frontiers | Research Topics

Mobile DNA element-driven evolution of bacterial pathogens

Edited by
Axel Cloeckaert, Michel Stanislas Zygmunt,
Filipa F. Vale and Eric Altermann

Published in
Frontiers in Microbiology

ISSN 1664-8714
ISBN 978-2-8325-6202-4
DOI 10.3389/978-2-8325-6202-4

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Mobile DNA element-driven evolution of bacterial pathogens

Topic editors

Axel Cloeckaert — Institut National de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), France

Michel Stanislas Zygmunt — Institut National de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), France

Filipa F. Vale — University of Lisbon, Portugal

Eric Altermann — Massey University, New Zealand

Citation

Cloeckaert, A., Zygmunt, M. S., Vale, F. F., Altermann, E., eds. (2025). *Mobile DNA element-driven evolution of bacterial pathogens*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6202-4

Table of contents

frontiers | Frontiers in Microbiology

TYPE Editorial
PUBLISHED 19 March 2025
DOI 10.3389/fmicb.2025.1583263

OPEN ACCESS

EDITED AND REVIEWED BY
Tao Li,
Chinese Academy of Agricultural Sciences, China

*CORRESPONDENCE
Axel Cloeckaert
✉ axel.cloeckaert@inrae.fr

RECEIVED 25 February 2025
ACCEPTED 04 March 2025
PUBLISHED 19 March 2025

CITATION
Cloeckaert A, Zygmunt MS, Vale FF and Altermann E (2025) Editorial: Mobile DNA element-driven evolution of bacterial pathogens. *Front. Microbiol.* 16:1583263. doi: 10.3389/fmicb.2025.1583263

Editorial: Mobile DNA element-driven evolution of bacterial pathogens

Axel Cloeckaert[1]*, Michel S. Zygmunt[1], Filipa F. Vale[2] and Eric Altermann[3,4]

[1]INRAE, Université de Tours, UMR, ISP, Nouzilly, France, [2]Pathogen Genomics and Translational Microbiology Lab, BioISI – Instituto de Biosistemas e Ciências Integrativas, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal, [3]Massey University, School of Veterinary Science, Palmerston North, New Zealand, [4]Blue Barn Life Sciences, Ltd, Feilding, New Zealand

KEYWORDS

insertion sequence, bacterial pathogen, genome evolution, chromosome, plasmid, antimicrobial resistance, virulence, adaptation

Editorial on the Research Topic
Mobile DNA element-driven evolution of bacterial pathogens

Mobile DNA elements, such as Insertion Sequence (IS) elements, transposons, integrative elements, and prophages are key players in bacterial adaptation and evolution (Ghaly and Gillings, 2018; Gomberg and Grossman, 2024; Harmer and Hall, 2024; Vos et al., 2024). They are currently classified in numerous families, the representatives of which may be genus- or species-specific. Through their mobility by transposition or integration they shape the bacterial genome and contribute to the adaptation of bacteria to survive changing environmental conditions or to adapt to animal or human hosts and evolve to a pathogenic status. They are also involved in the uptake of foreign DNA via horizontal gene transfer (HGT), ranging in size and function from single genes to pathogenicity islands, or islands providing new functional and metabolic characteristics, such as changes to the bacterial cell surface. The action of mobile DNA elements and associated HGT may therefore drive evolutionary pathogenic processes that include altered responses to inflammatory markers and the evasion of the host immune system.

Mobile DNA elements are also major contributors to the spread of antimicrobial resistance genes, some of which are known to mobilize intrinsic chromosomal genes to major carriers of antimicrobial resistance such as plasmids or genomic islands. In addition, IS elements may carry promoter sequences that can activate the expression of silent genes following their transposition, and through regulatory interference, contribute to enhanced resistance to environmental factors such as antimicrobials or other anti-bacterial factors within the host. IS-induced enhanced resistance may be the result of overexpression of efflux systems, decreased outer membrane permeability or modulated biofilm formation. High throughput sequencing has revealed numerous mobile DNA elements in the bacterial world, many of which remain to be functionally characterized.

The present Research Topic focused on mobile DNA elements and their contribution to pathogen evolution, including their acquisition of pathogenicity and/or antimicrobial resistance characteristics.

Twelve articles were published within this Research Topic in five different sections of Frontiers in Microbiology, namely "Antimicrobials, Resistance and Chemotherapy," "Infectious Agents and Disease," "Food Microbiology," "Microbiotechnology," and "Phage Biology." This Research Topic highlights the interest in studying mobile DNA as a common theme across a wide range of impact areas.

