

CENTRO UNIVERSITÁRIO SÃO CAMILO

CURSO DE BIOMEDICINA

IC-05

Sthefany Alves Pereira

CARACTERIZAÇÃO GENÔMICA DE PROFAGOS PROVENIENTES DE ENTEROBACTÉRIAS

MULTIDROGA RESISTENTES

GENOMIC CHARACTERIZATION OF PROPHAGES FROM MULTIDRUG-RESISTANT ENTEROBACTERIA

São Paulo

2025

Sthefany Alves Pereira

**CARACTERIZAÇÃO GENÔMICA DE PROFAGOS PROVENIENTES DE ENTEROBACTÉRIAS
MULTIDROGA RESISTENTES**

**GENOMIC CHARACTERIZATION OF PROPHAGES FROM MULTIDRUG-RESISTANT
*ENTEROBACTERIACEAE***

Trabalho de Conclusão de Curso apresentado ao curso de Biomedicina do Centro Universitário São Camilo, orientado pela Prof.^a Dr.^a Dyana Alves Henriques e coorientado pelo Dr. Willames M. B. S. Martins, como requisito parcial para obtenção do título de biomédica.

São Paulo

2025

RESUMO

A resistência antimicrobiana é um desafio crescente à saúde pública global, associada a aproximadamente 4,95 milhões de mortes anuais. Evidências sugerem que a emergência de clones bacterianos de alto risco pode estar relacionada à transferência horizontal de genes mediada por profagos. Nesse sentido, este estudo teve como objetivo caracterizar genomicamente profagos integrados aos clones de alto risco de *Enterobacter spp.*, *Escherichia coli* e *Klebsiella spp.* Foram utilizadas 57 enterobactérias multidroga resistentes, isolando profagos completos e questionáveis, e avaliando similaridade, classificação taxonômica, anotação funcional, predição de hospedeiro e ciclo de vida através de ferramentas de bioinformática. 242 profagos foram identificados, organizados em 21 agrupamentos filogenéticos distintos, com predominância do ciclo lisogênico, diversidade genética elevada e baixa similaridade com fagos em bancos de dados, sugerindo sequências inéditas. A maioria apresentava tamanho moderado, conteúdo guanina-citosina estável e predomínio de proteínas de montagem viral, infecção do hospedeiro, lise celular e hipotéticas. Ocorreram 50 genes associados à resistência e fatores de virulência, incluindo *NDM-1*, *AAC(6′)-Ib7*, *sul1*, *oqxAB*, *emrE* e *mdfA*, ligados a bombas de efluxo e enzimas inativadoras. Dentre os fatores de virulência, destacaram-se o sistema de secreção tipo VI, *pilus* comum de *Escherichia coli* e elementos flagelares, com indícios de disseminação de operons via profagos. Conclui-se que profagos desempenham um papel central na evolução bacteriana, atuando como vetores de inovação genética que ampliam a adaptabilidade, a patogenicidade e o potencial de resistência das enterobactérias frente às pressões seletivas, reforçando a importância da caracterização de profagos como ferramenta complementar no combate à resistência a antibióticos.

PALAVRAS-CHAVE: Profagos; Caracterização genômica; Clones de alto risco; Enterobactérias multidroga resistentes.

ABSTRACT

Antimicrobial resistance is a growing global public health challenge, associated with approximately 4.95 million deaths annually. Evidence suggests that the emergence of high-risk bacterial clones may be linked to horizontal gene transfer mediated by prophages. In this context, the aim of this study was to genomically characterize prophages integrated into high-risk clones of *Enterobacter spp.*, *Escherichia coli*, and *Klebsiella spp.* 57 multidrug-resistant *Enterobacteriaceae* were analyzed, isolating complete and questionable prophages and evaluating their similarity, taxonomic classification, functional annotation, host prediction, and life cycle using bioinformatics tools. A total of 242 prophages were identified and organized into 21 distinct phylogenetic clusters, with a predominance of the lysogenic cycle, high genetic diversity, and low similarity to phages available in databases, suggesting novel sequences. Most prophages exhibited moderate genome size, stable guanine-cytosine content, and predominantly encoded proteins related to viral assembly, host infection, cell lysis, and hypothetical functions. 50 genes associated with antimicrobial resistance and virulence factors were identified, including *NDM-1*, *AAC(6′)-Ib7*, *sul1*, *oqxAB*, *emrE*, and *mdfA*, which are linked to efflux pumps and inactivating enzymes. Among the virulence factors, genes related to the type VI secretion system, the common pilus of *Escherichia coli*, and flagellar elements were prominent, with evidence suggesting the dissemination of operons via prophages. In conclusion, prophages play a central role in bacterial evolution, acting as vectors of genetic innovation that enhance the adaptability, pathogenicity, and resistance potential of *Enterobacteriaceae* under selective pressures, underscoring the importance of prophage characterization as a complementary tool in the fight against antibiotic resistance.

KEYWORDS: Prophages; Genomic characterization; High-risk clones; Multidrug-resistant *Enterobacteriaceae*.

INTRODUÇÃO

Bacteriófagos, ou simplesmente fagos, são as entidades biologicamente mais abundantes na biosfera, estes vírus ubiqüitários são capazes de infectar bactérias ou arqueias, utilizando o metabolismo dos hospedeiros para sua replicação¹. Assim como para outros tipos de vírus, os bacteriófagos podem possuir DNA ou RNA como material genético, podendo ser de fita simples ou dupla, com comprimento entre aproximadamente 5 e 500.000 kbp, estando contidos em capsídeos de formato icosaédrico, helicoidal, pleomórfico ou conectados à uma cauda (Figura 1). Contudo, apesar da diversidade morfológica, os mais abundantes são os fagos caudados com DNA de dupla fita, classificados atualmente na classe *Caudoviricetes*^{2,3}. Em 2021, o Comitê Internacional de Taxonomia de Vírus (ICTV) propôs a abolição da ordem *Caudovirales* e das famílias *Myoviridae*, *Siphoviridae* e *Podoviridae*, instituindo uma nova classificação baseada em identidade genômica, sendo necessário compartilhar genes ortólogos para uma mesma família, com alinhamento entre nucleotídeos $\geq 70\%$ para o nível de gênero, e similaridade de sequência $\geq 95\%$ para a classificação como uma mesma espécie. Todavia, os fagos ainda estão sendo reclassificados, e ainda pode-se utilizar os termos “miovírus”, “sifovírus” e “podovírus” como identificadores morfológicos^{2,4}.

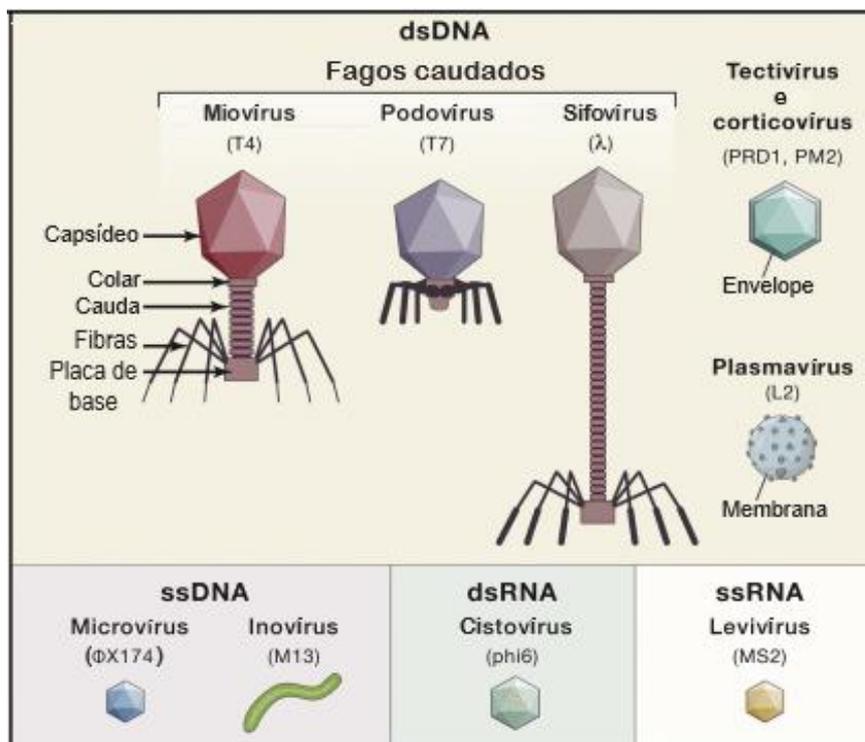


Figura 1. Diversidade de bacteriófagos com base no tipo de material genético e morfologia. Para cada identificador morfológico há um fago representativo entre parênteses. Adaptado de Ofir e Sorek (2018)⁵.

Os ciclos de vida lítico e lisogênico são os mais amplamente descritos para bacteriófagos caudados⁶ (Figura 2). Após a interação fago-receptor, um processo altamente específico e seletivo, ocorre a injeção do material genético no citoplasma bacteriano, através da formação de um canal ou poro na membrana. Uma vez no interior da célula, o material pode ser utilizado para replicação e montagem de progênes virais, além da síntese de proteínas de lise, ocasionando a degradação da membrana e parede celular bacteriana, e consequente liberação de toda a numerosa progênie madura, sendo este o estado lítico^{7,8,9}. Alternativamente, pode ocorrer inserção no cromossomo bacteriano, e, em alguns casos, em plasmídeos, permanecendo integrado sob a forma de profago (ciclo lisogênico). Neste estado de lisogenia, o genoma viral é replicado simultaneamente ao do hospedeiro, possibilitando a transferência do material genético fágico para todos os descendentes bacterianos¹⁰. Profagos utilizam repressores para inibir a expressão de genes líticos a fim de permanecerem latentes, até que ocorra uma situação de intenso dano ao DNA da bactéria, promovendo o desligamento do repressor, e indução da transição para o ciclo lítico¹¹. A lisogenia promove imunidade a superinfecções por outros fagos filogeneticamente semelhantes, através da transferência horizontal de genes de resistência a bacteriófagos para o hospedeiro bacteriano¹⁰.

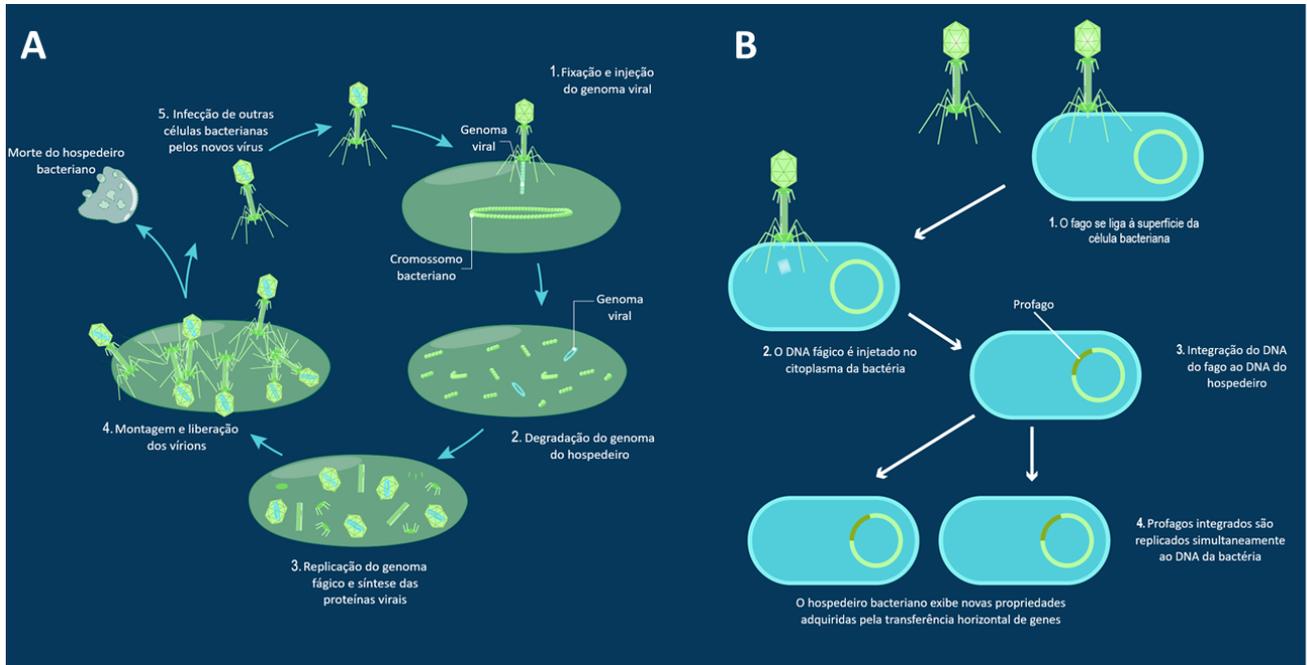


Figura 2. Uma visão geral acerca do ciclo lítico (A) e lisogênico (B) na infecção por bacteriófagos. Adaptado de Konieczka e Wolyniak (2022)¹².

Os fagos foram descritos pela primeira vez em 1915 e 1917 frente às observações de Frederick Twort¹³ e Félix d'Hérelle¹⁴, e apesar de terem sido conduzidos estudos iniciais de aplicação

terapêuticas, as pesquisas desaceleraram após o advento da penicilina e subsequente ascensão do uso de antibióticos¹⁵. Entretanto, atualmente, o número de casos de resistência antimicrobiana (RAM) tem crescido alarmantemente, com uma estimativa de 4,95 milhões de mortes por ano associadas a resistência aos antibióticos, afetando principalmente populações de países de baixa e média renda. Além disso, sem o desenvolvimento de novas terapias eficazes, há uma projeção que até 2050, o número de mortes anuais possa alcançar 10 milhões^{16,17}. A Organização Mundial de Saúde (OMS) categoriza as bactérias em grupos de prioridade crítica, alta e média, de acordo com a urgência de pesquisas e desenvolvimento, devido ao seu grande impacto para a saúde pública global na resistência antimicrobiana. Em 2024, foi divulgada a atualização da Lista de Patógenos Bacterianos Prioritários da OMS (OMS BPPL)¹⁸, evidenciando 15 famílias, com destaque para o grupo crítico, que compreende as enterobactérias multidroga resistentes. É neste cenário de crescimento da resistência antimicrobiana que as pesquisas utilizando bacteriófagos ascenderam novamente em todo o mundo¹⁹. Contudo, embora os esforços estejam principalmente concentrados no estudo de fagos líticos, inúmeras evidências demonstram o papel de profagos sob o predomínio de cepas bacterianas com vantagens evolutivas, pois estão inseridos em regiões que codificam proteínas associadas a virulência e promovem a transferência horizontal de genes de resistência a antibióticos²⁰. É descrito que estirpes patogênicas bacterianas carregam um número maior de profagos em comparação às não patogênicas^{21,22}. Foi demonstrado que *Staphylococcus aureus*, *Salmonella enterica* sorovar Typhimurium e *Escherichia coli* suscetíveis aos antibióticos se tornaram resistentes ao integrarem profagos transdutores generalizados²³. Assim como, observou-se que fagos lisogênicos são capazes de mediar a transdução de resistência à estreptomicina entre linhagens de *Streptococcus pyogenes*²⁴. Também foi descrito que, em regiões contendo profagos, há genes que codificam para as subunidades Stx1 e Stx2 da toxina Shiga em *E. coli*²⁵, bem como, a toxina da cólera em *Vibrio cholerae*²⁶, toxina diftérica em *Corynebacterium diphtheriae*²⁷, e ainda, contribui para a formação de biofilme, produção de citotoxinas, resistência à fagocitose e aumento da adesão as células epiteliais humanas em *Pseudomonas aeruginosa*²⁸.

A RAM é uma causa crescente de morbidade e mortalidade em países de baixa e média renda, portanto, a concentração de estudos voltados para novas alternativas terapêuticas é cada vez mais necessária²⁹. É crucial elucidar o papel dos profagos nesta ascensão evolutiva, evidenciado pelo desenvolvimento de vantagens por parte dos clones de alto risco que os carregam³⁰. Estes fatores destacam a necessidade de compreender melhor as interações evolutivas entre fagos e bactérias, evidenciando uma provável relação mutualística que deve ser explorada e caracterizada, como meio

de encontrar respostas para lacunas que ainda precisam ser preenchidas nos avanços da terapia fágica. Além disso, frente ao advento recente da reclassificação taxonômica de bacteriófagos baseada em similaridade genética, é evidente a demanda de estudos voltados para caracterização genômica, principalmente de profagos que estão constantemente interagindo com os genomas dos hospedeiros bacterianos. E ainda, são escassos os trabalhos envolvendo especificamente a caracterização de profagos em clones de alto risco de enterobactérias multidroga resistentes. Com base na problemática evidenciada e considerando o grupo crítico da Lista de Patógenos Prioritários Bacterianos da OMS de 2024, este trabalho tem como objetivo realizar uma análise genômica de profagos identificados em clones bacterianos multidroga resistentes de *Enterobacter* spp., *E. coli* e *Klebsiella* spp., isolados em países de baixa e média renda, com enfoque em determinar as características genômicas da coleção de bacteriófagos, investigar a diversidade nos diferentes clones de alto risco analisados e efetuar curadoria dos genes e respectivas proteínas produzidas, visando auxiliar na compreensão do papel dos fagos na evolução e disseminação de clones de alto risco.

