*, 2016; THE UNIPROT CONSORTIUM, 2021).

Posteriormente, todas as ORFs foram submetidas a uma reanotação funcional manual com base nas informações obtidas nas etapas anteriores. Cada ORF foi avaliada considerando

critérios como identidade superior a 30%, cobertura maior que 50% e um valor de *e-value* menor que  $1 \times 10^{-10}$ , conforme descrito para a anotação manual (SALISBURY *et al.*, 2019). Para essa tarefa, utilizamos o programa Ugene (OKONECHNIKOV *et al.*, 2012) a fim de assegurar a precisão e a acurácia das anotações funcionais das ORFs.

#### **4.2.6. Análise estatística**

Os testes de correlação foram conduzidos utilizando o coeficiente de correlação de Pearson, adotando um nível de significância de  $P < 0,05$  e um valor de  $r > 0,05$  como critério. Os resultados foram apresentados em formato de média  $\pm$  erro padrão médio. Para representar a presença de profagos em cada espécie, foram calculados valores de moda e média, devido à considerável variação no número de amostras de cada espécie.

Todas as análises estatísticas e a criação das figuras foram realizadas utilizando os seguintes *softwares*: IBM SPSS Statistics (IBM CORP, Armonk, NY, USA), Microsoft Excel (MICROSOFT CORPORATION, 2018), R (R CORE TEAM, 2021), R Studio (R STUDIO TEAM, 2020), pacote tidyverse (WICKHAM, 2017) e pacote ComplexHeatmap (GU *et al.*, 2016).

## 5. RESULTADOS E DISCUSSÃO

### 5.1 Sequenciamento

O sequenciamento do DNA das três bactérias utilizando a plataforma HiSeq 2500 da Illumina gerou dois arquivos, *paired-ends*, para cada biblioteca. No total foram sequenciados 3,32 Gb, distribuídos em 22.140.764 *reads* do tipo *paired-ends*, cada *read* com 150 pares de base (pb) de comprimento (Tabela 1). Todas as sequências tiveram a média de sua qualidade por bases com Phred scores superiores a 32 em todas as posições, tendo uma leve e esperada diferença na qualidade entre os pares da sequência, segundo o próprio manual do programa FastQC (ANDREWS *et al.*, 2010). Uma vez que a qualidade de todas as bibliotecas foi satisfatória, as *reads* foram submetidas à montagem *de novo*, usando os programas Edena, SPAdes e Unicycler. A partir da análise da qualidade da montagem usando o programa QUAST, o pipeline do Unicycler foi selecionado como o melhor para a montagem das três bibliotecas, levando em consideração as discrepâncias com relação ao tamanho do genoma, conteúdo GC e número de *contigs*.

Durante a identificação da espécie e linhagem usando o servidor TYGS foi observado que uma linhagem foi erroneamente identificada pelo trabalho prévio (STEINBERG, 2016). Através da comparação do gene *16S rDNA* pelo TYGS e pela sintonia dos *contigs* totais e o genoma da referência, foi possível reclassificar a linhagem GIR46L4 (VSF140) como sendo da espécie *Weissella confusa*. Os resultados apresentados para os outros dois genomas corroboram o encontrado por Alvim (201) e Sandes (2013), classificando a linhagem 1BM (VSF134) como pertencendo à espécie *Weissella cibaria* e classificando a linhagem V3B-11 (VSF142) como pertencendo à espécie *Weissella paramesenteroides* (Figura Suplementar S1).

**Tabela 1:** Painel de características das diferentes bibliotecas produzidas.

<b>Biblioteca</b>	<b>Nº total de <i>reads</i> (paired-ends)</b>	<b>%GC</b>	<b>Referência para montagem híbrida</b>
<b>VSF134</b>	3.733.284 + 3.733.284 = 7.466.568	44,8	<i>W. cibaria</i> CMS2
<b>VSF140</b>	3.907.257 + 3.907.257 = 7.814.514	45,0	<i>W. confusa</i> VTTE-90392
<b>VSF142</b>	3.429.841 + 3.429.841 = 6.859.682	38,2	<i>W. paramesenteroides</i> STCH-BD1

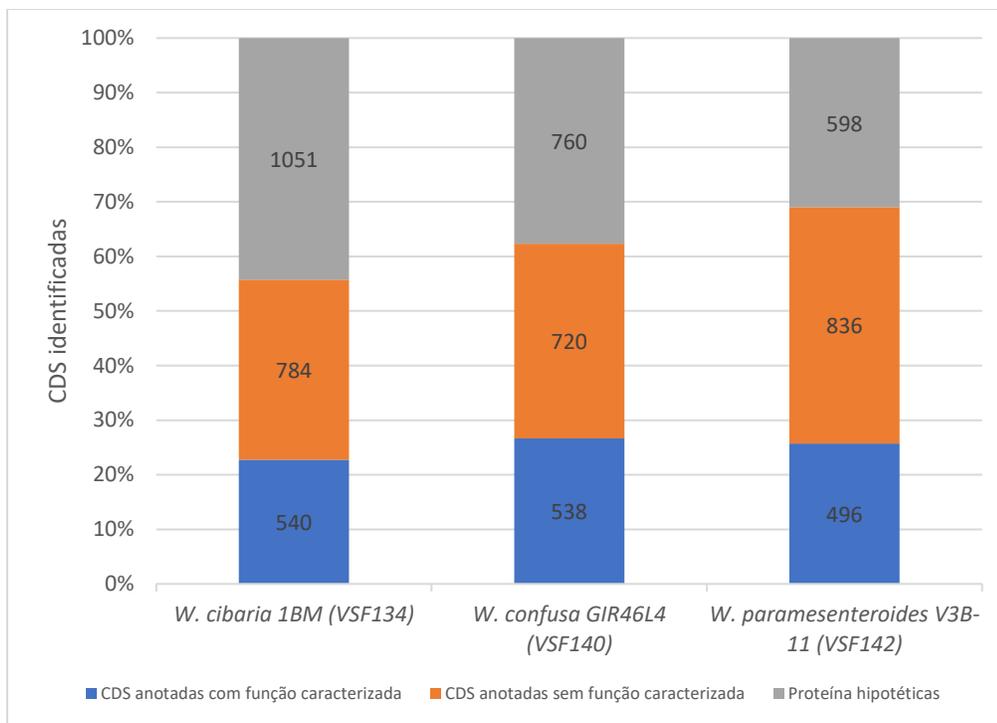
A ferramenta CONTIGuator ordenou os *contigs* gerados com diversas linhagens da espécie referência, a partir da melhor sintonia e maior cobertura, foi gerado dois arquivos com os *contigs* mapeados e os *contigs* não mapeados. Os não mapeados foram submetidos a uma análise manual. Afim de melhorar a qualidade da montagem, *contigs* não mapeados com tamanhos inferiores a 500 bases, com baixa cobertura que, ao serem analisados usando a ferramenta *BLASTn*, não possuam hit contra o banco de dados *nr* (*Non-redundant*) do NCBI foram excluídos da análise e classificados como artefatos de erro do sequenciamento. A montagem dos genomas das bibliotecas VSF134 e VSF140 ficou no nível de *scaffold* devido à dificuldade do fechamento de regiões repetitivas, como regiões de transposase e rRNA.

No caso da biblioteca VSF142, usando o mapeamento de *reads* pela ferramenta *Bowtie2*, e analisando o mapeamento usando a ferramenta *SamTools*, a profundidade de cobertura média do sequenciamento ficou em 514x, ou seja, cada base foi confirmada em média em 514 *reads* durante o sequenciamento. Ao final, foi possível identificar um cromossoma e dois plasmídeos, todos circularizados, não possuindo gaps e sem bases ambíguas (Ns), todos os *contigs* gerados na montagem foram mapeados. Das 6.859.622 *reads* obtidas no sequenciamento, 99,82% das *reads* foram mapeadas. Logo, a biblioteca VSF142 de *Weissella paramesenteroides* possui um cromossoma com 1.951.502 pb e dois plasmídeos circulares, de 33.263 pb e 10.954 pb. O conteúdo %GC e o tamanho do genoma, respectivamente 38,22% e 1.951.502 pb da linhagem

condiz com o esperado encontrado em outros membros da espécie depositados no NCBI, como no caso da linhagem STCH-BD1, 38,14% e 2.052.536 pb; e na linhagem FDAARGOS, 38,10% e 1.947.910 pb.

## 5.2 Anotação funcional

Os genomas montados foram anotados usando o programa Prokka e o servidor RAST. Não houve muita diferença na quantidade de CDS de cada programa, contudo, o número de CDS identificadas como proteínas (não hipotéticas) foi maior na anotação do servidor RAST. Os genomas *Weissella cibaria* 1BM (VSF134), *Weissella confusa* GIR46L4 (VSF140) e *Weissella paramesenteroides* V3B-11 (VSF142) possuem, respectivamente, 2375 CDS e 102 RNAs, 2020 CDS e 82 RNAs, e 1930 CDS e 94 RNAs, esses valores são subestimados uma vez que os genomas *W. cibaria* 1BM (VSF134) e *W. confusa* GIR46L4 (VSF140) estão em *draft* e podem ter sequências quebradas, contudo, os números de CDS das duas sequências estão próximos dos já reportados para *W. confusa*, de média 2078 no NCBI e *W. cibaria*, de média 2300 em outros trabalhos (LYNCH *et al.*, 2015). No caso da *W. paramesenteroides* V3B-11 (VSF142) o valor de CDS é próximo do encontrado nos outros membros da espécie no NCBI, que possuem média de aproximadamente 1868 CDS. Analisando a distribuição das possíveis funções da CDS de cada uma das sequências (Figura 9), podemos observar que a anotação conseguiu identificar somente algumas poucas proteínas relacionadas a profagos. Esse resultado era esperado, tendo em vista que existe uma grande dificuldade na anotação de genes de fagos por não apresentarem homólogos dentro dos bancos de dados usados para anotação, isso se deve também pela falta de informação sobre genes de fagos sendo a maioria classificado como função putativa ou hipotético (SALISBURY *et al.*, 2019).



**Figura 9. Distribuição das CDS em categorias para cada montagem.** Distribuição, em porcentagem, da quantidade de CDS identificadas em cada uma das montagens realizadas: No eixo X as linhagens bacterianas analisadas e no eixo Y a porcentagem das CDS identificadas em cada uma delas. Em azul, CDS anotadas com função predita por COGs; em laranja, CDS anotadas sem clareza sobre função, logo, não caracterizadas dentro de nenhuma COGs; em cinza CDS que foram anotadas como proteínas hipotéticas.

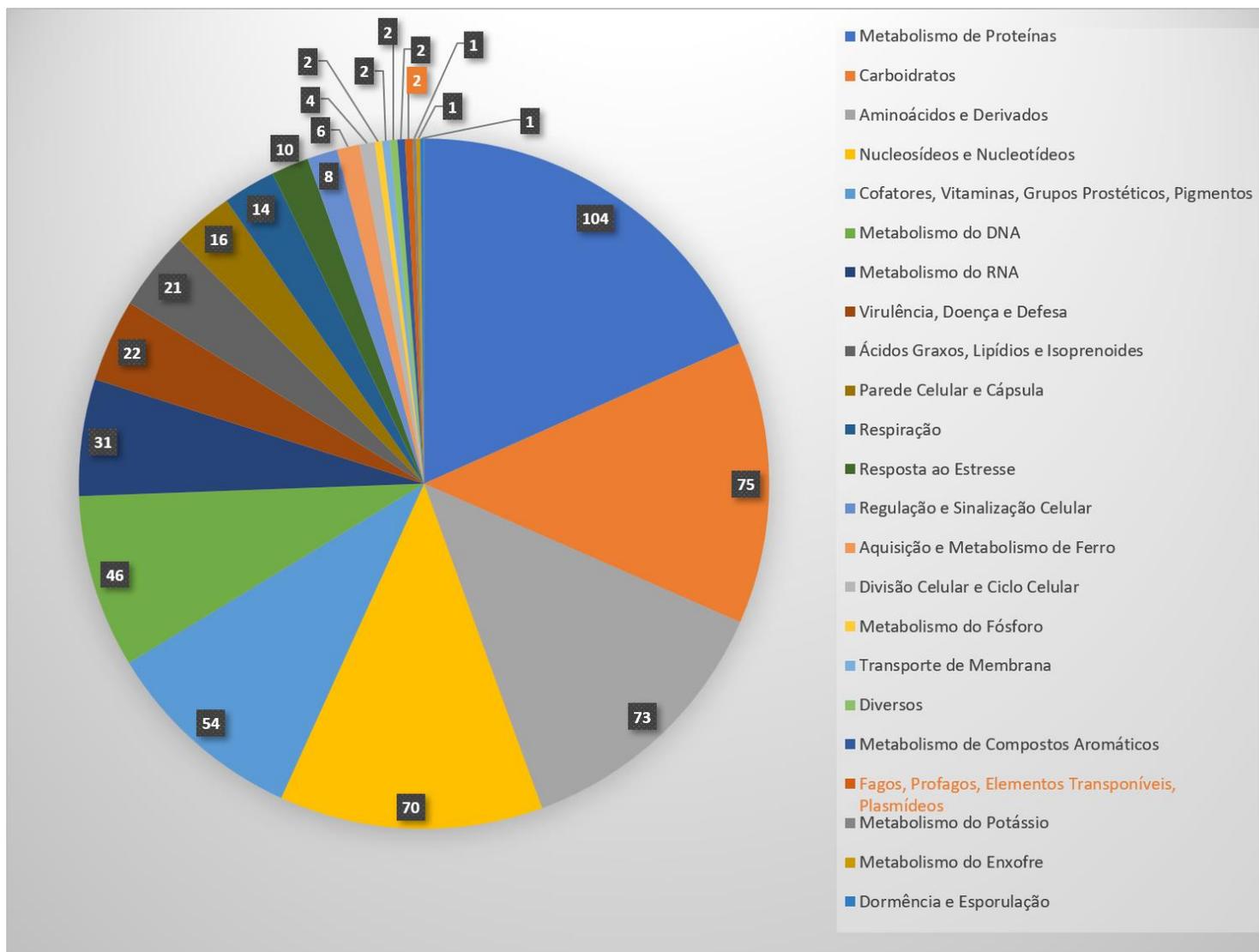
*Cluster of Orthologous Groups* (COGs) são grupos de ortólogos que são classificados em subsistemas, esses subsistemas são coleções de papéis funcionais que um anotador determinou que devem ser considerados como inter-relacionados. Frequentemente, os subsistemas representam o conjunto de papéis funcionais que compõem uma via metabólica, um complexo (como o ribossomo) ou uma classe de proteínas (por exemplo, proteínas de transdução de sinal de dois componentes em *Staphylococcus aureus*). A relação entre os ortólogos é definida e são atribuídos códigos de variante que descrevem quais das possíveis variantes do subsistema existem dentro de cada genoma (códigos especiais são usados para expressar a ausência total do subsistema ou incerteza sobre sua presença). A construção de COGs com subsistemas curados é um dos papéis centrais nos esforços de anotação realizados pelo SEED, *National Microbial Pathogen Data Resource* (NMPDR) e pelo *Bacterial and Viral Bioinformatics Resource Center* (BV-BRC) (OVERBEEK *et al.*, 2005; MCNEIL *et al.*, 2007; OLSON *et al.*, 2023).

No genoma da linhagem *W. cibaria* 1BM (VSF134), que foram indicadas 2375 CDS, cerca de 56% das CDS foram anotadas com genes conhecidos. Dentre essas CDS, aproximadamente 23% (540) apresentaram possíveis funções preditas por meio de COGs (*Cluster of Orthologous Groups*), com apenas 2 dessas CDS sendo classificadas como proteínas envolvidas na montagem de fagos. Por outro lado, 33% (784) das CDS receberam anotações gênicas, mas sem funções claramente definidas. Enquanto o restante, 44% (1.051) das CDS, foram catalogadas como proteínas hipotéticas (conforme Figura 9 e 10).

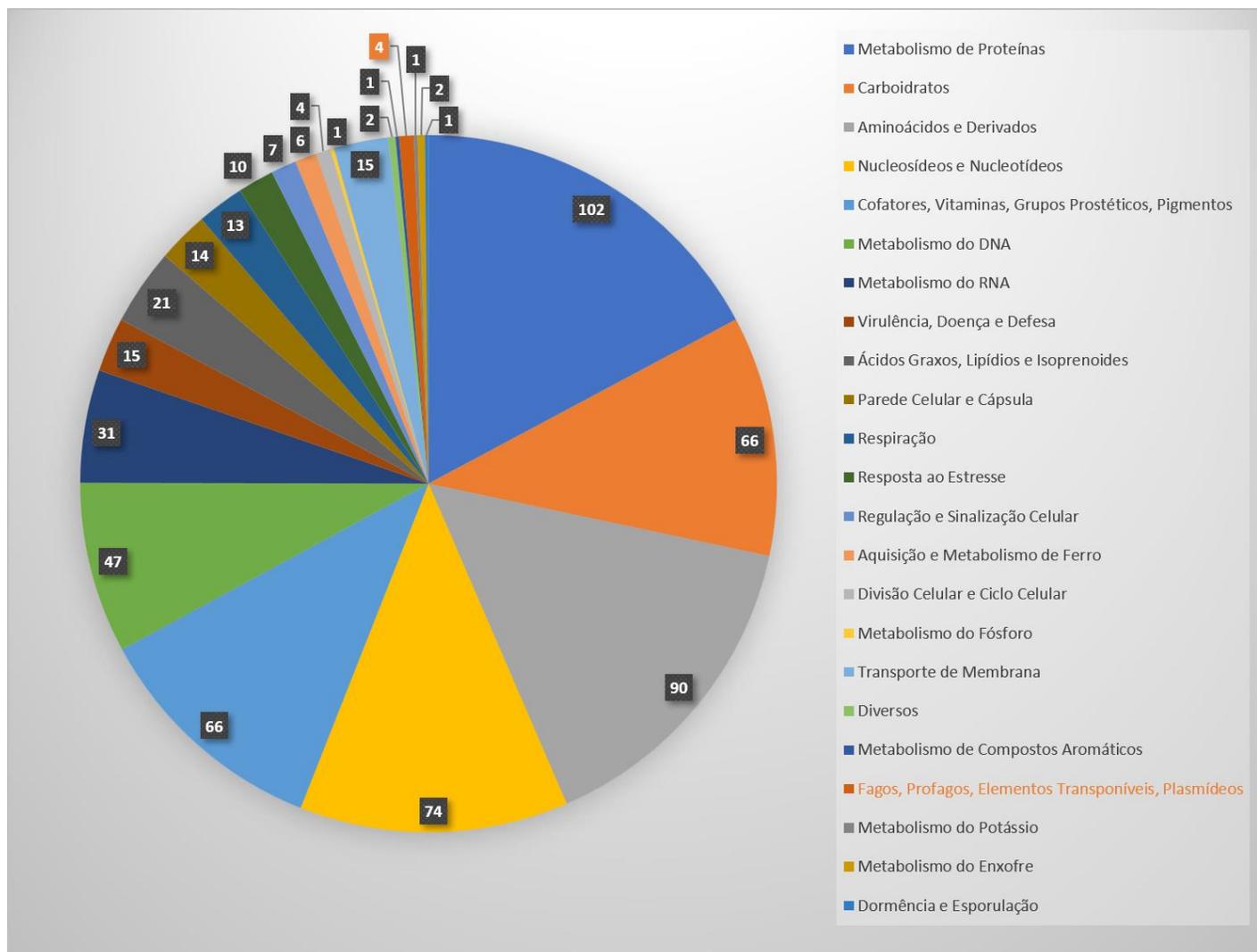
No genoma da linhagem *W. confusa* GIR46L4 (VSF140), que abrange um total de 2.020 CDS anotadas, aproximadamente 63% delas tiveram seus genes anotados. Cerca de 27% (538) das CDS totais tiveram funções possivelmente preditas por meio da utilização de COGs, com 4 CDS sendo classificadas como proteínas envolvidas na montagem de fagos. Adicionalmente, 35,6% (720) das CDS foram anotadas, mas sem funções claramente definidas, e o restante, 37,6% (760) das CDS, foi anotado como proteínas hipotéticas (conforme Figura 9 e 11).

No genoma da linhagem *W. paramesenteroides* V3B-11 (VSF142), composto por um total de 1.930 CDS, foi possível anotar cerca de 71,7% das CDS em genes. Aproximadamente 25,7% (496) dessas CDS totais possuíam funções identificadas com base em ortólogos conhecidos, usando COGs, incluindo 3 desses genes envolvidos com a montagem, estrutura e replicação de fagos. Adicionalmente, cerca de 43,3% (836) das CDS totais foram anotadas com genes, mas não tiveram funções preditas em COGs. Por fim, 31% (598) do total de CDS foram catalogados como proteínas hipotéticas (conforme Figura 9 e 12).

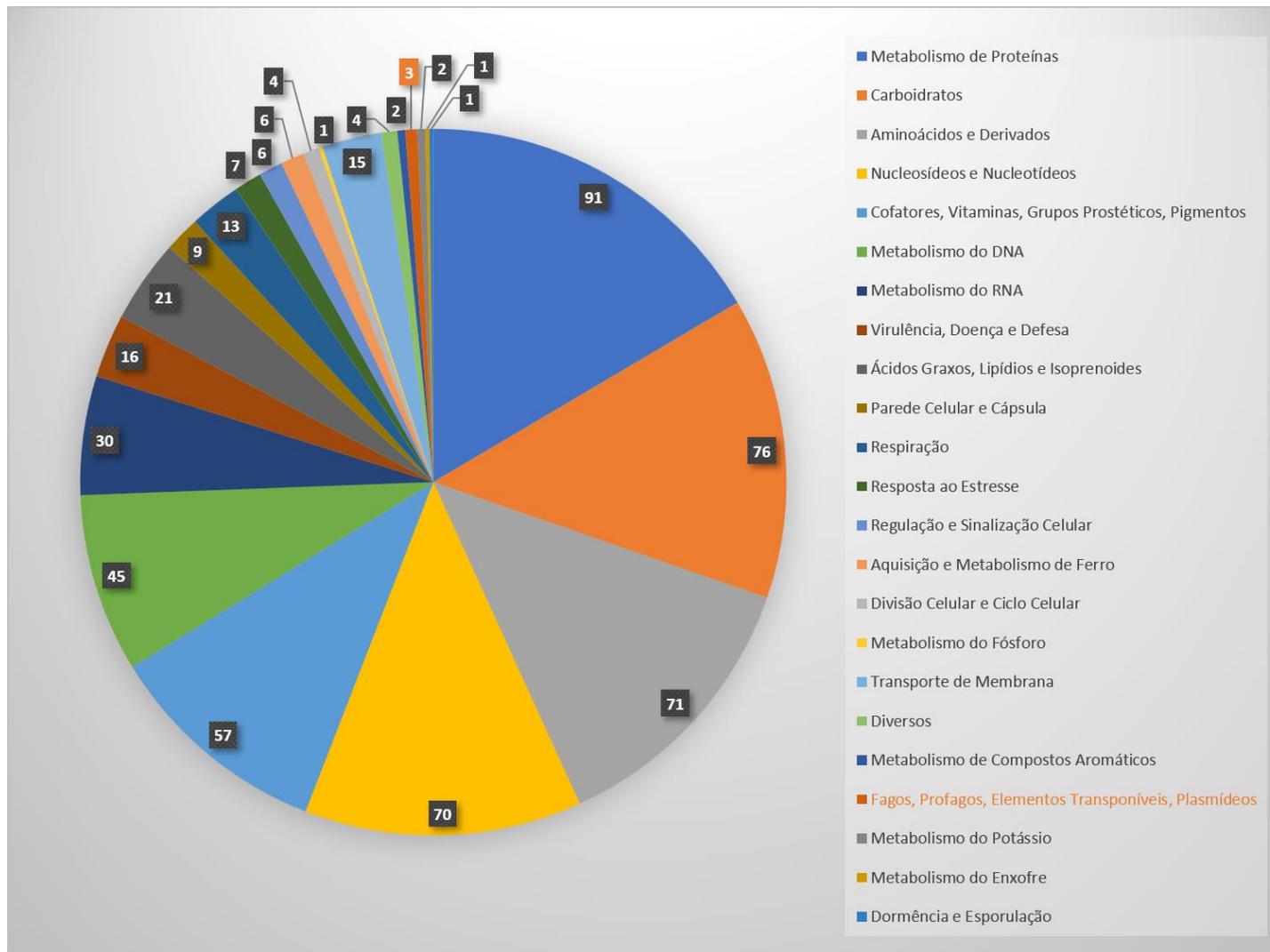
Considerando o perfil dos genes com funções inferidas pelos COGs, e dado que nosso sequenciamento abrange três espécies distintas dentro do mesmo gênero, é possível observar uma conservação grande das funções dentro do gênero, uma vez que todas as nossas sequências apresentaram percentualmente valores parecidos para cada função (Figura 10, 11 e 12). A linhagem de *W. paramesenteroides* V3B-11 (VSF142) teve a maior porcentagem de genes anotados, possivelmente por ser um genoma montado e classificado como completo (>99% das *reads* cobertas, plasmídeos e genomas circularizados, e sem *contigs* não mapeados).