The majority of articles included here focus on antimicrobial resistance, with six articles published in the section "Antimicrobials, Resistance and Chemotherapy," one published in the section "Infectious Agents and Disease," and one published in the section "Food Microbiology," again demonstrating the interest in this important theme across the different sections of Frontiers in Microbiology. The majority of articles address specific pathogens (at the genus or species level) and two aim at assessing the resistome in (i) non-pathogenic bacterial species used in food products and (ii) pathogenic or non-pathogenic species found in the aquatic environment.

The following articles have been published in the Antimicrobials, Resistance and Chemotherapy section. Wang, Zhu, et al. investigated class 1 integrons and multiple mobile genetic elements in clinical isolates of the *Klebsiella pneumoniae* complex from a tertiary hospital in Eastern China. Using whole genome sequencing of 167 isolates, the authors identified a total of 169 antibiotic resistance gene cassettes encoding 19 types of resistance genes, including important carbapenem and class D beta-lactamase genes. Of particular interest, a duplicated region of 19 kb on one plasmid carrying an IS*26*-Int1 complex multidrug resistance integron was identified, which constitutes a new structure of a mobile genetic element involved in the spread of antibiotic resistance. In another study, Wang, Shen, et al. reported the mobilization of the carbapenem resistance gene bla_{KPC-14} among heterogeneous plasmids in extensively drug-resistant hypervirulent *K. pneumoniae*. This resistance gene was located on an IncFII/IncR plasmid within a genetic structure, called the NTE_{KPC}-Ib element, consisting of the bla_{KPC-14} gene flanked by two IS elements, IS*Kpn27* and IS*Kpn6*. The authors assessed the horizontal transferability of the integrated NTE_{KPC}-Ib plasmids and concluded that bla_{KPC-14} is prone to integrate into other conjugative plasmids through this mobile DNA element. Also with regard to multidrug resistance in opportunistic bacteria, Mei et al. characterized a 522 kb mobilizable megaplasmid carrying a 93.5 kb multiple antibiotic resistance region, including *mer* operons conferring heavy metal mercury resistance, from a clinical *Pseudomonas aeruginosa* isolate. A bioinformatic analysis further revealed that many functional genes are flanked by IS elements that may have accumulated in the megaplasmid following multiple acquisition events. Focusing on enteric pathogens, Peng et al. reported the emergence of the fourth mobile sulfonamide resistance gene *sul4* in clinical *Salmonella enterica*. The authors showed that this resistance gene was carried by a complex chromosomally integrated hybrid plasmid. Regarding its mobilization, an IS*CR20*-like element was found to be associated with *sul4*. Of interest in the field of antibiotic resistance reservoirs, Kaszab et al. investigated the resistome of Lactobacilliales by analyzing whole genome sequences available in the NCBI RefSeq database. These bacteria are commonly used in food products and as probiotics in veterinary and human medicine. They are considered safe but may nevertheless carry antibiotic resistance genes (ARGs) that can be transferred to human or veterinary pathogens, raising veterinary and public health concerns. The authors screened the database for ARGs and assessed the possibility of their transmissibility by plasmid transfer or by linkage to integrative mobile genetic elements. The most prevalent transferable ARGs appeared to be *tet*M and *tet*W, which confer resistance to tetracycline. Although not as critical as the resistance genes found in pathogenic species cited in the other studies above, this study highlighted the One Health concept by demonstrating the potential for Lactobacilliales to serve as reservoirs for transferable ARGs. Regarding the aquatic environment resistome, Jiao et al. provided new insights into the microbiome, resistome, and mobilome of dental wastewater in the context of a heavy metal environment. Among hospital wastewater, dental wastewater contains heavy metals that may contribute to the development of antimicrobial resistance in this aquatic environment. The authors identified numerous ARGs, such as those that confer multidrug resistance and resistance to antibiotics that are frequently used in clinical practice. The main bacterial species identified as harboring these ARGs were *P. aeruginosa*, *Pseudomonas putida*, *Chryseobacterium indologenes*, and *Sphingomonas lateriae*. Along with the ARGs many mobile genetic elements were detected, IS elements and transposons, highlighting their potential role in the mobilization of ARGs as in the studies cited above.