MÉTODO

Para o presente projeto foram selecionados 57 isolados bacterianos advindos de três estudos que abordam a caracterização genômica de bactérias Gram-negativas multidroga resistentes oriundas de infecções da corrente sanguínea ou isoladas da microbiota de habitantes e animais de países de baixa e média renda^{31,32,33}, com predomínio de isolados advindos de estudo do grupo *Burden of Antibiotic Resistance in Neonates from Developing Societies* (BARNARDS)³⁴. Todos os isolados foram previamente sequenciados pela metodologia Illumina (*short reads*), sendo cada *read* trimada e alinhada para formar os *contigs* do genoma bacteriano. Para a realização deste estudo, sequências de nucleotídeos no formato FASTA (.fasta) foram obtidas através de download no banco de dados Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>)³⁵.

A seleção dos clones bacterianos seguiu como critério o perfil clonal de gêneros representativos da família *Enterobacteriaceae* que possuíam o fenótipo e genótipo de multidroga resistência, sendo selecionados isolados de *Enterobacter* spp. (n=19), *E. coli* (n=14) e *Klebsiella* spp. (n=24), em diferentes países de baixa e média renda (Figura 3). As informações acerca de cada isolado podem ser verificadas no [Anexo 1](#).

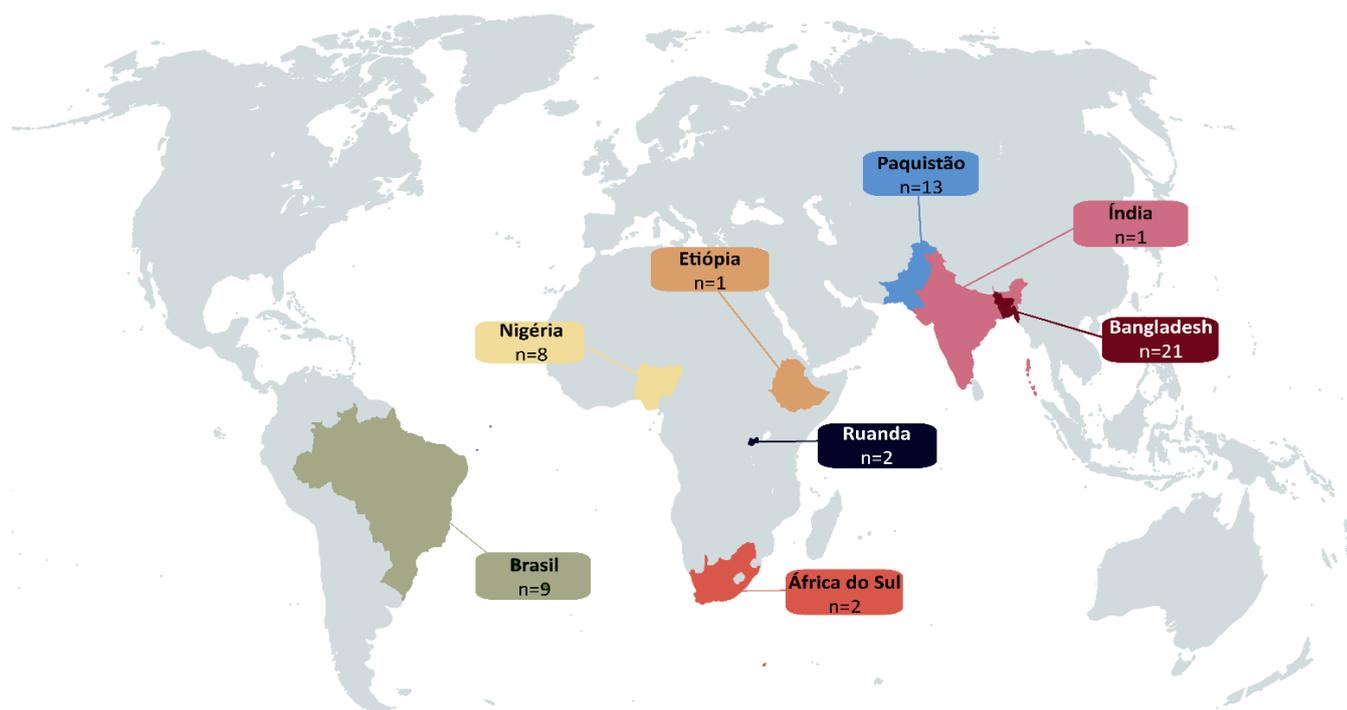


Figura 3. Distribuição dos 57 isolados bacterianos provenientes de diferentes países de baixa e média renda. Figura criada com MapChart (<https://www.mapchart.net/>)³⁶.

Cada um dos genomas bacterianos foi submetido a plataforma Phage Search Tool with Enhanced Sequence Translation (PHASTEST - <https://phastest.ca/>)³⁷, a fim de identificar a sequência de cada profago inserido no genoma bacteriano. Como critério de inclusão no estudo, foram utilizados apenas os fagos classificados como “completos” e “questionáveis”, aqueles classificados como “incompletos” não passaram para as próximas etapas, considerando os critérios de pontuação da plataforma (Figura 4). As sequências intactas e questionáveis foram copiadas, convertidas ao formato FASTA e renomeadas seguindo o padrão: *Prophage* + numeração (de acordo com a ordem crescente em que foram identificadas no genoma) + identificação do clone bacteriano.

- Método 1: Se o tamanho de determinado organismo fágico na tabela [NCBI virus](#) for maior ou igual a 100% do número total de regiões codificadoras (CDS) da região alinhada no genoma bacteriano, a região será marcada com pontuação total de 150. Se for menor que 100%, os métodos 2 e 3 serão usados.
- Método 2: Se o tamanho de determinado organismo da tabela for maior que 50% do número total de CDS da região, esse organismo é considerado o principal fago potencial para essa região; a porcentagem do tamanho do fago nesta tabela pelo número total de proteínas da região é calculada e então multiplicada por 100; a porcentagem do comprimento desse organismo pelo comprimento da região é calculada e então multiplicada por 50.
- Método 3: Se qualquer uma das palavras-chave específicas relacionadas ao fago (como 'capsídeo', 'cabeça', 'integrase', 'placa', 'cauda', 'fibra', 'revestimento', 'transposase', 'portal', 'terminase', 'protease' ou 'lisina') estiver presente, a pontuação será aumentada em 10 para cada palavra-chave encontrada. Se o tamanho da região for maior que 30 Kb, a pontuação será aumentada em 10. Se houver pelo menos 40 proteínas na região, a pontuação será aumentada em 10. Se todas as proteínas relacionadas ao fago e proteínas hipotéticas constituírem mais de 70% do número total de proteínas na região, a pontuação será aumentada em 10.

Comparando a pontuação total do método 2 com a pontuação total do método 3, a maior é escolhida como a pontuação total da região.

Se a pontuação total da região for inferior a 70, ela é marcada como incompleta; se estiver entre 70 e 90, ela é marcada como questionável; se for superior a 90, ela é marcada como intacta.

Figura 4. Critérios para pontuação de regiões pressupostas de profago como intactas, questionáveis ou incompletas. Adaptado de PHASTEST³⁷.

Devido ao grande número de bacteriófagos identificados, procedeu-se o alinhamento múltiplo de sequências por meio do plugin MAFFT v7 no software Geneious Prime® v2025.1.2, gerando a quantificação de similaridade entre os fagos através do valor de *Average Nucleotide Identity* (ANI). Os dados foram organizados em matriz *dot plot*, além de serem utilizados para construção de árvore filogenética pelo método Neighbor-Joining através do modelo de distância genética Jukes-Cantor com 100 repetições bootstrap. O resultado obtido foi visualizado e anotado por meio da ferramenta *online Interactive Tree of Life v7* (iTOL - <https://itol.embl.de/>)³⁸.

A classificação taxonômica foi conduzida seguindo os critérios do ICTV², considerando ortologia de genes para uma mesma família, alinhamento $\geq 70\%$ para o nível de gênero e $\geq 95\%$ para mesma espécie. Para tal, utilizou-se o servidor taxMyPhage (<https://ptax.ku.dk/>)³⁹, que compara a

sequência submetida ao banco de dados do ICTV (<https://ictv.global/msl>)⁴⁰, havendo complementação da análise utilizando a ferramenta BLAST® (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)⁴¹ com parâmetros padrão para megablast, obtendo alinhamento ao banco de dados NCBI³⁵.

Os arquivos FASTA foram utilizados para todas as etapas de caracterização dos profagos, sendo que cada parâmetro era avaliado em no mínimo duas diferentes plataformas para confirmação do dado gerado. Não havendo congruência entre os dados, plataformas adicionais era utilizadas como contraprova. Para obtenção do tamanho dos genomas em pares de base (bp) e o conteúdo de guanina+citosina (GC%), utilizou-se a plataforma baseada em inteligência artificial PhageAI (<https://app.phage.ai/>)⁴² e confirmação por meio do servidor PhaBOX (<https://phage.ee.cityu.edu.hk/phabox>)⁴³. O módulo PhaVIP em PhaBOX em conjunto ao módulo *ORFs prediction and protein classification*, do banco de dados interativo PhageScope (<https://phagescope.deepomics.org/>)⁴⁴ geraram a predição da quantidade de genes, além da anotação e classificação das proteínas em classes funcionais: tRNA, integração; imunidade; regulação; lise; empacotamento; replicação; montagem; infecção, proteínas hipotéticas e proteínas sem classificação. O módulo de predição de PhageScope utiliza Prodigal v2.6.3 para identificação de *open reading frames* (ORFs) e atribuição de ortologia com Egnog-mapper v2.1.10 e DIAMOND v2.1.11 transferindo as anotações para as sequências codificadoras, e ainda, para aquelas sem resultado satisfatório, emprega-se o método de homologia com MMseqs2 a partir do banco de dados PHROG. MMseqs2 também foi empregado para buscar homologia aos genes de resistência a antibióticos descritos em *Comprehensive Antibiotic Resistance Database* (CARD - <https://card.mcmaster.ca/>)⁴⁵ e fatores de virulência de *Virulence Factor Database* (VFDB - <https://www.mgc.ac.cn/VFs/>)⁴⁶, considerando identidade >80% e cobertura >40%.

A partir da seleção de um fago representativo para cada gênero de bactérias, sendo completo, de tamanho, conteúdo GC e número de genes moderados, foi gerado o arquivo GBK (.gbk) através de Pharokka v1.7.0⁴⁷, sendo utilizados para gerar os mapas genômicos por meio de LoVis4u v0.0.12⁴⁸, um software que se baseia na utilização de MMseqs2 para atribuir homologia entre proteínas, gerando uma análise comparativa entre os genomas submetidos.

A ferramenta CHERRY parte integrante de PhaBOX⁴⁰, elencou o provável hospedeiro para cada bacteriófago submetido, sendo confirmado pelo módulo *host assignment* de PhageScope⁴¹, baseado em alinhamento BLASTN v2.5.0, considerando o menor *e-value* abaixo do limiar 1e-5, com emprego de DeepHost sob a ausência de *hit* via BLASTN.

Para previsão de predileção de ciclo de vida, as sequências FASTA foram submetidas a cinco plataformas, tendo sido elencadas as mais usuais para esta finalidade: PhageAI³⁹, PhageLeads (<https://phageleads.dk/>)⁴⁹, BACPHLIP v0.9.3⁵⁰, PhaBOX⁴⁰ e PhageScope⁴¹. As classificações como lítico ou lisogênico foram anotadas, bem como casos em que não houve classificação da sequência submetida. Em caso de incongruências entre as plataformas, levou-se em consideração o maior número de uma mesma classificação, bem como, a presença de genes no fago relacionados a lisogenia, como de integração, tRNA e prevenção de superinfecção por outros fagos.

Todos os dados gerados foram tabulados, sendo que a análise e apresentação desses dados foi realizada através de estatística descritiva e a correlação entre variáveis analisada por meio do método de Pearson. As análises estatísticas foram performadas utilizando o software GraphPad Prism v8.0.1.

RESULTADOS

1. Distribuição de bacteriófagos integrados aos genomas de clones de alto risco

Ao todo, foram identificados 504 profagos nos 57 isolados bacterianos analisados através da plataforma PHASTEST, sendo que, de acordo com a completude do genoma, 150 foram classificados como intactos, 92 como questionáveis e 262 como incompletos. 227 foram identificados nos clones de *Klebsiella* spp., 132 em *E. coli* e 145 em *Enterobacter* spp., havendo uma predominância de fagos incompletos em comparação aos completos e questionáveis para os três grupos, mas principalmente para as espécies de *Klebsiella* (Figura 1).

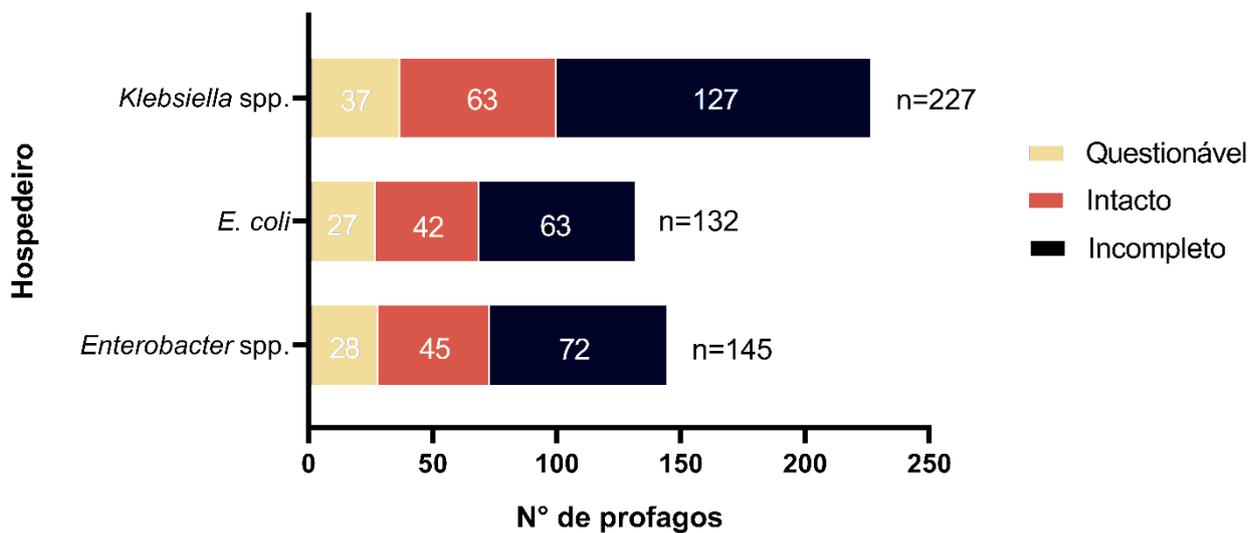


Figura 5. Número de profagos identificados em *Enterobacter* spp., *E. coli* e *Klebsiella* spp. de acordo com a classificação de completude do genoma. Gráfico autoral.

Considerando apenas os profagos completos e questionáveis, 100 genomas virais foram isolados a partir do cromossomo dos clones de alto risco em *Klebsiella* spp., 69 em *E. coli* e 73 em *Enterobacter* spp., sendo, portanto, 242 fagos submetidos as análises subsequentes. A média de distribuição de profagos por genoma foi semelhante entre as três enterobactérias, com cerca de 4-5 fagos por isolado. Interessantemente, os clones de *Klebsiella* spp. demonstraram uma maior variação na quantidade de profagos por genoma. O maior número identificado em um único genoma (n=8) foi para *Klebsiella pneumoniae* sequence type (ST) 11 e ST258 e em *E. coli* ST131 isolados a partir de bacteremia, enquanto o menor número (n=1) em *K. pneumoniae* ST307 e *Enterobacter hormoachei* STXIA provenientes de swab de colonização (Figura 1).