**Figura 10.** Distribuição funcional e contagem dos genes em COGs (*Cluster of Orthologous Groups*) para o genoma *Weissella cibaria* 1BM (VSF134). No lado esquerdo a distribuição em gráfico de pizza a distribuição proporcional dos genes do genoma de *W. cibaria* 1BM (VSF134) em diferentes categorias funcionais. Cada fatia do gráfico de pizza representa uma categoria específica, e a área das fatias reflete a proporção de genes atribuídos a cada categoria. O enfoque é dado à cor laranja, que representa a categoria que inclui genes relacionados a fagos, profagos, elementos transponíveis e plasmídeos. No lado direito é encontrado a legenda das cores que correspondem às diferentes categorias funcionais representadas no gráfico de pizza. Novamente, como mencionado, enfoque na categoria com escrito em laranja.



**Figura 11. Distribuição funcional e contagem dos genes em COGs (Cluster of Orthologous Groups) para o genoma *Weissella confusa* GIR46L4 (VSF140).** No lado esquerdo a distribuição em gráfico de pizza a distribuição proporcional dos genes do genoma de *W. confusa* GIR46L4 (VSF140) em diferentes categorias funcionais. Cada fatia do gráfico de pizza representa uma categoria específica, e a área das fatias reflete a proporção de genes atribuídos a cada categoria. O enfoque é dado à cor laranja, que representa a categoria que inclui genes relacionados a fagos, profagos, elementos transponíveis e plasmídeos. No lado direito é encontrado a legenda das cores que correspondem às diferentes categorias funcionais representadas no gráfico de pizza. Novamente, como mencionado, enfoque na categoria com escrito em laranja.



**Figura 12.** Distribuição funcional e contagem dos genes em COGs (*Cluster of Orthologous Groups*) para o genoma *Weissella paramesenteroides* V3B-11 (VSF142). No lado esquerdo a distribuição em gráfico de pizza a distribuição proporcional dos genes do genoma de *W. paramesenteroides* V3B-11 (VSF142) em diferentes categorias funcionais. Cada fatia do gráfico de pizza representa uma categoria específica, e a área das fatias reflete a proporção de genes atribuídos a cada categoria. O enfoque é dado à cor laranja, que representa a categoria que inclui genes relacionados a fagos, profagos, elementos transponíveis e plasmídeos. No lado direito é encontrado a legenda das cores que correspondem às diferentes categorias funcionais representadas no gráfico de pizza. Novamente, como mencionado, enfoque na categoria com escrito em laranja.

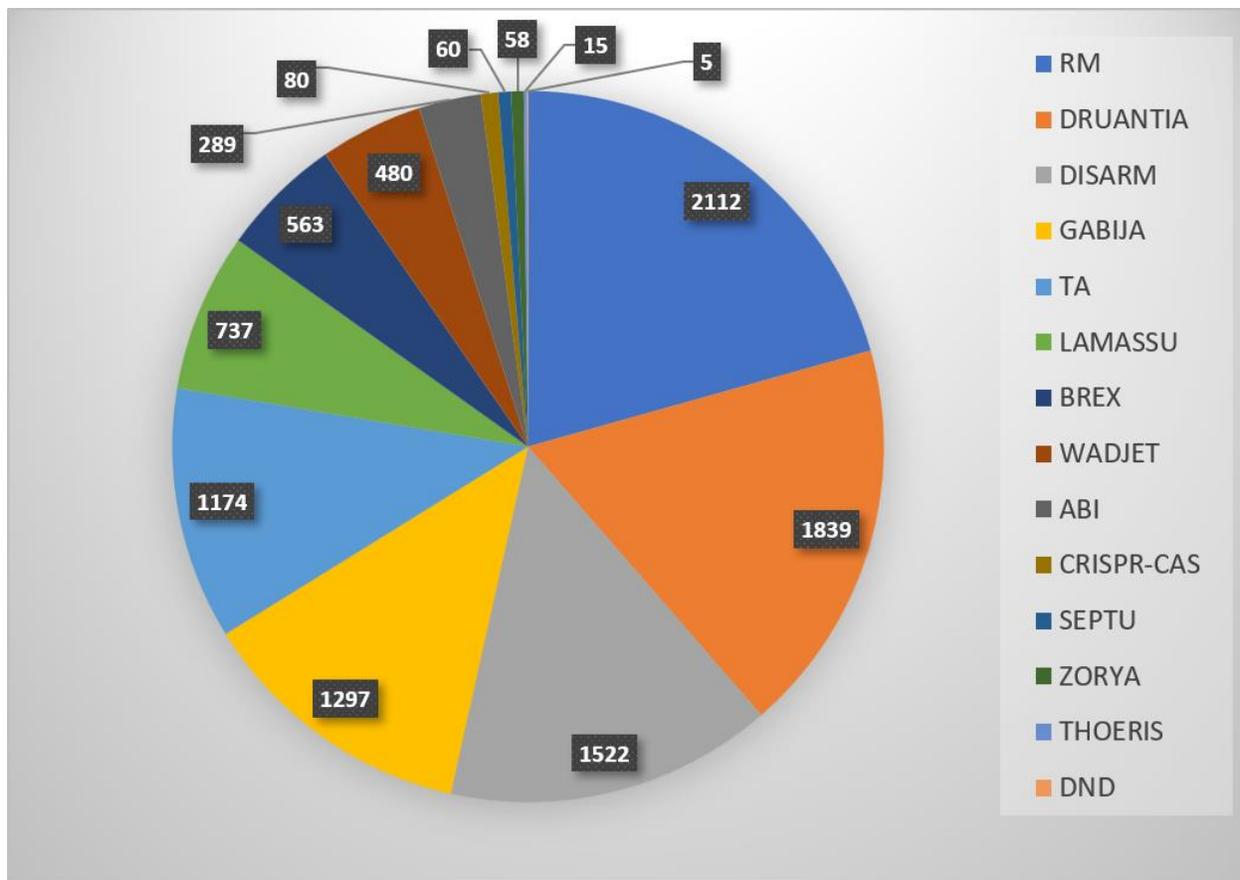
### 5.3 Sistema de defesa contra fagos

Leigh Van Valen, em 1973, nos apresentou a hipótese da Rainha Vermelha e que já foi amplamente discutida e apresentada no contexto da evolução dos fagos e das bactérias (VALEN, 1973; HEILMANN *et al.*, 2010; PATERSON *et al.*, 2010; STERN *et al.*, 2011; KOONIN *et al.*, 2012; PAPKOU *et al.*, 2019; LOH *et al.*, 2020; HO e HARSHEY, 2021; MAHADEVAN *et al.*, 2023; RAMOS-BARBERO *et al.*, 2023). A hipótese da Rainha Vermelha tenta explicar o *status quo* da relação entre o parasita e o hospedeiro, ou predador e presa, onde ambos coevoluem e se selecionam em um ciclo sem fim de evolução, aumentando a diversidade de mecanismos de defesa e de ataque de ambos os organismos. Nosso estudo se propôs a melhor elucidar esses mecanismos de defesa no gênero *Weissella*, bem como identificar a prevalência ou ausência de alguns deles no gênero, uma vez que foi observado nos genomas sequenciados genes dentro de COGs relacionados a fagos, profagos, elementos transponíveis e plasmídeos.

Para identificar formas de proteção bacteriana contra fagos, todos os genomas (316) do gênero *Weissella* foram submetidos à procura usando o banco de dados PADS ARSENAL (*Prokaryotic antiviral defense systems*), que detectou a presença 190.665 genes com identidade maior 30%, e com função possivelmente conservada; com identidade maior que 70%, 14.003 genes; e muito conservados com identidade maior que 95%, 10.231 genes. Abordando somente os muito conservados, esses genes foram divididos em 14 categorias de defesa contra fagos (Figura 13). Levando em consideração a ação dessas categorias, elas se dividem em dois grupos, os sistemas de indução de dormência ou morte celular programada do grupo infectado e os sistemas de imunidade.

O primeiro grupo é representado pelas categorias: Sistema de infecção abortiva (ABI) que bloqueia a multiplicação viral e promove a morte celular, limitando a dispersão da progênie viral dentro da população bacteriana (CHOPAN *et al.*, 2005), esta categoria representou um total de 2,82% dos genes encontrados; e o Sistema toxina-antitoxina (TA), que pode produzir toxinas, proteínas que reduzem o metabolismo, ou antitoxinas, que podem ser proteínas ou RNA que inibem a produção das toxinas, assim como em plasmídeos é capaz de inibir a propagação de fagos (SONG *et al.*, 2020). Sendo a quinta categoria mais abundante dentro da espécie, ela corresponde a 11,47% dos genes encontrados.

O segundo grupo é representado pelas categorias: Sistema de exclusão de bacteriófagos (BREX) que é um sistema de 6 genes que estão envolvidos na metilação do DNA da célula e bloqueio da replicação do DNA viral (GOLDFARB *et al.*, 2015), essa categoria apresentou 5,5% dos genes encontrados; o sistema de ilhas de defesa associadas a restrição/modificação (DISARM) esse sistema é composto por cinco genes, com uma DNA metilase e quatro helicases e trabalha junto com o sistema restrição/modificação (RM); o sistema RM é o sistema mais estudado em defesa de profagos, tendo a *EcoRI* como a mais utilizada, esse sistema funciona através do reconhecimento de regiões específicas de um fago levando a clivagem dessa região, essas regiões são chamadas de sítios de restrição. No gênero *Weissella* o sistema DISARM representou 14,7% dos genes de defesa, enquanto o sistema RM representou 20,5%, esses dados revelam que dentro da espécie o principal sistema de defesa contra fagos é por RM; *Druantia* foi o segundo em número de genes identificados, tendo 17,8% do total, esse sistema é caracterizado por codificar proteína enormes (~1800 a 2100 aminoácidos), que contêm domínios não caracterizados, ação de helicases e sítios de ligação de ATP, com pedaços de proteínas que lembram o sistema DISARM, apesar de novo e pouco estudado, confirma mais uma vez a hegemonia dos sistemas envolvendo RM dentro do gênero. Dentro deste segundo grupo também inclui sistemas recentemente descobertos, mas já apontadas com genes de proteção contra fagos, como os sistemas Septu, Hachiman, Lamassu, Wadjet, Gabija Zorya, Thoeris e DND, contudo, os mecanismos de ação ainda não foram completamente esclarecidos (DORON *et al.*, 2018; XIONG, *et al.*, 2020).



**Figura 13: Distribuição da quantidade de genes encontrados entre os membros do gênero *Weissella* para cada tipo de sistema de defesa contra fagos.** Esse gráfico de pizza, as fatias representam a proporção de genes associados a cada categoria de sistema de defesa encontrado nos 316 genomas analisados dentro do gênero *Weissella*. Cada fatia do gráfico está rotulada com uma legenda que indica o tipo específico de sistema de defesa. Essa representação visual permite uma rápida compreensão da diversidade e distribuição dos sistemas de defesa no gênero *Weissella*, destacando quais sistemas são mais prevalentes ou menos comuns na amostra de genomas analisados. Isso pode fornecer informações valiosas sobre as estratégias de defesa genética presentes nas espécies do gênero e como esses sistemas podem contribuir para a adaptação e sobrevivência desses organismos em seu ambiente.

O sistema CRISPR-Cas desempenha um papel importante na defesa contra fagos, pois evita infecções quando os fagos apresentam sequências semelhantes às dos espaçadores presentes no sistema CRISPR. Estudos anteriores já demonstraram que genomas com um maior número de profagos tendem a ter menos espaçadores CRISPR (TOUCHON *et al.*, 2016). Com o objetivo de compreender melhor a presença dos genes (80) relacionados ao sistema CRISPR-Cas dentro do gênero, os genomas bacterianos foram analisados utilizando a ferramenta CRISPRCasFinder. Essa ferramenta permitiu quantificar os espaçadores CRISPR, bem como a presença de proteínas Cas, e em conjunto, apontar o possível tipo de sistema CRISPR-Cas. Dos 316 genomas analisados, 295 apresentaram algum espaçador CRISPR, variando de 1 a 4 espaçadores por genoma. Quanto às

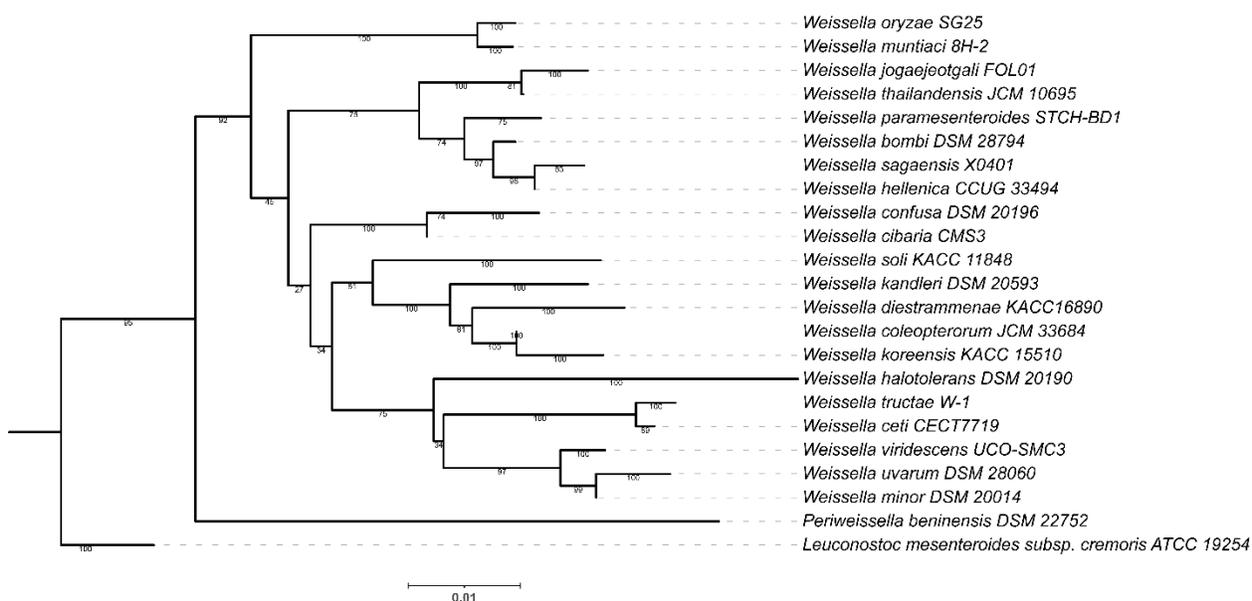
proteínas Cas, somente 44 genomas continham Cas, sendo 43 deles o sistema CRISPR-Cas subtipo IIA e 2 o subtipo IIIA. Curiosamente a linhagem *Weissella soli* DB-2 apresentou os dois tipos de sistemas e o subtipo IIIA foi exclusivo da espécie *Weissella soli*. A maioria dos espaçadores CRISPR estava associada ao subtipo IIA, com destaque para a espécie *Weissella cibaria*. No entanto, é importante notar que a maioria desses espaçadores foi classificada como nível 1, o que sugere uma alta probabilidade de serem falsos espaçadores (POURCEL *et al.*, 2020). Além disso, a presença do subtipo IIA estava restrita a alguns membros do gênero, incluindo *W. cibaria*, *W. confusa*, *W. soli*, *W. hellenica*, *W. halotolerans* e *W. kandleri*. O subtipo IIIA, por sua vez, parece ter sido recentemente integrado em *Weissella soli*.

Esses resultados indicam que o gênero *Weissella* não costuma apresentar um sistema CRISPR-Cas robusto, com um grande número de espaçadores, como observado em outros gêneros bacterianos (POURCEL *et al.*, 2020). A análise filogenética das espécies dentro do gênero, com base no gene *16S rRNA* (Figura 14), corroborou os resultados anteriores, mostrando que a presença do sistema CRISPR-Cas do subtipo IIA estava associada a membros próximos na árvore filogenética. A espécie *Weissella soli*, que apresentava dois subtipos, ocupa uma posição próxima ao ramo encontrado do subtipo IIA, mas também se encontra em um nó único para o *16S rRNA* e para o subtipo IIIA, uma vez que foi a única que apresentou tal subtipo.

Em conclusão, as bactérias do gênero *Weissella* apresentam diversas formas de defesa contra fagos, os resultados sugerem uma grande participação de sistemas semelhantes que envolvem restrição e modificação de sequências como nos sistemas RM/DRUANTIA/DISARM. Esses sistemas conseguem agir tanto em fagos temperados como em líticos, fazendo com que a coevolução ocorra tanto na bactéria hospedeira quando no fago e a até a fixação de profagos e genes virais dentro de seus genomas (DEDRICK *et al.*, 2017; PLESKA *et al.*, 2018; BROOKS *et al.*, 2020). Levando em conta o sistema CRISPR-Cas, observamos que não é muito desenvolvido e nem presente em todos os seus membros. Ainda não é claro os mecanismos de evasão presentes nos profagos, para escapar desses sistemas de defesa presentes nas bactérias *Weissella*, levando em consideração que nenhum sistema é garantido de proteção contra fagos. Fagos possuem diferentes formas de contra atacar esses sistemas de defesa, seja, mudando os sítios de ligação,

modificando proteínas virais ou cooperando com outros fagos para burlar o sistema CRISPR-Cas (LANDSBERGER, 2018).

Considerando que elementos parecidos com profagos são encontrados em diversas bactérias (CANCHAYA *et al.*, 2003), e que vírus são ubíquos, existe a probabilidade desses elementos carregarem fatores que podem melhorar o *fitness* bacteriano, característica que pode ser interessante para bactérias de interesse biotecnológico. Claro que é importante ressaltar a necessidade de saber a estabilidade desses elementos. O rearranjo entre os genomas dos fagos lisogênicos e seus hospedeiros pode levar a emergência de novas espécies de fagos e novas espécies de bactérias.

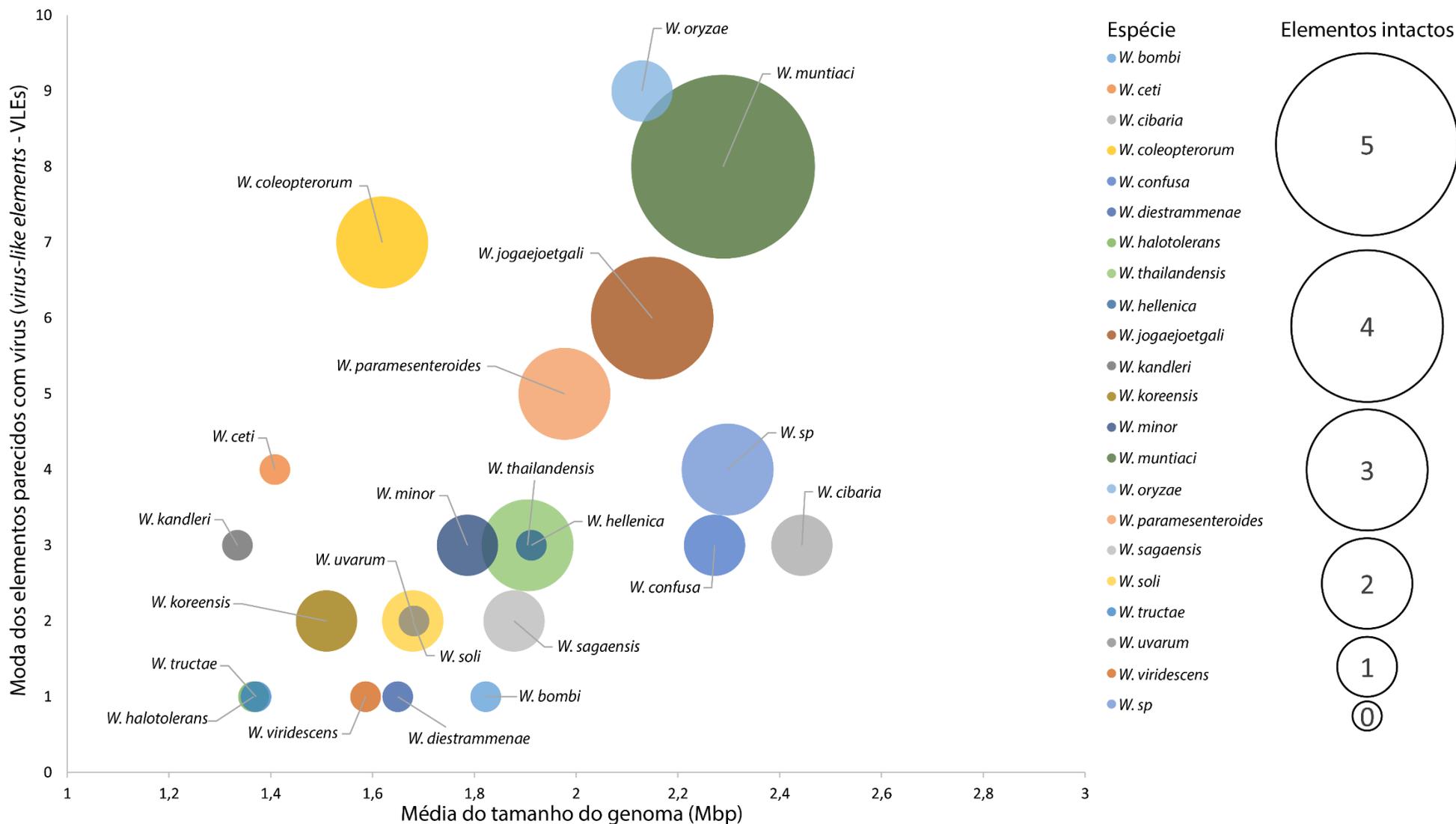


**Figura 14: Árvore filogenética dos membros do gênero *Weissella*.** Árvore filogenética de todas as sequências de referência de cada espécie do gênero *Weissella* e dois membros outgroup (*Periweissella* (um antigo membro do gênero) e *Leuconostoc* (um membro da mesma família *Lactobacillaceae*)). A árvore foi construída extraindo o 16S *rDNA* usando RNAmmer, alinhando-os usando MAFFT, RAXML foi usado para inferência da árvore filogenética por maximum-likelihood (ML) e bootstrap de 100.