In the Infectious Agents and Disease section, Glambek et al. reported on antimicrobial resistance patterns in *Streptococcus dysgalactiae* from a One Health perspective. This bacterial species is an important pathogen in both humans and a wide range of animal species and is therefore of interest from a One Health point of view. The authors investigated, using whole genome sequencing and antimicrobial susceptibility testing, the zoonotic potential of *S. dysgalactiae* and the exchange of antimicrobial resistance traits between different host populations carrying this pathogen in Norway. The authors provided evidence for niche specialization with respect to the distribution of resistance genes and mobile genetic elements, associated with a specific phylogenetic distribution, in isolates from infected humans ($n = 274$, bloodstream infections) and from infected animals ($n = 133$). For example, the erythromycin resistance gene *erm*(A) appeared dominant in human isolates, whereas *erm*(B) and *lsa*(C) were only identified in animal isolates. The *tet*(O) tetracycline resistance gene was located on distinct mobile elements between animal and human isolates. Common mobile elements were observed in only four isolates from different host species including one human, among the total of 407 isolates investigated. In conclusion, this study suggests that *S. dysgalactiae* has evolved into host-adapted populations and niche specialization, and direct exchange of strains or genetic elements from different ecological niches appears to be rare, at least in the geographical region of Norway investigated.

In the Food Microbiology section, Bartsch et al. characterized multidrug-resistant (MDR) *Salmonella enterica* serovar Agona isolates from a dietary supplement in Germany. This study involved

serovar Agona isolates that appeared phylogenetically distinct from others available in databases and aimed thus to find a potential reservoir of this MDR strain and associated mobile genetic elements conferring MDR. Whole genome sequencing revealed the presence of 23 different ARGs conferring resistance to 12 different classes of antibiotics, together with genes conferring resistance to six different heavy metals. A large plasmid of 295 kb belonging to the IncHI2 plasmid family was shown to carry 16 ARGs, organized in two clusters. Each ARG was associated with putative composite transposons. A database search further revealed that similar plasmids are found in *Salmonella* isolates from a wide variety of livestock and in other bacterial genera from different geographical origins and isolation sources. In other words, the host range of this MDR plasmid appears to be broad and has already spread into different bacterial populations, highlighting the need for continuous surveillance of MDR foodborne pathogens such as *Salmonella* spp.

In addition to the articles on antimicrobial resistance cited above, two articles published in the Infectious Agents and Disease section examined the role of mobile DNA in bacterial physiology, fitness, or virulence. Kopkowski et al. studied the effect of DNA-binding proteins on the transposition of the IS element upstream of the *bgl* operon in *Escherichia coli*. This operon, which is normally not expressed, is required for the uptake and metabolism of β-glucosides. Insertion of either IS*1* or IS*5* upstream of the *bgl* promoter activates expression of the operon only when the cell is starved in the presence of a β-glucoside, resulting in increased transcription and allowing the cell to survive and support growth using this carbon source. The authors provided evidence that the DNA-binding proteins Crp and IHF exert a positive effect on insertional *bgl* mutations. Their experimental study indicates that through its binding, IHF may exert its effect by altering the DNA conformation of IS*1* and IS*5* at their native locations, rather than by directly influencing transposase gene expression. On the other hand, the cAMP-Crp complex binds upstream of the promoter and presumably alters the local DNA into a conformation that enhances IS insertion. The study of Hussain et al. aimed to remove mobile genetic elements from the genome of *Clostridioides difficile* and to assess the implications of this removal on the biology of the organism. The genome of this pathogen is highly variable and contains mobile DNA elements such as transposons and prophages that influence its biology. Using allele replacement methodology facilitated by CRISPR-Cas9, the authors succeeded in deleting the following DNA elements from two *C. difficile* strains: Tn*5397* (21 kb) and φ027 (56 kb). The growth characteristics of the deleted strains were only altered in minimal medium. The impact of the deletion on conjugal transfer and phage sensitivity was also investigated. The created deletants will be further investigated for the contribution of the targeted mobile DNA elements to the bacterial host's virulence, fitness, and physiology. Related to this study, in the Phage Biology section, Shüler et al. published novel insights into the phage biology of the pathogen *C. difficile* based on the active virome. The authors examined active prophages from different *C. difficile* strains by sequencing and characterizing phage particle-protected DNA following standard cultivation or cultivation under prophage-inducing conditions. Spontaneous prophage release was demonstrated to be common in this pathogen. Fourteen different phages were identified. In addition, the authors showed that enveloped DNA mapped to genomic regions with characteristics of mobile DNA other than prophages, suggesting DNA mobility mechanisms that have not been fully studied in *C. difficile*. Moreover, phage-mediated lateral transduction of bacterial DNA was detected for the first time in this species. Thus, this study contributed to new knowledge regarding prophage activity and phage biology in *C. difficile*.

In the more general field of DNA binding and modification, Helbrecht et al. published in the Microbiotechnology section the characterization of winged helix domain fusion endonucleases as N6-methyladenine-dependent type IV restriction systems. The authors showed that the role of the winged helix domain as a sensor of adenine methylation is widespread in prokaryotes and other potential sensors in modified DNA are also discussed.