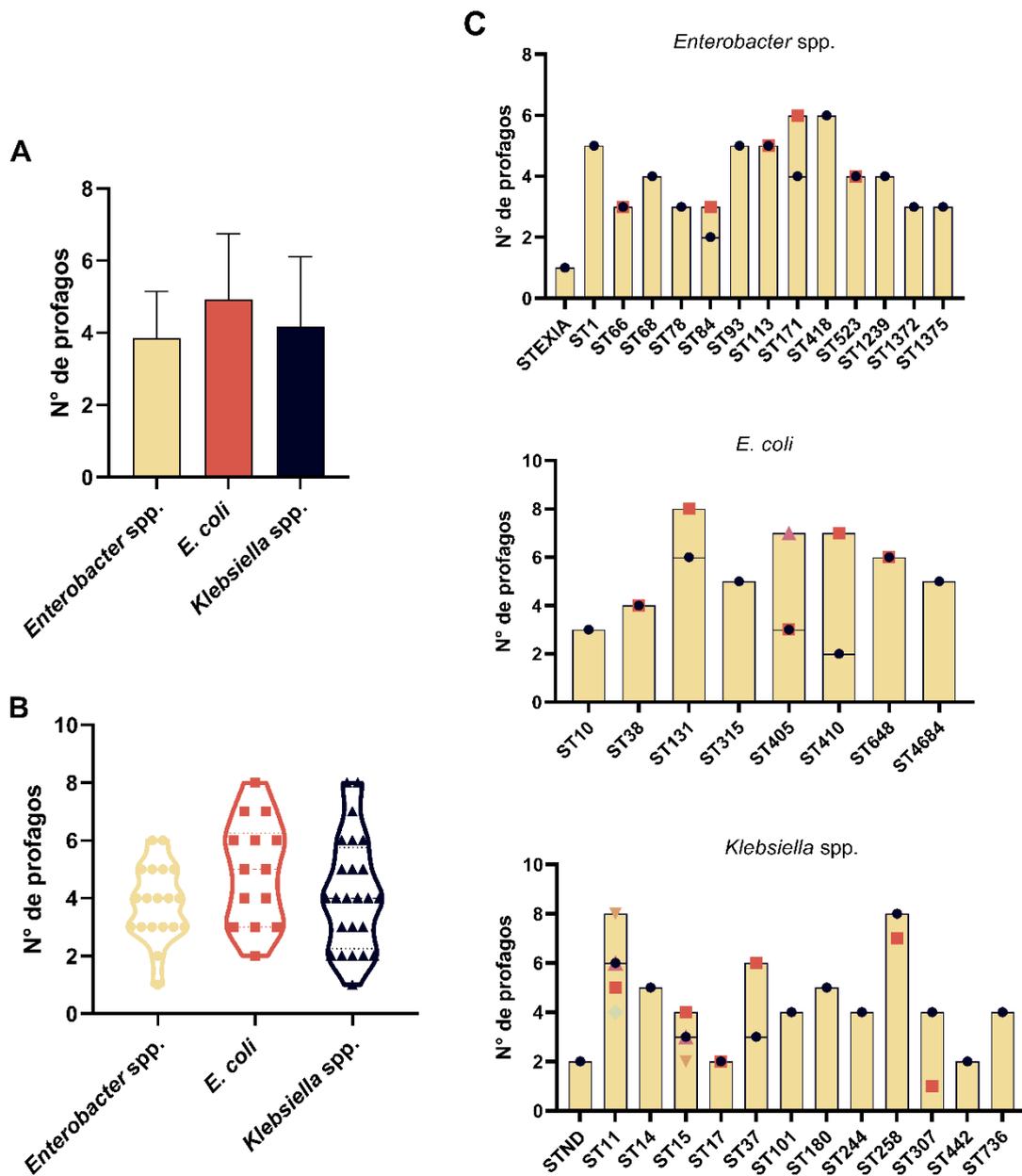


Figura 6. A) Comparativo da quantidade de profagos por genoma entre os três grupos de enterobactérias. Barras representam a média±desvio padrão do número de profagos por gênero analisado. B) Distribuição do número de profagos para clones de alto risco de *Enterobacter* spp., *E. coli* e *Klebsiella* spp. C) Quantidade de profagos por *sequence type* dos clones analisados. Figura autoral.

2. Agrupamento dos profagos com base em similaridade de sequência

A construção de árvore filogenética a partir do alinhamento múltiplo de sequências utilizando o MAFFT foi essencial para realização do agrupamento de fagos semelhantes. Com base nisto, visualizou-se uma grande diversidade genética entre os profagos isolados a partir das enterobactérias multidroga resistentes. 229 genomas foram atribuídos a 21 grupos diversos, nomeados de A até U, sendo: A ($n= 5$), B ($n= 11$), C ($n= 2$), D ($n= 37$), E ($n= 2$), F ($n= 18$), G ($n= 4$), H ($n= 8$), I ($n= 4$), J ($n= 7$), K ($n= 14$), L ($n= 10$), M ($n= 14$), N ($n= 14$), O ($n= 10$), P ($n= 23$), Q ($n= 19$), R ($n= 2$), S ($n= 8$), T ($n= 6$) e U ($n= 11$). Os 13 profagos restantes não puderam ser agrupados, portanto, foram tidos como *singleton* (Figura 7).

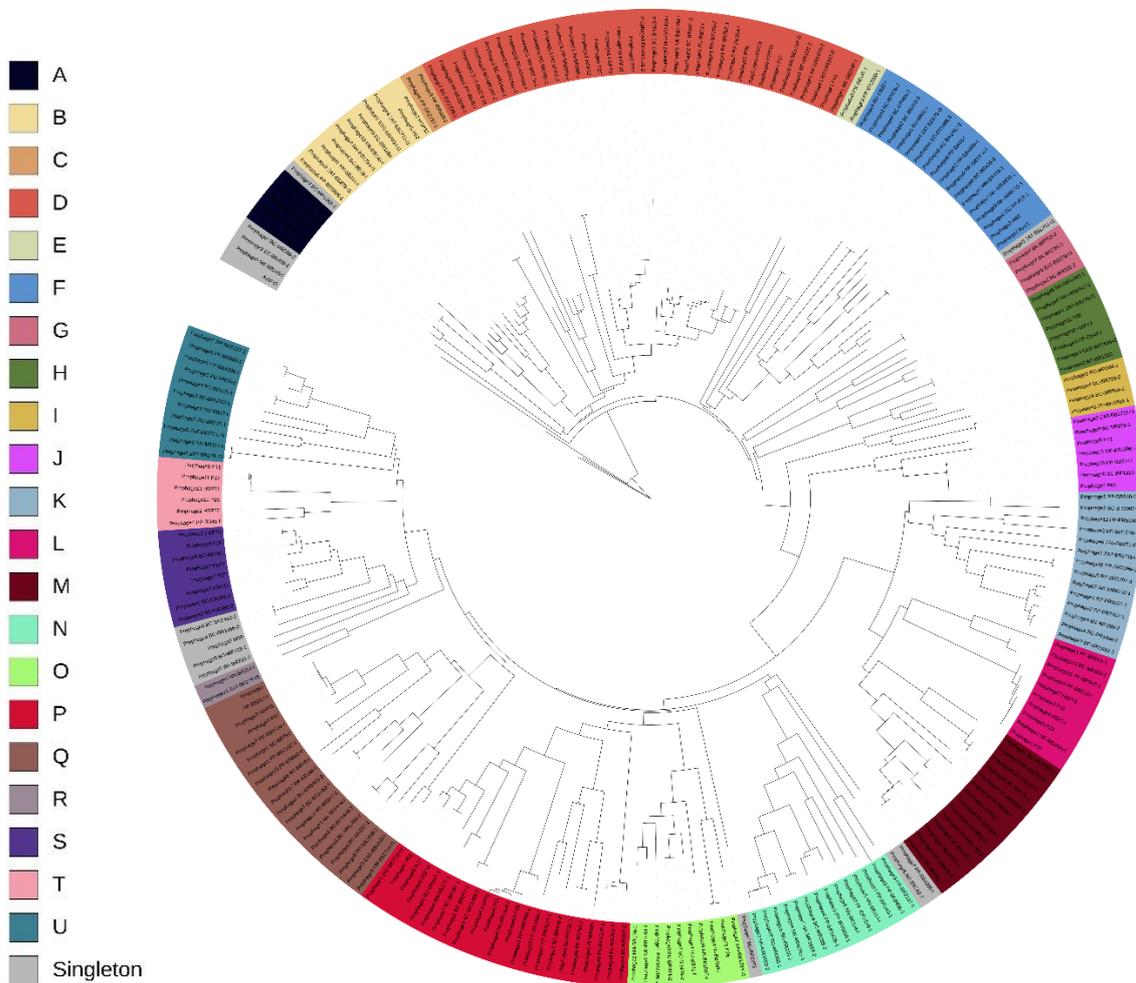


Figura 7. Construção de árvore filogenética através de alinhamento dos 242 genomas de profagos. Foram identificados 21 grupos baseados em similaridade de sequência. As sequências foram alinhadas através de MAFFT v7 e o resultado utilizado para construção da árvore pelo método Neighbor-Joining, sendo visualizada via iTOL v7.

Os profagos associados compartilhavam um valor de ANI superior a 40%, sendo o percentual de identidade representado através da intensidade de coloração em matriz *dot plot* (Figura 8). É possível visualizar uma predominância dos valores de identidade inferiores a 30%, evidenciando a variedade genética entre as sequências comparadas, enquanto a maior similaridade é visualizada principalmente no canto superior esquerdo da figura, atribuída aos *clusters* B, C, G, J, L, S e T. Algumas sequências atingiram um ANI de 100% quando comparadas, e aparecem plotadas ao redor da linha diagonal principal, estas apresentaram plena identidade de tamanho do genoma, conteúdo GC e número de genes, levando a inferência de que se trate do mesmo profago infectando cepas divergentes. Essa coexistência predominou para hospedeiros de mesmo ST, todavia, ST divergentes também foram identificados (Quadro 1).

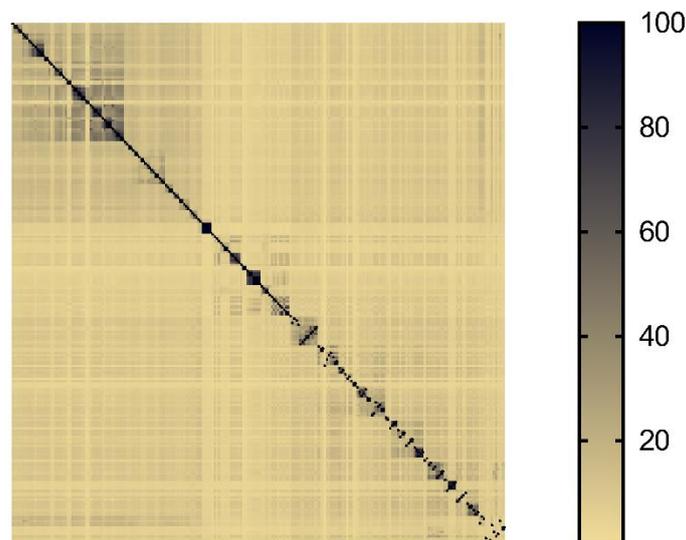


Figura 8. Matriz *dot plot* de comparação dos valores ANI entre as sequências alinhadas via MAFFT. Cores mais intensas demonstram percentuais mais altos de identidade e cores mais claras indicam baixa similaridade entre os profagos plotados naquela região. Figura autoral.

<i>Enterobacter spp.</i>		<i>E. coli</i>	
ID fago	ST	ID fago	ST
Prophage2_BK-BR752-3	113	Prophage3_NK-MR6742-1	315
Prophage6_BK-BR755-2		Prophage4_NK-MR1658-2	648
Prophage7_BK-BR752-3	113	Prophage1_NK-MR1658-2	648
Prophage7_BK-BR755-2		Prophage1_BC-MR388-2	
Prophage3_BC-BR476-2	171	Prophage2_BC-BR1462-2	38
Prophage4_RU-BB32-I		Prophage3_BC-BR1466-2	
Prophage5_PC-BR56-2	84	Prophage8_BC-BR1462-2	38
Prophage6_PC-BR295-1		Prophage8_BC-BR1466-2	
Prophage3_BC-BR1468-3	66		
Prophage3_BC-MR1459-2			
<i>Klebsiella spp.</i>			
ID fago	ST	ID fago	ST
Prophage8_BC-BR1283	14	Prophage2_P41	11
Prophage1_P02	15	Prophage7_HSP73	
Prophage3_PP-BB314-I	11		
Prophage5_P02	15	Prophage8_P41	11
Prophage3_HSP32		Prophage11_P27	258
Prophage1_HSP32	15	Prophage6_P26	11
Prophage1_IN-BB598-I		Prophage1_HSP73	
Prophage1_NK-BB1596-I	11	Prophage10_P26	11
Prophage1_P41		Prophage13_HSP31	258
Prophage2_NK-BB1596-I	11	Prophage4_P27	258
Prophage2_P26	11	Prophage5_HSP31	
Prophage1_P27	258		
Prophage4_HSP31	258		
Prophage5_NK-BB1596-I	11	Prophage9_P27	258
Prophage5_P41		Prophage10_HSP31	

Quadro 1. Enumeração dos bacteriófagos idênticos em caracterização e com valor de ANI igual a 100%. Os fagos de mesma identidade são apresentados em uma mesma caixa, bem como a descrição da classificação ST do hospedeiro originário. Quadro autoral.

3. Classificação taxonômica da coleção de bacteriófagos

Todos os profagos apresentaram material genético do tipo DNA de dupla fita (dsDNA) e foram classificados dentro da classe *Caudoviricetes*. Entretanto, apenas 1,4% do total de fagos completos ou questionáveis possuiu similaridade de sequência $\geq 70\%$ aos genomas de referência do ICTV ou os depositados no NCBI, portanto, somente uma pequena porcentagem pode ser classificada a nível de família, gênero e/ou espécie, sendo a maioria dos fagos classificados provenientes dos genomas de *E. coli*. Além disso, todos os profagos de *Enterobacter* spp. não apresentaram similaridade de sequência a nenhum outro fago descrito, e, portanto, não receberam classificação. Dos 34 profagos com correspondência as sequências dos bancos de dados, 16 obtiveram similaridade de sequência $\geq 95\%$, indicativo de identidade total ao fago correspondente ou de uma nova cepa da espécie descrita. Entretanto, do total, somente 16 apresentaram classificação a um ou mais dos três níveis taxonômicos. 12 fagos distribuídos nos grupos B, D ou P, são da família *Peduviridae*, de acordo com os critérios atuais de taxonomia do ICTV, sendo que os demais ainda não foram classificados a este nível, além disso, foram identificados 9 gêneros de fagos e 5 diferentes espécies (Quadro 2). O grupo D apresentou diversidade taxonômica a nível de gênero, com profagos *Elveevirus*, *Gegavirus*, *Inibicivirus*, *Peduvirus* e *Wadgaonvirus*. O gênero *Reipivirus* foi identificado no grupo B, além de *Reginaelenavirus*, enquanto *Punavirus* no grupo K e *Bievrevirus* em N.

SIMILARIDADE DE SEQUÊNCIA				CLASSIFICAÇÃO TAXONÔMICA		
Nome	Correspondência	SS	Grupo	Família	Gênero	Espécie
<i>Klebsiella</i> spp.						
Prophage6_BC-BB79-I	Klebsiella phage ST101-KPC2phi6.3	99%	Q	SC	SC	NE
Prophage5_P02	Klebsiella phage ST16-OXA48phi5.4	74%	B	Peduviridae	Reipivirus	NE
Prophage1_HSP32	Klebsiella phage ST15-OXA48phi14.1	75%	D	Peduviridae	Gegavirus	NE
Prophage6_HSP32	Klebsiella phage vB_Kpn_1825- KPC53	100%	M	SC	SC	NE
Prophage1_IN-BB598-I	Klebsiella phage ST15-OXA48phi14.1	75%	D	Peduviridae	Gegavirus	NE

Prophage2_IN-BB598-I	Klebsiella phage vB_Kpn_1825- KPC53	77%	M	SC	SC	NE
Prophage2_PP-BB1096-I	Klebsiella phage pJN2-26	98%	M	SC	SC	NE
Prophage1_ESS-BB0434-I	Klebsiella phage ST16-OXA48phi5.4	83%	B	Peduoviridae	Reipvivirus	NE
Prophage3_ZAT-BB2711-I1	Klebsiella phage ST974- OXA48phi18.2	87%	Sing.	SC	SC	NE
Prophage4_ZAT-BB2711-I1	Klebsiella phage ST16-OXA48phi5.4	73%	B	Peduoviridae	Reipvivirus	NE
Prophage5_ZAT-BB2711-I1	Caudoviricetes sp. isolate ctQMb4	90%	U	SC	SC	NE
Prophage12_P26	Caudoviricetes sp. isolate cto2z2	76%	H	SC	SC	NE
Prophage5_P27	Klebsiella phage ST11-VIM1phi8.2	99%	P	SC	SC	NE
Prophage6_HSP31	Klebsiella phage ST11-VIM1phi8.2	99%	P	SC	SC	NE
Prophage4_NN-BB1794-I1	Klebsiella phage 3LV2017	96%	B	Peduoviridae	Reginaelenavirus	Reginaelenavirus rv3LV2017
Prophage6_NN-BB1794-I1	Klebsiella phage 4LV2017	100%	D	Peduoviridae	Elveevirus	Elveevirus 4LV2017
<i>E. coli</i>						
Prophage5_NN-BB114-I	Escherichia phage 4A7	82%	N	SC	Bievrevirus	NE
Prophage6_NN-BB114-I	Escherichia phage vB_EcoS-689R5	99%	N	SC	SC	NE
Prophage9_NN-BB114-I	Escherichia phage CMS-2020 ^a	99%	M	SC	SC	NE
Prophage3_BC-BR1278-1	Escherichia phage vB_EcoM-613R3	100%	D	Peduoviridae	Inibicivirus	Inibicivirus 613R3
Prophage5_BC-BR1278-1	Escherichia phage vB_EcoS-666R9	100%	P	Peduoviridae	SC	NE
Prophage3_PP-BR1826-1	Yersinia phage vB_YpM_41	87%	I	SC	SC	NE

Prophage2_BC-MR366-2	Escherichia phage P2_2H4	73%	D	Peduoviridae	Peduovirus	NE
Prophage4_BC-MR366-2	Salmonella phage SJ46	74%	K	SC	Punavirus	NE
Prophage6_BC-MR366-2	Escherichia phage vB_EcoM-613R2	89%	I	SC	SC	NE
Prophage8_BC-MR366-2	Escherichia phage vB_EcoS-666R9	100%	P	SC	SC	NE
Prophage1_BC-BR228-2	Escherichia phage 500465-1	95%	D	Peduoviridae	Wadgaonvirus	Wadgaonvirus wv5004651
Prophage6_NK-MR1658-2	Escherichia phage Cyrano	86%	M	SC	SC	NE
Prophage2_BC-MR388-2	Escherichia phage EC.W1-9	89%	S	SC	SC	NE
Prophage4_BC-MR388-2	Escherichia phage vB_EcoS-683R6	95%	P	SC	SC	NE
Prophage2_BC-BR1462-2	Escherichia phage SIAC10	95%	D	Peduoviridae	Peduovirus	Peduovirus SIAC10
Prophage3_BC-BR1466-2	Escherichia phage SIAC10	95%	D	Peduoviridae	Peduovirus	Peduovirus SIAC10
Prophage4_BC-BR1466-2	Escherichia phage P1	82%	K	SC	Punavirus	NE
Prophage2_PP-BR2707-3	Escherichia phage P1	87%	K	SC	Punavirus	NE

Quadro 2. Classificação taxonômica de acordo com a similaridade de sequência aos fagos depositados em bancos de dados. Abreviações: SS= Similaridade de sequência; Sing.= Singleton; NE= Nova espécie; SC= Sem classificação. Quadro autoral.