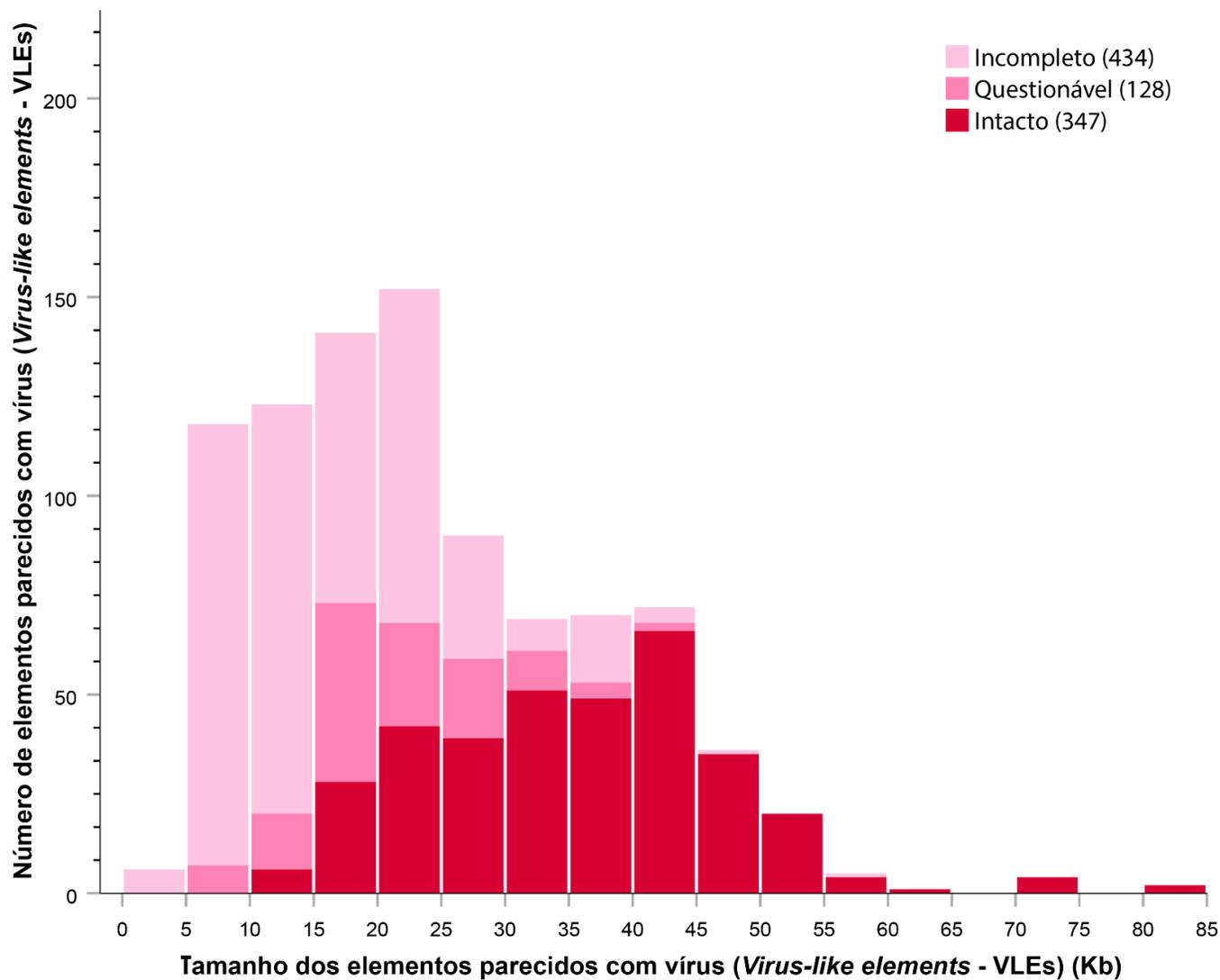
#### 5.4 Detecção *in silico* e análise de distribuição de elementos semelhantes a profagos em espécies de *Weissella*

Utilizamos a ferramenta PHASTER com argumentos *default* para identificar sistematicamente profagos dentro de 313 genomas de *Weissella* disponíveis publicamente, obtidos do banco de dados do NCBI. Além disso, adicionamos três genomas recentemente sequenciados, cuidadosamente selecionados por seu potencial biotecnológico, conforme descrito anteriormente (ALVIM *et al.*, 2015; SANDES *et al.*, 2020; STEINBERG *et al.*, 2022).

Nossa análise revelou um total de 908 elementos semelhantes a vírus (*virus-like elements* - VLEs) dentro desses genomas. Dentre esses, 347 foram categorizados como profagos intactos, 127 como questionáveis e 434 como incompletos. Essas classificações foram determinadas com base nos critérios do programa PHASTER, como tamanho da região, locais de ligação, número, composição e proximidade de genes semelhantes a vírus. De todas as bactérias (316), 27,5% dos genomas não apresentavam nenhum profago classificado como íntegro. Notavelmente, apenas *W. viridescens* NJ100, *W. confusa* BCC 4255, *W. confusa* LMG 14040 e *W. cibaria* VSF134 não apresentaram elementos semelhantes a fagos, indicando que a maioria das bactérias é lisogênica (43%) (contendo um profago integrado) ou, variando de 2 a 5 profagos íntegros, polilicogênica (29,4%) (abrangendo mais de um profago integrado) (Figuras 15 e 16). Essa alta frequência de profagos incompletos em genomas bacterianos tem sido descrita para outros gêneros, tanto ao trabalhar exclusivamente com genomas completos quanto com genomas incompletos (BOBAY *et al.*, 2014; TOUCHON *et al.*, 2014; COSTA *et al.*, 2018; CRISPIM *et al.*, 2018; ZHANG *et al.*, 2019). A antecipação de detectar um alto número de profagos degenerados está fundamentada no fato de que profagos íntegros geralmente estão sujeitos a intensas pressões seletivas impostas pela bactéria hospedeira, que pode empregar uma variedade de mecanismos de defesa contra fagos (BOBAY *et al.*, 2014; ZHANG *et al.*, 2019).



**Figura 15: Número de elementos parecidos com vírus (Virus-like elements – VLEs), VLEs íntegros e tamanho do genoma das espécies de *Weissella*.** Nesta figura, o eixo X exibe o tamanho médio do genoma para cada espécie de *Weissella*, enquanto o eixo Y representa a moda de VLEs encontrados em cada espécie. O tamanho de cada ponto de dados em forma de bolha corresponde à contagem média de VLEs íntegros descobertos em cada espécie. As setas e as cores das bolhas indicam os nomes das espécies.

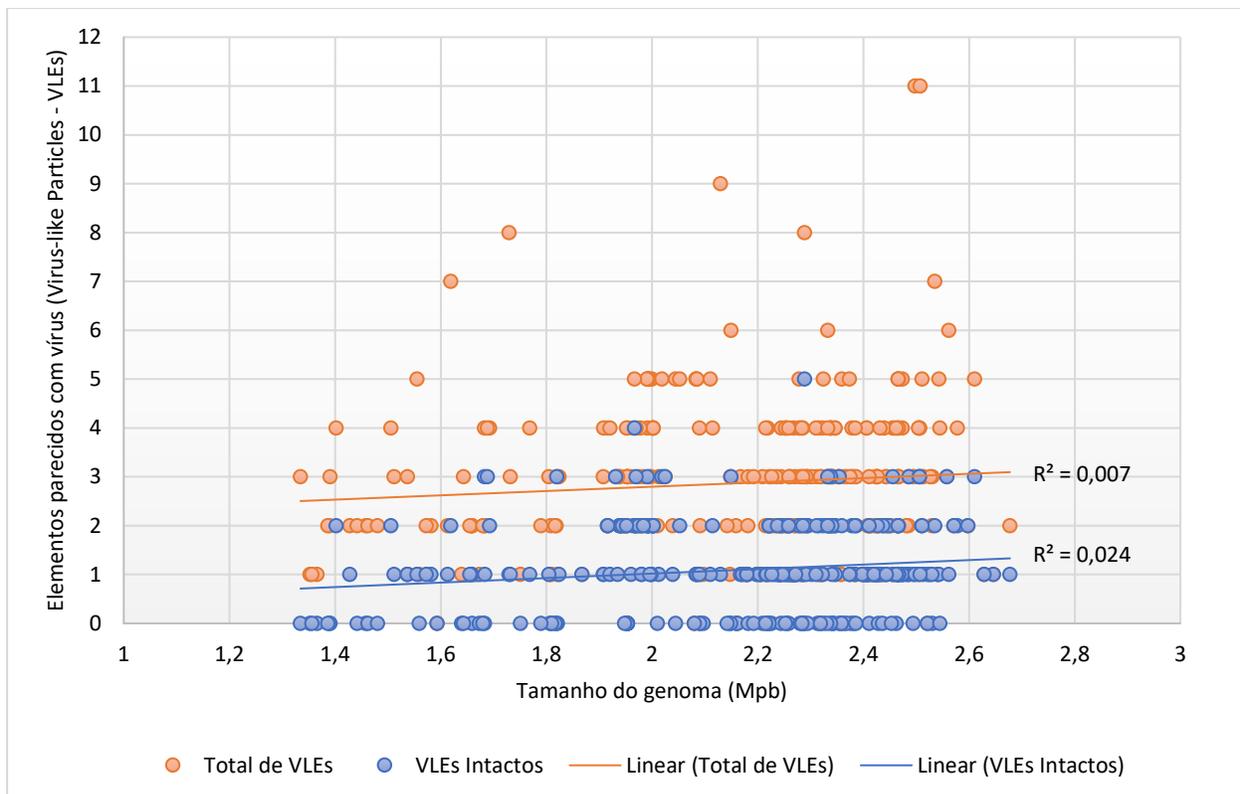


**Figura 16: Distribuição do comprimento e quantidade de elementos parecidos com vírus (*Virus-like elements* – VLEs) dentro do Gênero *Weissella*.** Esta figura representa a distribuição de VLEs categorizados pelo seu tamanho dentro do gênero *Weissella*. O eixo X exibe o tamanho dos VLEs em quilobases (Kb), e o eixo Y mostra o número de VLEs. Os dados são apresentados como barras sobrepostas: barras vermelhas representam VLEs íntegros; barras rosas escuras representam VLEs questionáveis; e barras rosas claras representam VLEs incompletos. O tamanho da amostra para cada categoria de elementos também é fornecido na legenda em parênteses.

No ambiente, as bactérias estão constantemente sujeitas a ataques de fagos, com alguns fagos se integrando ao genoma bacteriano e estabelecendo uma relação irreversível entre o fago e o hospedeiro, o que pode resultar na perda de determinadas funções no fago incorporado (BOBAY *et al.*, 2014; WANG *et al.*, 2010). Apesar desse fator, um número substancial de profagos íntegros (347) foi identificado em genomas incompletos (204), indicando provavelmente que esses profagos foram adquiridos relativamente recentemente, uma vez que seu estado íntegro sugere que as pressões seletivas bacterianas tiveram influência mínima sobre eles e permaneceram inalterados ou não quebrados (BOBAY *et al.*, 2014).

Os profagos classificados como questionáveis e incompletos, totalizando 59,5% (62+274), foram categorizados como profagos degenerados, enquanto os íntegros foram designados como profagos intactos, representando 40,5% (228). O tamanho médio do genoma dos profagos foi de 24,3 kb. Entre os profagos íntegros, o tamanho médio do genoma foi de 35,5 kb; o menor foi encontrado em *W. viridescens* MSK.22.110 e *W. viridescens* MSK.22.125, com um tamanho de 12,9 kb, enquanto o maior foi identificado em *W. ceti* CECT 7719, medindo 83 kb. Entre aqueles considerados incompletos ou questionáveis, o menor foi observado em *W. cibaria* SP7 com 3,5 kb e o maior em *W. paramesenteroides* SJ21, medindo 55 kb.

A hipótese de que genomas maiores podem abrigar mais profagos foi examinada anteriormente em outras bactérias. Bobay *et al.* (2014) estabeleceram uma clara correlação entre o tamanho do genoma e profagos em *Escherichia coli* e *Salmonella enterica*. Costa *et al.* (2018) relataram um aumento leve no número de profagos com a expansão do genoma bacteriano em *Acinetobacter baumannii*. No entanto, Touchon *et al.* (2014) observaram essa correlação principalmente em genomas bacterianos pequenos (1-5 Mb), sem tal correlação evidente em genomas grandes (5-7 Mb). No caso do gênero *Weissella*, uma análise do gráfico de dispersão para os 316 genomas de *Weissella* investigados neste estudo, examinando o tamanho do genoma e o número de elementos semelhantes a fagos, não indica uma correlação entre essas variáveis. Essa observação é ainda apoiada pelo teste de correlação de Pearson, que resultou em um valor de  $R^2 = 0,007$  (Figura 17), indicando a ausência de correlação significativa. Da mesma forma, não há correlação entre o tamanho do genoma e o número de profagos íntegros, com um coeficiente de correlação de Pearson  $R^2 = 0,024$  (Figura 17).



**Figura 17: Gráfico de dispersão da correlação entre o número de elementos semelhantes a fagos (*Virus-like elements* – VLEs) e o tamanho do genoma em genomas de *Weissella*.** Neste gráfico de dispersão, cada ponto de dados corresponde a um genoma de *Weissella*. No eixo X, indicamos o tamanho do genoma em Mpb (milhões de pares de bases), enquanto no eixo Y, representamos a contagem total de elementos semelhantes a fagos identificados em cada genoma. Os pontos de dados laranja representam o total de VLEs, e os pontos de dados azuis representam o total de VLEs intactos. A linha horizontal no gráfico indica o padrão de melhor ajuste nos dados. No final dessa linha, a equação de regressão linear é apresentada, acompanhada pelo valor de  $R^2$ , indicando o coeficiente de determinação para a relação linear.

Dentre todos os elementos semelhantes a fagos, *W. confusa* apresentou o maior número (287), seguido por *W. cibaria* (242) e *W. paramesenteroides* (164). Esses valores eram esperados, uma vez que essas espécies têm uma quantidade significativa de genomas depositados no banco de dados. No entanto, o número médio de elementos semelhantes a fagos dentro das linhagens de cada espécie mostrou considerável variabilidade. *Weissella oryzae* teve o maior valor médio com nove elementos, enquanto alguns membros de *W. cibaria* e *W. confusa* não apresentaram nenhum elemento semelhante a fagos.

É importante destacar que oito elementos semelhantes a fagos foram detectados nos plasmídeos anotados de *Weissella*; alguns genomas estão em forma de rascunho, enquanto outros estão completos. Em casos de genomas em *draft*, torna-se desafiador identificar inequivocamente um *contig* como um elemento semelhante a fagos em vez de um plasmídeo, devido às suas características sobrepostas. Consequentemente, é impossível determinar se os plasmídeos adquiriram esses profagos por meio de recombinação homóloga com o cromossomo bacteriano ou se foram integrados diretamente nos plasmídeos. A incorporação de profagos em plasmídeos pode levar a várias formas de troca de material genético, resultando potencialmente na aquisição de diferentes genes tanto pelo plasmídeo quanto pelo profago (SALMOND *et al.*, 2015).

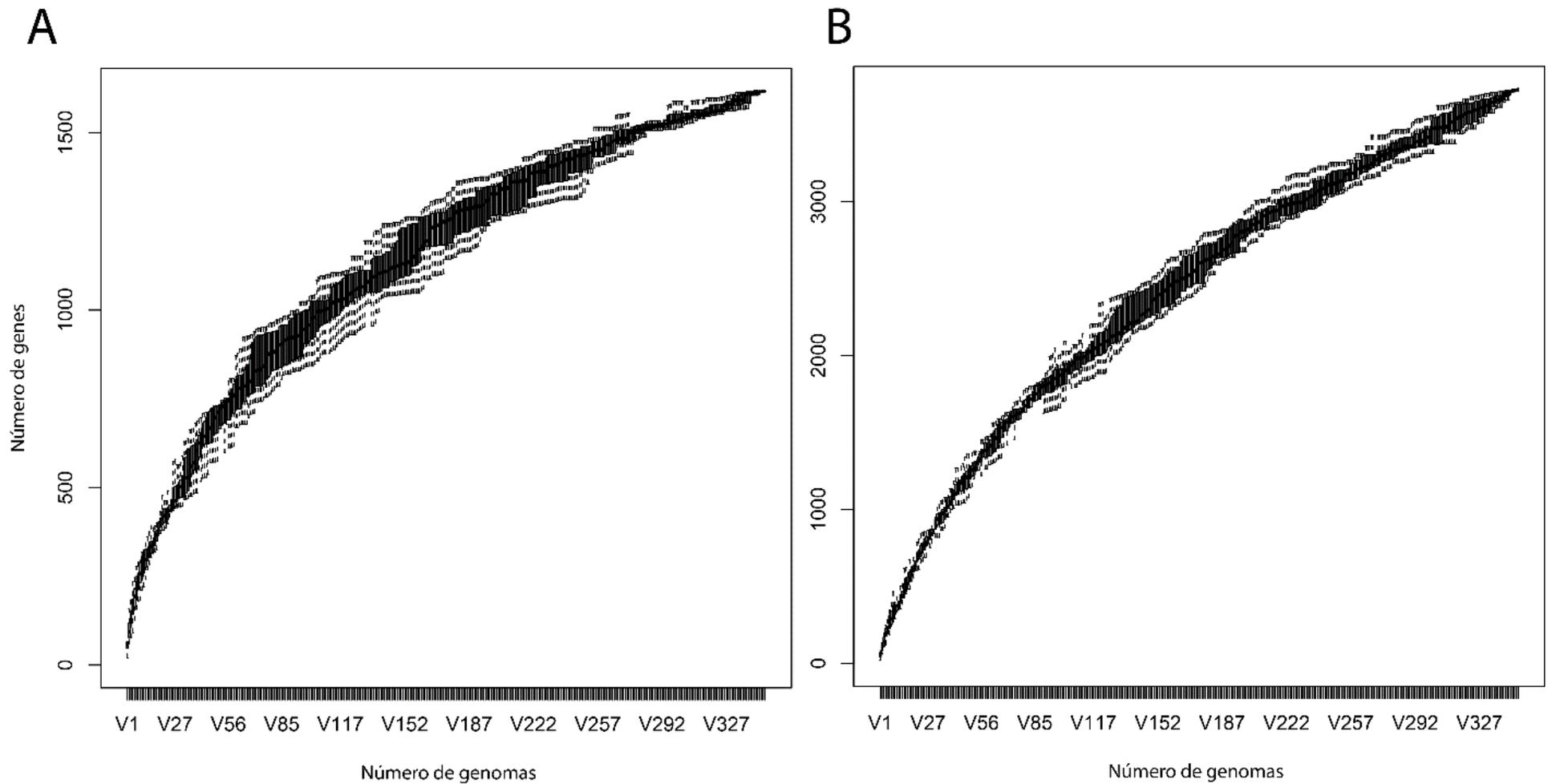
### 5.5 Características dos Profagos de *Weissella*

A presença potencial de fagos conhecidos dentro do gênero *Weissella* foi verificada alinhando todos os elementos semelhantes a fagos com as sequências dos fagos líticos de *Weissella* previamente isolados e caracterizados, bem como as sequências disponíveis no banco de dados GenBank. Isso incluiu o fago phiYS61 de *Weissella cibaria*, com morfologia de podovírus, e os fagos PWc de *Weissella ceti* e WCP30 de *Weissella cibaria*, ambos com morfologia de sifovírus (KLEPPEN *et al.*, 2012; LEE *et al.*, 2017; HOAI *et al.*, 2018). No entanto, não foi possível identificar correspondências com uma identidade superior a 5% ou cobertura superior a 20% entre esses fagos e todos os elementos semelhantes a fagos identificados.

Todas as sequências de profagos intactos foram alinhadas, revelando um baixo grau de identidade de sequência entre elas. Uma matriz foi gerada para facilitar a identificação de grupos de genomas semelhantes. Essa abordagem analítica foi escolhida considerando a substancial diversidade interespecífica dentro de fagos e a possibilidade de algumas sequências serem truncadas devido a mecanismos de defesa bacteriana. Posteriormente, utilizamos sequências de genes estruturais virais para prever, *in silico*, a morfologia viral. A maioria dos profagos apresentou morfologia de sifovírus (254; 73,2%), enquanto 85 (24,5%) foram classificados com morfologia de miovírus. Profagos de miovírus têm tamanhos de genoma variando de 23,4 kb a 83 kb, codificando de 25 a 96 proteínas. Os profagos de sifovírus têm tamanhos de genoma variando de 12,9 a 57,5 kb, codificando de 16 a 71 proteínas. Notavelmente, 8 (2,3%) profagos permaneceram morfologicamente não classificados, com tamanho de genoma variando de 14,4

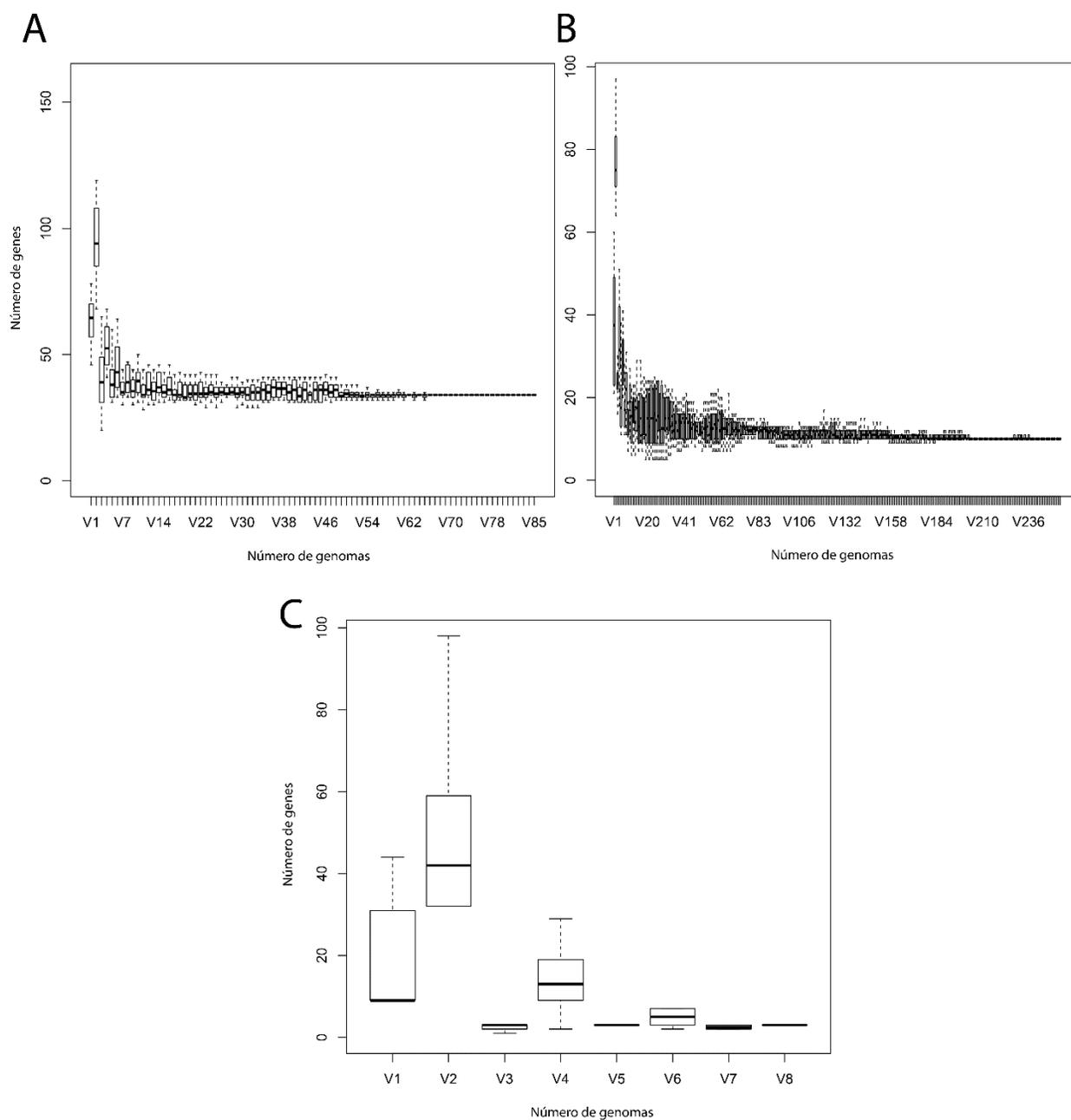
kb a 41,3 kb, codificando de 22 a 35 proteínas. Essa falta de classificação pode ser atribuída à ausência de genes estruturais específicos ou à completa ausência desses genes dentro desses genomas. Esses profagos não classificados foram distinguíveis por um alto grau de variabilidade e foram caracterizados por um excedente de genes de integrases e transposases. Uma possibilidade intrigante é que esses profagos não classificados exibem características semelhantes aos “*pirate phages*”, que são conhecidos por não ter genes estruturais, mas utilizam as estruturas de outros fagos (*helper*) durante a montagem nos ciclos líticos. Esse processo efetivamente permite que eles se apropriem da estrutura do vírion para seus propósitos (MITARAI, 2020). Eles também podem pertencer a bacteriófagos de dsDNA sem cauda (não membros de *Caudoviricetes*) ou bacteriófagos de ssDNA, já que este perfil de HMM não possui alguns genes estruturais de outras classes.