In summary, this Research Topic provides a collection of Original Research articles on mobile DNA elements and their contribution to bacterial evolution, such as the acquisition of novel features by their bacterial host to resist environmental or *in vivo* conditions, such as antimicrobial resistance.

Author contributions

AC: Writing – original draft, Writing – review & editing. MZ: Writing – original draft, Writing – review & editing. FV: Writing – original draft, Writing – review & editing. EA: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. FV was funded by Fundação para a Ciência e a Tecnologia (FCT) through project grant (project PTDC/BTM-TEC/3238/2020), alongside UID/00100, BioISI (DOI: 10.54499/UIDB/04046/2020) Centre grant from FCT, Portugal (to BioISI).

Conflict of interest

EA was employed Blue Barn Life Sciences, Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Ghaly, T. M., and Gillings, M. R. (2018). Mobile DNAs as ecologically and evolutionarily independent units of life. *Trends Microbiol.* 26, 904–912. doi: 10.1016/j.tim.2018.05.008

Gomberg, A. F. S., and Grossman, A. D. (2024). It's complicated: relationships between integrative and conjugative elements and their bacterial hosts. *Curr. Opin. Microbiol.* 82:102556. doi: 10.1016/j.mib.2024.102556

Harmer, C. J., and Hall, R. M. (2024). IS*26* and the IS*26* family: versatile resistance gene movers and genome reorganizers. *Microbiol. Mol. Biol. Rev.* 88:e0011922. doi: 10.1128/mmbr.00119-22

Vos, M., Buckling, A., Kuijper, B., Eyre-Walker, A., Bontemps, C., Leblond, P., et al. (2024). Why do mobile genetic elements transfer DNA of their hosts? *Trends Genet.* 40, 927–938. doi: 10.1016/j.tig.2024.07.008

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 06 March 2023
DOI 10.3389/fmicb.2023.985102

OPEN ACCESS

EDITED BY
Benjamin Andrew Evans,
University of East Anglia,
United Kingdom

REVIEWED BY
Ulises Garza-Ramos,
National Institute of Public Health (Mexico),
Mexico
Jonathan Rodriguez-Santiago,
Center for Research on Infectious Diseases
(CISEI), Mexico

*CORRESPONDENCE
Teng Xu
✉ xuteng@wmu.edu.cn
Jialei Liang
✉ hnqyljl@126.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to
Antimicrobials, Resistance and Chemotherapy,
a section of the journal
Frontiers in Microbiology

RECEIVED 03 July 2022
ACCEPTED 07 February 2023
PUBLISHED 06 March 2023

CITATION
Wang L, Zhu M, Yan C, Zhang Y, He X, Wu L, Xu J, Lu J, Bao Q, Hu Y, Xu T and Liang J (2023) Class 1 integrons and multiple mobile genetic elements in clinical isolates of the *Klebsiella pneumoniae* complex from a tertiary hospital in eastern China.
Front. Microbiol. 14:985102.
doi: 10.3389/fmicb.2023.985102

Class 1 integrons and multiple mobile genetic elements in clinical isolates of the *Klebsiella pneumoniae* complex from a tertiary hospital in eastern China

Lan Wang[1†], Mei Zhu[2†], Chunxia Yan[1], Yanfang Zhang[1], Xuying He[1], Lin Wu[1], Jiefeng Xu[1], Junwan Lu[1], Qiyu Bao[1,3], Yunliang Hu[3], Teng Xu[4*] and Jialei Liang[1,3*]

[1]Medical Molecular Biology Laboratory, School of Medicine, Jinhua Polytechnic, Jinhua, China, [2]Department of Clinical Laboratory, Zhejiang Hospital, Hangzhou, Zhejiang, China, [3]The Second Affiliated Hospital and Yuying Children's Hospital, Wenzhou Medical University, Wenzhou, China, [4]Institute of Translational Medicine, Baotou Central Hospital, Baotou, China

Background: The emergence of highly drug-resistant *K. pneumoniae*, has become a major public health challenge. In this work, we aim to investigate the diversity of species and sequence types (STs) of clinical *Klebsiella* isolates and to characterize the prevalence and structure of class 1 integrons.

Methods: Based on the whole genome sequencing, species identification was performed by 16S rRNA gene homology and average nucleotide identity (ANI) analysis. STs were determined in accordance with the international MLST schemes for *K. pneumoniae* and *K. variicola*. Integron characterization and comparative genomic analysis were performed using various bioinformatic tools.