4. Caracterização genômica dos agrupamentos de profagos

Os tamanhos de genoma variaram entre 3-108 kbp, sendo que a maioria dos fagos possuía comprimento entre 28-35 kbp. Os grupos taxonômicos apresentaram homogeneidade, exceto por K e M, devido a abrigarem tanto os maiores profagos (superiores a 70 kb), quanto alguns dos genomas moderados (25-70 kbp) e pequenos (inferiores a 25 kbp), além disso, os grupos J, O, P, R e T também apresentaram certa heterogeneidade de comprimento, sendo que J, L, T e Singleton demonstraram predominância de genomas pequenos (figura 9A). Seguindo o mesmo padrão, o número de genes,

que compreendeu valores entre 6 e 132, também foi parcialmente uniforme, apresentando-se heterogêneo para os mesmos grupos referidos anteriormente, com adição do grupo E (figura 9B), isto ocorre devido a correlação moderada positiva entre o tamanho do genoma e número de genes (figura 9D). Em contrapartida, o conteúdo GC variou entre 43,45% e 57,07%, com grande uniformidade para um mesmo grupo, mas também entre distintos (figura 9C), havendo correlação nula entre esta variável e o tamanho do genoma (figura 9E).

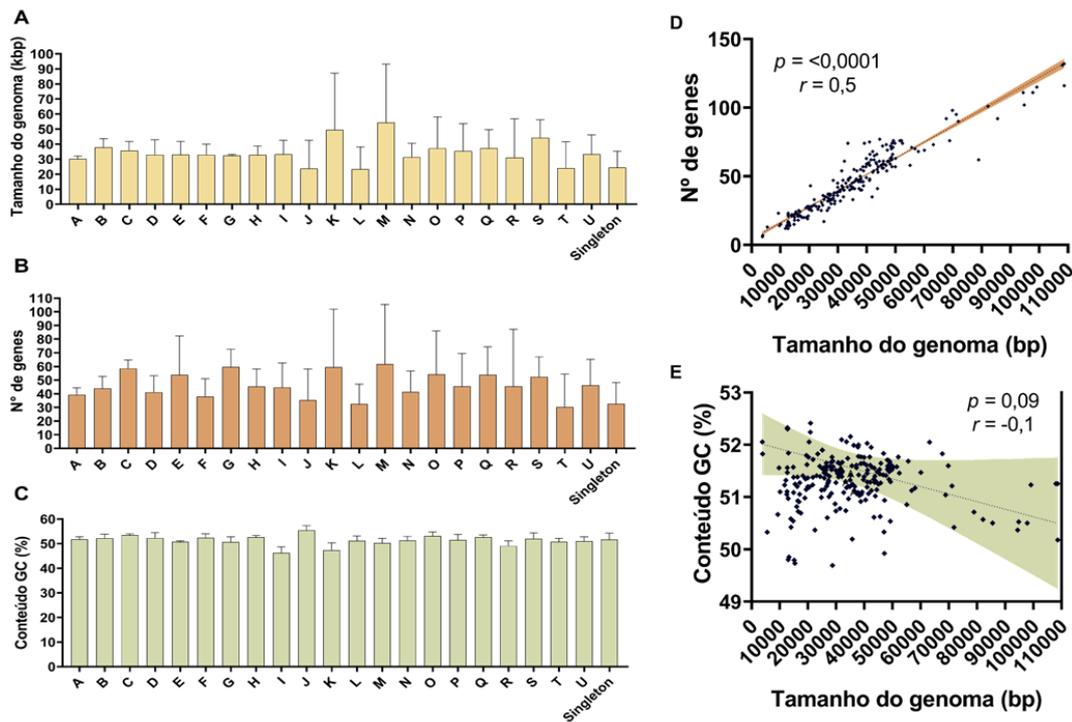


Figura 9. Características genômicas dos diferentes grupos taxonômicos identificados. A) Gráfico do tamanho dos genomas em kbp. B) Gráfico do número de genes previstos. C) Gráfico do percentual de conteúdo de guanina+citosina (GC). D) Correlação linear entre tamanho do genoma e número de genes para as sequências analisadas. E) Correlação linear entre tamanho do genoma e conteúdo GC para as sequências analisadas. Figura autoral.

Por meio da anotação genômica, foi possível verificar uma predominância de proteínas estruturais associadas à montagem de novas partículas virais (14%), como *portal protein*, *major* e *minor head protein*, *tail-collar fibre protein*, *tape measure protein*, *major capsid protein*, entre outras, além das cruciais para infecção do hospedeiro (16%), como *minor tail protein*, *tail tube protein* e *baseplate hub*. As proteínas associadas à regulação do processo de replicação representaram 7% do total (fatores transcricionais, *antitermination proteins*, proteínas regulatórias),

enquanto de empacotamento (*terminase small e large subunit*, endonucleases, exonucleases) e replicação do material genético (polimerases, helicases, primases, ligases, *DNA binding*) foram 6% das proteínas anotadas. Verificou-se ainda a presença em todos os genomas de proteínas essenciais para o processo de lise do hospedeiro, principalmente lisinas, holinas e lisozimas. Entretanto, o predomínio maior esteve entre as proteínas hipotéticas (15%) e aquelas que ainda não possuem classificação funcional (20%), sendo incluídas nesta classe, as de função putativa atribuída ao hospedeiro.

Em 16% dos profagos não havia genes codificando para proteínas associadas ao processo de integração ao material genético do hospedeiro (como integrases, recombinases, repressores e excisionases), sendo que desse número, 21% também não possuía proteínas associadas a imunidade, ou seja, relacionadas a inibição da infecção do hospedeiro por outros fagos filogeneticamente semelhantes e/ou evasão do sistema de restrição-modificação (R-M) das bactérias (*superinfection exclusion proteins* e transferases), apresentando mesmo percentual de ausência de tRNA. Essas ausências são indicativas de predição de ciclo lítico ou perda de regiões gênicas relevantes para o ciclo lisogênico.

Dentre os agrupamentos, o grupo J demonstrou a mais baixa caracterização do perfil genômico, com predominância de proteínas hipotéticas. Alguns grupos, como C, E, G, K, M e R, tiveram uma caracterização moderada, com grande número de hipotéticas e não classificadas, mas um número razoável de outras proteínas de funções previstas. Os demais tiveram médias de identificação de proteínas semelhantes para as classes. De modo geral, houve uma média baixa de identificação de proteínas associadas a imunidade com exceção do grupo C (Figura 10).

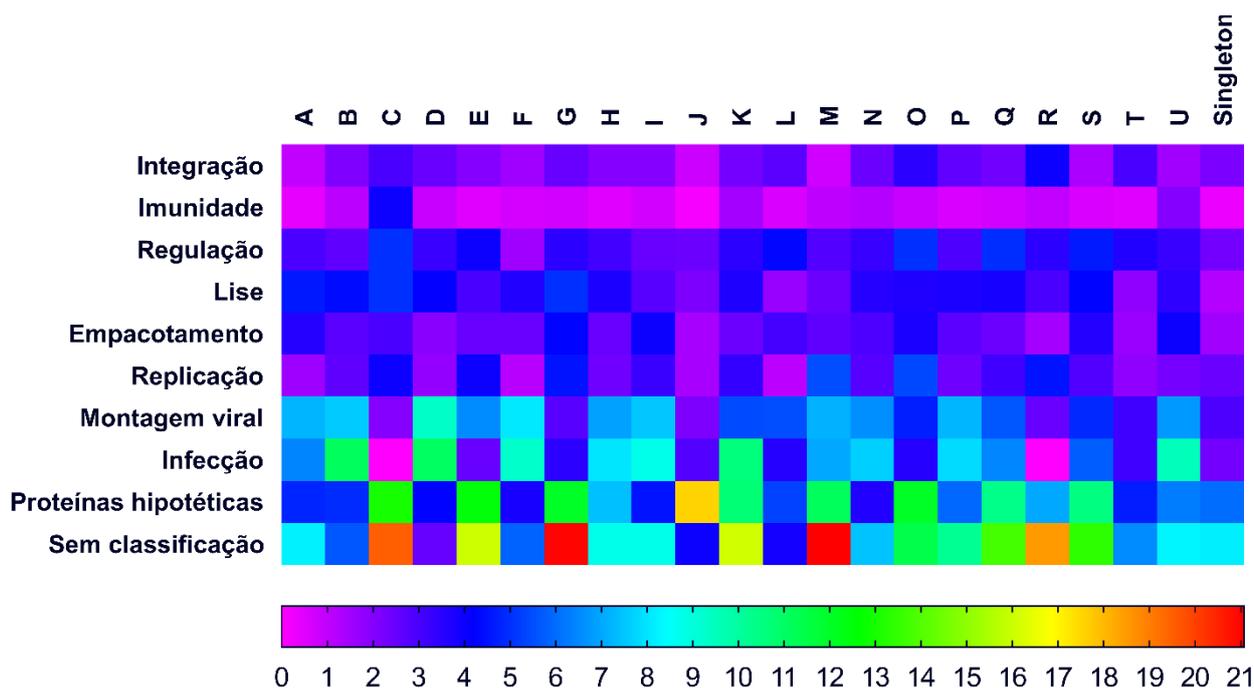


Figura 10. Média de distribuição de proteínas por classes funcionais para os agrupamentos de fagos. A intensidade de cores reflete proporcionalmente o valor da média. Figura autoral.

A análise genômica comparativa de 3 profagos de baixa similaridade, portanto, de grupos diferentes, e selecionados especificamente para representar cada um dos gêneros de bactérias avaliados neste estudo, evidenciou uma baixa similaridade nucleotídica entre as sequências, especialmente para o profago representativo de *Enterobacter* em comparativo aos outros. A única homologia compartilhada entre as três sequências ocorreu para as proteínas associadas ao processo de lise, especificamente entre a holina, codificada pelo profago do gênero *Enterobacter*, *Rz-like spanin* do fago de *E. coli* e holina e endolisina do bacteriófago de *Klebsiella* spp. O tRNA codificado pelos fagos de *Klebsiella* spp. e *E. coli* apresentaram similaridade de sequência, bem como, destaca-se também para estes dois isolados, as semelhanças nas estruturas principalmente de proteínas hipotéticas, além de fatores transcricionais (*CII-like transcriptional activator* e *transcriptional repressor*), uma proteína de função putativa do hospedeiro (*gam-like host nuclease inhibitor*); e a presença de uma sequência de genes que codificam o regulador transcricional da família MarR, 2 proteínas hipotéticas e a proteína flagelar, sugerindo a homologia para um operon do hospedeiro (Figura 11).

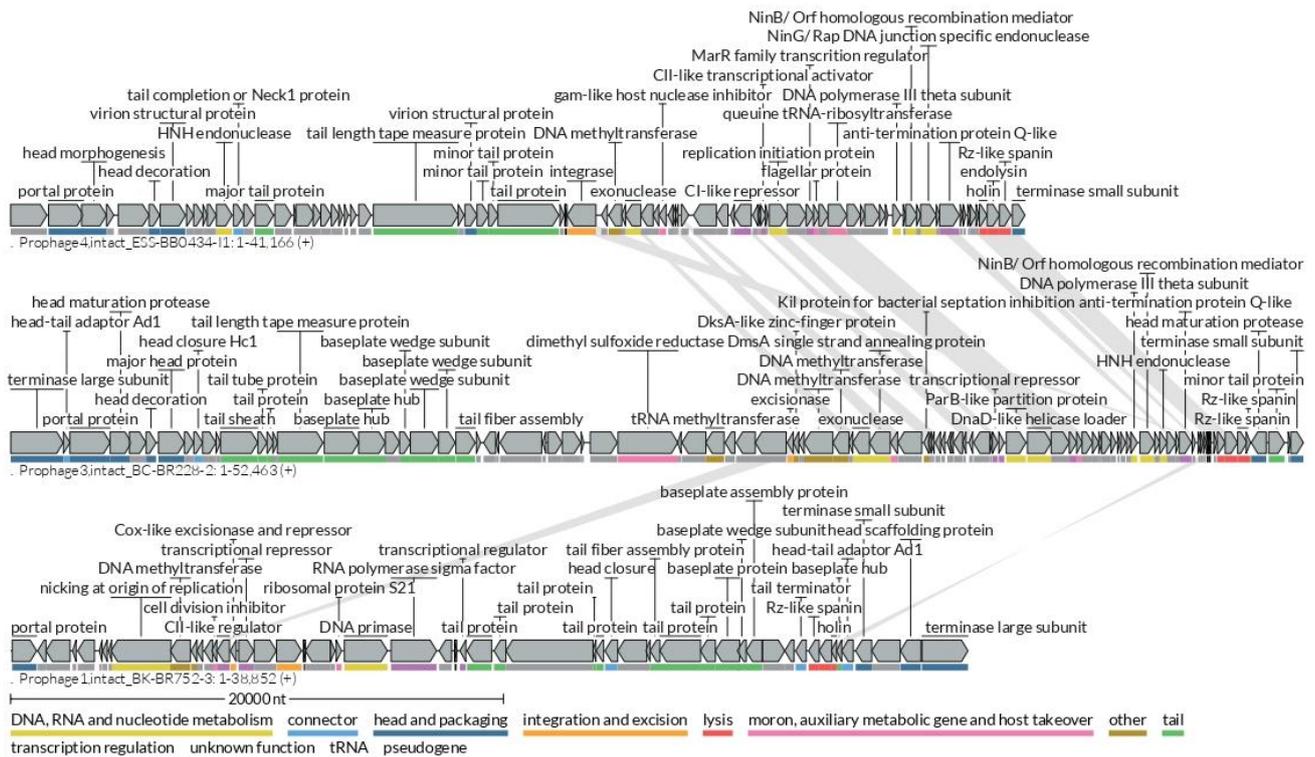


Figura 11. Análise comparativa dos mapas genômicos de três profagos de baixa similaridade e isolados de gêneros bacterianos distintos. As setas apontam para o sentido de replicação dos genes, sendo a função atribuída demonstrada pelas diferentes cores apresentadas. Os nomes das proteínas codificadas por cada gene são mostrados acima de cada seta. As regiões conectadas por linhas cinza entre os genomas indicam homologia de proteínas através de MMSeqs2. Prophage4_ESS-BB0434-11 pertence ao grupo H e representa *Klebsiella* spp. Prophage3_BC-BR228-2 pertencente ao grupo J e representa a espécie *E. coli*; Prophage1_BK-BR752-3 pertencente ao grupo D e é representativo do gênero *Enterobacter*. Mapas gerados a partir de LoVis4u version 0.0.12.

5. Genes de funções putativas ao hospedeiro

Conforme descrito anteriormente, a maior parte dos genes não possuiu classificação, e em sua grande maioria isso ocorreu devido a função da proteína estar atribuída ao hospedeiro. A maior ocorrência desses genes esteve atrelada ao metabolismo, transcrição e tradução, e transportadores de membrana, principalmente do tipo ABC.

Ao todo, foram identificados 50 genes associados a resistência a antibióticos ou fatores de virulência em 19 diferentes profagos, com completude do genoma variável (Quadro 2). Dentre os diferentes mecanismos de resistência a antibióticos identificados, houve predominância de bombas

de efluxo e porinas, como para Prophage1_PP-BB1096-I e Prophage3_ESS-BB0434-I, em que foi descrita a presença da bomba de efluxo multidrogas da família RND codificada pelo complexo OqxAB. Houve também a presença de enzimas inativadoras de antibióticos, sobressaindo-se principalmente Prophage3_PP-BR5996-1, apresentando as metalobetalactamases codificadas pelos genes NDM-1 e ble(MBL) que conferem resistência aos carbapenêmicos.