Em relação à diversidade genética dentro desses profagos, realizamos uma análise detalhada usando a ferramenta ROARY, seguindo a metodologia descrita por Yahara *et al.* (2019). Empregamos várias combinações de parâmetros relacionados à identidade de sequência para considerar a considerável diversidade observada nos fagos. A análise do pan-genoma revelou que os profagos de *Weissella* abrangiam 1617 genes únicos (com um limiar de BLASTp de >40%) ou 3728 genes únicos (com um limiar de BLASTp de >90%) (Figura 18). Por outro lado, na análise do core-genoma, não identificamos nenhum gene central, e sabíamos com a anotação PHASTER que quase todos esses profagos tinham genes semelhantes, como integrase/replicase. Levando em consideração a substancial diversidade genômica observada comumente em profagos, conforme apoiado por estudos anteriores sobre pan-genomas de profagos, e o considerável número de sequências de profagos identificadas em nossa análise, utilizamos o limiar de 40% para BLASTp para compilar o pan-genoma de profagos (YAHARA, 2019; MARQUES, 2021).



**Figura 18: Pan-genoma dos profagos no gênero *Weissella*.** A: Pan-genoma dos profagos analisados no gênero *Weissella* com BLASTp>40%. B: Pan-genoma dos profagos analisados no gênero *Weissella* com BLASTp>90%. Esta figura ilustra o pan-genoma dos fagos *Weissella*. O eixo X representa o número de genomas de profagos, e o eixo Y representa o número de genes únicos. O gráfico é representado pela média e desvio padrão de cada gene único em todos os genomas de profagos analisados. Diferentes valores de BLASTp foram atribuídos a cada gráfico, >40% e >90%.

O termo parálogo, nesse contexto, refere-se a genes com um códon de parada e um códon de iniciação ocorrendo no meio do gene, em vez de serem caracterizados como duplicações verdadeiras (YAHARA *et al.*, 2019). É bem estabelecido que os genomas de fagos e profagos exibem compacidade, frequentemente contendo ORFs sobrepostas (LANG *et al.*, 2017; DION *et al.*, 2020). Esse resultado era esperado devido à alta diversidade observada entre esses genomas. Para aprofundar essa investigação, procedemos à categorização da morfologia do profago em miovírus, sifovírus e aqueles sem classificação. Posteriormente, uma análise dos *core genes*, definidos como aqueles compartilhados por pelo menos 50% das sequências, revelou 34 *core genes* no profago de miovírus, dez *core genes* no sifovírus e três *core genes* no grupo sem inferência de morfologia (Figura 19). Como descrito anteriormente, os miovírus eram os maiores, enquanto os profagos sem classificação morfológica eram os menores em tamanho de genoma e número de proteínas. Essa observação está alinhada com a análise de *core genes*, onde os miovírus possuem mais genes potencialmente relacionados. Em contraste, os profagos sem classificação morfológica exibem menos genes e tamanhos de proteínas e genômicos menores do que as outras duas morfologias.



**Figura 19: Core-genoma das sequências de profagos do gênero *Weissella* com a mesma morfologia predita. A:** Core-genoma de sequências com morfologia predita de miovírus. B: Core-genoma de sequências com morfologia predita de sifovírus. C: Core-genoma de sequências com morfologia predita não classificada. O eixo X representa o número de sequências analisadas, enquanto o eixo Y representa o número de *core*-genes classificados, com média e desvio padrão. Em A (miovírus), observamos 34 *core*-genes; para B (sifovírus), dez *core*-genes; e em C (não classificado), apenas três *core*-genes foram identificados.

Os genes centrais identificados dentro desses grupos de profagos se referem predominantemente a funções virais. No caso dos miovírus, os genes centrais abrangem não apenas componentes estruturais, como esperado, mas também genes associados à lisogenia, lise celular e replicação de DNA. Pelo contrário, com apenas dez genes centrais, os sifovírus exibem mais genes centrais associados a funções estruturais. Esse achado pode ser atribuído às muitas sequências dentro dos sifovírus, o que poderia representar desafios ao determinar os genes centrais dentro de um grupo tão diversificado. Quanto aos profagos não classificados, os genes centrais envolvem recombinação/replicação e empacotamento de DNA. Essas observações estão em consonância com os resultados de análises de genoma central de profagos em outras espécies bacterianas, conforme demonstrado em estudos anteriores (YAHARA *et al.*, 2019).

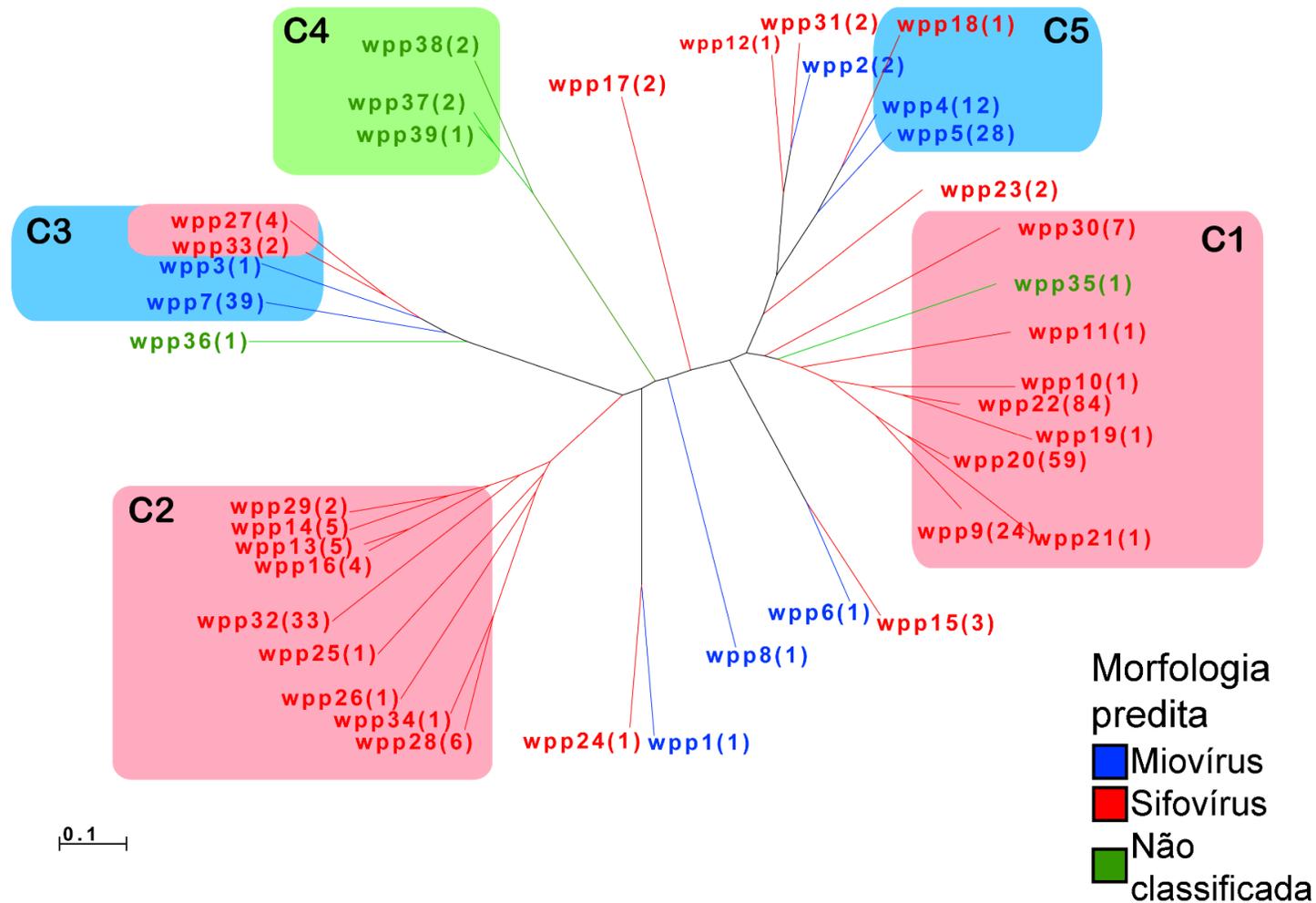
Dada a ampla variedade de genomas bacterianos, diferentes tecnologias de sequenciamento e abordagens algorítmicas variadas para montagem, buscamos elucidar o estilo de vida dos profagos. Consideramos que alguns fagos/profagos (em um ciclo lítico) podem estar incluídos na população bacteriana, erroneamente atribuídos dentro do genoma ou tratados como *contigs* não mapeados no cromossomo. Utilizamos o PHACTS para prever o estilo de vida do profago para abordar essa preocupação. Os resultados indicaram que, com uma probabilidade superior a 50%, 299 profagos intactos foram classificados como temperados, enquanto 48 foram categorizados como líticos. Apesar do número relativamente alto de profagos líticos, essa observação pode ser atribuída a alguns genomas que exibem um alto desvio padrão em suas previsões. Além disso, genomas específicos podem estar truncados, faltando alguns genes normalmente associados a fagos temperados, como integrases.

### **5.6 Montagem de fagos em grupos e filogenética**

Fagos que infectam hospedeiros diferentes geralmente mostram pouca ou nenhuma similaridade de sequência, e até mesmo aqueles que infectam um único hospedeiro podem exibir variações substanciais de sequência (MAVRICH & HATFULL, 2017). Ao infectar um único hospedeiro, eles tendem a gerar mosaicos de regiões semelhantes. No entanto, nem toda essa diversidade de mosaicos pode ser demonstrada e pode existir como um espectro de diversidade genômica (POPE *et al.*, 2018). Profagos foram classificados como profagos semelhantes quando mostraram mais de 50% de identidade e mais de 60% de cobertura (COSTA *et al.*, 2018). A sequência de referência

foi selecionada com base no maior tamanho e na maior cobertura em cada grupo de profagos intactos, atendendo a esses critérios. Um novo nome foi atribuído para abranger todos os profagos intactos dentro desse grupo de fagos semelhantes (Tabela Suplementar). Essa abordagem foi implementada para acomodar a diversidade interespecífica dos fagos, a possibilidade de sequências mais curtas devido a mecanismos de defesa bacteriana e a diversidade de genes acessórios.

Consequentemente, identificamos 39 grupos de sequências exclusivas que atenderam a esses critérios de similaridade, compreendendo 15 profagos únicos para montagens únicas. Eles receberam nomes de “wpp” (*Weissella* ProPhage) e foram numerados aleatoriamente de 1 a 39, seguindo a morfologia. Entre esses, quatro foram categorizados como tendo um estilo de vida lítico. Uma árvore filogenética foi construída para obter insights mais profundos sobre as relações entre esses grupos de profagos (Figura 20).



**Figura 20: Árvore filogenética de genomas completos de grupos similares de fagos encontrados em *Weissella*.** Com base na sequência de referência, uma árvore filogenética de genomas completos foi inferida usando Máxima Verossimilhança (*Maximum Likelihood* - ML) com 100 replicatas de *bootstrap*. Os nomes representam grupos de fagos semelhantes (identidade e cobertura >50%), com o número de membros de cada grupo indicado entre parênteses. A morfologia predita é indicada pela cor do nome: azul para miovírus, vermelho para sifovírus e verde para morfologia não classificada. *Clusters* de morfologias predita predominantes (mais de 4 membros) foram separados em C1-5.

Nesta árvore, os grupos de profagos são nomeados com base no número de profagos que compõem, e a cor do nome representa sua morfologia. Essa abordagem permitiu a criação de clusters de grupos com morfologias semelhantes. Os resultados sugerem que profagos classificados como semelhantes compartilham conservação em genes relacionados à estrutura do capsídeo viral, conforme evidenciado pelos Clusters C1-5. No entanto, é importante observar que, em alguns casos, esses clusters são polifiléticos, mais representados no Cluster 3 (C3), mas também nos Clusters 1 (C1) e 5 (C5), com um membro do grupo tendo apenas um profago. Notavelmente, sifovírus compreendem a maioria dessas sequências selecionadas (Figura 20).

No entanto, há uma diversidade significativa entre os profagos intactos, com alguns membros do grupo de miovírus sendo mais intimamente relacionados aos sifovírus do que outros miovírus, como mostra o Cluster 3. Além disso, alguns profagos anteriormente sem classificação de estrutura, como wpp35 (mostra-se próximo a membros sifovírus) no Cluster 1 (C1) e wpp36 (mostra-se próximo a membros de miovírus e sifovírus) no Cluster 3 (C3), mostraram compartilhar algumas semelhanças com outros grupos. Da mesma forma, wpp37, wpp38 e wpp39, todos no Cluster 4 (C4), também compartilham semelhanças, provavelmente devido ao fato, já bem estabelecido na literatura, de que várias avaliações independentes indicaram que essas classificações com base em morfologia são polifiléticas e não representam adequadamente linhagens evolutivas compartilhadas. Isso destaca por que inferências taxonômicas contemporâneas muitas vezes se baseiam em várias fontes de dados além das características morfológicas (AIEWSAKUN *et al.*, 2018; BARYLSKI *et al.*, 2020; TURNER *et al.*, 2023). Classificar esses fagos em famílias ou classes depende de mais do que apenas morfologia e composição genômica devido à falta de um marcador genético conservado e à vasta diversidade genômica de fagos. No entanto, nossos resultados sugerem que esses membros provavelmente pertencem à classe *Caudoviricetes*, já que compartilham alguns genes com membros da classe. É concebível que os fagos encontrados no Cluster 4 (C4) sejam mais propensos a pertencer a uma família diferente ou até a outra classe de vírus de DNA de fita dupla. A ausência de genes estruturais nesses fagos impede até mesmo sua classificação como *Caudoviricetes*, pois eles carecem de proteínas essenciais da cauda viral. No entanto, vale ressaltar que o banco de dados usado pelo PHASTER para identificar profagos e classificá-los como intactos é composto principalmente por

genes de *Caudoviricetes*, pois esses são os fagos mais extensivamente estudados e sequenciados. A classificação do PHASTER de profagos leva mais do que apenas genes estruturais como elementos para uma sequência de profago intacta, levando em conta também genes de replicação, lise, integração e regiões de inserção (ARNDT *et al.*, 2016).

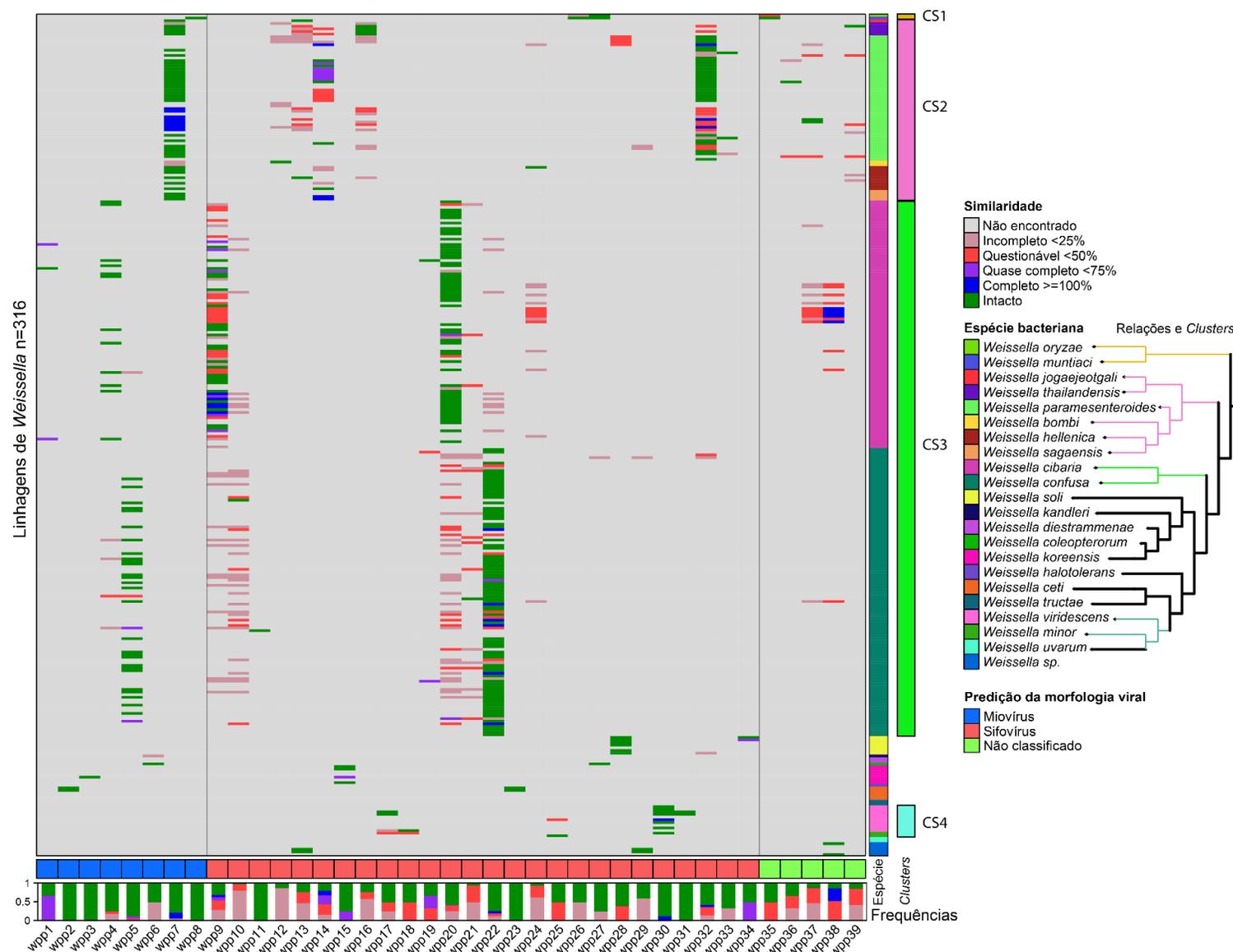
Para mitigar um desafio significativo associado à utilização de montagens de genomas, onde profagos intactos podem estar fragmentados em vários *contigs*, levando o PHASTER a classificá-los como múltiplos profagos incompletos ou questionáveis, realizamos alinhamentos rigorosos com BLASTn dessas sequências fragmentadas em relação às 39 sequências únicas. Aplicamos um critério de classificação designando sequências como componentes desses profagos semelhantes se exibissem uma similaridade superior a 50%, e atribuímos um valor percentual de cobertura relativo à sequência única correspondente.

Esta análise revelou que algumas sequências de profagos incompletos e questionáveis representavam fragmentos de uma sequência única. Essa fragmentação pode ser atribuída à sua disposição em *contigs* distintos, à exposição a pressões evolutivas seletivas ou a mecanismos de defesa bacterianos, enquanto retêm genes virais e acessórios. Como resultado, essa abordagem refinada identificou 562 sequências (compostas por 128 profagos incompletos e 434 questionáveis), com 350 sendo categoricamente designadas como constituintes dessas sequências análogas de profagos. As 212 sequências restantes não foram submetidas a investigação adicional. No entanto, é digno de nota que profagos fragmentados podem não apresentar similaridade com aqueles identificados anteriormente neste estudo entre essas sequências não examinadas. Algumas dessas sequências podem, de fato, representar profagos completos. Consequentemente, é essencial reconhecer que essas sequências podem abrigar genes acessórios com implicações para a aptidão bacteriana.

A presença de grupos de profagos semelhantes, tanto intactos quanto fragmentados, dentro de cada espécie, é ilustrada nas Figuras 21 e 22. Como observado, grupos específicos de profagos demonstram similaridade e são compartilhados entre diferentes espécies. Podemos usar as relações filogenéticas entre essas espécies para identificar agrupamentos de espécies com profagos semelhantes designados como CS1, CS2, CS3 e CS4. Espécies que não se encaixam nesses agrupamentos possuem profagos distintos. Embora apenas três genomas em nossa

análise possam representar CS1, é digno de nota que ambas as espécies dentro desse agrupamento ainda compartilham os mesmos profagos intactos conservados, como wpp26, wpp27 e wpp35 (Figura 21).





**Figura 22: Mapa de calor ilustrando a presença de grupos similares de profagos em cada espécie e linhagem de *Weissella*.** O eixo X representa os grupos de profagos similares, enquanto o eixo Y representa a presença de seqüências dentro dos 316 genomas analisados. Cada linha corresponde a um genoma de cepa de *Weissella*, com a posição da linha à direita indicando a espécie e aglomerados de profagos similares. Na parte superior, o mapa de calor exibe a frequência de cada grupo de fagos nas amostras analisadas. A legenda fornece informações sobre a codificação de cores usada para cada linha: branco para 'não encontrado', diferentes tons de cinza para 'fragmentos', preto para 'seqüências completas' e vermelho para 'seqüências intactas'. Além disso, a legenda inclui as espécies bacterianas e suas relações determinadas a partir de uma árvore construída usando as seqüências de *rDNA 16S* das espécies, gerada usando Máxima Verossimilhança (*Maximum likelihood* - ML) com 100 replicatas de *bootstrap*, com a cor dos ramos indicando aglomerados de profagos similares dentro das espécies. Por fim, as cores atribuídas aos grupos de fagos no eixo X representam a morfologia predita de cada grupo.

Embora *W. paramesenteroides* seja mais abundantemente representado em sequências genômicas do que outros membros de CS2, eles ainda compartilham quase todos os profagos semelhantes, com alguns mostrando conservação excepcionalmente alta, como wpp7. No entanto, alguns profagos são intactos em algumas espécies e fragmentados em outras, como wpp12, wpp13, wpp14, wpp16, wpp32 e wpp39, e estão dispersos ao longo das espécies dentro deste agrupamento. Curiosamente, em *W. paramesenteroides*, aonde há um sequencias intactas do grupo wpp33, há apenas sequencias parciais de wpp32. Essa discrepância pode ser atribuída a um problema no processo de montagem do genoma, ou pode sugerir a presença de uma nova variante que divergiu em dois grupos distintos. Investigações adicionais são necessárias para elucidar essa observação e determinar se é um artefato de anotação ou uma variação genética genuína dentro de *W. paramesenteroides* (Figura 22).

Do ponto de vista filogenético, *W. cibaria* e *W. confusa* parecem estar intimamente relacionados e compartilham profagos semelhantes em grupos. No entanto, esses profagos estão fragmentados, e as duas espécies não compartilham profagos intactos, apenas fragmentos. Uma hipótese para esse fenômeno é que mutações nessas sequências podem ter causado a não conformidade com nossos critérios de agrupamento. Essas sequências podem ter se alterado e criado variantes ao longo do tempo por meio de mosaicismo, duplicação e mutação. Por exemplo, wpp9 e wpp20 estão apenas fragmentados em *W. confusa*, e wpp22 está apenas em *W. cibaria*.

Embora CS4 tenha menos sequências, o que implica menor cobertura e diversidade de sequências, eles compartilham similaridades em profagos intactos e fragmentos. Isso sugere que elementos conservados dentro deste grupo de profagos podem estar presentes em vários genomas, apesar da menor cobertura e diversidade geral de sequências. Embora CS4 pareça ser um grupo parafilético, com *W. uvarum* estando intimamente relacionado fora do grupo, isso provavelmente se deve à menor representação de genomas de *W. uvarum*. Com a adição de novos depósitos de sequências, análises adicionais poderiam determinar a extensão e a significância desses elementos compartilhados dentro dos profagos de CS4, e esse conceito pode ser aplicado a CS1 e a todas as outras espécies.