Results: Species identification showed that the 167 isolates belonged to four species: *K. pneumoniae*, *K. variicola* subsp. *variicola*, *K. quasipneumoniae* and *K. aerogenes*. Thirty-six known and 5 novel STs were identified in *K. pneumoniae*, and 10 novel STs were identified in *K. variicola* subsp. *variicola*. Class 1 integrons were found in 57.49% (96/167) of the isolates, and a total of 169 resistance gene cassettes encoding 19 types of resistance genes, including carbapenem resistance gene (bla_{IPM-4}) and class D β-lactamases gene (bla_{OXA-1} and bla_{OXA-10}), were identified. Among the 17 complete genomes, 29 class 1 integrons from 12 groups were found, only 1 group was encoded on chromosomes. Interestingly, one plasmid (pKP167-261) carrying two copies of approximately 19-kb IS*26*-Int1 complex resistance region that contains an integron and a multidrug resistance gene fragment.

Conclusion: The results of this work demonstrated that the species and STs of the clinical *Klebsiella* isolates were more complex by the whole genome sequence analysis than by the traditional laboratory methods. Finding of the new structure of MGEs related to the resistance genes indicates the great importance of deeply exploring the molecular mechanisms of bacterial multidrug resistance.

KEYWORDS

***Klebsiella* species identification, multilocus sequence typing, integron, antimicrobial resistance, mobile genetic element**

Introduction

Since all members of the *Klebsiella pneumoniae* species complex (KpSC) show overlapping biochemical and phenotypic characteristics, classification of the genomic features of clinical isolates identified as *K. pneumoniae* based on biochemical assays or mass spectrometry (MALDI-TOF) could result in misclassification (Long et al., 2017; Rodriguez-Medina et al., 2019). Whole-genome sequencing (WGS) has clarified that these multiple related species and subspecies share 95–96% average nucleotide identity (ANI) with *K. pneumoniae* but only 90% ANI with other *Klebsiella* species (Wyres et al., 2020; Lam et al., 2021). *K. pneumoniae* (Kp1) is very common in clinical collections and usually accounts for approximately 85% of the isolates identified as *K. pneumoniae*. *K. variicola* and *K. quasipneumoniae* are relatively common pathogens in hospital-acquired infections (10–20% of the incidence of *K. pneumoniae*; Holt et al., 2015). Along with several other bacteria, *K. pneumoniae* has shown a dramatic increase in antibiotic resistance in recent decades (Paczosa and Mecsas, 2016; Navon-Venezia et al., 2017; Effah et al., 2020). Through horizontal gene transfer mediated by plasmid and mobile genetic elements (MGEs), more than 400 antimicrobial resistance genes were found in *K. pneumoniae* (Navon-Venezia et al., 2017; Wyres and Holt, 2018). A recent study estimating the current known *K. pneumoniae* "pangenome" demonstrated that the pangenome is "open," indicating that these species have a high horizontal gene transfer rate (Martin and Bachman, 2018).

Mobile genetic elements, such as insert sequences (ISs), transposons (Tns) and integrons, play an important role in increasing antibiotic resistance (Partridge et al., 2018). Bacteria share MGEs and their associated resistance genes with other bacterial species *via* horizontal gene transfer (HGT), which has promoted the accumulation and dissemination of antibiotic resistance genes (ARGs) in bacteria (Tao et al., 2022). Integrons constitute an important and near-ubiquitous class of genetic elements. An integron is generally defined by the presence of an *int* gene encoding an integrase of the tyrosine recombinase family, an *attI* recombination site and a promoter (Domingues et al., 2015). As classified by the sequence encoding integrase, five classes of integrons associated with drug resistance have been found (Cambray et al., 2010). The typical structure of a class 1 integron is composed of two conserved segments (5′ CS and 3′ CS) and a variable region with one or more antimicrobial resistance gene cassettes (Hall and Collis, 1995; Fluit and Schmitz, 1999). Class 1 integrons are the most common and widespread among clinical gram-negative bacteria, including *Escherichia coli*, *Klebsiella*, *Salmonella*, *Shigella*, *Yersinia* and other disease-causing bacteria, because of their close association with transposons, often embedded within conjugative plasmids (Goldstein et al., 2001; Lima et al., 2014). Class 1 integrons function as a genetic platform for antimicrobial resistance gene cassette capture. At least 200 different gene cassettes have been identified from class 1 integrons, most of which are antibiotic resistance gene cassettes, including the genes conferring resistance to the quaternary ammonium compound family, aminoglycosides, sulfonamides, quinolones, chloramphenicol, fosfomycin, trimethoprim, β-lactams, and other clinically relevant antibiotics (Mazel, 2006; Cambray et al., 2010; Deng et al., 2015).