Foi visualizado que para alguns dos profagos houve mais de um gene identificado associado a fatores de virulência, e nestes casos, as múltiplas proteínas codificadas estavam funcionalmente relacionadas. Destacam-se Prophage1_PP-BR2096-1 com proteínas de síntese e funcionamento de flagelos peritríquios; a presença do operon *ecpRABCDE* em Prophage3_BC-MR388-2 codificando para o pilus comum (ECP) em *E. coli*; o sistema de dois componentes BaeSR promotor do controle da expressão de porinas e bombas de efluxo em resposta ao estresse ambiental em Prophage2_BC-MR366-2; e para Prophage4_HSP73, múltiplas proteínas do sistema de secreção do tipo 6 (T6SS), um aparato molecular de atividade antibacteriana e modulador da resposta do hospedeiro. Identificou-se ainda o *locus rfb* em Prophage6_PP-BR2137-1, determinante para a síntese do antígeno O de LPS, atrelado a variedade antigênica entre sorotipos de enterobactérias.

ID fago	Completo	Gene	Proteína	Mecanismo de resistência/virulência
Enterobacter spp.				
Prophage3_RU-BB32-I	Intacto	<i>gtrB</i>	Bactoprenol glicosiltransferase	Síntese da parede celular
Prophage3_BC-BR491-2	Intacto	<i>gtrB</i>	Bactoprenol glicosiltransferase	Síntese da parede celular
Prophage2_NN-BR176-1	Intacto	<i>gtrB</i> <i>gtrA</i>	Bactoprenol glicosiltransferase	Síntese da parede celular
Prophage1_PP-BR2096-1	Intacto	1. <i>gtrA</i> 2. <i>fliA</i> 3. <i>fliF</i> 4. <i>gtrB</i> 5. <i>fliI</i> 6. <i>fliG</i> 7. <i>ppdD</i> 8. <i>fliY</i>	1. Bactoprenol glicosiltransferase 2. Fator sigma da biossíntese flagelar (flagelos peritríquios) 3. Proteína do anel MS flagelar (flagelos peritríquios) 4. Bactoprenol glicosiltransferase	1 e 4. Síntese da parede celular 2, 3, 5, 6 e 8. Motilidade Adesão

			<p>5. ATP sintase específica do flagelo (flagelos peritríquios)</p> <p>6. Proteína G do interruptor motor flagelar (flagelos peritríquios)</p> <p>7. Subunidade principal da pilina</p> <p>8. Subunidade transportadora de cistina (flagelos peritríquios)</p>	
Prophage2_ PP-BR2096-1	Questionável	<i>FosA2</i>	Glutaciona transferase FosA	Inativação de antibióticos
Prophage6_ PP-BR2137-1	Questionável	<i>LPS rfb locus</i>	Proteínas diversas associadas a síntese de antígeno O do lipopolissacarídeo (LPS)	Variedade antigênica de sorotipos
Prophage1_ PP-BR663-2	Intacto	<i>Ecol_mdfA/cmlA/cmr</i>	Transportador multidrogas MdfA	Efluxo de antibióticos
<i>E. coli</i>				
Prophage5_ NN-BB114-I	Intacto	<ol style="list-style-type: none"> <i>Ecol_emrE</i> <i>nmpC</i> 	<ol style="list-style-type: none"> Transportador multidrogas EmrE Suposta proteína porina da membrana externa NmpC 	Efluxo de antibióticos
Prophage1_ ZAT-BB279-I3	Questionável	<i>nmpC</i>	Suposta proteína porina da membrana externa NmpC	Porina
Prophage2_ BC-MR366-2	Intacto	<ol style="list-style-type: none"> <i>yegO/mdtC</i> <i>baeS</i> <i>baeR</i> 	<ol style="list-style-type: none"> Proteína de multidroga resistência MdtC Transdução de sinal histidina-proteína quinase BaeS Proteína reguladora transcricional BaeR 	<ol style="list-style-type: none"> Efluxo de antibióticos e 3. Controle da expressão de bomba de efluxos e porinas
Prophage3_ BC-BR228-2	Intacto	<i>FlhE</i>	Proteína flagelar (flagelos peritríquios)	Motilidade

Prophage4_ BC-BR228-2	Questionável	1. <i>Ecol_emrE</i> 2. <i>nmpC</i>	1. Transportador de multidrogas EmrE 2. Suposta proteína porina da membrana externa NmpC	1. Efluxo de antibióticos 2. Porina
Prophage3_ BC-MR388-2	Intacto	1. <i>yagX/ecpC</i> 2. <i>yagW/ecpD</i> 3. <i>ykgK/ecpR</i> 4. <i>yagZ/ecpA</i> 5. <i>yagY/ecpB</i> 6. <i>ecpE</i>	1. EcpC, portadora do pilus comum de <i>E. coli</i> 2. EcpD, portador do pilus comum de <i>E. coli</i> 3. Proteína reguladora EcpR 4. Subunidade estrutural do pilus comum de <i>E. coli</i> EcpA 5. Chaperona do pilus comum de <i>E. coli</i> EcpB 6. Proteína fimbrial putativa de <i>E. coli</i> EcpE	Adesão
Prophage3_ PP-BR5996-1	Questionável	1. <i>NDM-1</i> 2. <i>rmtC</i> 3. <i>AAC(6')-Ib7</i> 4. <i>BRP(MBL)/ble/blmA/blmS/BRP</i> 5. <i>sul1</i> 6. <i>qacEdelta1</i>	1. Metalo-beta-lactamase tipo 2 2. 16S rRNA (guanina(1405)-N(7))-metiltransferase 3. Aminoglicosídeo N(6')-acetiltransferase tipo Ib 4. Proteína de resistência à bleomicina 5. Di-hidropteroato sintase tipo 1 6. QacEdelta1	1. Inativação de antibióticos 2. Alteração do alvo do antibiótico 3. Inativação de antibióticos 4. Inativação de antibióticos 5. Substituição do alvo do antibiótico 6. Efluxo de agentes antissépticos
<i>Klebsiella spp.</i>				
Prophage8_ BC-BB79-1	Intacto	<i>rcaA</i>	RcsAB	Ativador transcricional para biossíntese de cápsula

Prophage2_ HSP32	Intacto	<i>mdtK</i>	Proteína de multidroga resistência MdtK	Efluxo de antibióticos
Prophage1_ PP-BB1096-I	Intacto	<i>oqxA</i> <i>oqxB</i>	Subunidade adaptadora periplásmica OqxA e OqxB do transportador RND de efluxo de múltiplas drogas	Efluxo de antibióticos
Prophage3_ ESS-BB0434- I	Intacto	<i>oqxA</i> <i>oqxB</i>	Subunidade adaptadora periplásmica OqxA e OqxB do transportador RND de efluxo de múltiplas drogas	Efluxo de antibióticos
Prophage4_ HSP73	Intacto	1. <i>hcp/tssD</i> 2. <i>vgrG/tssI</i> 3. <i>clpV/tssH</i> 4. <i>dotU/tssL</i> 5. <i>vipB/tssC</i> 6. <i>vipA/tssB</i> 7. <i>ompA</i> 8. <i>vasE/tssK</i>	1. Proteína do sistema de secreção tipo VI, família Hcp; 2. Proteína da ponta do sistema de secreção tipo VI VgrG 3. ATPase TssH do sistema de secreção tipo VI 4. Proteína do sistema de secreção tipo VI, família DotU/TssL 5. Subunidade grande VipB da bainha contrátil do sistema de secreção tipo VI 6. Subunidade pequena VipA da bainha contrátil do sistema de secreção tipo VI 7. Proteína A da membrana externa/Domínio de ligação ao peptidoglicano de TagL 8. Subunidade TssK da placa de base do sistema de secreção tipo VI	Sistema de Secreção do Tipo VI

Quadro 3. Descrição dos genes identificados em profagos associados a resistência a antibióticos e fatores de virulência no hospedeiro. A completude do genoma do fago, bem como as proteínas

codificadas pelos genes e suas funções previstas nos bancos de dados CARD e VFDB estão atribuídas nas colunas. Quadro autoral.

6. Previsão de hospedeiros e predição de ciclo de vida

Ao analisar cada fago acerca do gênero bacteriano do qual foi originalmente isolado, descobriu-se que não há um padrão preciso de hospedeiros para cada grupo. Embora A e T tenham apresentado apenas fagos isolados de *Klebsiella* spp., I e R somente de *E. coli* e C unicamente de *Enterobacter* spp., a grande maioria dos grupos apresentava dois ou os três gêneros, demonstrando variações de predomínio, mas sem padrões de exclusividade de hospedeiro (Figura 8).

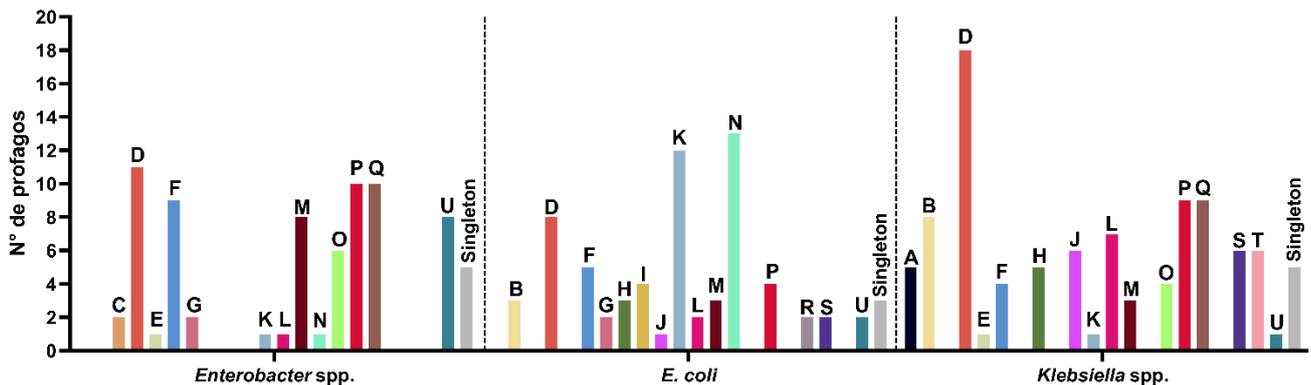


Figura 12. Ocorrência dos agrupamentos de profagos de acordo com o gênero bacteriano do qual foram originalmente isolados. Figura autoral.

Analisando a predição de provável hospedeiro individualmente para cada fago por meio das plataformas elencadas, observou-se uma alta concordância entre o resultado obtido e o hospedeiro original de isolamento, com 71% de concordância neste parâmetro. Nos casos de divergência, o hospedeiro previsto também fazia parte da família *Enterobacteriaceae*, entretanto, foram descritas espécies diferentes da originária (Figura 13). A exceção ocorreu em Prophage2_BC-BR1468-3 e Prophage2_BC-MR1459-2 originalmente de *Enterobacter cloacae*, classificados no grupo Q, que curiosamente tiveram a predição de hospedeiro como *Listeria monocytogenes*.

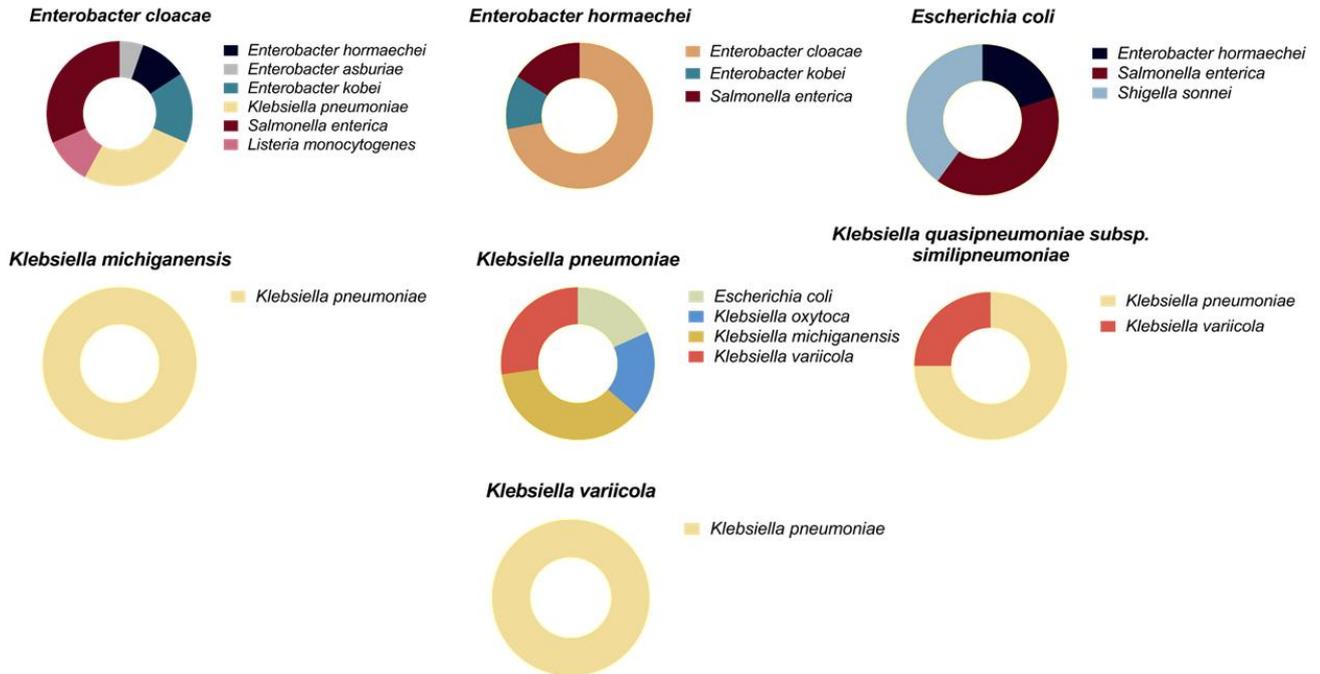


Figura 13. Divergências identificadas entre os hospedeiros originais de isolamento e os previstos para os profagos utilizados no estudo. Cada gráfico representa a espécie bacteriana originária de isolamento, sendo as partes do gráfico representativas da previsão de hospedeiros para os profagos. Figura autoral.

Quanto ao ciclo de vida, diversas incongruências foram encontradas entre as plataformas, tendo sido necessário avaliar em cada genoma, a presença de genes associados à lisogenia para categorizar a predileção de ciclo de vida. 93% dos fagos foram classificados como lisogênicos, distribuídos de modo homogêneo entre os grupos, com exceção do grupo J, no qual os fagos foram predominantemente líticos (Figura 14). PhageScope obteve o melhor resultado em concordância com a predileção final, categorizando 93% para o ciclo lisogênico, seguido de PhageTYP com 77%; 74% para PhageAI; 67% em BACPHLIP; e por fim, PhageLeads com apenas 55%, sendo a plataforma com a maior discrepância. PhageLeads e PhageTYP foram as únicas plataformas que apresentaram erros de submissão da sequência em 13% e 14% dos casos, respectivamente.

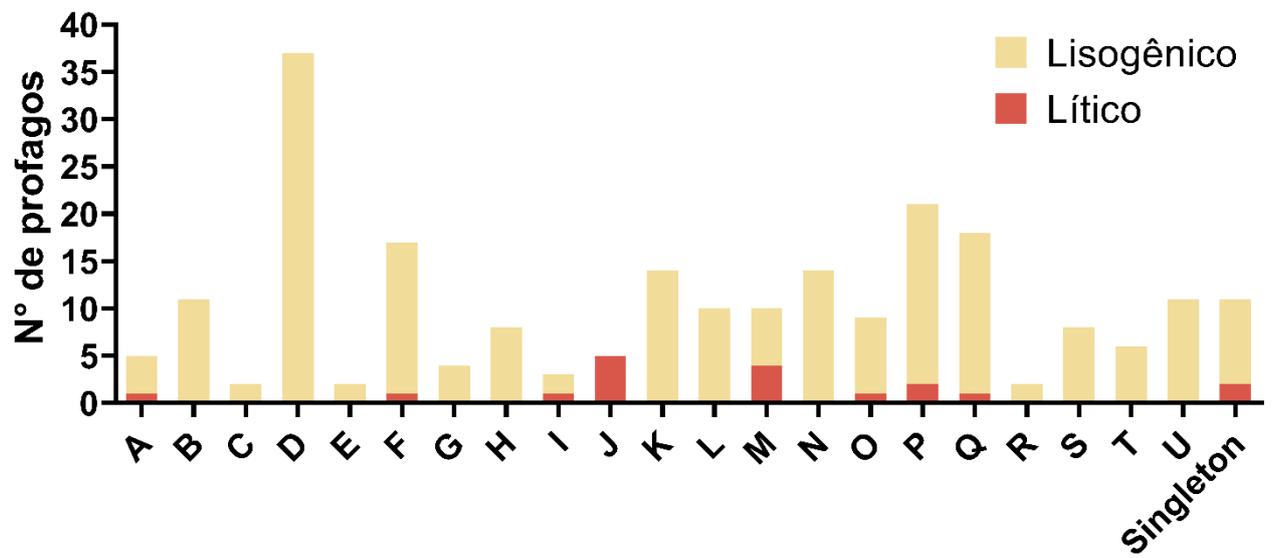


Figura 14. Predominância entre os grupos de profagos das predições de ciclo de vida em lisogênico ou lítico. Figura autoral.