### 5.7 A contribuição da genética de profagos para a resistência bacteriana e virulência em *Weissella*

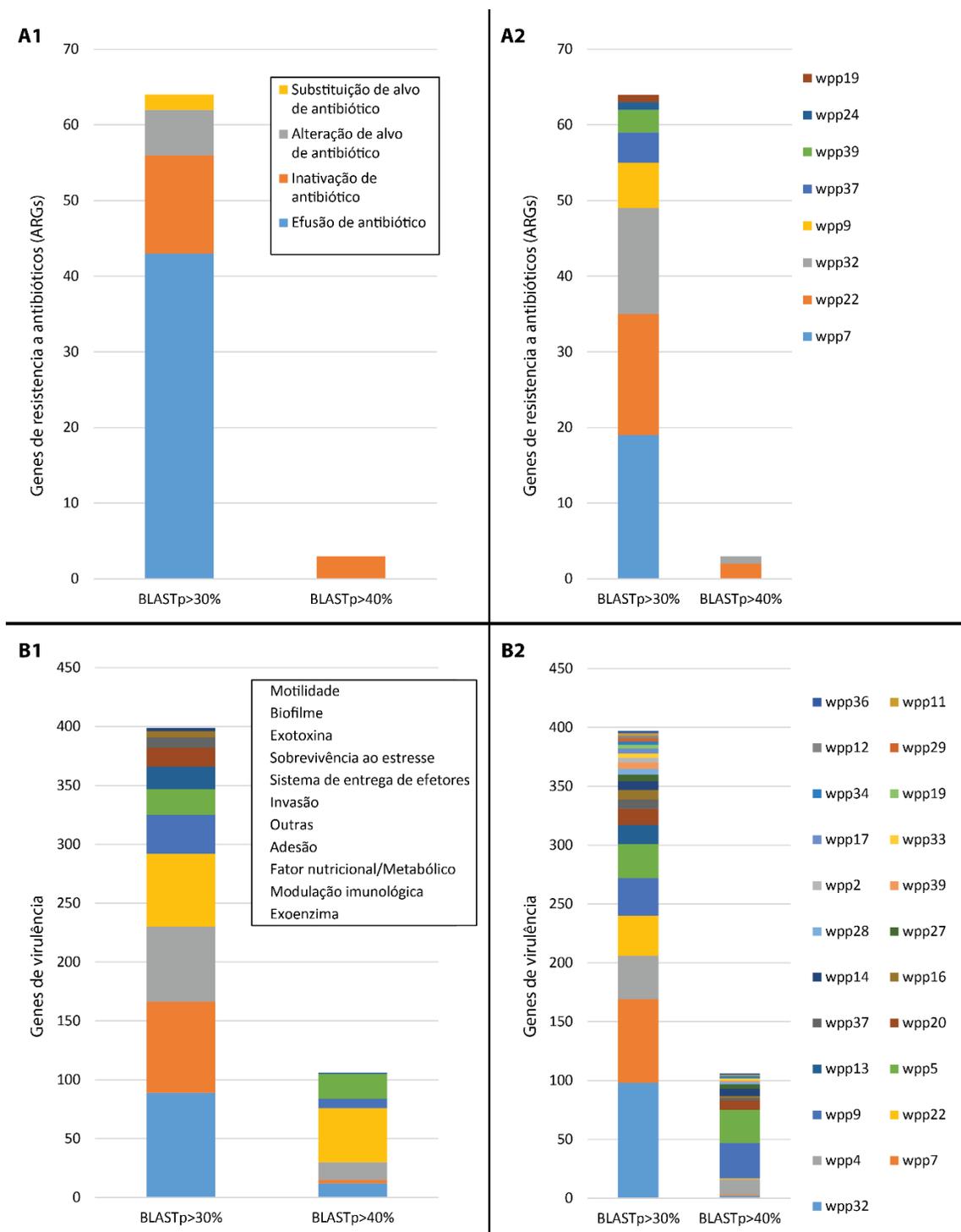
Profagos podem codificar vários genes que beneficiam a bactéria hospedeira, incluindo genes para resistência a antibióticos, funções metabólicas e fatores de virulência. Para compreender melhor os efeitos benéficos que esses profagos podem ter na aptidão bacteriana, procuramos identificar genes de resistência e virulência em seus genes auxiliares nos profagos do gênero *Weissella*.

Como é bem sabido, os genomas de fagos frequentemente exibem sobreposição de genes, genomas compactos e carregam genes bacterianos quando entram na fase de indução (RANCUREL *et al.*, 2009; TREANGEN *et al.*, 2011; SABATH *et al.*, 2012), o que pode dificultar a identificação das regiões de sequência codificante (CDS) dos profagos. O Prodigal é a ferramenta mais comumente usada para essa tarefa em bactérias, mas quando aplicado a essa condição apresentada por fagos, nem sempre é preciso. Estudos recentes mostraram que usar múltiplos algoritmos de detecção de CDS é uma abordagem eficaz para a anotação de profagos (POPE *et al.*, 2018; SALISBURY *et al.*, 2019; LAZEROFF *et al.*, 2020).

Com isso em mente, identificamos regiões de CDS usando o programa PhageCommander e usamos o consenso e a melhor cobertura de nove algoritmos diferentes de detecção de CDS para determinar as regiões de CDS de todos os 347 profagos intactos encontrados nos genomas de *Weissella*. A análise desses CDS com os bancos de dados CARD e VFDB revelou que esses genes estão comumente situados no início ou no final da sequência do profago. Essa observação está de acordo com as expectativas, pois profagos frequentemente abrigam genes bacterianos próximos ao local de ligação (*attachment* - att), que podem ser incorporados durante a excisão. Como diretriz geral, a avaliação de genes de resistência e virulência como confiáveis estava condicionada à identidade de sequência e à cobertura superando o limite de 80%, em conformidade com os critérios estabelecidos para a identificação de fatores de resistência e virulência (CHEN *et al.*, 2015; PEARSON *et al.*, 2016; WANG *et al.*, 2020; QIAO *et al.*, 2020; ZHOU *et al.*, 2021). A ausência de profagos que possuem tais CDS que atendam a esses critérios definidos implica um risco limitado, potencialmente ausente, de transferência horizontal de genes, pois esses casos envolvem principalmente fragmentos de genes. No entanto, permanece

plausível empregar critérios alternativos para explorar a presença potencial de fragmentos de genes dentro desses profagos, mesmo que não sejam identificáveis com confiança como confiáveis ou funcionalmente ativos.

Foram identificados 64 genes associados a mecanismos de resistência e categorizados com base em seus respectivos atributos de função. Em particular, 13 desses genes estavam relacionados à inativação de antibióticos, 43 genes com efluxo de antibióticos, dois com substituição do alvo do antibiótico e seis com alteração do alvo do antibiótico (Figura 23 - A1). Uma proporção substancial desses genes de resistência foi associada a profagos únicos específicos, com o wpp7, representando 30% dos hits, wpp22 com 25% dos hits, wpp32 com 22% dos hits, e wpp9, wpp37, wpp39, wpp19 e wpp24 contribuindo coletivamente com 23% dos hits restantes (Figura 23 - A2).



**Figura 23: Quantidade de genes e grupos de profagos associados à resistência e virulência presente nos fagos de *Weissella* analisados. A1: Número de genes de resistência a antibióticos (ARGs) presentes nos fagos de *Weissella* analisados. A figura exhibe barras sobrepostas representando a quantidade de genes de resistência a antibióticos (ARGs) com ações indicadas por cor na legenda para identidades superiores a 30% e 40%. A2: Número de genes de virulência presentes em cada grupo semelhante de fagos. Esta figura ilustra a quantidade de genes de virulência em cada grupo de fagos semelhantes para identidades superiores a 30% e 40%. A legenda indica os grupos de fagos por cor. B1: Número de genes de virulência presentes nos fagos de *Weissella* analisados. A figura apresenta barras sobrepostas representando o número de genes de virulência, com suas funções listadas na legenda por cor para identidades superiores a 30% e 40%. B2: Número de genes de virulência presentes em cada grupo semelhante de fagos. Esta figura mostra a quantidade de genes de virulência em cada grupo de fagos semelhantes para identidades superiores a 30% e 40%. A legenda indica os grupos de fagos por cor.**

Esse resultado destaca a presença de mecanismos de resistência dentro da diversidade dos profagos que eles representam. No entanto, é essencial reconhecer que essa associação existe dentro de um subconjunto relativamente limitado, pois apenas 20% (8) dessas sequências únicas de profagos exibiram tais mecanismos. É importante observar que os critérios para a detecção de homologia empregados nesta análise foram relativamente rigorosos, exigindo um limite de identidade superior a 30% e um limite de *e-value* inferior a  $1 \times 10^{-5}$ , com um comprimento mínimo de hit superior a 50 aminoácidos. Notavelmente, a maioria desses genes caiu na "twilight zone" dos alinhamentos de sequências de proteínas, com identidades de BLASTp variando de 30% a 40% (ROST, 1999; MAVRICH & HATFULL, 2017). Notavelmente, apenas três genes apresentaram identidades superiores a 40%, especificamente *vatH* no profago de *W. paramesenteroides* STCH-BD1, *catB8* no profago de *W. confusa* TM115-1155 e *catB3* no profago de *W. confusa* 1477.

*vatH* é uma acetiltransferase, tipicamente mediada por plasmídeos, e está envolvida na inativação de antibióticos da classe estreptogramina em *Enterococcus faecium* (JUNG *et al.*, 2010). Acetiltransferases também foram identificadas em profagos como agentes resistentes a antibióticos em *Acinetobacter baumannii* (COSTA *et al.*, 2018). Os *catB* são cloranfenicol acetiltransferases (CATs); os CATs do tipo B também foram chamados de acetiltransferases de xenobióticos (XATs) na literatura; conferem baixa resistência ao cloranfenicol e não são homólogos estruturais ou de sequência dos CATs do tipo A, mas são estruturalmente relacionados ao *vatA* e *vatD* (JUNG *et al.*, 2010; ALCALA, 2020). No entanto, é crucial destacar que, apesar dessas identidades mais altas, todos os hits exibiram valores de *e-value* variando de  $1 \times 10^{-5}$  a  $1 \times 10^{-10}$ , juntamente com comprimentos inferiores a 100 aminoácidos, mais curtos que os outros genes identificados.

Relacionado a todos os genes encontrados, incluindo hits na *twilight zone*, o gene mais prevalente encontrado foi o *macB* (55%), presente em profagos de *W. cibaria*, *W. confusa*, *W. paramesenteroides* e *W. thailandensis*. Esse gene faz parte de um transportador de ATP (ABC) e desempenha um papel fundamental na captação de nutrientes, exportação de toxinas, regulação de canais de potássio e contribui para a resistência contra medicamentos macrolídeos por meio de uma bomba de efluxo. Allen *et al.* (2011) mostraram a predominância de genes de resistência a bombas de efluxo, incluindo transportadores ABC e transportadores RND (Resistência-

Nodulação-Divisão) nos viromas fecais de porcos após a indução de fagos por antibióticos. Subirats *et al.* (2016), por meio de estudos metagenômicos de esgoto hospitalar, destacaram o papel dos fagos como reservatórios de genes de resistência, demonstrando que a abundância de leituras contendo genes de transportadores ABC era maior em leituras de fagos do que em leituras bacterianas. Wang *et al.* (2021) mostraram uma presença aumentada do gene *macB* em todos os grupos de metagenômica do fagoma de esgoto e muitos outros ARGs com uma baixa representação (*vatH* (16%), *vanXF* (9%), *qacH* (6%), *golS* (5%), *mecl* (3%) e *catB* (2%)).

Uma análise abrangente de genes de virulência resultou em 389 genes distintos sistematicamente categorizados com base em seus atributos funcionais. Esta categorização revelou 89 genes associados a exoenzimas, nove genes associados a exotoxinas, 62 genes relacionados à aderência, 64 genes ligados a fatores nutricionais/metabólicos, 16 genes envolvidos em mecanismos de sobrevivência ao estresse, 78 genes associados à modulação imunológica, três genes associados à mobilidade, 22 genes relacionados à invasão, 19 genes envolvidos em sistemas de entrega de efetores, cinco genes associados à formação de biofilme e 33 genes relacionados a transportadores ABC de reciclagem, T6SS (Sistema de Secreção Tipo VI) e peptidase de sinal (Figura 23 - B1). Entre essas 39 sequências únicas, descobrimos que 59% exibiam algum tipo de gene de virulência (Figura 23 - B2). Este resultado está de acordo com as expectativas, pois genes específicos relacionados a funções virais (como componentes estruturais, lise, replicação, integrase e transposases) apresentam homologia com alguns genes de virulência em bactérias, como o T6SS, que tem homologia com a cauda de fagos do tipo miovírus (CHEN *et al.*, 2015).

A maioria desses genes de virulência (287) estava situada na *twilight zone* de similaridade de sequência, enquanto 102 genes exibiam identidades superiores a 40%. Após uma análise mais aprofundada desses 106 genes, eles foram classificados em várias categorias funcionais, incluindo fatores nutricionais/metabólicos (15 genes), aderência (46 genes), exoenzimas (12 genes), invasão (21 genes), modulação imunológica (3 genes), sistemas de entrega de efetores (1 gene) e outros (8 genes). Notavelmente, nenhum gene associado a exotoxinas, sobrevivência ao estresse, mobilidade ou formação de biofilme foi encontrado fora da *twilight zone*. Além disso, genes de invasão e aderência foram proporcionalmente menos afetados por esse limite de

similaridade de sequência, mostrando que esses genes são mais propensos a homologia com função preservada. Como mencionado anteriormente, genes provavelmente relacionados à função de infecção viral. No entanto, não está descartado que esses genes possam adquirir novas funções, como demonstrado com o T6SS (CHEN *et al.*, 2015). Em relação aos genes de virulência, embora a maioria seja categorizada sob funções de invasão/exoenzima/aderência, com identidade abaixo de 70%, alguns também estão associados a genes estruturais virais, o que está alinhado com a compreensão de que fagos/profagos são parasitas e que alguns genes de virulência bacteriana evoluíram a partir de elementos estruturais virais (CHEN *et al.*, 2015).

É importante cautela ao atribuir funções aos genes de resistência e virulência identificados. Embora muitos genes de resistência tenham sido identificados, suas funções específicas, se houver, permanecem altamente questionáveis. Em relação aos genes de virulência, é essencial considerar que eles estão associados a elementos genéticos móveis, frequentemente carregando genes de virulência necessários para sua mobilidade.

A presença de genes dentro da *twilight zone* de similaridade de sequência, que normalmente não são considerados homólogos e podem não funcionar de maneira semelhante aos seus homólogos, oferecendo uma percepção maior sobre a evolução de fagos (ROST, 1999; MAVRICH & HATFULL, 2017). Fagos e profagos estão sujeitos a modificações genéticas contínuas, e a presença de genes dentro desse limite de similaridade de sequência sugere que esses fagos podem ter carregado genes de resistência e virulência em seu repertório de genes acessórios, conferindo papéis significativos na aptidão tanto do fago quanto da bactéria que abriga esses profagos. No entanto, esses genes parecem ter sofrido modificações genéticas, especialmente em profagos sujeitos a mecanismos de defesa bacteriana (SRIKANT, 2022). Além disso, é digno de nota que alguns desses genes estão presentes em sequências únicas; no entanto, nem todos os profagos representados por essas sequências únicas necessariamente possuem esses genes como parte de seu conteúdo genético acessório.

## 6. CONCLUSÃO

Neste estudo, foram realizados o sequenciamento, anotação gênica e análise funcional dos genomas de três novas linhagens de *Weissella* com potencial probiótico – *Weissella cibaria* 1BM, *Weissella confusa* GIR46L4 e *Weissella paramesenteroides* V3B-11 – evidenciando a presença de uma notável quantidade de genes associados a sistemas de restrição e modificação de sequências, tais como os sistemas RM/DRUANTIA/DISARM. Estes sistemas desempenham um papel crucial na defesa contra fagos, sejam eles temperados ou líticos, e desencadeiam processos de coevolução entre as bactérias e os fagos, propiciando a integração de profagos e genes virais em seus genomas, bem como a modificação dos profagos previamente incorporados.

A presença de genes acessórios nos profagos emerge como um suporte fundamental para a sobrevivência destas bactérias em ambientes adversos, possibilitando uma eficaz competição com outras espécies bacterianas. Em relação ao sistema CRISPR-Cas, o gênero *Weissella* não possui um sistema CRISPR-Cas altamente desenvolvido, caracterizado por um número limitado de espaçadores e poucas proteínas Cas. Uma exceção é a espécie *Weissella soli*, que apresenta dois subtipos distintos desse sistema, ainda que com um número limitado de espaçadores.

Ademais, certos profagos específicos estão preservados em diferentes espécies, evidenciando a riqueza e complexidade da diversidade genética presente. É importante ressaltar que o mecanismo de ativação dos profagos permanece sujeito a incerteza, demandando, portanto, investigações laboratoriais adicionais para uma compreensão completa. Quanto à resistência a antibióticos, não foram identificados genes com alta identidade que representem um significativo risco de transferência genética horizontal. Por outro lado, no contexto dos genes de virulência, embora sejam identificados em grande quantidade, a maioria compartilha semelhança com genes de natureza estrutural de vírus, não sendo identificado nenhum gene de virulência de relevância com alta identidade e cobertura. Ressalta-se, ainda, a identificação de proteínas virais de considerável importância, como as endolisinas, que detêm um promissor potencial em aplicações biomédicas.

Este estudo enriquece nosso conhecimento acerca dos genes presentes em profagos de *Weissella*, destacando, sobretudo, o significativo papel desempenhado por essas entidades na

evolução bacteriana do gênero. Observa-se uma expressiva quantidade de profagos íntegros, sendo que a detecção de profagos em estado degenerado era esperada. Contudo, é relevante considerar que a quantidade de genomas bacterianos incompletos pode ter influenciado os resultados obtidos. Além disso, é pertinente mencionar o possível viés introduzido por ferramentas de bioinformática no processo de identificação de profagos.

Neste contexto, este estudo representa uma valiosa contribuição para a compreensão dos profagos no âmbito do gênero *Weissella*, que até então permanecia relativamente sub-explorado na pesquisa científica. Além disso, as evidências sugerem que o gênero *Weissella* ostenta uma relativa segurança em relação à presença de profagos, respaldando a visão de que tais elementos não constituem um substancial risco para a disseminação de genes de relevância clínica em ambientes naturais.

## 7. REFERÊNCIAS

- ABDELRAHMAN, F., *et al.* ***Pseudomonas* Phage ZCPS1 Endolysin as a Potential Therapeutic Agent.** *Viruses*, 15(2), 520. (2023)
- ARBIOUEL, H., Lavilla, L., Muñoz, M. C., Pérez, B., Kabisch, J., Rohtraud, P., Cho, G. S., Neve, H., Vincenzina, F., Charles, M. A. P. F., Gálvez, A., & Benomar, N. **The controversial nature of the *Weissella* genus: technological and functional aspects versus whole genome analysis-based pathogenic potential for their application in food and health.** *Frontiers in Microbiology*, 6.(2015)
- ACKERMANN, H.-W., DuBow, M.S. **Viruses of Prokaryotes**, CRC Press: Boca Raton, FL, USA. 2. (1987)
- ACKERMANN, H. **Frequency of morphological phage descriptions in the year 2000.** *Arch. Virol.* 2001, 146, 843–857. (2001)
- ACKERMANN, H. **Classification of bacteriophages.** In *The Bacteriophages*, Calendar, R., Ed., Oxford University Press: New York, NY, USA, pp. 8–17. (2006)
- ACKERMANN, H. **Phage classification and characterization.** *Methods in molecular biology* (Clifton, N.J.), v. 501, p. 127–140. (2009)
- ACKERMANN, H., Prangishvili, D. **Prokaryote viruses studied by electron microscopy.** *Arch. Virol.* 157, 1843–1849. (2012)
- ADRIAENSSENS, E. M. *et al.* **Taxonomy of prokaryotic viruses: 2018-2019 update from the ICTV Bacterial and Archaeal Viruses Subcommittee.** *Archives of Virology*, v. 165, n. 5, p. 1253–1260, (2020)
- ADRIAENSSENS, E.M., Van Vaerenbergh, J., Vandenheuvel, D., Dunon, V., Ceysens, P.J., *et al.* **T4-Related Bacteriophage LIMEstone Isolates for the Control of Soft Rot on Potato Caused by ‘*Dickeya solani*’.** *PLOS ONE* 7(3): e33227. (2012)
- AIEWSAKUN, P., *et al.* **Evaluation of the genomic diversity of viruses infecting bacteria, archaea and eukaryotes using a common bioinformatic platform: steps towards a unified taxonomy.** *Journal of General Virology*, 99, 1331–1343. (2018)
- ALCALA, A., Ramirez, G., Solis, A., Kim, Y., Tan, K., Luna, O., Nguyen, K., Vazquez, D., Ward, M., Zhou, M., Mulligan, R., Maltseva, N., Kuhn, Ml. **Structural and functional characterization of three Type B and C chloramphenicol acetyltransferases from *Vibrio* species.** *Protein Sci.*, 2020 Mar;29(3):695-710. (2020)
- ALCOCK *et al.* 2023. **CARD 2023: Expanded Curation, Support for Machine Learning, and Resistome Prediction at the Comprehensive Antibiotic Resistance Database.** *Nucleic Acids Research*, 51, D690-D699.
- ALEDO M. G., Blasco L, Lopez M, Ortiz-Cartagena C, Bleriot I, Pacios O, Hernández-García M, Cantón R, Tomas M. **Prophage identification and molecular analysis in the genomes of *Pseudomonas aeruginosa* strains isolated from critical care patients.** *mSphere*. 2023 Aug 24;8(4):e0012823. (2023)

- ALLEN, H.K., Looft, T., Bayles, D.O., Humphrey, S., Levine, U.Y., Alt, D., Stanton, T.B. **Antibiotics in feed induce prophages in swine fecal microbiomes.** *mBio*. 2011 Nov 29,2(6):e00260-11. (2011).
- AL-SHAYEB, B. *et al.* **Clades of huge phages from across Earth's ecosystems.** *Nature*, v. 578, n. 7795, p. 425–431. (2020)
- ALVIM, L. B. (2011). **Identificação molecular e seleção de bactérias lácticas com potencial probiótico isoladas de diferentes mucosas de suínos.** Dissertação apresentada ao Departamento de Biologia Geral do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito para obtenção de grau de mestre em genética. (2011)
- ALVIM, L. B., Sandes, S. H. C., Silva, B. C., Steinberg, R. S., Campos, M. H. A., Acurcio, L. B., Nunes, Á. C. **Weissella paramesenteroides WpK4 reduces gene expression of intestinal cytokines, and hepatic and splenic injuries in a murine model of typhoid fever.** *Beneficial Microbes*, 7(1), 61–73. (2015)
- ANDREWS, S. **FastQC: a quality control tool for high throughput sequence data.** Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (2010)
- APPELT, S. *et al.* **Viruses in a 14th- century coprolite.** *Appl. Environ. Microbiol.* 80, 2648–2655 (2014).
- ARNDT, D. *et al.* **PHASTER: a better, faster version of the PHAST phage search tool.** *Nucleic Acids Res.* 44, W16–W21 (2016).
- ASADULGHANI, M., Ogura, Y., Ooka, T., Itoh, T., Sawaguchi, A., *et al.* **The Defective Prophage Pool of *Escherichia coli* O157: Prophage–Prophage Interactions Potentiate Horizontal Transfer of Virulence Determinants.** *PLOS Pathogens* 5(5): e1000408. (2009)
- AZIZ RK, *et al.* **The RAST Server: Rapid Annotations using Subsystems Technology.** *BMC Genomics* Feb 8,9:75. 2008. (2008)
- BAE, H., Cho, Y. **Complete genome sequence of *Pseudomonas aeruginosa* podophage MPK7, which requires type IV pili for infection.** *Genome Announc.* 1, e00744–e00813. (2013)
- BAE, T., Baba, T., Hiramatsu, K., Schneewind, O. **Prophages of *Staphylococcus aureus* Newman and their contribution to virulence.** *Molecular Microbiology*, 62: 1035-1047. (2006)
- BAJAJ, J.S., Rodriguez, M.P., Fagan, A., McGeorge, S., Sterling, R.K., & Lee, H. **Impact of bacterial infections and spontaneous bacterial peritonitis prophylaxis on phage-bacterial dynamics in cirrhosis.** *Hepatology*, 76, 1723–1734. (2022)
- BANKEVICH, A., *et al.* **SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing** *Journal of Computational Biology*. May 2012. 455-477. (2012)
- BARR, J.J., Auro, R., Furlan, M., Whiteson, K.L., Erb, M.L., Pogliano, J., Stotland, A., Wolkowicz, R., Cutting, A.S., Doran, K.S., *et al.* **Bacteriophages adhering to mucus provide a non host-derived immunity.** *PNAS* 2013, 110, 10771–10776. (2013)
- BARRELL, B. G., Air, G. M., & Hutchison, C. A. **Overlapping genes in bacteriophage phiX174.** *Nature*, 264, 34–41. (1976)