Recently, the emergence of multidrug-resistant (MDR) *K. pneumoniae* has become a serious issue in healthcare settings worldwide. In China, a close relationship between MDR *K. pneumoniae* strains and the presence of integrons has been demonstrated (Li et al., 2013; Xu et al., 2017; Liao et al., 2020a). Therefore, understanding the molecular characterization of class 1 integrons in *K. pneumoniae* is essential for the implementation of intervention strategies. In this work, we investigated the species and sequence type (ST) diversity and drug resistance profiles of clinical *Klebsiella* isolates and characterized the structure of resistance gene-related class 1 integrons. Notably, we identified a double IS*26*-Int1 complex resistance region in an IncFIB (K) plasmid for the first time.

Materials and methods

Sample collection and bacterial identification

A total of 167 clinical *Klebsiella* isolates were collected from patients in different wards in Zhejiang Hospital in Hangzhou, Zhejiang, China, in 2019. Zhejiang Hospital, one of the largest public hospitals in Zhejiang Province, is a Grade III, Class A general hospital integrating medical treatment, teaching, research, prevention and health care. The largest fraction of the specimens were sputum specimens (44.91%, 75/167), followed by urine (19.76%, 33/167), blood (10.18%, 17/167), throat swabs (7.78%, 13/167), feces (4.79%, 8/167), pus (4.19%, 7/167), alveolar lavage fluid (1.80%, 3/167), duodenal drainage (1.80%, 3/167), prostatic fluid (0.60%, 1/167), catheter specimens (0.60%, 1/167), subglottic secretions (0.60%, 1/167), wound secretions (0.60%, 1/167), perianal secretions (0.60%, 1/167), and other secretions (1.80%, 3/167). All of the isolates were initially identified using the Vitek-60 microorganism auto analysis system (BioMerieux Corporate, Craponne, France). Further species identification was performed by 16S rRNA gene homology comparison (Clarridge, 2004) and ANI analysis using FastANI (Jain et al., 2018). According to previous publications, seven strains representative of the KpSC (*K. pneumoniae*: CP003200, *K. quasipneumoniae* subsp. *quasipneumoniae*: AYIC00000000, *K. quasipneumoniae* subsp. *similipneumoniae*: CP084787, *K. quasivariicola*: AKYX00000000, *K. variicola* subsp. *tropica*: CP084767, *K. variicola* subsp. *variicola*: CP072130, and *K. africana*: CP084874) and one reference strain (*K. aerogenes*, FKIV00000000) were selected for ANI analysis (Goris et al., 2007; Wyres et al., 2020).

Antibiotic susceptibility testing

Minimum inhibitory concentrations (MICs) were determined using the agar dilution method following the guidelines of the Clinical and Laboratory Standards Institute (CLSI), and the susceptibility patterns were interpreted according to the CLSI breakpoint criteria (CLSI, 2019). Multidrug-resistant (MDR) strains were defined as those that were unsusceptible to ≥1 agent in each of >3 antimicrobial categories (Magiorakos et al., 2012). The antimicrobials tested in this work included aminoglycosides (gentamicin and amikacin), cephalosporins (cefepime and

ceftazidime), quinolones (nalidixic acid), monobactams (aztreonam), carbapenems (meropenem), phosphonic acids (fosfomycin), tetracyclines (tetracycline), glycylcyclines (tigecycline) and phenicols (chloramphenicol). *Escherichia coli* ATCC 25922 was used as a reference strain for quality control. The interpretive criteria for tigecycline susceptibility (≤2 μg/ml, susceptible; 4 μg/ml, intermediate; ≥8 μg/ml, resistant) were based on the breakpoints established by the Food and Drug Administration[1].