DISCUSSÃO

A resistência aos antimicrobianos é um cenário de preocupação global e tem sido alvo de intensas pesquisas nas últimas décadas. Soluções inovadoras são constantemente buscadas⁵¹, entretanto, grandes contribuições podem ser encontradas ao se voltar para o passado, reavendo estudos acerca da terapia fágica¹⁹, principalmente ao estudar o contexto das relações entre fago-hospedeiro. É essencial compreender o papel dos profagos na perpetuação de patógenos, em especial, as enterobactérias multidroga resistentes, que fazem parte do grupo crítico da Lista de Patógenos Prioritários da OMS.

No presente estudo, realizou-se a análise de 57 clones de alto risco de *Enterobacter* spp., *E. coli* e *Klebsiella* spp., obtidos a partir de isolados clínicos ou da microbiota de seres humanos e animais de 8 diferentes países. A identificação de ao menos um profago por genoma, entra em concordância com o descrito para estirpes patogênicas⁵², além disso, os dados aqui descritos reafirmam que patógenos da família *Enterobacteriaceae* tendem a apresentar abundância na quantidade de profagos, sugerindo que a patogenicidade de uma cepa pode estar atrelada ao número de profagos que carrega⁵³. A distribuição heterogênea de fagos por genoma visualizada em *Klebsiella* spp. pode ser indicativa de trajetórias evolutivas diferentes entre os isolados. Todavia, a principal limitação do estudo está envolvida no fator de que os mecanismos de identificação de profagos ainda passa por refinamentos, e embora já existam ferramentas de identificação como PHASTEST, pode haver uma subestimação ou até uma superestimação do número de sequências identificadas que são de fato profagos⁵⁴.

Destaca-se que os isolados utilizados para os experimentos fazem parte do grupo de bactérias de elevada prioridade global, portanto, são fornecidas novas evidências de que a inserção de bacteriófagos nos genomas pode estar contribuindo para o surgimento e ascensão de clones de elevado risco. Essa hipótese é amparada ao verificar que as cepas que carregavam as maiores quantidades de profagos foram isoladas a partir de infecção de corrente sanguínea e pertencem aos ST de elevada incidência mundial. Estes são amplamente difundidos em ambientes hospitalares, principalmente em unidades de terapia intensiva (UTIs), e estão associadas à uma elevada mortalidade, além de terem sido marcados por uma rápida disseminação após as primeiras descrições na década de 2000⁵⁵. O ST11 e ST258 pertencem ao complexo clonal 258, e fazem parte das cepas de *Klebsiella pneumoniae* produtoras de carbapenemases (CP-Kp)⁵⁶, enquanto *E. coli* ST131 predomina como linhagem extraintestinal caracteristicamente hipervirulenta, produtora de

beta-lactamases de espectro estendido (ESBL) e resistência a fluoroquinolonas^{57,58}. Em contrapartida, os menores números de profagos foram provenientes de ST menos incidentes e foram identificados em isolados de colonização, portanto, mesmo que as cepas tenham um potencial patogênico, ainda assim não provêm de um quadro infeccioso, reforçando a hipótese.

Outro ponto é a predominância evidente de fagos incompletos para os três gêneros bacterianos, isso está diretamente associado a pressão seletiva sofrida por esses vírus, que passam por eventos de degradação por parte das defesas do hospedeiro e mutações ao longo do tempo, bem como, ao se inserirem no material genético, podem perder regiões codificantes significativas^{59,60}. Entretanto, embora profagos incompletos (descritos na literatura como crípticos) percam o potencial de “saltar” do genoma para um ciclo lítico, esses elementos defectivos permanecem permanentemente integrados ao genoma, conferindo ainda as vantagens genéticas ao hospedeiro⁶¹.

Todos os aspectos destacados demonstram a contribuição direta da distribuição de fagos no genoma para diversidade genética clonal, bem como o papel relevante da conversão lisogênica para o surgimento de clones hipervirulentos e de acelerada propagação, ou seja, mais fagos inseridos no genoma conferem mais características de adaptabilidade a diferentes condições, e portanto, favorecem o aspecto evolutivo do hospedeiro bacteriano, para esse sentido, ao conferir um fenótipo patogênico para uma bactéria, há um benefício direto para os bacteriófagos, que tem seus hospedeiros sendo mais perpetuados na natureza^{62,63}.

Através dos resultados obtidos, ampliou-se ainda mais a percepção de diversidade dentro da classe *Caudoviricetes*, pois, embora todas as sequências tenham sido classificadas taxonomicamente como tal e muitas tenham apresentado certa similaridade nucleotídica entre si, ainda sim, 21 *clusters* distintos filogeneticamente foram identificados, com predominância de ANI <30% e poucos grupos de alta similaridade, evidenciando uma grande variabilidade genética. Os grupos de maior similaridade apontam para uma maior relação evolutiva entre seus membros⁶⁰.

Demonstrou-se um número muito reduzido de fagos com similaridade de sequência aos genomas depositados nos principais bancos de dados de referência, explicitando que a coleção descrita apresenta sequências que até onde se sabe ainda não foram descritas, apontando para a inovação dos dados apresentados, além de indicar para uma baixa caracterização e escassez de dados sobre fagos do gênero *Enterobacter*. Houve ausência de classificação taxonômica a nível de família, gênero e espécie para maioria dos genomas, expressando a diversificação de fagos dsDNA caudados. É relevante ressaltar que as sequências obtidas e anotadas neste projeto podem ser

utilizadas para posteriores estudos direcionados, classificando por meio de genômica comparativa e homologia de proteínas com fagos representativos catalogados até o momento pelo ICTV⁴⁰, ou ainda para proposição de novas famílias, gêneros ou espécies, contribuindo diretamente para reclassificação taxonômica.

Embora a pluralidade de agrupamentos identificados e uma grande amplitude dos comprimentos de genomas, o intervalo de tamanho é consistente com o descrito para os fagos de enterobactérias, com a maioria dos grupos apresentando tamanho moderado^{64,65}. Há uma tendência de decaimento nos valores dos genomas, tendo sido identificados somente alguns profagos grandes localizados em apenas dois *clusters*, seguido por comprimentos cada vez menores distribuídos entre os grupos, algo que corrobora com os fenômenos já descritos anteriormente de degradação por múltiplos eventos ao longo da escala evolutiva⁵⁹, é esperado que esses bacteriófagos pequenos, ainda que possuam completude questionável ou intacta pela presença de algumas proteínas funcionalmente relevantes, em um futuro evolutivo próximo, se tornem profagos crípticos. Isso é sustentado quando se observa que os grupos de predominância de genomas pequenos também apresentaram um menor número de proteínas associadas a replicação, lise e empacotamento, que são descritos como significativos para a perda da capacidade de mudança para um ciclo lítico⁶¹. É interessante ressaltar que os *clusters* K e M que continham os maiores genomas, também continham alguns dos menores, levantando a hipótese de que o ancestral comum ao infectar diferentes hospedeiros e ser exposto a diversos eventos, como mutações pontuais, rearranjos genômicos, invasão por outros elementos móveis de DNA e deleção maciça, originou cepas virais heterólogas, porém ainda com certas características conservadas conferindo similaridade relevante entre esses fagos. Essa inferência pode ser descrita para outros grupos em que há ocorrência de genomas moderados e pequenos em conjunto.

Tal como descrito para outras enterobactérias, existe uma correlação positiva entre a quantidade de genes previstos e o tamanho do genoma⁶⁶, indicando que a capacidade de codificar proteínas está diretamente associada à quantidade total de DNA e, portanto, a uma elevada densidade gênica. Apresentam também homogeneidade do conteúdo GC dentro dos agrupamentos, com baixa variação entre os *clusters*, corroborando com o pressuposto que geralmente os valores de GC guardam maiores semelhanças dentre grupos filogeneticamente relacionados e entre membros que vivem em ambientes semelhantes, além disso, os fagos aqui descritos apresentaram valores moderados de GC, o que pode ser indicativo de possuírem uma certa estabilidade do material genético e resistência as pressões externas do ambiente, contudo, sendo ainda

moderadamente suscetíveis a esses fatores^{67,68}. Anteriormente, uma correlação entre menores valores de GC e os maiores comprimentos de genoma foi descrita, algo não observado para este conjunto de dados, mas, como no presente estudo houve um n menor e baixa variação em GC em relação ao estudo citado, não é possível afirmar com certeza que há uma oposição ao descrito⁶⁹.

Toda a análise da anotação genômica conduzida para a coleção de profagos, evidenciou a elevada variação dos genomas desses vírus. As diferenças quantitativas e qualitativas de tipos de genes entre cada grupo, reafirma a presença de diferentes histórias evolutivas de trocas horizontais genéticas entre fagos⁷⁰. A classe *Caudoviricetes* é marcada por genes de função conservada entre seus membros, o que permite o agrupamento em módulos funcionais, essencialmente codificando *major capsid protein*, *portal protein*, *terminase small* e *large subunit*⁷¹. Em concordância com o descrito para essa taxonomia, das proteínas com classificação atribuída, há predomínio das de montagem viral em todos os grupos, demonstrando uma tendência funcional conservativa para essa classe, entretanto, há uma variabilidade quantitativa e qualitativa intergrupos, devido as diferenças morfológicas apresentadas pelos fagos⁷³. Esse aspecto conservativo funcional se repete para as classes de replicação e empacotamento, embora exista uma variedade de mecanismos descritos para esses processos em *Caudoviricetes*^{74,75}, sugerindo que essa variação de mecanismos está muito mais atrelada a modificações da estrutura das proteínas a sua quantidade por genoma.

O número reduzido de genes de defesa contra superinfecção por outros fagos e aos sistemas de R-M do hospedeiro em quase todos os grupos, pode ajudar a explicar a diversidade observada para a coleção de bacteriófagos descrita, pois sob a redução destes mecanismos, um único hospedeiro pode ser infectado por múltiplos fagos de características diversas, sendo que esses vírus podem sofrer rearranjos entre eles, o que acarreta em fagos com novas constelações gênicas⁷⁶, bem como, pode ocorrer modificações significativas de extensão e perfil gênico nos fagos que estão mais propensos a ação do sistemas de R-M⁷⁷. Em contrapartida, o grande número de proteínas por grupo associadas a infecção da bactéria, pode exercer influência direta na predição de hospedeiros. Sejam líticos ou lisogênicos, os fagos infectam seus hospedeiros com elevada especificidade, tendo sido costumeiramente denominados como “espécie-específicos”, havendo descrições de especificidade em diferentes níveis, de uma única cepa a vários membros de uma mesma família, havendo raras descrições de fagos que infectam famílias diferentes, uma característica diretamente relacionada a infecção, quando durante a adsorção as proteínas fágicas interagem com receptores específicos da célula bacteriana⁷⁸. Essa especificidade é sugerida em primeiro momento pela ausência de uma espécie de profago infectando diferentes gêneros bacterianos, além disso, profagos idênticos

localizados em cepas diferentes de uma mesma bactéria, possuíam predominância de pertencimento a ST relacionados filogeneticamente ou a um mesmo ST. Analisando a predição individual para cada fago, a maioria demonstrou ser gênero-específica, entretanto, para os agrupamentos, não houve tendência de especificidade de hospedeiro para os grupos de fagos. Considerando as similaridades entre alguns fagos e a possibilidade de eventos de recombinação e mutações nos genomas⁷⁹, é possível que profagos de um mesmo *cluster* apresentem um espectro ampliado para hospedeiros de uma mesma família bacteriana, algo que pode estar refletido no grande número de proteínas de infecção identificado. Alguns dos fagos identificados já parecem possuir um espectro mais amplo nesse sentido, levando ao direcionamento de possíveis alterações moleculares nas proteínas de infecção que permitam o perfil fenotípico de interação com uma gama maior de hospedeiros⁸⁰. A presença de *Listeria monocytogenes* como *host* predito para dois profagos de *Enterobacter cloacae* pode sugerir uma transferência horizontal do profago entre as espécies, embora seja algo incomum, visto que são distantes filogeneticamente, ou ainda que há um ancestral comum muito antigo dessas duas espécies que coevoluiu com o profago⁸¹.

Há uma sugestão de predominância de ciclo de vida lisogênico, devido ao número elevado de genes associados à integração no genoma e a avaliação nas plataformas de predição, este que já era um dado esperado, pois, parece haver uma tendência de predomínio da lisogenia na natureza para *Caudoviricetes*, sendo que os genes dessa classe aparecem em diferentes grupos filogenéticos, sugerindo a evolução conservada dessa característica em diferentes linhagens⁸⁰. Embora seja possível destacar que há uma consistência da presença de proteínas de lise do hospedeiro entre todos os *clusters*, sugerindo passagem anterior pelo ciclo lítico ou aparecendo também como uma característica conservada evolutivamente. Essa última hipótese parece ser a mais provável quando é verificada a análise comparativa entre três genomas de profagos distintos, em que especificamente essas proteínas aparecem apresentando homologia, um indicativo de origem em um ancestral comum^{81,82}. A utilização dessas enzimas de lise codificadas pode ser uma estratégia promissora no desenvolvimento de novas soluções para clones de alto risco⁸³, sendo os genes identificados aqui, potenciais contribuintes para estudos futuros nesse sentido.

Essa análise comparativa conduzida corroborou com o já reforçado nesta discussão: a diversidade gênica entre profagos de enterobactérias é muito grande, refletida também na ínfima homologia de proteínas identificada entre fagos de grupos distintos. Mesmo que as quantidades de proteínas entre determinadas classes funcionais apresentem fortes indícios de serem conservadas entre diferentes linhagens, evolutivamente, essas proteínas são análogas. Aqui se evidencia o

resultado da intensa pressão seletiva sobre as populações de bacteriófagos, demonstrando que as proteínas inerentes a funcionalidade fágica não parecem convergir de um ancestral comum, e sim serem resultados de adaptações semelhantes sofridas entre linhagens evolutivas diferentes⁸⁴.

A presença de homologia para um provável operon do hospedeiro reforça o contexto de transferência horizontal de genes entre fago-fago e fago-hospedeiro⁸⁵. Para além da transdução generalizada de fatores de virulência e genes de resistência em clones de alto risco, os profagos parecem desempenhar um papel muito mais complexo na evolução de seus hospedeiros. As análises demonstraram a presença de elementos fágicos para vários contextos fisiológicos básicos das bactérias, incluindo replicação, divisão celular e metabolismo, o que pode indicar a essencialidade dos profagos no favorecimento seletivo dos espécimes bacterianos na natureza em um contexto muito mais antigo na escala evolutiva⁸⁶.

Observa-se um forte indicativo de que profagos conferem mecanismos de resistência aos antibióticos e fatores de virulência para as cepas bacterianas, sendo observado esse fenômeno tanto para as sequências intactas quanto para as questionáveis (com maior potencial de se tornarem profagos crípticos). Parece haver uma tendência de genes codificando para diferentes tipos de bombas de efluxo multidrogas, uma classe de transportadores de membrana associada a resistência intrínseca e adquirida aos antibióticos⁸⁷, tendo sido identificados também reguladores da expressão desses transportadores, como no caso da identificação do gene *mdtC* e os dois componentes do seu complexo regulador BaeS/BaeR em um dos profagos de *E. coli*, estando associado a resistência a temocilina, costumeiramente utilizado como alternativa aos antibióticos de última escolha⁸⁸. Foram identificadas diferentes superfamílias de transportadores, sendo relacionados ao efluxo de uma gama extremamente ampla de antibióticos não quimicamente relacionados e de elevada relevância clínica⁸⁹, ao exemplo de transportadores altamente promíscuos como no caso de EmrE⁹⁰. O transportador MdfA identificado em um dos profagos de *Enterobacter* spp., tido como relevante para o efluxo de anfenicois, aminoglicosídeos e fluoroquinolonas, é costumeiramente descrito na literatura para *E. coli*, o que reforça a hipótese da transdução generalizada de genes de resistência entre enterobactérias mediada pela inserção de profagos no genoma⁹¹. O complexo OqxAB, determinante para resistência a quinolonas e nitrofuranos, localizado em dois profagos diferentes de *Klebsiella* não agrupados⁹², demonstra eventos de recombinação e mutação em porções de fagos que podem culminar na conferência de características de resistência ao hospedeiro. Associado a presença de bombas de efluxo, alguns profagos codificavam em conjunto porinas, proteínas que

alteram a permeabilidade de membrana, fornecendo uma relação de sinergismo que pode induzir ainda mais um perfil de multirresistência⁹³.