- BARYLSKI, J., *et al.* **Analysis of Spounaviruses as a Case Study for the Overdue Reclassification of Tailed Phages.** *Systematic Biology*, 69, 110–123. (2020)
- BELDA, E., Pedrola, L., Peretó, J., Martínez-Blanch, J.F., Montagud, A., *et al.* **Microbial Diversity in the Midguts of Field and Lab-Reared Populations of the European Corn Borer *Ostrinia nubilalis*.** *PLOS ONE* 6(6): e21751. (2011)
- BENLER, S., Yutin, N., Antipov, D. *et al.* **Thousands of previously unknown phages discovered in whole-community human gut metagenomes.** *Microbiome* 9, 78. (2021)
- BENSING, B., Siboo, I., Sullam, P. **Proteins PblA and PblB of *Streptococcus mitis*, Which Promote Binding to Human Platelets, Are Encoded within a Lysogenic Bacteriophage.** *Infection and Immunity*, pag 6186-6192, vol69, 2001 (2001)
- BENSING, B.A., Rubens, C.E., Sullam, P.M. **Genetic loci of *Streptococcus mitis* that mediate binding to human platelets.** *Infect Immun.* 2001 Mar, 69(3):1373-80. (2001)
- BERNHEIM, A. & SOREK, R. **The pan-immune system of bacteria: antiviral defence as a community resource.** *Nat Rev Microbiol* 18, 113–119 (2020).
- BLERIOT, I., Trastoy, R., Blasco, L., *et al.* **Genomic analysis of 40 prophages located in the genomes of 16 carbapenemase-producing clinical strains of *Klebsiella pneumoniae*.** *Microb Genom.* 2020,6(5):e000369. (2020)
- BOBAY, L-M., Rocha, E.P.C, Touchon, M. **The adaptation of temperate bacteriophages to their host genomes.** *Mol Biol Evol* 2013 30 737 51 (2013)
- BOBAY, L.-M., Touchon, M., Rocha, E. P. C. **Pervasive domestication of defective prophages by bacteria.** *Proc Natl Acad Sci USA* 111, 12127–12132 (2014)
- BONDY-DENOMY, J., Davidson, A.R. **When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness.** *J Microbiol.* 2014 Mar,52(3):235-42. (2014)
- BOYD, E. F. **Bacteriophage-Encoded Bacterial Virulence Factors and Phage-Pathogenicity Island Interactions.** 1. ed. [s.l.] Elsevier Inc., 2012. v. 82 (2012)
- BRANDT, M.J. **Starter cultures for cereal based foods.** *Food Microbiol.* 2014, 37, 41–43. (2014)
- BRIAN, A. *et al.* **CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database,** *Nucleic Acids Research*, Volume 48, Issue D1, 08 January 2020, Pages D517–D525, (2020)
- BRIANDET, R., Lacroix-Gueu, P., Renault, M., Lecart, S., Meylheuc, T., Bidnenko, E., *et al.* **Fluorescence correlation spectroscopy to study diffusion and reaction of bacteriophages inside biofilms.** *Appl Environ Microb* 74: 2135–2143. (2008)
- BRISTER, J.R., Ako-Adjei, D., Bao, Y., Blinkova, O. **NCBI viral genomes resource.** *Nucleic Acids Res.* 2015 Jan,43(Database issue):D571-7. (2015)
- BROWN CL, Mullet J, Hindi F, Stoll JE, Gupta S, Choi M, Keenum I, Vikesland P, Pruden A, Zhang L. **mobileOG-db: a Manually Curated Database of Protein Families Mediating the Life Cycle of Bacterial Mobile Genetic Elements.** *Appl Environ Microbiol.* 2022 Aug. (2022)

- BROWN-JAQUE, M., Muniesa, M., Navarro, F. **Bacteriophages in clinical samples can interfere with microbiological diagnostic tools.** *Sci Rep* 6, 33000. (2016)
- BRÜGGEMANN H, Lood R. **Bacteriophages infecting *Propionibacterium acnes*.** *Biomed Res Int.* 2013;2013:705741. (2013)
- BRUM, J. R., Schenck, R. O., Sullivan, M. B. **Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses.** *ISME J.* 7, 1738–1751 (2013).
- BRÜSSOW, H., Hendrix, R.W. **Phage genomics: small is beautiful.** *Cell* 2002, 108:13–6, PMID:11792317. (2002)
- BRÜSSOW, H., CANCHAYA, C., HARDT, W.-D. **Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion.** *Microbiology and Molecular Biology Reviews*, v. 68, n. 3, p. 560–602. (2004)
- BUCHFINK, B, Reuter, K, Drost, HG. **Sensitive protein alignments at tree-of-life scale using DIAMOND,** *Nature Methods* **18**, 366–368. (2021)
- CALLANAN, J., Stockdale, S.R., Shkoporov, A., Draper, L.A., Ross, R.P., Hill, C. **RNA phage biology in a metagenomic era.** *Viruses* 10:386 (2018)
- CALLANAN, J., Stockdale, S.R., Shkoporov, A., Draper, L.A., Ross, R.P., Hill, C. **Expansion of known ssRNA phage genomes: from tens to over a thousand.** *Sci Adv* 6(6):eaay5981. (2020)
- CANCHAYA, C., Proux, C., Fournous, G., Bruttin, A., Brüssow, H. **Prophage genomics.** *Microbiol. Mol. Biol. Rev.* 67, 238–276. (2003)
- CAO, Z., Zhang, J., Niu, Y.D., Cui, N., Ma, Y., Cao, F., Xu, Y. **Isolation and characterization of a phiKMV-like bacteriophage and its therapeutic effect on mink hemorrhagic pneumonia.** *PLoS One* 10 (1), e0116571. (2015)
- CARROLO, M., Frias, M. J., Pinto, F. R., Melo-Cristino, J., & Ramirez, M. **Prophage spontaneous activation promotes DNA release enhancing biofilm formation in *Streptococcus pneumoniae*.** *PLoS One*, 5(6), e15678. (2010)
- CASJENS, S. **Prophages and bacterial genomics: what have we learned so far?** *Mol. Microbiol.* 2003;49:277–300. (2003)
- CASJENS, S. R., Gilcrease, E. B., Winn-Stapley, D. A., Schickmaier, P., Schmieger, H., Pedulla, M. L., *et al.* **The generalized transducing *Salmonella* bacteriophage es18: complete genome sequence and DNA packaging strategy.** *J. Bacteriol.* 187, 1091–1104. (2005)
- CHAN, B.K., Turnet, P.E., Kim, S., Mojibian, H.R., Eleftheriades, J.A., Narayan, D. **Phage treatment of an aortic graft infected with *Pseudomonas aeruginosa*.** *Evol. Med. Public Health* 2018, 1, 60–66 (2018)
- CHANYOUNG LEE, HYEONGSOON KIM & SANGRYEOL RYU. **Bacteriophage and endolysin engineering for biocontrol of food pathogens/pathogens in the food: recent advances and future trends.** *Critical Reviews in Food Science and Nutrition.* (2022)

- CHATURONGAKUL S., Ounjai P. **Phage–host interplay: examples from tailed phages and Gram-negative bacterial pathogens.** *Front. Microbiol.* 5:442. (2014)
- CHEAITO, A. *et al.*, **Meningitis due to *Weissella confusa***, *The American Journal of Emergency Medicine*, Volume 38, Issue 6, 2020, Pages 1298.e1-1298.e3, ISSN 0735-6757. (2020)
- CHEN, L., Zou, Y., She, P., Wu, Y. **Composition, function, and regulation of T6SS in *Pseudomonas aeruginosa***, *Microbiological Research*, Volume 172, 2015, Pages 19-25, ISSN 0944-5013. (2015)
- CHEN Y, Guo X, Wu J, Jin M, Zeng R. **A novel deep-sea bacteriophage possesses features of  $\beta$ -like viruses and prophages.** *Arch Virol.* 2020 May;165(5):1219-1223. (2020)
- CHOPIN, M, Chopin, A, Bidnenko, E, **Phage abortive infection in *lactococci*: variations on a theme.** *Current Opinion in Microbiology*, Volume 8, Issue 4, 2005, Pages 473-479, ISSN 1369-5274, (2005)
- CHRISTIE, G, CALENDAR, R. **INTERACTIONS BETWEEN SATELLITE BACTERIOPHAGE P4 AND ITS HELPERS.** *Annual Review of Genetics.* Vol. 24:465-490 (Volume publication date December (1990)
- CHUNG, Y.J., Saier, M.H Jr. **Overexpression of the *Escherichia coli sugE* gene confers resistance to a narrow range of quaternary ammonium compounds.** *J Bacteriol.* 2002,184(9):2543-2545. (2002)
- CLARK, J.R., March, J.B. **Bacteriophages and biotechnology: vaccines, gene therapy and antibacterials.** *Trends Biotechnol.* 24 (5), 212–218. (2006)
- COLEMAN, D., Knights, J., Russell, R., Shanley, D., Birkbeck, T.H., Dougan, G., *et al.* **Insertional inactivation of the *Staphylococcus aureus*  $\beta$ -toxin by bacteriophage phi 13 occurs by site- and orientation-specific integration of the phi 13 genome.** *Mol Microbiol* 1991,5:933-9. (1991)
- COLLINS, J. W., La Ragione, R. M., Woodward, M. J., Searle, L. E. J. **Prebiotics and Probiotics Science and Technology** in, eds. D. Charalampopoulos and R. A. Rastall (New York, NY: Springer New York), 1123–1192. (2009)
- COLLINS, M.D., Samelis, J., Metaxopoulos, J., Wallbanks, S. **Taxonomic studies on some *leuconostoc*-like organisms from fermented sausages: description of a new genus *Weissella* for the *Leuconostoc paramesenteroides* group of species.** *J. Appl. Bacteriol.* 75:595-603. (1993)
- CORREA, A.M.S., Howard-Varona, C., Coy, S.R. *et al.* **Revisiting the rules of life for viruses of microorganisms.** *Nat Rev Microbiol* 19, 501–513 (2021)
- COSTA, A. R., Monteiro, R., Azeredo, J. **Genomic analysis of *Acinetobacter baumannii* prophages reveals remarkable diversity and suggests profound impact on bacterial virulence and fitness.** *Sci. Rep.* 8:15346. (2018)

- COUVIN, D., *et al.* **CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins**, *Nucleic Acids Research*, Volume 46, Issue W1, Pages W246–W251 (2018)
- CREASY, A., Rosario, K., Leigh, B.A., Dishaw, L.J., Breitbart, M. **Unprecedented Diversity of ssDNA phages from the family *Microviridae* detected within the gut of a protochordate model organism (*Ciona robusta*)**. *Viruses* 10:404 (2018)
- CRISPIM, J.S., Dias, R.S., Vidigal, P.M.P., *et al.* **Screening and characterization of prophages in *Desulfovibrio* genomes**. *Sci Rep* **8**, 9273. (2018)
- CZAJKOWSKI, R. **May the phage be with you? Prophage-like elements in the genomes of soft rot *Pectobacteriaceae*: *Pectobacterium* spp. and *Dickeya* spp.** *Frontiers in Microbiology*, v. 10, n. (2019)
- D'HERELLE, F. **Sur un microbe invisible antagoniste des bacilles dysentériques**. *CR Acad Sci Paris*. (1917)
- DALMASSO, M., Hill, C., Ross, R.P. **Exploiting gut bacteriophages for human health**. *Trends Microbiol.* Jul,22(7):399-405. (2014)
- DARRIBA, D., Taboada, G.L., Doallo, R., Posada, D. **jModelTest 2: more models, new heuristics and parallel computing**. *Nature Methods* 9(8), 772. (2012)
- DAVE, M., Higgins, P.D., Middha, S., *et al.* **The human gut microbiome: current knowledge, challenges, and future directions**. *Translational Research*, 160:246–57. (2012)
- DAVID Couvin, Aude Bernheim, Claire Toffano-Nioche, Marie Touchon, Juraj Michalik, Bertrand Néron, Eduardo P C Rocha, Gilles Vergnaud, Daniel Gautheret, Christine Pourcel. **CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins**. *Nucleic Acids Research*, 46(W1), W246–W251. (2018)
- DECAESTECKER, E., & King, K. **“Red Queen Dynamics.”** Reference Module in Earth Systems and Environmental Sciences 3: 185–92. (2019)
- DI LALLO, G., Evangelisti, M., Mancuso, F., Ferrante, P., Marcelletti, S., Tinari, A., Superti, F., D'Addabbo, P., Frezza, D., Scortichini, M., *et al.* **Isolation and partial characterization of bacteriophages infecting *Pseudomonas syringae* pv. *Actinidiae*, causal agent of kiwifruit bacterial canker**. *J. Basic Microbiol.* 54, 1210–1221. (2014)
- DIARD, M., Bakkeren, E., Cornuault, J.K., Moor, K., Hausmann, A., Sellin, M.E., Loverdo, C., Aertsen, A., Ackermann, M., De Paepe, M., Slack, E., Hardt, W.D. **Inflammation boosts bacteriophage transfer between *Salmonella* spp.** *Science* 355:1211–1215. (2017)
- DODD, I.B., Shearwin, K.E., Egan, J.B. **Revisited gene regulation in bacteriophage lambda**. *Curr. Opin. Genet. Dev.* 15, 145–152. (2005)
- DORON, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., *et al.* **Systematic discovery of antiphage defense systems in the microbial pangenome**. *Science*. 359(6379):eaar4120. (2018)

- DOYLE, J.J., Doyle, J.L. **A rapid DNA isolation procedure for small quantities of fresh leaf tissue** PHYTOCHEMICAL BULLETIN v.19(1):11-15, (1987)
- DRI, A.M., Moreau, P.L. **Control of the LexA regulon by pH: evidence for a reversible inactivation of the LexA repressor during the growth cycle of *Escherichia coli***. Mol Microbiol 12:621–629. (1994)
- DRILLING, A.J., Ooi, M.L., Miljkovic, D., James, C., Speck, P., Vreugde, S., Clark, J., Wormald, P.J. **Long-term safety topical bacteriophage application to the frontal sinus region**. Front. Cell Infect. Microbiol. (2017)
- DUERKOP, B.A., Clements, C.V., Rollins, D., Rodrigues, J.L.M., Hooper, L.V. **A composite bacteriophage alters colonization by an intestinal commensal bacterium**. Proc Natl Acad Sci U S A 2012, 109:17621-6, PMID:23045666, (2012)
- DUERKOP, B.A. **Bacteriophages shift the focus of the mammalian microbiota**. PLOS Pathogens 14(10): e1007310. (2018)
- EFSA (European Food Safety Authority) **Opinion of the Scientific Committee on a request from EFSA related to a generic approach to the safety assessment by EFSA of microorganisms used in food/feed and the production of food/feed additives** (2005)
- ENDO, A., OKADA, S. **Reclassification of the genus *Leuconostoc* and proposals of *Fructobacillus fructosus* gen. nov., comb. nov., *Fructobacillus durionis* comb. nov., *Fructobacillus ficulneus* comb. nov. and *Fructobacillus pseudoficulneus* comb. nov.** International Journal of Systematic and Evolutionary Microbiology, v. 58, n. 9, p. 2195–2205. (2008)
- EREZ, Z., Steinberger-Levy, I., Shamir, M, *et al.* **Communication between viruses guides lysis–lysogeny decisions**. Nature 541, 488–493. (2017)
- EWING, B., Hillier, L., Wendl, M.C., Green, P. **Base-calling of automated sequencer traces using phred. I. Accuracy assessment**. *Genome Research*. 8 (3): 175–185. (1998)
- FAO/WHO **Working Group Report on Drafting Guidelines for the Evaluation of Probiotics in Food** London, Ontario, Canada, (2005)
- FASTREZ, J. **Phage lysozymes**. EXS.75:35-64. doi: 10.1007/978-3-0348-9225-4\_3. PMID: 8765293. (1996)
- FDA (Food and Drugs Administration). **Federal Register Notice – the GRAS Final Rule** (81 FR 54960 – August 17, 2016)
- FESSARD, A, & Remize, F. **Why Are *Weissella* spp. Not Used as Commercial Starter Cultures for Food Fermentation?** Fermentation. 3(3):38. (2017)
- FIDDES, J. C., & Godson, G. N. **Evolution of the three overlapping gene systems in G4 and phi X174**. Journal of Molecular Biology, 133, 19–43. (1979)
- FIGUEROA-BOSSI, N., Bossi, L., **Inducible prophages contribute to *Salmonella* virulence in mice**. Molecular Microbiology, 33: 167-176. (1999)

- FILLOL-SALOM, A., Martínez-Rubio, R., Abdulrahman, R.F., Chen, J., Davies, R., Penadés, J.R. **Phage-inducible chromosomal islands are ubiquitous within the bacterial universe.** ISME J. Sep,12(9):2114-2128. (2018)
- FLEMMING, HC., Wingender, J. **The biofilm matrix.** *Nat Rev Microbiol* **8**, 623–633. <https://doi.org/10.1038/nrmicro2415> (2010)
- FLOROU-PANERI, P., Christaki, E., Bonos, E. **Lactic Acid Bacteria as Source of Functional Ingredients.** Lactic Acid Bacteria - R & D for Food, Health and Livestock Purposes. (2013)
- FOKINE, A., Rossmann, M. G. **Molecular architecture of tailed double-stranded DNA phages.** *Bacteriophage* **4**, e28281. (2014)
- FONSECA, J. F., Alvim, L. B., Nunes, Á. C., Oliveira, F. M. S., Amaral, R. S., Caliar, M. V, Gomes, M. A. **Probiotic effect of *Bifidobacterium longum* 51A and *Weissella paramesenteroides* WpK4 on gerbils infected with *Giardia lamblia*.** *Journal of Applied Microbiology*, **127**(4), 1184–1191. (2019)
- FORTIER, L. Chapter Five - **The Contribution of Bacteriophages to the Biology and Virulence of Pathogenic *Clostridia***, Editor(s): Sima Sariaslani, Geoffrey Michael Gadd, *Advances in Applied Microbiology*, Academic Press, Volume 101, 2017, Pages 169-200, ISSN 0065-2164, ISBN 9780128120460. (2013)
- FORTIER, L., Sekulovic, O. **Importance of prophages to evolution and virulence of bacterial pathogens.** *Virulence*. 2013,4(5):354-365. (2013)
- FURFARO, L.L., Chang, B.J., Payne, M.S. **Applications for bacteriophage therapy and the perinatal period.** *Front. Microbiol.* 2018, **11**, 2660. (2018)
- FUSCO, V. *et al.* **The genus *Weissella*: Taxonomy, ecology and biotechnological potential.** *Frontiers in Microbiology*, v. 6, n. MAR, 2015. (2015)
- GALARDINI, M., Biondi, E., Bazzicalupo, M., Mengoni, A. **CONTIGSuator: a bacterial genomes finishing tool for structural insights on draft genomes** *Source Code for Biology and Medicine* 2011, **6**:11 (2011)
- GARNEAU, J. E., Moineau, S. **Bacteriophages of lactic acid bacteria and their impact on milk fermentations.** *Microb. Cell Fact.* **10**, 1-10. (2011)
- GOLDFARB, T., Sberro, H, Weinstock, E., Cohen, O., Doron, S., Charpak-Amikam, Y., Afik, S., Ofir, G. Rotem Sorek. **BREX is a novel phage resistance system widespread in microbial genomes.** *EMBO J* **34**:169-183 (2015)
- GONG, Z., Liang, Y., Wang, M., Jiang, Y., Yang, Q., Xia, J., Zhou, X., You, S., Gao, C., Wang, J., He, J., Shao, H., McMinn, A. **Viral diversity Taxonomy of prokaryotic viruses 13 and its relationship with environmental factors at the surface and deep sea of Prydz Bay, Antarctica.** *Front Microbiol* **9**:2981 (2018)
- GU, Z., Eils, R., & Schlesner, M. **Complex heatmaps reveal patterns and correlations in multidimensional genomic data.** *Bioinformatics*. Sep 15;32(18):2847-9. (2016)

- HALTER, M.C., Zahn, J.A. **Characterization of a novel lytic bacteriophage from an industrial *Escherichia coli* fermentation process and elimination of virulence using a heterologous CRISPR–Cas9 system.** J Ind Microbiol Biotechnol 45: 153-163.
- HAMPTON, H.G., Watson, B.N.J. & Fineran, P.C. **The arms race between bacteria and their phage foes.** Nature 577, 327–336 (2020). <https://doi.org/10.1038/s41586-019-1894-8>
- HARADA, L.K., Silva, E.C., Campos, W.F., Del Fiol, F.S., Vila, M., Dąbrowska, K., Krylov, V.N., Balcão, V.M. **Biotechnological applications of bacteriophages: State of the art.** Microbiol Res. 2018 Jul-Aug,212-213:38-58. (2018)
- HEILMANN, S., Sneppen, K., & Krishna, S. **Sustainability of virulence in a phage-bacterial ecosystem.** Journal of Virology, 84(6), 3016-3022. (2010)
- HEILMANN, S., Sneppen, K., & Krishna, S. **Coexistence of phage and bacteria on the boundary of self-organized refuges.** Proceedings of the National Academy of Sciences, 109(31), 12828-12833. (2012)
- HEIR E, *et al.* 1998. **The *Staphylococcus qacH* gene product: a new member of the SMR family encoding multidrug resistance.** FEMS Microbiol. Lett. 163(1):49-56 (1998)
- HERNANDEZ, D., Francois, P., Farinelli, L., Osteras, M., Schrenzel, J. **De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer.** Genome Res. 2008 May,18(5):802-9. (2008)
- HILL, C, Guarner, F., Reid, G., Gibson, G.R., Merenstein, D.J., Pot, B *et al.* **Expert consensus document: the international scientific association for probiotics and prebiotics consensus statement on the scope and appropriate use of the term probiotic.** Nat Rev Gastroenterol Hepatol 11:506–514 (2014)
- HOAI, T.D., Nishiki, I., Yoshida, T., Nakai, T. **Host range and influence of a cell capsule on the phage efficacy of three *Lactococcus garvieae* lytic phages.** Dis Aquat Org 128:81-86. (2018)
- HOOPER, L. V., Macpherson, A. J. **Immune adaptations that maintain homeostasis with the intestinal microbiota.** Nat.Rev.Immunol. 10, 159–169. (2010)
- HOU, Ruiyan, Wang, Lida, Wu, Yi-Jun. **Predicting ATP-Binding Cassette Transporters Using the Random Forest Method.** Frontiers in Genetics vol.11 pag. 156 (2020)
- HUANG, J.X., Bishop-Hurley, S.L., Cooper, M.A. **Development of anti-infectives using phage display: biological agents against bacteria, viruses and parasites.** Antimicrob. Agents Chemother. 56 (9), 4569–4582. (2012)
- HUELSENBECK, J. P., Ronquist F. **MRBAYES: Bayesian inference of phylogeny.** Bioinformatics 17:754-755. (2001)
- HYATT, D., Chen, GL., LoCascio, P.F. *et al.* **Prodigal: prokaryotic gene recognition and translation initiation site identification.** BMC Bioinformatics 11, 119 (2010).
- HYMAN, P. **Bacteriophages and nanostructured materials.** Adv. Appl. Microbiol. 78, 55–73. (2012)
- IBM Corp. **IBM SPSS Statistics for Windows, Version 28.0.** Armonk, NY: IBM Corp. (2021)

ICTV report, 2023 – ICTV Master Species List 2022 MSL38 V3.

IWANO, H., Inoue, Y., Takasago, T., Kobayashi, H., Furusawa, T., Taniguchi, K., Fujiki, J., Yokota, H., Usui, M., Tanji, Y., *et al.* **Bacteriophage  $\Phi$ SA012 has a broad host range against *Staphylococcus aureus* and effective lytic capacity in a mouse mastitis model.** *Biology* 2018, 7, pii:E8 (2018)

Ji, Y.L., Xi, H.Y., Zhao, Z., Jiang, Q.J., Chen, C., Wang X.W., Li, F.Y., Li, N., Sun, C.J., Feng, X., Lei, L.C., & Han, W.Y. **Metagenomics analysis reveals potential pathways and drivers of piglet gut phage-mediated transfer of ARGs.** *Sci. Total Environ.*, 859, Article 160304. (2023)

JIANG, N., *et al.* **Development of Beet necrotic yellow vein virus-based vectors for multiple-gene expression and guide RNA delivery in plant genome editing.** *Plant Biotechnology Journal*, 17, 1302–1315. (2019)

JONES, D.T., Shirley, M., Wu, X., Keis, S. **Bacteriophage infections in the industrial Acetone Butanol (AB) Fermentation Process.** *J Mol Microbiol Biotechnol* 2: 21–26. (2000)

JUNG, Y. H., *et al.* **Characterization of two newly identified genes, *vgaD* and *vatG*, conferring resistance to streptogramin A in *Enterococcus faecium*.** *Antimicrobial Agents and Chemotherapy*, 54(11), 4744-4749. (2010)

KAIDI, P., Koubaa, M., Bals, O., Vorobiev E. **Recent insights in the impact of emerging technologies on lactic acid bacteria: A review.** *Food Research International*, Volume 137, 2020, 109544, ISSN 0963-9969. (2020)

KAUFFMAN, K. M. *et al.* **A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria.** *Nature* 554, 118–122 (2018).