Genome sequencing, assembly, annotation and bioinformatic analysis

The whole-genome DNA of 167 *Klebsiella* strains was extracted using the AxyPrep Bacterial Genomic DNA Miniprep kit (Axygen Biosciences, Union City, CA, United States). WGS of 167 isolates was performed using the Illumina HiSeq 2,500, of which the 17 isolates with the widest resistance spectra, the highest MIC levels and the most resistance genes were further sequenced by PacBio RS II platforms by Shanghai Personal Biotechnology Co., Ltd. (Shanghai, China). The Illumina short reads and PacBio long reads were initially assembled by SPAdes v3.14.1, Canu v2.1 and Unicycler v0.8 (Bankevich et al., 2012; Koren et al., 2017; Wick et al., 2017). Further correction was conducted by using Pilon and SAMtools to improve assembly quality by mapping short reads to the draft of the whole-genome assembly (Li et al., 2009; Walker et al., 2014). The open reading frames (ORFs) were predicted and annotated using Prokka v1.14.0 and further annotated by DIAMOND against the UniProtKB/Swiss-Prot and NCBI nonredundant protein databases with an e-value threshold of 1e-5 (Seemann, 2014; Buchfink et al., 2015). Identification of resistance genes was performed using Resistance Gene Identifier (RGI) v4.0.3 in the Comprehensive Antibiotic Resistance Database (CARD; Silva et al., 2011). Identification of ISs and integrons was performed using ISfinder and INTEGRALL, respectively (Siguier et al., 2006; Moura et al., 2009). A phylogenetic tree was inferred from the Mash distances of the 167 whole genome sequences. Pairwise distances were calculated using Mash v2.1.11 and used to infer a phylogenetic tree with iToL (Ondov et al., 2016; Letunic and Bork, 2021). Gview was used to construct basic genomic features[2]. Easyfig was used to generate a figure showing structural comparisons and the nucleotide identities between several segments in a linear fashion (Sullivan et al., 2011). Comparisons of the nucleotide sequences were performed using BLASTN. Molecular types were determined in accordance with the international MLST schemes for *K. pneumoniae* (Diancourt et al., 2005) and *K. variicola* (Barrios-Camacho et al., 2019).

Statistical analysis

Statistical analysis was performed by Fisher's exact test using SPSS (version 22.0), and a p value of $p < 0.05$ was considered to indicate statistical significance.

1 https://www.fda.gov/drugs/development-resources/antibacterial-susceptibility-test-interpretive-criteria

2 https://server.Gview.ca/

Results and discussion

Molecular identification of *Klebsiella* isolates

According to the Vitek-60 microorganism auto analysis system, the 167 clinical isolates were all identified as *K. pneumoniae*. Homology analysis of the 16S rRNA gene revealed that the 167 isolates belonged to two species, 166 *K. pneumoniae* isolates and 1 *K. aerogenes* isolate (Supplementary Table S1). However, the results of the ANI analysis of these strains showed that 152, 11, 3 and 1 isolates were *K. pneumoniae*, *K. variicola* subsp. *variicola*, *K. quasipneumoniae* subsp. *similipneumoniae*, and *K. aerogenes*, respectively. They all showed ANIs of more than 98% with the reference strains of the corresponding species (Supplementary Table S2). Whole-genome-based tree showing the phylogenetic relationships between those 167 isolates, their close relatives in the *K. pneumoniae* species complex and *K. aerogenes* (Figure 1).

MLST analysis revealed that 147 of the 152 *K. pneumoniae* isolates could be assigned to 36 known sequence types (STs). ST11 was the most prevalent, accounting for more than half of the total (51.32%, 78/152), followed by ST23 (8.55%, 13/152) and ST412 (5.92%, 9/152; Figure 2; Supplementary Table S3; Supplementary Figure S1). Five isolates could not be assigned to any of the existing STs. According to the MLST criteria, these isolates represented 5 new STs, designated ST6013 (id: 21814), ST6023 (id: 21835), ST6026 (id: 21838), ST6254 (id: 22343), and ST6255 (id: 22344; Supplementary Table S3). Clonal relatedness analysis of the ST11 strains showed that clonal dissemination occurred among the different departments within the hospital (such as clusters B and D of Figure 3), and clonal outbreaks appeared in some departments, such as the Intensive Care Unit (A) and Hematology Department (B) (clusters A and C of Figure 3). Among the 11 *K. variicola* subsp. *Variicola* isolates, none matched the known ST profiles. We therefore submitted data to the *K. variicola* MLST system and obtained 10 new STs (with one ST represented by two isolates; Table 1).

Antibiotic susceptibility and resistance genes

Antibiotic susceptibility tests revealed that the 167 isolates had a high prevalence (≥ 50%) of resistance to 5 of the 11 antimicrobials tested, including nalidixic acid (64.07%), cefepime (61.68%), aztreonam (60.48%), ceftazidime (59.28%), and meropenem (55.69%). The prevalence of resistance for the remaining 6 antimicrobials was 46.71% (both fosfomycin and tetracycline), 43.71% (gentamicin), 35.93% (amikacin), 33.53% (chloramphenicol) and 24.55% (tigecycline; Table 2; Supplementary Tables S4, S5; Figure 4).