Talvez o isolado mais relevante tenha sido Prophage3_PP-BR5996-1, que apresentou 6 genes do tipo diferentes, que quando em conjunto, podem conferir ao hospedeiro um perfil de resistência extensiva, evidenciados pela resistência aos aminoglicosídeos mediada por *rmtC* e *AAC(6')-Ib7*, bem como, as sulfonamidas através de *sul1*, mas principalmente, aos carbapenêmicos mediada por genes das metalobetalactamases *NDM-1* e *ble(MBL)*, sendo conhecidamente como de extrema relevância global na disseminação de resistência^{94,95}. Além disso, há a presença de no genoma desse fago do marcador genético de resistência de antissépticos, contribuindo para disseminação de cepas bacteriano carregando esse profago em ambientes hospitalares⁹⁶. Esses fatores demonstram a relevância da descoberta desse provável bacteriófago.

Preocupantemente, todos esses genes identificados em regiões de profagos, estão fortemente relacionadas a resistência a uma ampla gama de antibióticos, evidenciando o alarmante papel desses parasitas intragenômicos como elemento genético móvel da transmissão de múltiplos genes de resistência entre enterobactérias. E conforme aqui demonstrado, os fagos de enterobactérias possivelmente se disseminam entre diferentes membros da família *Enterobacteriaceae*, ou seja, pode haver um acúmulo de profagos contendo esses genes, contribuindo para o surgimento de clones extensivamente e até mesmo panresistentes.

Os fatores de virulência foram mais predominantemente encontrados nos genomas, havendo frequência de um único profago apresentando vários desses genes, mas nestes casos, houve um claro padrão de relação funcional entre eles, sugerindo que os profagos podem influenciar diretamente na organização e disseminação de operons entre bactérias, algo que até onde se sabe ainda não foi sugerido na literatura até o momento. Houve descrição da presença do operon completo *ecpRABCDE*, que tem como produto gerado o *pilus* comum (ECP) de *E. coli*, onipresente em cepas comensais e patogênicas⁹⁷, e ainda, de vários genes que controlam a estrutura e função dos flagelos peritríquios de *Enterobacter* spp. associados a motilidade⁹⁸, bem como, concordantemente com o já descrito para profagos de *Klebsiella* spp.⁶⁰, a presença de diversos componentes do Sistema de Secreção do Tipo VI⁹⁹ que favorece a interação com as células eucarióticas e beneficia no predomínio sobre outras bactérias nas competições por nichos ambientais. Além disso, houve presença de determinantes de variabilidade de sorotipos através da biossíntese do antígeno O mediado pelo *locus* LPS *rfb* e bactoprenol glicosiltransferase, conferindo vantagens de interação e evasão das bactérias para com as defesas dos hospedeiros eucariotos^{100,101}.

Todos esses fatores de virulência identificados evidenciam o papel essencial dos profagos na adaptação e nas vantagens evolutivas por parte de seus hospedeiros, favorecendo a sobrevivência e competição das enterobactérias em diferentes ambientes, respaldando fortemente o contexto da relação simbiótica mutualística que parece existir entre profagos e bactérias⁶².

Por fim, a presença de muitas proteínas hipotéticas em todos os *clusters* identificados, é algo dentro do esperado, sendo um fenômeno muito observado em análises genômicas, sendo resultado direto da carência de evidências experimentais, que por consequência afeta a atualidade dos bancos de dados proteômicos¹⁰². Entretanto, dada a frequência de aparecimento nos genomas fágicos, sua caracterização deve ser alvo de prioridade em futuros estudos, uma vez que podem estar associadas a mecanismos ainda não descritos de conferência de vantagens evolutivas ao hospedeiro, ou ainda, proteínas com potencial biotecnológico.

CONCLUSÃO

A caracterização genômica de profagos integrados ao genoma de clones de alto risco de *Enterobacter* spp., *E. coli* e *Klebsiella* spp. demonstrou um cenário de ampla diversidade genética e complexidade evolutiva, refletindo a intensa interação simbiótica entre fagos e seus hospedeiros bacterianos. A ocorrência de perfis gênicos altamente variáveis, mesmo entre fagos agrupados filogeneticamente, e a existência de genes de possível origem hospedeira, apontam para uma história evolutiva marcada por eventos recorrentes de recombinação, seleção e coevolução. A baixa similaridade entre os genomas virais e as sequências em bancos de dados internacionais demonstra que, mesmo com crescentes estudos, a classe *Caudoviricetes* ainda é pouco caracterizada. Bem como, a detecção de genes de resistência a antibióticos e fatores de virulência, mesmo em fagos com genomas parciais, reforça a hipótese de que os profagos atuam como importantes vetores na amplificação do potencial patogênico e na adaptação ambiental dessas bactérias. Além disso, a elevada frequência de proteínas hipotéticas e sem anotação funcional revela um território inexplorado, cuja elucidação pode fornecer novas evidências evolutivas.

Diante disso, os resultados deste estudo não apenas ampliam a compreensão sobre o papel dos profagos na dinâmica genômica de enterobactérias multidroga resistentes, como também fornecem fundamentos para futuras investigações taxonômicas, funcionais e aplicadas no desenvolvimento de terapias fágicas ou estratégias de controle da resistência antimicrobiana. A exploração e o aproveitamento desse potencial genético fágico podem representar um passo decisivo no enfrentamento da crise global da resistência aos antibióticos. Conclui-se que profagos desempenham um papel central na evolução bacteriana, atuando como vetores de inovação genética que ampliam a adaptabilidade, a patogenicidade e o potencial de resistência das enterobactérias frente às pressões seletivas, reforçando a importância da caracterização de profagos como ferramenta complementar no combate à resistência a antibióticos.

REFERÊNCIAS

1. White HE, Orlova EV. Bacteriophages: Their Structural Organisation and Function [Internet]. IntechOpen. 2019. Available from: <https://www.intechopen.com/chapters/66740>
2. Turner D, Kropinski AM, Adriaenssens EM. A Roadmap for Genome-Based Phage Taxonomy. *Viruses* [Internet]. 2021 Mar 18 [cited 2024 Jun 17];13(3):506. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8003253/>
3. Marquet M, Marz M, Pletz MW, Viehweger A, Makarewicz O, Ehricht R, *et al.* What the Phage: a scalable workflow for the identification and analysis of phage sequences. *GigaScience* [Internet]. 2022 Nov 18 [cited 2024 Jun 17];11(1). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9673492/>
4. Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM, Alfenas-Zerbini P, *et al.* Recent changes to virus taxonomy ratified by the International Committee on Taxonomy of Viruses (2022). *Archives of Virology*. 2024 Jun 17;167(11):2429–40.
5. Ofir G, Sorek R. Contemporary Phage Biology: From Classic Models to New Insights. *Cell*. 2018 Mar;172(6):1260–70.
6. Mäntynen S, Laanto E, Oksanen HM, Poranen MM, Díaz-Muñoz SL. Black box of phage–bacterium interactions: exploring alternative phage infection strategies. *Open Biology* [Internet]. 2021 Sep [cited 2024 Jun 15];11(9):210188. Available from: <https://helda.helsinki.fi/bitstream/handle/10138/334575/rsob.210188.pdf?sequence=1>
7. Straka M, Dubinová M, Liptáková A. Phascinating Phages. *Microorganisms* [Internet]. 2022 Jul 6 [cited 2024 Jun 15];10(7):1365. Available from: <https://www.mdpi.com/2076-2607/10/7/1365>
8. Hibstu Z, Belew H, Akelew Y, Mengist HM. Phage Therapy: A Different Approach to Fight Bacterial Infections. *Biologics: Targets and Therapy* [Internet]. 2022 Oct [cited 2024 Jun 16];16(1):173–86. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9550173/>
9. Chang Y. Bacteriophage-Derived Endolysins Applied as Potent Biocontrol Agents to Enhance Food Safety. *Microorganisms*. 2020 May 13;8(5):724.
10. Strathdee SA, Hatfull GF, Mutalik VK, Schooley RT. Phage therapy: From biological mechanisms to future directions. *Cell* [Internet]. 2023 Jan 5 [cited 2024 Jun 16];186(1):17–31. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9827498/%3E>.
11. Silpe JE, Bassler BL. A Host-Produced Quorum-Sensing Autoinducer Controls a Phage Lysis-Lysogeny Decision. *Cell*. 2019 Jan;176(1-2):268-280.e13.
12. Konieczka TA, Wolyniak M. The Isolation and Characterization of Novel Bacteriophages. 2022.
13. Twort FW. An investigation on the nature of ultra-microscopic viruses. *The Lancet* [Internet]. 1915 Dec [cited 2024 Jun 16];186(4814):1241–3. Available from: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(01\)20383-3/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(01)20383-3/fulltext)
14. d’Hérelles F. Sur un microbe invisible antagoniste des bacilles dysentériques. 1st ed. Vol. 165.

Half Leather.; 1917.

15. Aranaga C, Pantoja LD, Martínez EA, Falco A. Phage Therapy in the Era of Multidrug Resistance in Bacteria: A Systematic Review. *International Journal of Molecular Sciences* [Internet]. 2024 Jun 15 [cited 2022 May 31];23(9):4577. Available from: <https://www.mdpi.com/1422-0067/23/9/4577>

16. Okeke IN, Kraker MEA, Boeckel TPV, Kumar CK, Schmidt H, Gales AC. The scope of the antimicrobial resistance challenge. *Lancet* [Internet]. 2024 May 1 [cited 2024 Jun 16];403(10442). Available from: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(24\)00876-6/abstract](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(24)00876-6/abstract)

17. World Health Organization. Global Burden of Bacterial Antimicrobial Resistance in 2019: a Systematic Analysis. *The Lancet* [Internet]. 2022 Feb 12 [cited 2024 Jun 15];399(10325):629–55. Available from: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(21\)02724-0/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)02724-0/fulltext)

18. World Health Organization. WHO Bacterial Priority Pathogens List, 2024: bacterial pathogens of public health importance to guide research, development and strategies to prevent and control antimicrobial resistance.

19. Jin Y, Li W, Zhang H, Ba X, Li Z, Zhou J. The Post-Antibiotic Era: A New Dawn for Bacteriophages. *Biology* [Internet]. 2023 May 4 [cited 2024 Jun 16];12(5):681–1. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10215799/>

20. Kondo K, Kawano M, Sugai M. Distribution of Antimicrobial Resistance and Virulence Genes within the Prophage-Associated Regions in Nosocomial Pathogens. *mSphere* [Internet]. 2021 Aug 25 [cited 2024 Jun 17];6(4):e0045221. Available from: <https://pubmed.ncbi.nlm.nih.gov/34232073/>

21. Gummalla VS, Zhang Y, Liao YT, Wu VCH. The Role of Temperate Phages in Bacterial Pathogenicity. *Microorganisms* [Internet]. 2023 Feb 21 [cited 2024 Jun 16];11(3):541. Available from: <https://pubmed.ncbi.nlm.nih.gov/36985115/>

22. Davies EV, Winstanley C, Fothergill JL, James CE. The role of temperate bacteriophages in bacterial infection. *FEMS Microbiology Letters* [Internet]. 2016 Mar 1 [cited 2024 Jun 16];363(5). Available from: <https://academic.oup.com/femsle/article/363/5/fnw015/2570168>

23. Fillol-Salom A, Alsaadi A, de Sousa JAM, Zhong L, Foster KR, Rocha EPC, *et al.* Bacteriophages benefit from generalized transduction. *PLoS Pathogens* [Internet]. 2019 Jul 5 [cited 2024 Jun 16];15(7). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6636781/>

24. McShan WM, Nguyen SV. *The Bacteriophages of Streptococcus pyogenes*. 2nd ed. Ferretti JJ, Stevens DL, Fischetti VA, editors. PubMed. Oklahoma City (OK): University of Oklahoma Health Sciences Center; 2022.

25. Chen Y, Yang L, Yang D, Song J, Wang C, Sun E, *et al.* Specific Integration of Temperate Phage Decreases the Pathogenicity of Host Bacteria. *Frontiers in Cellular and Infection Microbiology* [Internet]. 2020 Feb 4 [cited 2024 Jun 16];10(1):14. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7010805/>

26. Fan F, Kan B. Survival and proliferation of the lysogenic bacteriophage CTX Φ in *Vibrio cholerae*. *Virologica Sinica* [Internet]. 2015 Jan 20 [cited 2024 Jun 16];30(1):19–25. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8200927/>
27. Muthuirulandi DP, Subramanian N, Pragasam AK, Inbanathan FY, Gupta P, Johnson J, *et al.* Insights to the diphtheria toxin encoding prophages amongst clinical isolates of *Corynebacterium diphtheriae* from India. *Indian Journal of Medical Microbiology* [Internet]. 2019 Jul 1 [cited 2024 Jun 17];37(3):423–5. Available from: <https://pubmed.ncbi.nlm.nih.gov/32003344/>
28. Johnson G, Banerjee S, Putonti C. Diversity of *Pseudomonas aeruginosa* Temperate Phages. Young VB, editor. *mSphere* [Internet]. 2022 Feb 23 [cited 2024 Jun 16];7(1). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8865926/>
29. MacNair CR, Rutherford ST, Tan MW. Alternative therapeutic strategies to treat antibiotic-resistant pathogens. *Nature Reviews Microbiology* [Internet]. 2023 Dec 11 [cited 2024 Jun 16];22(1):1–14. Available from: <https://www.nature.com/articles/s41579-023-00993-0#Sec9>
30. Yu P, Liao H, Wang N, Goel R. Editorial: Bacteria-phage coevolution in antimicrobial resistance. *Frontiers in Microbiology* [Internet]. 2023 Jun 15 [cited 2024 Jun 16];14(1):1221867. Available from: <https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1221867/full>
31. Andrey DO, Dantas PP, Martins WBS, Carvalho FM de, Paula de Almeida LG, Sands K, *et al.* An Emerging Clone, *Klebsiella pneumoniae* Carbapenemase 2–Producing K. pneumoniae Sequence Type 16, Associated With High Mortality Rates in a CC258-Endemic Setting. *Clinical Infectious Diseases*. 2019 Nov 12;71(7):e141–50.
32. Martins WBS, Nicolás MF, Yu Y, Li M, Dantas PP, Sands K, *et al.* Clinical and Molecular Description of a High-Copy IncQ1 KPC-2 Plasmid Harbored by the International ST15 *Klebsiella pneumoniae* Clone. *mSphere*. 2020 Oct 28;5(5).
33. Sands K, Carvalho MJ, Portal E, Thomson K, Dyer C, Akpulu C, *et al.* Characterization of antimicrobial-resistant Gram-negative bacteria that cause neonatal sepsis in seven low- and middle-income countries. *Nature Microbiology* [Internet]. 2021 Apr 1 [cited 2024 Jun 16];6(4):512–23. Available from: <https://www.nature.com/articles/s41564-021-00870-7>
34. Ineos Oxford Institute for Antimicrobial Research. Burden of antibiotic resistance in neonates from developing societies (BARNARDS) [Internet]. Available from: <https://www.ineosoxford.ox.ac.uk/barnards/>
35. National Center for Biotechnology Information. GenBank Overview [Internet]. 2013. Available from: <https://www.ncbi.nlm.nih.gov/genbank/>
36. MapChart. MapChart - Create Your Own Custom Map [Internet]. Available from: <https://www.mapchart.net/>
37. Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, *et al.* PHASTEST: faster than PHASTER, better than PHAST. *Nucleic Acids Research* [Internet]. 2023 May 17 [cited 2024 Aug 10];51(1):443–50. Available from: <https://academic.oup.com/nar/article/51/W1/W443/7167344>

38. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research*. 2024 Apr 13;52(1).
39. Millard A, Denise R, Lestido M, Thomas MT, Webster D, Turner D, *et al.* taxMyPhage: Automated Taxonomy of dsDNA Phage Genomes at the Genus and Species Level. *PHAGE*. 2025 Mar 1;6(1):5–11.
40. International Committee on Taxonomy of Viruses (ICTV). Master Species Lists | ICTV [Internet]. Available from: <https://ictv.global/msl>
41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. *Journal of Molecular Biology*. 1990 Oct;215(3):403–10.
42. Tynecki P, Guziński A, Kazimierczak J, Jadczyk M, Dastyk J, Onisko A. PhageAI - Bacteriophage Life Cycle Recognition with Machine Learning and Natural Language Processing. *bioRxiv*. 2020 Jul 12;
43. Shang J, Peng C, Liao H, Tang X, Sun Y. PhaBOX: a web server for identifying and characterizing phage contigs in metagenomic data. *Bioinformatics Advances*. 2023 Jan 1;3(1).
44. Han Wang R, Yang S, Liu Z, Zhang Y, Wang X, Xu Z, *et al.* PhageScope: a well-annotated bacteriophage database with automatic analyses and visualizations. *Nucleic Acids Research* [Internet]. 2023 Oct 30 [cited 2024 Oct 5];52(1). Available from: <https://academic.oup.com/nar/article/52/D1/D756/7334092?login=false>
45. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya A, Wlodarski MA, *et al.* CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Research*. 2022 Oct 20;51(1):D690–9.
46. Zhou S, Liu B, Zheng D, Chen L, Yang J. VFDB 2025: an integrated resource for exploring anti-virulence compounds. *Nucleic Acids Research*. 2024 Oct 29;53(1).
47. Bouras G, Nepal R, Houtak G, Psaltis AJ, Wormald PJ, Vreugde S. Pharokka: a fast scalable bacteriophage annotation tool. *Bioinformatics (Oxford, England)* [Internet]. 2022 Dec 1 [cited 2024 Nov 5];39(1):btac776. Available from: <https://pubmed.ncbi.nlm.nih.gov/36453861/>
48. Egorov A, Atkinson G. LoVis4u: a locus visualization tool for comparative genomics and coverage profiles. *NAR Genomics and Bioinformatics* [Internet]. 2025 Jan 7 [cited 2025 Mar 3];7(1). Available from: <https://academic.oup.com/nargab/article/7/1/lqaf009/8039213>
49. Yukgehnaish K, Rajandas H, Parimannan S, Manickam R, Marimuthu K, Petersen B, *et al.* PhageLeads: Rapid Assessment of Phage Therapeutic Suitability Using an Ensemble Machine Learning Approach. *Viruses*. 2022 Feb 8;14(2):342.
50. Hockenberry AJ, Wilke CO. BACPHLIP: predicting bacteriophage lifestyle from conserved protein domains. *PeerJ* [Internet]. 2021 May 6 [cited 2024 Nov 18]9: e11396. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8106911/>
51. Chinemerem Nwobodo D, Ugwu MC, Oliseloke Anie C, Al-Ouqaili MTS, Chinedu Ikem J, Victor Chigozie U, *et al.* Antibiotic resistance: The challenges and some emerging strategies for tackling a

global menace. *Journal of Clinical Laboratory Analysis* [Internet]. 2022 Aug 10 [cited 2025 Apr 15];36(9). Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9459344/>

52. Touchon M, Bernheim A, Rocha EP. Genetic and life-history traits associated with the distribution of prophages in bacteria. *The ISME Journal*. 2016 Mar 25;10(11):2744–54.