KILIÇ, A., Pavlova, S, Ma, W, Tao, L. **Analysis of *Lactobacillus* phages and bacteriocins in American dairy products and characterization of a phage isolated from yogurt.** *Applied and Environmental Microbiology*, p. 2111–2116 (1996).

KIM, M., Kim, S., Park, B., and Ryu, S. **Core lipopolysaccharide-specific phage SSU5 as an auxiliary component of a phage cocktail for *Salmonella* biocontrol.** *Appl. Environ. Microbiol.* 80, 1026–1034. (2014)

KLEPPEN, H. P. *et al.* **Novel *Podoviridae* family bacteriophage infecting *Weissella cibaria* isolated from kimchi.** *Applied and Environmental Microbiology*, v. 78, n. 20, p. 7299–7308, 2012. (2012)

KLEPPEN, H. P., NES, I. F., HOLO, H. **Characterization of a *Leuconostoc* bacteriophage infecting flavor producers of cheese starter cultures.** *Applied and Environmental Microbiology*, v. 78, n. 18, p. 6769–6772, 2012. (2012)

KONG, S. J., PARK, J. H. **Acid tolerance and morphological characteristics of five *Weissella cibaria* bacteriophages isolated from kimchi.** *Food Science and Biotechnology*, v. 29, n. 6, p. 873–878. (2020)

- KOONIN, E.V., & Wolf, Y.I. **Evolution of microbes and viruses: a paradigm shift in evolutionary biology.** *Front. Cell Infect. Microbiol*, 2, 119. (2012)
- KOONIN, E.V., Makarova, K. S. & Wolf, Y. I. **Evolutionary genomics of defense systems in archaea and bacteria.** *Annu. Rev. Microbiol.* **71**, 233–261 (2017).
- KOT, W., Neve, H., Heller, K. J., and Vogensen, F. K. **Bacteriophages of *Leuconostoc*, *Oenococcus*, and *Weissella*.** *Front. Microbiol.* 28, 1-9. (2014)
- KOURILSKY, P. **Lysogenization by bacteriophage lambda. I. Multiple infection and the lysogenic response.** *Mol. Gen. Genet.* 122:183–95 (1973)
- KOURILSKY, P., Knapp, A. **Lysogenization by bacteriophage lambda. III. Multiplicity dependent phenomena occurring upon infection by lambda.** *Biochimie* 56:1517–23 (1974)
- KRISHNAMURTHY, S.R., Janowski, A.B., Zhao, G., Barouch, D., Wang, D. **Hyperexpansion of RNA bacteriophage diversity.** *PLoS Biol* 14:e1002409 (2016)
- KRONHEIM, S. *et al.* **A chemical defence against phage infection.** *Nature* 564, 283–286 (2018).
- KROPINSKI, A.M., Borodovsky, M., Carver, T.J., Cerdeno-Tarraga, A.M., Darling, A., Lomsadze, A., Mahadevan, P., Stothard, P., Seto, D., Van Domselaar, G., *et al.* **In Silico identification of genes in bacteriophage DNA.** In *Bacteriophages: Methods and Protocols*, Volume 2: Molecular and Applied Aspects, Kropinski, A.M., Clokie, M.R.J., Eds., Humana Press: New York, NY, USA, 2009, Volume 502, pp. 57–89. (2009)
- KRUPOVIC, M. & ICTV Report Consortium. **ICTV virus taxonomy profile: *Plasmaviridae*.** *J. Gen. Virol.* 99, 617–618 (2018)
- KUTATELADZE, M., Adamia, R. **Phage therapy experience at the Eliava Institute.** *Med. Mal. Infect.* 38 (8), 426–430. (2008)
- LANDSBERGER, M., Gandon, S., Meaden, S *et al.* **Anti-CRISPR phages cooperate to overcome CRISPR-Cas immunity.** *Cell* 174(4):908–916 (2018)
- LANG, A.S., Westbye, A.B., Beatty, J.T. **The Distribution, Evolution, and Roles of Gene Transfer Agents in Prokaryotic Genetic Exchange.** *Annu Rev Virol.* 2017 Sep 29,4(1):87-104. (2017)
- LANGERHANS, R.B., *Coevolution*, Editor(s): Sven Erik Jørgensen, Brian D. Fath, **Encyclopedia of Ecology.** Academic Press, 2008, Pages 644-648, ISBN 9780080454054. (2008)
- LAVERGNE, S., Hamilton, T., Biswas, B., Kumaraswamy, M., Schooley, R.T., Wooten, D. **Phage therapy for a multidrug-resistant *Acinetobacter baumannii* craniectomy site infection.** *Open Forum Infect. Dis.* 2018, 5, ofy064. (2018)
- LAZEROFF, M., Harris, S., Tsourkas, P. **Phage Commander, a software tool for rapid annotation of bacteriophage genomes using multiple programs 2020.** *bioRxiv* 2020.11.11.378802. (2020)
- LEBEER, S., Vanderleyden, J., and De Keersmaecker, S. C. J. **Genes and molecules of *Lactobacilli* supporting probiotic action.** *Microbiol. Mol. Biol. Rev.* 72, 728–764. (2008)

- LEE, H.L., Shen, H., Hwang, I.Y., Ling, H., Yew, W.S., Lee, Y.S., Chang, M.W. **Targeted approaches for in situ gut microbiome manipulation**. *Genes* 2018, 9, 351 (2018)
- LEE, K., Park, J., Jeong, H., Heo, H., Han, N., Kim, J. **Probiotic properties of *Weissella* strains isolated from human faeces**. *Anaerobe* Volume 18, Issue 1, 2012, Pages 96-102, ISSN 1075-9964 (2012)
- LEE, S., Park, J. H. **Characteristics on host specificity, infection, and temperature stability of *Weissella* phages from watery kimchi**. *Food Science and Biotechnology*, v. 30, n. 6, p. 843–851. (2021)
- LEE, Y. D., Park, J. H. **Genomic analysis of WCP30 phage of *Weissella cibaria* for dairy fermented foods**. *Korean Journal for Food Science of Animal Resources*, v. 37, n. 6, p. 884–888. (2017)
- LEPLAE, R., Lima-Mendez, G., Toussaint, A. **ACLAME: A CLAssification of Mobile genetic Elements**, update 2010, *Nucleic Acids Research*, Volume 38, Issue suppl\_1, 1 January 2010, Pages D57–D61, (2010)
- LEROY, F., De Vuyst, L. **Lactic acid bacteria as functional starter cultures for the food fermentation industry**. *Trends in Food Science & Technology*, 15(2), 67–78. (2004)
- LETUNIC, I., & Bork, P. **Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation**. *Nucleic Acids Research*, 49(W1), W293–W296. (2021)
- LI H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and 1000 Genome Project Data Processing Subgroup (2009) **The Sequence alignment/map (SAM) format and SAMtools**. *Bioinformatics*, 25, 2078-9. (2009)
- LIAO, H., Li, H., Duan, C.S., Zhou, X.Y., An, X.L., Zhu, Y.G., & Su, J.Q. **Metagenomic and viromic analysis reveal the anthropogenic impacts on the plasmid and phage borne transferable resistome in soil**. *Environ. Int.*, 170, Article 107595. (2022)
- LIM, E. S. *et al.* **Early life dynamics of the human gut virome and bacterial microbiome in infants**. *Nat. Med.* 21, 1228–1234 (2015).
- LINDQVIST, B. H., Dehò, G., & Calendar, R. (1993). **Mechanisms of genome propagation and helper exploitation by satellite phage P4**. *Microbiological Reviews*, 57(3), 683-702.
- LISCHER, H.E.L., Shimizu, K.K. **Reference-guided *de novo* assembly approach improves genome reconstruction for related species**. *BMC Bioinformatics* 18, 474. (2017)
- LIU, B., Zheng, D.D., Jin, Q., Chen, L.H., Yang, J. **VFDB 2019: a comparative pathogenomic platform with an interactive web interface**. *Nucleic Acids Res.* 2019 Jan 8,47(D1):D687-D692. (2019)
- LIU, D., *et al.* **Heritable base-editing in *Arabidopsis* using RNA viral vectors**. *Plant Physiology*, 189, 1920–1924. (2022)
- LIU, X., Li, Y., Guo, Y., Zeng, Z., Li, B., Wood, T.K., Cai, X., Wang, X. **Physiological function of Rac prophage during biofilm formation and regulation of Rac excision in *Escherichia coli* K-12**. *Sci. Rep.* 5, 16074. (2015)

- LIU, Z., Deng, Y., Ji, M., Sun, W., & Fan, X. **Prophages domesticated by bacteria promote the adaptability of bacterial cells.** *BIOCELL*. (2020)
- ŁOBOCKA, M. B. *et al.* **Genome of Bacteriophage P1.** *Journal of Bacteriology*. 186 (21): 7032–7068. (2004)
- LOH, B. *et al.* **A Biological Inventory of Prophages in *A. baumannii* Genomes Reveal Distinct Distributions in Classes, Length, and Genomic Positions.** *Frontiers in Microbiology*, v. 11, n. December, p. 1–13. (2020)
- LOH, B., & Leptihn, S. **A call for a multidisciplinary future of phage therapy to combat multi-drug resistant bacterial infections.** *Infectious Microbes & Diseases*, 2(1), 1-2. (2020)
- LOPES, A., Tavares, P., Petit, M.A., Guérois, R., Zinn-Justin, S. **Automated classification of tailed bacteriophages according to their neck organization.** *BMC Genomics*. 15:1027. (2014)
- LÓPEZ-LEAL, G., Santamaria, R. I., Cevallos, M. Á, Gonzalez, V., and Castillo-Ramírez, S. **Prophages encode antibiotic resistance genes in *Acinetobacter baumannii*.** *Microb. Drug Resist.* 26, 1275–1277. (2020)
- LU, Z. *et al.* **Bacteriophage ecology in a commercial cucumber fermentation.** *Applied and Environmental Microbiology*, v. 78, n. 24, p. 8571–8578 (2012)
- LÜBKE, M., *et al.* **Identification of HCMV-derived T cell epitopes in seropositive individuals through viral deletion models.** *Journal of Experimental Medicine*, 217(3). (2020)
- LUNDE, M., Aastveit, A.H., Blatny, J.M., Nes IF. **Effects of diverse environmental conditions on {phi}LC3 prophage stability in *Lactococcus lactis*.** *Appl Environ Microbiol* 71:721–727. (2005)
- LWOFF, A. **Lysogeny.** *Bacteriol. Rev.* 17, 269–337. (1953)
- LWOFF, A. **Interaction among virus, cell, and organism.** *Science* 152, 1216–1220 (1966).
- LYNCH, K.M, Lucid, A., Arendt, E.K., Sleator, R.D., Lucey, B., Coffey, A. **Genomics of *Weissella cibaria* with an examination of its metabolic traits.** *Microbiology (Reading)*. 2015 Apr,161(Pt 4):914-30. (2015)
- MAGAZINER, S., Zeng, Z., Chen, B., & Salmond, G. **The Prophages of *Citrobacter rodentium* Represent a Conserved Family of Horizontally Acquired Mobile Genetic Elements Associated with Enteric Evolution towards Pathogenicity.** *Journal of Bacteriology*, 201. <https://doi.org/10.1128/JB.00638-18>. (2019).
- MAGILL, D.J., Kucher, P.A., Krylov, V.N. *et al.* **Localised genetic heterogeneity provides a novel mode of evolution in dsDNA phages.** *Sci Rep* 7, 13731 (2017).
- MAHADEVAN, A., & Fisher, D. **High-Diversity Coevolutionary Dynamics in a Model of Interacting Bacteria and Phage.** *Bulletin of the American Physical Society*. (2023)
- MAKAROVA, K.S., Wolf, Y.I., Iranzo, J. *et al.* **Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and Derived variants.** *Nat Rev Microbiol* 18, 67–83. (2020)
- MALIK, D.J., Sokolov, I.J., Vinne, G.K., Mancuso, F., Cincuerrui, S., Vladisavjevic, G.T., Clokie, M.R.J., Garton, N.J., Stapley, A.G.F., Kirpichnikova, A. **Formulation, stabilisation and**

- encapsulation of bacteriophage for phage therapy.** *Adv. Colloid. Interface Sci.* 2017, 249, 100–133. (2017)
- MANRIQUE, P. *et al.* **Healthy human gut phageome.** *Proc. Natl Acad. Sci. USA* 113, 201601060 (2016).
- MARCÓ, M.B, Moineau, S., Quiberoni, A. **Bacteriophages and dairy fermentations.** *Bacteriophage* 2: 149–158. (2012)
- MARIANO, D. **SIMBA: uma ferramenta Web para gerenciamento de montagens de genomas bacterianos.** Dissertação de Mestrado – Programa de Pós-Graduação em Bioinformática da UFMG, Belo Horizonte. (2015)
- MARQUES, A.T., *et al.* **Genomic Analysis of Prophages from *Klebsiella pneumoniae* Clinical Isolates.** *Microorganisms*, 9(10), 2252. (2021)
- MARTHA, R.J. Clokie *et al.* (eds.), **Bacteriophages: Methods and Protocols, Volume 3, Methods in Molecular Biology**, vol. 1681, (2017)
- MARTI, R., Zurfluh, K., Hagens, S., Pianezzi, J., Klumpp, J., and Loessner, M. J. **Long tail fibres of the novel broad-host-range T-even bacteriophage S16 specifically recognize *Salmonella* OmpC.** *Mol. Microbiol.* 87, 818–834. (2013)
- MATOS, R.C, Lapaque, N., Rigottier-Gois, L., Debarbieux, L., Meylheuc, T *et al.* ***Enterococcus faecalis* Prophage Dynamics and Contributions to Pathogenic Traits.** *PLOS Genetics* 9(6): e1003539\_(2013)
- MAVRICH, T., & HATFULL, G. **Bacteriophage evolution differs by host, lifestyle and genome.** *Nature Microbiology*, 2, 17112. (2017)
- MCCALLUM, S.J., Gallagher, M.J., Corkill, J.E., Hart, C.A., Ledson, M.J., Walshaw, M.J., 2002. **Spread of an epidemic *Pseudomonas aeruginosa* from a patient with cystic fibrosis (CF) to non-CF relatives.** *Thorax* 57, 559. (2002)
- MCCARVILLE, J.L., Caminero, A., Verdu, E.F. **Novel perspectives on therapeutic modulation of the gut microbiota.** *Ther. Adv. Gastroenterol.* 9, 580–593. (2016)
- MCKENNA, R. *et al.* **Atomic structure of single- stranded dna bacteriophage  $\phi$ x174 and its functional implications.** *Nature* 355, 137–143 (1992)
- MCNAIR, K., Bailey, B.A., & Edwards, R.A. **PHACTS, a computational approach to classifying the lifestyle of phages.** *Bioinformatics*, 28(5), 614-618. (2012)
- MCNEIL, L.K., *et al.* **The National Microbial Pathogen Database Resource (NMPDR): a genomics platform based on subsystem annotation.** *Nucleic Acids Research*, 35, D347-53. (2007)
- MEADEN, S., Koskella, B. **Exploring the risks of phage application in the environment.** *Front. Microbiol.* 4:358. (2013)
- MEIER-KOLTHOFF J.P., Göker M. **TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy.** *Nat Commun* 10:2182. (2019)

- MEIER-KOLTHOFF, J.P., Sardà Carbasse, J., Peinado-Olarte, R.L., & Göker, M. **TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes.** *Nucleic Acid Res*, 50(D), D801–D807. (2022)
- MENGYU Shen, Yuhui Yang, Wei Shen, Lujia Cen, Jeffrey S. McLean, Wen Yuan Shi, Shuai Le, Xuesong He, Charles M. Dozois. **A Linear Plasmid-Like Prophage of *Actinomyces odontolyticus* Promotes Biofilm Assembly (2018).** *Applied and Environmental Microbiology* pag.e01263-18 vol 84 (2018)
- MERCANTI, D. J. *et al.* **Genomic diversity of phages infecting probiotic strains of *Lactobacillus paracasei*.** *Applied and Environmental Microbiology*, v. 82, n. 1, p. 95–105 (2016)
- MICROSOFT Corporation. **Microsoft Excel.** (2018)
- MIKHEENKO, Alla, Andrey Prjibelski, Vladislav Saveliev, Dmitry Antipov, Alexey Gurevich, **Versatile genome assembly evaluation with QUASt-LG.** *Bioinformatics* 34 (13): i142-i150. (2018)
- MINOT, S., Bryson, A., Chehoud, C., Wu, G., Lewis, J., Bushman, F. **Rapid evolution of the human gut virome.** *PNAS* July 23, 2013 110 (30) 12450-12455. (2002)
- MITARAI N. **How pirate phage interferes with helper phage: Comparison of the two distinct strategies.** *J Theor Biol.* 2020 Feb 7, 486:110096. (2020)
- MOÏRA, B., Oechslin, F., Moineau, S. **Phage diversity, genomics and phylogeny.** *Nature Reviews Microbiology* volume 18, pages125–138 (2020)
- MORARU, C., Varsani, A., Kropinski, A.M. **VIRIDIC – a novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses.** *Viruses* 12(11). (2020)
- MOSIMANN. W., Ritter, W. **Bacteriophages as cause of loss of aroma in butter cultures.** *Schweiz. Milchz.* 72:211–212. (1946)
- MURRAY, ELLEN, LORRAINE A. DRAPER, R. PAUL ROSS, and COLIN HILL. **The Advantages and Challenges of Using Endolysins in a Clinical Setting.** *Viruses*, 13(4), 680. (2021)
- NAKAMURA, T., *et al.* **Parallelization of MAFFT for large-scale multiple sequence alignments.** *Bioinformatics*, 34(14), 2490-2492. (2018)
- NANDA, A. M., THORMANN, K., FRUNZKE, J. **Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions.** *Journal of Bacteriology*, v. 197, n. 3, p. 410–419. (2015)
- NEPAL, R. (2023). **Role of Prophages in *Staphylococcus aureus* Virulence and Pathogenicity.** Tese apresentada na University of Adelaide, Adelaide Medical School, como requisito para obtenção de grau de doutor. (2023)
- NAYFACH, S., Páez-Espino, D., Call, L., Low, S., Sberro, H., Ivanova, N., Proal, A., Fischbach, M., Bhatt, A., Hugenholtz, P., & Kyrpides, N. (2021). **Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome.** *Nature Microbiology*, 6, 960 - 970. <https://doi.org/10.1038/s41564-021-00928-6>.

- NEVES, P.R., McCulloch J.A., Mamizuka, E.M., Lincopan, N., **In: PSEUDOMONAS | *Pseudomonas aeruginosa*, Encyclopedia of Food Microbiology (Second Edition)**, Academic Press, 2014, Pages 253-260, ISBN 9780123847331. (2014)
- NIGRO, O. D. *et al.* **Viruses in the oceanic basement.** *mBio* 8, 1–15 (2017).
- OKONECHNIKOV, K., Golosova, O., Fursov, M., *et al.* **Unipro UGENE: a unified bioinformatics toolkit .** *Bioinformatics* 2012 28: 1166-1167. (2012)
- OLSON, R.D., *et al.* **Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR.** *Nucleic Acids Research*, 51(D1), D678-D689. (2023)
- OPPENHEIM, A. B., Kobilier, O., Stavans, J., Court, D. L., Adhya, S. **Switches in bacteriophage lambda development.** *Annu. Rev. Genet.* 39, 409–429. (2005)
- ORAZI, Giulia, Alan J. Collins, and Rachel J. Whitaker. **Prediction of prophages and their host ranges in pathogenic and commensal *Neisseria* species.** *Msystems* 7.3 (2022)
- OREN, A., & Garrity, G. M. **Notification list. Notification that new names and new combinations have appeared in volume 71, part 10 of the IJSEM.** *International Journal of Systematic and Evolutionary Microbiology*, 72, 5165. (2022)
- OREN, A., & Garrity, G. M. **Valid publication of the names of forty-two phyla of prokaryotes.** *International Journal of Systematic and Evolutionary Microbiology*, 71, 5056. (2021)
- OREN, A., *et al.* **International code of nomenclature of prokaryotes. Prokaryotic code (2022 revision).** *International Journal of Systematic and Evolutionary Microbiology*, 73(5), 005585. (2023)
- OUWEHAND, A.C., Salminen, S., Isolauri, E. **Probiotics: an overview of beneficial effects.** *Antonie Van Leeuwen* 82:279–289 (2002)
- OVERBEEK, R., *et al.* **The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes.** *Nucleic Acids Research*, 33(17), 5691-5702. (2005)
- OVERBEEK, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., Stevens, R. **The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST).** *Nucleic Acids Res.* 2014 Jan,42(Database issue):D206-14. (2014)
- PABST, O., Slack, E. **IgA and the intestinal microbiota: the importance of being specific.** *Mucosal Immunol* 13, 12–21. (2020)
- PAEZ-ESPINO D, *et al.*, **Uncovering Earth’s virome.** *Nature* 536:425–430. (2016)
- PAGE, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A., Parkhill, J. **Roary: rapid large-scale prokaryote pan genome analysis.** *Bioinformatics*, 31(22), 3691–3693. (2015)
- PANTHEE, S., Atmika, P., Jochen, B., Hiroshi, H., Kazuhisa, S. **Complete Genome Sequence of *Weissella hellenica* 0916-4-2 and Its Comparative Genomic Analysis.** *Frontiers in Microbiology*, volume 10, pag 1619 (2019)

- PATERSON, S., *et al.* **Antagonistic coevolution accelerates molecular evolution.** *Nature*, 464(7286), 275-278. (2010)
- PAVESI, A. **Origin, Evolution and Stability of Overlapping Genes in Viruses: A Systematic Review.** *Genes (Basel)*, 12(6), 809. (2021)
- PAWLUK, A., Staals, R.H.J., Taylor, C., Watson, B.N.J., Saha, S., Fineran, P.C., Maxwell, K.L., Davidson, A.R., 2016b. **Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species.** *Nat. Microbiol.* 1, 16085. (2016)
- PAWLUK, A., Shah, M., Mejdani, M., Calmettes, C., Moraes, T.F., Davidson, A.R., Maxwell, K.L. **Disabling a type I-E CRISPR-Cas nuclease with a bacteriophage-encoded anti-CRISPR protein.** *MBio* 8, e01751-17. (2017)
- PEARSON, J.S., GIOGHA, C., LUNG, T., & HARTLAND, E.L. **The genetics of enteropathogenic *Escherichia coli* virulence.** *Annual Review of Genetics*, 50, 493–513. (2016)
- PENADÉS, J.R., Christie, G.E. **The Phage-Inducible Chromosomal Islands: A Family of Highly Evolved Molecular Parasites.** *Annu Rev Virol.* 2015 Nov,2(1):181-201. (2015)
- PERALTA, B. *et al.* **Mechanism of membranous tunnelling nanotube formation in viral genome delivery.** *PLoS Biol.* 11, e1001667 (2013).
- PICKARD, D., Toribio, A. L., Petty, N. K., de Tonder, A., Yu, L., Goulding, D. *et al.* (2010). **A conserved acetyl esterase domain targets diverse bacteriophage to the Vi capsular receptor of *Salmonella enterica* serovar Typhi.** *J. Bacteriol.* 192, 5746–5754. (2010)
- PIRES, D., Sillankorva, S., Faustino, A., Azeredo J. **Use of newly isolated phages for control of *Pseudomonas aeruginosa* PAO1 and ATCC 10145 biofilms.** *Res Microbiol* 162: 798–806. (2011)
- POPE, W.H., & JACOBS-SERA, D. **Annotation of Bacteriophage Genome Sequences Using DNA Master: An Overview.** In: *Bacteriophages.* Methods in Molecular Biology, vol 1681. (2018)
- POURCEL, C., Touchon, M., Villeriot, N *et al.* **CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download and query lists of repeats and spacers.** *Nucleic Acids Res.* 2020,48(D1):D535-D544. (2020)
- PRADO, G. K. S., Torrinha, K. C., Cruz, R. E., Gonçalves, A. B. B., Silva, C. A. V, Oliveira, F. M.S., Caliari, M. V. (2020). ***Weissella paramesenteroides* WpK4 ameliorate the experimental amoebic colitis by increasing the expression of MUC-2 and the intestinal epithelial regeneration.** *J Appl Microbiol.* 2020 Dec,129(6):1706-1719. (2020)
- PRINGSULAKA, O. *et al.*, **Isolation and characterisation of a novel Podoviridae-phage infecting *Weissella cibaria* N 22 from Nham, a Thai fermented pork sausage,** *Food Microbiology*, Volume 28, Issue 3, 2011, Pages 518-525, ISSN 0740-0020, (2011)
- PUJATO, S., Quiberoni, A., Guglielmotti, D. **Technological performance of spontaneous phage-resistant derivatives of *Leuconostoc mesenteroides* and *Leuconostoc pseudomesenteroides* during milk fermentation,** *International Dairy Journal*, Volume 78, 2018, Pages 46-52, ISSN 0958-6946. (2018)