53. López-Leal G, Camelo-Valera LC, Hurtado-Ramírez JM, Verleyen J, Castillo-Ramírez S, Reyes-Muñoz A. Mining of Thousands of Prokaryotic Genomes Reveals High Abundance of Prophages with a Strictly Narrow Host Range. *mSystems* [Internet]. 2022 Jul 26 [cited 2025 Apr 3];7(4). Available from: <https://journals.asm.org/doi/10.1128/msystems.00326-22>

54. Arndt D, Marcu A, Liang Y, Wishart DS. PHAST, PHASTER and PHASTEST: Tools for finding prophage in bacterial genomes. *Briefings in Bioinformatics*. 2017 Sep 25;20(4):1560–7.

55. Mathers AJ, Peirano G, Pitout JDD. The Role of Epidemic Resistance Plasmids and International High-Risk Clones in the Spread of Multidrug-Resistant Enterobacteriaceae. *Clinical Microbiology Reviews* [Internet]. 2015 Apr 29 [cited 2025 Apr 8];28(3):565–91. Available from: <https://cmr.asm.org/content/28/3/565.short>

56. Lee CR, Lee JH, Park KS, Kim YB, Jeong BC, Lee SH. Global Dissemination of Carbapenemase-Producing *Klebsiella pneumoniae*: Epidemiology, Genetic Context, Treatment Options, and Detection Methods. *Frontiers in Microbiology* [Internet]. 2016 Jun 13 [cited 2025 Apr 18];7(1). Available from: <https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2016.00895/full>

57. Stoesser N, Sheppard AE, Pankhurst L, Maio N, Moore CE, Sebra R, *et al*. Evolutionary History of the Global Emergence of the *Escherichia coli* Epidemic Clone ST131. Rasko DA, Keim PS, editors. *mBio*. 2016 May 4;7(2).

58. Shaik S, Ranjan A, Tiwari SK, Hussain A, Nandanwar N, Kumar N, *et al*. Comparative Genomic Analysis of Globally Dominant ST131 Clone with Other Epidemiologically Successful Extraintestinal Pathogenic *Escherichia coli* (ExPEC) Lineages. Rappuoli R, editor. *mBio*. 2017 Nov 8;8(5).

59. Khan A, Burmeister AR, Wahl LM. Evolution along the parasitism-mutualism continuum determines the genetic repertoire of prophages. McHardy AC, editor. *PLOS Computational Biology*. 2020 Dec 4;16(12):e1008482.

60. Marques AT, Tanoeiro L, Duarte A, Gonçalves L, Vítor JMB, Vale FF. Genomic Analysis of Prophages from *Klebsiella pneumoniae* Clinical Isolates. *Microorganisms* [Internet]. 2021 Oct 28 [cited 2025 Apr 23];9(11):2252. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8617712/>

61. Bobay LM, Touchon M, Rocha EPC. Pervasive domestication of defective prophages by bacteria. *Proceedings of the National Academy of Sciences*. 2014 Aug 4;111(33):12127–32.

62. Fortier LC, Sekulovic O. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence*. 2013 Jul;4(5):354–65.

63. Canchaya C, Fournous G, Brüßow H. The impact of prophages on bacterial chromosomes.

Molecular Microbiology. 2004 Jun 4;53(1):9–18.

64. Gao R, Naushad S, Moineau S, Levesque R, Goodridge L, Ogunremi D. Comparative genomic analysis of 142 bacteriophages infecting *Salmonella enterica* subsp. *enterica*. BMC Genomics [Internet]. 2020 May 26 [cited 2024 Apr 16];21(1). Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC7251866/#Sec9>

65. Khan A, Wahl LM. Quantifying the forces that maintain prophages in bacterial genomes. Theoretical Population Biology [Internet]. 2020 Jun [cited 2025 Apr 16]; 133:168–79. Available from: <https://www.sciencedirect.com/science/article/abs/pii/S0040580919301844?via%3Dihub>

66. Canchaya C, Proux C, Fournous G, Bruttin A, Brüssow H. Prophage Genomics. Microbiology and Molecular Biology Reviews [Internet]. 2003 Jun 1 [cited 2025 Apr 16];67(2):238–76. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC156470/>

67. Das R, Rahlff J. Phage genome architecture and GC content: Structural genes and where to find them. bioRxiv (Cold Spring Harbor Laboratory) [Internet]. 2024 Jun 5 [cited 2025 Apr 23]; Available from: <https://www.biorxiv.org/content/10.1101/2024.06.05.597531v1.full>

68. Simón D, Cristina J, Musto H. Nucleotide Composition and Codon Usage Across Viruses and Their Respective Hosts. Frontiers in Microbiology. 2021 Jun 28;12(1).

69. Almpanis A, Swain M, Gatherer D, McEwan N. Correlation between bacterial G+C content, genome size and the G+C content of associated plasmids and bacteriophages. Microbial Genomics [Internet]. 2018 Apr 10 [cited 2025 Apr 23];4(4). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5989581/>

70. Casjens SR, Thuman-Commike PA. Evolution of mosaically related tailed bacteriophage genomes seen through the lens of phage P22 virion assembly. Virology [Internet]. 2011 Mar [cited 2024 Apr 18];411(2):393–415. Available from: <https://www.sciencedirect.com/science/article/pii/S0042682210008147?via%3Dihub>

71. Koonin EV, Dolja VV, Krupovic M, Varsani A, Wolf YI, Yutin N, *et al.* Global Organization and Proposed Megataxonomy of the Virus World. Microbiology and Molecular Biology Reviews [Internet]. 2020 Mar 4 [cited 2024 Apr 18];84(2). Available from: <https://mibr.asm.org/content/mibr/84/2/e00061-19.full.pdf>

72. Lopes A, Tavares P, Petit MA, Guérois R, Zinn-Justin S. Automated classification of tailed bacteriophages according to their neck organization. BMC Genomics. 2014;15(1):1027.

73. Gulyaeva A, Garmaeva S, Kurilshikov A, Vich Vila A, Riksen NP, Netea MG, *et al.* Diversity and Ecology of Caudoviricetes Phages with Genome Terminal Repeats in Fecal Metagenomes from Four Dutch Cohorts. Viruses [Internet]. 2022 Oct 20 [cited 2025 Apr 21];14(10):2305. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9610469/>

74. Chen W, Xiao H, Wang X, Song S, Han Z, Li X, *et al.* Structural changes of a bacteriophage upon DNA packaging and maturation. Protein & Cell [Internet]. 2020 Apr 7 [cited 2025 Apr 19];11(5):374–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/32266588/>

75. Roos WH, Ivanovska IL, Evilevitch A, Wuite GJL. Viral capsids: Mechanical characteristics, genome packaging and delivery mechanisms. *Cellular and Molecular Life Sciences* [Internet]. 2007 Jun 1 [cited 2025 Apr 17];64(12):1484–97. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2771126/>
76. Sneppen K, Semsey S, Seshasayee ASN, Krishna S. Restriction modification systems as engines of diversity. *Frontiers in Microbiology* [Internet]. 2015 Jun 2 [cited 2025 Apr 19];6(1). Available from: <https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2015.00528/full>
77. Stone E, Campbell K, Grant I, McAuliffe O. Understanding and Exploiting Phage–Host Interactions. *Viruses*. 2019 Jun 18;11(6):567.
78. Piel D, Bruto M, Labreuche Y, Blanquart F, Goudenège D, Barcia-Cruz R, *et al.* Phage–host coevolution in natural populations. *Nature Microbiology* [Internet]. 2022 Jul 1 [cited 2025 Apr 21];7(7):1075–86. Available from: <https://www.nature.com/articles/s41564-022-01157-1>
79. Dutta C, Pan A. Horizontal gene transfer and bacterial diversity. *Journal of Biosciences*. 2002 Feb;27(1):27–33.
80. Weinheimer AR, Ha AD, Aylward FO. Towards a unifying phylogenomic framework for tailed phages. Malik HS, editor. *PLOS Genetics* [Internet]. 2025 Feb 5 [cited 2025 Apr 23]; 21(2):e1011595. Available from: <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1011595#sec002>
81. Gründling A, Bläsi U, Young R. Biochemical and Genetic Evidence for Three Transmembrane Domains in the Class I Holin, λ S. *Journal of Biological Chemistry*. 2000 Jan 14;275(2):769–76.
82. Prosdocimi F, Cerqueira G, Binneck E, Silva A, Reis A, Junqueira A, *et al.* *Bioinformática: Manual do usuário*. 1st ed. Vol. 5. *Revista Bioinformática, ciência e desenvolvimento*; 2001.
83. Gonçalves T, Marques AT, Manageiro V, Tanoeiro L, Vital JS, Duarte A, *et al.* Antimicrobial activity of prophage endolysins against critical Enterobacteriaceae antibiotic-resistant bacteria. *International journal of pharmaceuticals* [Internet]. 2024 Feb 1 [cited 2024 Apr 16]; 651:123758–8. Available from: <https://www.sciencedirect.com/science/article/pii/S0378517323011808?via%3Dihub>
84. Krishna SS, Grishin NV. Structurally Analogous Proteins Do Exist! *Structure*. 2004 Jul;12(7):1125–7.
85. Brussow H, Canchaya C, Hardt WD. Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiology and Molecular Biology Reviews* [Internet]. 2004 Sep 1 [cited 2025 Apr 20];68(3):560–602. Available from: <https://mmb.asm.org/content/68/3/560>
86. Liu Z, Deng Y, Ji M, Sun WW, Fan X. Prophages domesticated by bacteria promote the adaptability of bacterial cells. *BIOCELL* [Internet]. 2020 Apr 27 [cited 2025 Apr 19];44(2):157–66. Available from: <https://www.techscience.com/biocell/v44n2/39266/>
87. Hernando-Amado S, Blanco P, Alcalde-Rico M, Corona F, Reales-Calderón JA, Sánchez MB, *et al.*

Multidrug efflux pumps as main players in intrinsic and acquired resistance to antimicrobials. *Drug Resistance Updates*. 2016 Sep; 28:13–27.

88. Pérez-Palacios P, Rodríguez-Ochoa JL, Velázquez-Escudero A, Rodríguez-Baño J, Rodríguez-Martínez JM, Pascual Á, *et al.* Implications of two-component systems *EnvZ/OmpR* and *BaeS/BaeR* in *in vitro* temocillin resistance in *Escherichia coli*. *The Journal of antimicrobial chemotherapy*/Journal of antimicrobial chemotherapy [Internet]. 2024 Feb 2 [cited 2025 Mar 18];79(3). Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10904727/>

89. Huang L, Wu C, Gao H, Xu C, Dai M, Huang L, *et al.* Bacterial Multidrug Efflux Pumps at the Frontline of Antimicrobial Resistance: An Overview. *Antibiotics* [Internet]. 2022 Apr 13 [cited 2025 Apr 20];11(4):520. Available from: <https://www.mdpi.com/2079-6382/11/4/520>

90. Robinson AE, Thomas NE, Morrison EA, Balthazor BM, Henzler-Wildman KA. New free-exchange model of EmrE transport. *Proceedings of the National Academy of Sciences* [Internet]. 2017 Nov 7 [cited 2025 Apr 23];114(47). Available from: <https://www.pnas.org/doi/full/10.1073/pnas.1708671114>

91. Edgar R, Bibi E. MdfA, an *Escherichia coli* multidrug resistance protein with an extraordinarily broad spectrum of drug recognition. *Journal of Bacteriology* [Internet]. 1997 [cited 2025 Apr 18];179(7):2274–80. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC178964/>

92. Wong MHY, Chan EWC, Chen S. Evolution and Dissemination of OqxAB-Like Efflux Pumps, an Emerging Quinolone Resistance Determinant among Members of Enterobacteriaceae. *Antimicrobial Agents and Chemotherapy* [Internet]. 2015 Jun [cited 2025 Apr 20];59(6):3290–7. Available from: <https://pubmed.ncbi.nlm.nih.gov/25801572/>

93. Fernandez L, Hancock REW. Adaptive and Mutational Resistance: Role of Porins and Efflux Pumps in Drug Resistance. *Clinical Microbiology Reviews*. 2012 Oct 1;25(4):661–81.

94. Moellering RC. NDM-1 — A Cause for Worldwide Concern. *New England Journal of Medicine*. 2010 Dec 16;363(25):2377–9.

95. Elshamy AA, Saleh SE, Aboshanab KM, Aboulwafa MM, Hassouna NA. Transferable IncX3 plasmid harboring bla_{NDM-1}, ble_{MBL}, and aph(3')-VI genes from *Klebsiella pneumoniae* conferring phenotypic carbapenem resistance in *E. coli*. *Molecular Biology Reports* [Internet]. 2023 Apr 20 [cited 2025 Apr 17];50(6):4945–53. Available from: <https://link.springer.com/article/10.1007/s11033-023-08401-9>

96. Kücken D, Feucht HH, Kaulfers PM. Association of qacE and qacE Δ 1 with multiple resistance to antibiotics and antiseptics in clinical isolates of Gram-negative bacteria. *FEMS Microbiology Letters*. 2000 Feb;183(1):95–8.

97. Garnett JA, Martínez-Santos VI, Saldaña Z, Pape T, Hawthorne W, Chan J, *et al.* Structural insights into the biogenesis and biofilm formation by the *Escherichia coli* common pilus. *Proceedings of the National Academy of Sciences* [Internet]. 2012 Feb 21 [cited 2025 Apr 18];109(10):3950–5. Available from: <https://www.pnas.org/doi/full/10.1073/pnas.1106733109>

98. Maayer P, Pillay T, Coutinho TA. Flagella by numbers: comparative genomic analysis of the

supernumerary flagellar systems among the Enterobacterales. *BMC Genomics* [Internet]. 2020 Sep 29 [cited 2025 Apr 22];21(1). Available from: <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-020-07085-w>

99. Shrivastava S, Mande SS. Identification and Functional Characterization of Gene Components of Type VI Secretion System in Bacterial Genomes. Hannenhalli S, editor. *PLoS ONE* [Internet]. 2008 Aug 13 [cited 2025 Apr 19];3(8):e2955. Available from: <https://pubmed.ncbi.nlm.nih.gov/18698408/>

100. Artyszuk D, Izdebski R, Maciejewska A, Kaszowska M, Herud A, Szijártó V, *et al.* The Impact of Insertion Sequences on O-Serotype Phenotype and Its O-Locus-Based Prediction in *Klebsiella pneumoniae* O2 and O1. *International Journal of Molecular Sciences* [Internet]. 2020 Sep 8 [cited 2025 Apr 23];21(18):6572–2. Available from: <https://www.mdpi.com/1422-0067/21/18/6572>

101. Ardiccioni C, Clarke OB, Tomasek D, Issa HA, C. von Alpen D, Pond HL, *et al.* Structure of the polyisoprenyl-phosphate glycosyltransferase GtrB and insights into the mechanism of catalysis. *Nature Communications* [Internet]. 2016 Jan 5 [cited 2025 Apr 18];7(1). Available from: <https://www.nature.com/articles/ncomms10175>

102. Amatore Z, Gunn SC, Harris L. An Educational Bioinformatics Project to Improve Genome Annotation. *Frontiers in Microbiology* [Internet]. 2020 Dec 7 [cited 2025 Apr 23];11. Available from: <https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2020.577497/full>