- PUJATO, S., Quiberon, A., Mercanti, D. **Bacteriophages on dairy foods**. J Appl Microbiol 126: 14-30. (2019)
- QIAO, L., *et al.* **Distribution of the microbial community and antibiotic resistance genes in farmland surrounding gold tailings: a metagenomics approach**. Sci. Total Environ., 779, Article 146502. (2021)
- QUAISER, A., Dufresne, A., Ballaud, F., Roux, S., Zivanovic, Y., Colombet, J., Sime-Ngando, T., Francez A.J. **Diversity and comparative genomics of *Microviridae* in Sphagnum-dominated peatlands**. Front Microbiol 6:375. (2015)
- R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. (2021)
- RAKONJAC, J., Bennett, N. J., Spagnuolo, J., Gagic, D., Russel, M. **Filamentous bacteriophage: biology, phage display and nanotechnology applications**. Curr. Issues Mol. Biol. 13, 51–76. (2011)
- RAMOS-BARBERO, M.D., *et al.* **Characterization of crAss-like phage isolates highlights *Crassvirales* genetic heterogeneity and worldwide distribution**. Nature Communications, 14(1), 4295. (2023)
- RANCUREL, C., Khosravi, M., Dunker, A. K., Romero, P. R., & Karlin, D. **Overlapping genes produce proteins with unusual sequence properties and offer insight into de novo protein creation**. Journal of Virology, 83, 10719–10736. (2009)
- RAVIN, N.V. **N15: the linear phage-plasmid**. Plasmid. 2011 Mar,65(2):102-9. (2011)
- REKHA CR, Vijayalakshmi G. **Bioconversion of isoflavone glycosides to aglycones, mineral bioavailability and vitamin B complex in fermented soymilk by probiotic bacteria and yeast**. J Appl Microbiol. 2010 Oct,109(4):1198-208. (2010)
- REYES, A., Wu, M, McNulty, N P., Rohwer, F L., Gordon, J I. **Gnotobiotic mouse model of phage–bacterial host dynamics in the human gut**. Proceedings of the National Academy of Sciences Dec 2013, 110 (50) 20236-20241. (2013)
- RICCI, V., and Piddock, L. J. V. **Exploiting the role of TolC in pathogenicity: identification of a bacteriophage for eradication of Salmonella serovars from poultry**. Appl. Environ. Microbiol. 76, 1704–1706. (2010)
- RICE, S., Tan, C., Mikkelsen, P., *et al.* **The biofilm life cycle and virulence of *Pseudomonas aeruginosa* are dependent on a filamentous prophage**. ISME J 3, 271–282. (2009)
- ROHWER, F. & Edwards, R. **The phage proteomic tree: a genome- based taxonomy for phage**. J. Bacteriol. 184, 4529–4535. (2002)
- RONCA, R., Benzoni, P., DeLuca, A., Crescini, E., Dell’Era, P., 2012. **Phage displayed peptides/antibodies recognizing growth factors and their tyrosine kinase receptors as tools for anti-cancer therapeutic**. Int. J. Mol. Sci. 13 (4), 5254–5277. (2012)
- RONQUIST, F. and J. P. Huelsenbeck. 2003. **MRBAYES 3: Bayesian phylogenetic inference under mixed models**. Bioinformatics 19:1572-1574. (2003)

- ROST, Burkhard. **Twilight zone of protein sequence alignments**. Protein Engineering, Design and Selection, Volume 12, Issue 2, February 1999, Pages 85–94
- ROOBAB, U., *et al.* **Sources, formulations, advanced delivery and health benefits of probiotics**, Current Opinion in Food Science, Volume 32, 2020, Pages 17-28, ISSN 2214-7993. (2020)
- ROUX, S., Brum, J.R., *et al.* **Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses**. Nature 537:689–693 (2016)
- ROUX, S. *et al.* **Cryptic inoviruses are pervasive in bacteria and archaea across Earth's biomes**. Nat. Microbiol. 4, 1895–1906 (2019).
- ROVIK, A., Aziz, S, Pramono H. **Isolation and selection of *Bacillus cereus* specific phages from hospital wastewater**. Biodiversitas 21: 2871-2877 (2020)
- RSTUDIO TEAM. **RStudio: Integrated Development for R**. RStudio, PBC, Boston, MA. (2020)
- RUTBERG, L. **Temperate bacteriophages of *Bacillus subtilis*** in The Molecular Biology of Bacilli Vol. 1 *Bacillus subtilis* (ed. Dubnau, D. A.), 247–268 (1982)
- DOSS, R., PALMER, M., MEAD, D, HEDLUND, B. **Functional biology and biotechnology of thermophilic viruses**. Essays Biochem, 67(4), 671–684. (2023)
- SABATH, N., Wagner, A., & Karlin, D. **Evolution of viral proteins originated de novo by overprinting**. Molecular Biology and Evolution, 29, 3767–3780. (2012)
- SAHU, R., Singh, A. K., Kumar, A., Singh, K., & Kumar, P. **Bacteriophages Concept and Applications: A Review on Phage Therapy**. Current Pharmaceutical Biotechnology, 24(10), 1245-1264. (2023)
- SALIH, H., Karaynir, A., & Yalcin, M. *et al.* **Metagenomic analysis of wastewater phageome from a University Hospital in Turkey**. Arch Microbiol, 204, 353. (2022)
- SALISBURY, A., Tsourkas, P.K. **A Method for Improving the Accuracy and Efficiency of Bacteriophage Genome Annotation**. *Int J Mol Sci*. 2019,20(14):3391. (2019)
- SALMOND, G., Fineran, P. **A century of the phage: past, present and future**. Nat Rev Microbiol 13, 777–786. (2015)
- SANDES, S. (2013). **Seleção de bactérias lácticas com potencial probiótico para uso como promotor de crescimento ou como adjuvante imune em vacinas de mucosa na pecuária bovina**. Dissertação apresentada ao Departamento de Biologia Geral do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito para obtenção de grau de mestre em genética. (2013)
- SANDES, S., Figueiredo, N., Pedroso, S, Sant'Anna, F, Acurcio, L., Junior, M. A., Barros, P., Oliveira, F., Cardoso, V., Generoso, S., Caliari, M., Nicoli, J., Neumann, E., Nunes, A. ***Weissella paramesenteroides* WpK4 plays an immunobiotic role in gut-brain axis, reducing gut permeability, anxiety-like and depressive-like behaviors in murine models of colitis and chronic stress**, Food Research International, Volume 137, 2020, 109741, ISSN 0963-9969 (2020)

- SAVADAGO, A., Ouattara CAT, Bassol IHN, Traore SA (2006) **Bacteriocins and lactic acid bacteria—a mini review**. *Afr J Biotec* 5:678–683 (2006)
- SCHACKART, K.E. III, Graham, J.B., Ponsero, A.J., & Hurwitz, B.L. **Evaluation of computational phage detection tools for metagenomic datasets**. *Front. Microbiol*, 14, 1078760. (2023)
- SCHNEEBERGER, K., Ossowski S, Ott F, Klein JD, Wang X, Lanz C, Smith LM, Cao J, Fitz J, Warthmann N, *et al.* **Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes**. *Proc Natl Acad Sci U S A*. 2011,108(25):10249–54. (2011)
- SEED, K., Lazinski, D., Calderwood, S. *et al.* **A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity**. *Nature* 494, 489–491 (2013).
- SEEMANN, T. **Prokka: rapid prokaryotic genome annotation**. *Bioinformatics*. 30(14):2068-9. (2014)
- SHARMA, N., GUPTA, D., & PARK, Y.S. **Genome analysis revealed a repertoire of oligosaccharide utilizing CAZymes in *Weissella confusa* CCK931 and *Weissella cibaria* YRK005**. *Food Science and Biotechnology*, 32, 553–564. (2023)
- SHIN, H., Lee, J., Kim, H., Choi, Y., Heu, S., and Ryu, S. (2012). **Receptor diversity and host interaction of bacteriophages infecting *Salmonella enterica* serovar Typhimurium**. *PLoS ONE* 7:e43392. (2012)
- SHIN, J., Jiang, F., Liu, J.-J., Bray, N.L., Rauch, B.J., Baik, S.H., Nogales, E., Bondy-Denomy, J., Corn, J.E., Doudna, J.A. **Disabling Cas9 by an anti-CRISPR DNA mimic**. *Sci. Adv.* 3, e1701620. (2017)
- SHKOPOROV, A.N, Clooney AG, Sutton TDS, Ryan FJ, Daly KM, Nolan JA, McDonnell SA, Khokhlova EV, Draper LA, Forde A, Guerin E, Velayudhan V, Ross RP, Hill C. **The human gut virome is highly diverse, stable, and individual specific**. *Cell Host Microbe* 26:527–541 (e525) (2019)
- SILVEIRA, C., Rohwer, F. **Piggyback-the-Winner in host-associated microbial communities**. *npj Biofilms Microbiomes* 2, 16010 (2016)
- SIMMONS, M., Drescher, K., Nadell, C. *et al.* **Phage mobility is a core determinant of phage–bacteria coexistence in biofilms**. *ISME J* 12, 531–543. (2018)
- SIMS, D., Sudbery, I., Ilott, N. *et al.* **Sequencing depth and coverage: key considerations in genomic analyses**. *Nat Rev Genet* 15, 121–132 (2014).
- SINGHANIA, R.R., Patel, A.K., Sukumaran, R.K., Larroche, C., Pandey, A. **Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production**. *Bioresour Technol.* 2013,127:500–7. (2013)
- SIRINGAN, P., Connerton, P.L., Payne, R.J., Connerton, I.F., 2011. **Bacteriophage-mediated dispersal of *Campylobacter jejuni* biofilms**. *Appl. Environ. Microbiol.* 77 (10), 3320–3326. (2011)
- SOMMERS, P., Fontenele, R.S., Kringen, T., Kraberger, S., Porazinska, D.L., Darcy, J.L., Schmidt, S.K., Varsani, A. **Single-stranded DNA viruses in antarctic cryoconite holes**. *Viruses* 11:1022 (2019)

- SONG, S., Wood, T.K. A **Primary Physiological Role of Toxin/Antitoxin Systems Is Phage Inhibition**. *Front Microbiol.* 2020,11:1895. (2020)
- SOUSA, Jorge A. Moura de; Pfeifer, Eugen; Touchon, Marie; Rocha, Eduardo P. C. **Causes and Consequences of Bacteriophage Diversification via Genetic Exchanges across Lifestyles and Bacterial Taxa**. *Molecular Biology and Evolution*. v. 38, n. 6, jun. 2021, p. 2497–2512. (2021)
- SPIEGELHAUER, M.R., Yusibova, M, Rasmussen IKB, Fuglsang KA, Thomsen K, Andersen LP. **A case report of polymicrobial bacteremia with *Weissella confusa* and comparison of previous treatment for successful recovery with a review of the literature**. *Access Microbiol.* 2020 Mar 26,2(5):acmi000119. (2020)
- SRIKANT S., Chantal K. Guegler, Michael T. Laub. **The evolution of a counter-defense mechanism in a virus constrains its host range**. *eLife*, 11, e79549. (2022)
- SRIVIDHYA, K.V., Rao G.V., Raghavenderan L., Mehta P., Prilusky J., Manicka S., Sussman J.L., Krishnaswamy S. **Database and comparative identification of prophages**. In: Huang D-S, Li K, Irwin GW, editors. *Intelligent Control and Automation, Lecture Notes in Control and Information Sciences*. Vol. 344. Berlin: Springer, 2006. pp. 863–868. (2006)
- SRIVIDHYA, K.V., *et al.* **Identification of prophages in bacterial genomes by dinucleotide relative abundance difference**. *PLoS One*. 2007 Nov 21, 2(11):e1193. (2007)
- STAMATAKIS, A. **RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies**. *Bioinformatics*, 30(9), 1312-1313. (2014)
- STERN, A., & Sorek, R. **The phage-host arms race: shaping the evolution of microbes**. *Bioessays*, 33, 43–51. (2011)
- STEINBERG, S. (2016). **Caracterização da microbiota da glândula mamária bovina saudável e com mastite subclínica e seleção de bactérias potencialmente probióticas isoladas deste ecossistema**. Tese apresentada ao Departamento de Biologia Geral do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito para obtenção de grau de doutor em genética. (2016)
- STRÖMSTEN, N. J., Benson, S. D., Burnett, R. M., Bamford, D. H., & Bamford, J. K. H. (2003). **The *Bacillus thuringiensis* Linear Double-Stranded DNA Phage Bam35, Which Is Highly Similar to the *Bacillus cereus* Linear Plasmid pBClin15, Has a Prophage State**. *Journal of Bacteriology*, 185(23), 6985-6989.
- SU, Y., Zhang, W., Liang, H., Liu, Y., Zheng, K., Liu, Z. **Identification and genomic analysis of temperate *Holomonas* bacteriophage vB\_HmeY\_H4907 from the surface sediment of the Mariana Trench at a depth of 8,900m**. *Microbiology Spectrum* Vol. 11, No. 5 (2023)
- SUBIRATS, J., Sánchez-Melsió, A., Borrego, C. M., Balcázar, J. L., and Simonet, P. **Metagenomic analysis reveals that bacteriophages are reservoirs of antibiotic resistance genes**. *Int. J. Antimicrob. Agents* 48, 163–167. (2016)
- SULAKVELIDZE, A, Alavidze, Z., Morris JG Jr. **Bacteriophage therapy**. *Antimicrob Agents Chemother.* 2001 Mar,45(3):649-59. (2001)

- SULLIVAN, M.J, Petty, N.K, Beatson, S.A. **Easyfig: a genome comparison visualizer.** *Bioinformatics*. 2011 Apr 1,27(7):1009-10. (2011)
- SUTHERLAND, I., Hughes, K., Skillman, L., Tait, K. **The interaction of phage and biofilms,** *FEMS Microbiology Letters*, Volume 232, Issue 1, March 2004, Pages 1–6\_(2004)
- SUTTLE, C.A. **Viruses in the sea.** *Nature* 2005, 437:356- 61, (2005)
- SUTTLE, C.A. **Marine viruses--major players in the global ecosystem.** *Nat Rev Microbiol* 2007, 5:801-12, PMID:17853907, (2007)
- SZAFRAŃSKI, S. P., WINKEL, A., STIESCH, M. **The use of bacteriophages to biocontrol oral biofilms.** *Journal of Biotechnology*, v. 250, p. 29–44. (2014)
- TAMANG, J.P., Watanabe, K., Holzapel, W.H. Review: **Diversity of microorganisms in global fermented foods and beverages.** *Front. Microbiol.* 2016, 7, 377. (2016)
- TAYLOR, V.L., Fitzpatrick AD, Islam Z, Maxwell KL. **The Diverse Impacts of Phage Morons on Bacterial Fitness and Virulence.** *Adv Virus Res.* 2019,103:1-31. (2019)
- TEIXEIRA, C.G., Fusieger, A., Milião, G.L. *et al.* **Weissella: An Emerging Bacterium with Promising Health Benefits.** *Probiotics & Antimicro. Prot.* 13, 915–925. (2021)
- THE UNIPROT CONSORTIUM, **UniProt: the universal protein knowledgebase in 2021,** *Nucleic Acids Research*, Volume 49, Issue D1, 8 January 2021, Pages D480–D489, (2021)
- TIWARI, R., Singh, P.K., Singh, S. *et al.* **Bioprospecting of novel thermostable  $\beta$ -glucosidase from *Bacillus subtilis* RA10 and its application in biomass hydrolysis.** *Biotechnol Biofuels* 10, 246 (2017).
- TOUCHON, M., Hoede C Tenaillon O *et al.* **Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths** *PLoS Genet* 2009 5 e1000344 (2009)
- TOUCHON, M., BERNHEIM, A., ROCHA, E. P. C. **Genetic and life-history traits associated with the distribution of prophages in bacteria.** *ISME Journal*, v. 10, n. 11, p. 2744–2754. (2016)
- TOUCHON, M., Bobay, L. M., and Rocha, E. P. **The chromosomal accommodation and domestication of mobile genetic elements.** *Curr. Opin. Microbiol.* 22, 22–29. (2017)
- TREANGEN, T. J., & Rocha, E. P. C. **Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes.** *PLoS Genetics*, 7(7), e1001284. (2011)
- TURNER, D., Kropinski, A., Adriaenssens, E. **A Roadmap for Genome-Based Phage Taxonomy** *Viruses* 13, no. 3: 506. (2021)
- TURNER, D., Shkoporov, A.N., Lood, C. *et al.* **Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee.** *Arch Virol*, 168(1), 74. (2023)
- TWORT, F. W. **An investigation on the nature of ultra-microscopic viruses.,** *The Lancet*, Volume 186, Issue 4814, 1915, Pages 1241-1243, ISSN 0140-6736 (1915)
- UC-MASS, A., Loeza, E.J., la Garza, de, M., Guarneros, G., Hernández-Sánchez, J., Kameyama, L. 2004. **An orthologue of the cor gene is involved in the exclusion of temperate lambdaoid**

- phages.** Evidence that Cor inactivates FhuA receptor functions. *Virology* 329, 425–433. (2004)
- VALEN, L.V. **A new evolutionary law.** *Evol. Theory*, 1, 1–30. (1973)
- VAN BELKUM, A., *et al.* **Phylogenetic distribution of CRISPR-Cas systems in antibiotic-resistant *Pseudomonas aeruginosa*.** *MBio* 6, e01796-15. (2015)
- VAN HOUTE, S., **Buckling, A. & Westra, E. R. Evolutionary ecology of prokaryotic immune mechanisms.** *Microbiol. Mol. Biol. Rev.* 80, 745–763. (2016).
- VENTURA, M, Canchaya, C, Pridmore, D, Berger, B, Brüßow, H. **Integration and distribution of *Lactobacillus johnsonii* prophages.** *J Bacteriol* 2003, 185:4603- 8, (2003)
- VILLION, M., Chopin, M.C., Deveau, H., Ehrlich, S.D., & Moineau, S. **P087, a lactococcal phage with a morphogenesis module similar to an *Enterococcus faecalis* prophage.** *Virology*, 388, 49-56. (2009)
- WALTER, E., Thomas C M, Ibbotson J P, Taylor D E. **Transcriptional analysis, translational analysis, and sequence of the *kilA*-tellurite resistance region of plasmid RK2Ter.** *Journal of Bacteriology* vol.173 n.3(1991)
- WAN, X., *et al.* **Comparative genomics of 40 *Weissella paramesenteroides* strains.** *Frontiers in Microbiology*, 14, 1128028. (2023)
- WANG, X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, Wood TK. **Cryptic prophages help bacteria cope with adverse environments.** *Nat Commun.* 2010;1:147. (2010)
- WANG, F., *et al.* **Fifteen-year application of manure and chemical fertilizers differently impacts soil ARGs and microbial community structure.** *Front. Microbiol*, 11(62). (2020)
- WANG, H, Ling Y, Shan T, Yang S, Xu H, Deng X, Delwart E, Zhang W (2019) **Gut virome of mammals and birds reveals high genetic diversity of the family *Microviridae*.** *Virus Evol* 5:vez013 (2019)
- WANG, I.N, Smith DL, Young R. **Holins: the protein clocks of bacteriophage infections.** *Annu Rev Microbiol.* 2000, 54():799-825. (2000)
- WANG, M., Sun, Y., Zeng, Z., & Wang, Z. **Metagenomics of wastewater phageome identifies an extensively cored antibiotic resistome in a swine feedlot water treatment environment.** *Ecotoxicology and Environmental Safety*, 222, 112552. (2021)
- WANG, X., & WOOD, T. K. Cryptic prophages as targets for drug development. *Drug Resistance Updates*, 27, 30-38. (2016)
- WANG, X., Kim, Y., Ma, Q. *et al.* **Cryptic prophages help bacteria cope with adverse environments.** *Nat Commun* 1, 147. (2010)
- WEI, B., Peng, Z., Huang, T. *et al.* **Stability of potential prophages in commercial strain *Lactobacillus plantarum* NCU116 under various stressors.** *Arch Microbiol* 202, 1241–1250 (2020).

- WHELAN, K., Colleran E, Taylor D E. **Phage inhibition, colicin resistance, and tellurite resistance are encoded by a single cluster of genes on the IncHI2 plasmid R478.** Journal of Bacteriology vol.177 n.17 p.5016-5027. (1995)
- WHITMAN, W.B., *et al.* **Prokaryotes: The unseen majority,** Proc. Natl Acad. Sci., 1998, vol. 95 (pg. 6578-6583) (1998)
- WICK, R., Judd L, Gorrie C, Holt Ke. **Unicycler: Resolving Bacterial Genome Assemblies From Short And Long Sequencing Reads.** Plos Comput Biol 2017. (2017)
- WICKHAM, H., *et al.* **Welcome to the tidyverse.** Journal of Open Source Software, 4(43), 1686. (2019)
- WOESE, C. R. **Bacterial evolution.** Microbiol. Rev. 51:221-271. (1987)
- WOESE, C. R., and G. E. Fox. **Phylogenetic structure of the prokaryotic domain: the primary kingdoms.** Proc. Natl. Acad. Sci. USA 74:5088-5090 (1977)
- WOMMACK, K.E., Colwell R.R.. **Virioplankton: Viruses in Aquatic Ecosystems,** Microbiol. Mol. Biol. Rev., 2000, vol. 64 (pg. 69-114) (2000)
- XIONG, X., Wu, G., Wei, Y. *et al.* **SspABCD–SspE is a phosphorothioation-sensing bacterial defence system with broad anti-phage activities.** *Nat Microbiol* 5, 917–928 (2020).
- XU Y, *et al.* **Crystal structure of the periplasmic region of MacB, a noncanonic ABC transporter.** Biochemistry 48(23): 5218-5225 (2009)
- YAHARA, Koji; LEHOURS, Philippe; VALE, Filipa F. **Analysis of genetic recombination and the pan-genome of a highly recombinogenic bacteriophage species.** Microbial genomics, v. 5, n. 8, 2019. (2019)
- YAMAMOTO, T., *et al.* **SP10 infectivity is aborted after bacteriophage SP10 infection induces nonA transcription on the prophage SP $\beta$  region of the *Bacillus subtilis* genome.** Journal of Bacteriology, 196, 693–706. (2014)
- YEN, M., Cairns, L.S., Camilli, A. **A cocktail of three virulent bacteriophages prevents *Vibrio cholerae* infection in animal models.** Nat. Commun. 2017, 8, 14187. (2017)
- ZABURLIN, D., MERCANTI, D. J., QUIBERONI, A. **A fast PCR-based method for the characterization of prophage profiles in strains of the *Lactobacillus casei* group.** Journal of Virological Methods, v. 248, p. 226–233. (2017)
- ZENG, L. *et al.* **Decision making at a subcellular level determines the outcome of bacteriophage infection.** Cell 141, 682–691 (2010).
- ZHANG. Y, Zhang Z, Zhang H, Zhao Y, Zhang Z, Xiao J. **PADS Arsenal: a database of prokaryotic defense systems related genes.** Nucleic Acids Res. Jan 8,48(D1):D590-D598. (2020)
- ZHENG, J. *et al.* **A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus beijerinck* 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*.** International Journal of Systematic and Evolutionary Microbiology, v. 70, n. 4, p. 2782–2858. (2020)

- ZHONG, X, Guidoni B., Jacas L, Jacquet S. **Structure and diversity of ssDNA *Microviridae* viruses in two peri-alpine lakes (Annecy and Bourget, France)**. Res Microbiol 166:644–654 (2015)
- ZHOU, T., *et al.* **Silencing Specific Genes in Plants Using Virus-Induced Gene Silencing (VIGS) Vectors**. In Plant Virology, Springer: New York, NY, USA. (2022)