Based on the whole genome sequencing of all the isolates, we identified 143 types of drug resistance genes (≥80% similarity to the functionally characterized drug resistance genes), including 43 types of β-lactamase-encoding genes, such as bla_{TEM}, bla_{SHV}, bla_{OXA}, bla_{OKP}, bla_{NDM}, bla_{LEN}, bla_{LAP}, bla_{KPC}, bla_{IMP}, bla_{FONA}, bla_{DHA}, bla_{CTX-M} and *ampC*-type genes (Supplementary Table S6). Among the β-lactamase genes, 5 types were carbapenemases, which included 102 genes: 86 bla_{KPC-2} (84.31%, 86/102), 8 bla_{IMP-4} (7.84%, 8/102), 5 bla_{NDM-1} (4.90%, 5/102), 2 bla_{NDM-5} (1.96%, 2/102), and 1 bla_{KPC-3}

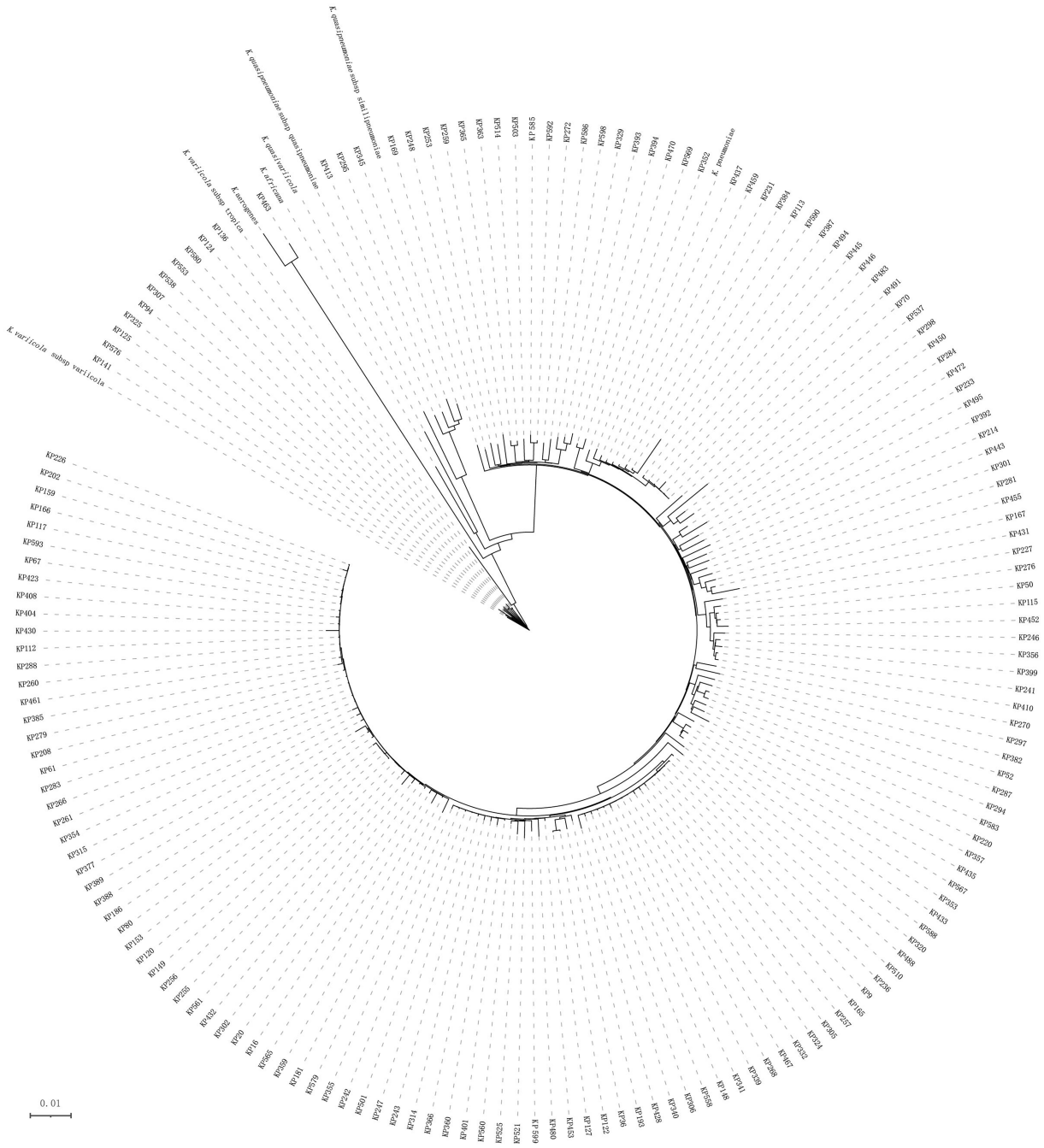

FIGURE 1
Phylogenetic tree of 167 *Klebsiella* isolates. This whole-genome-based tree shows the phylogenetic relationships between those 167 isolates, their close relatives in the seven *Klebsiella pneumoniae* species complex (*K. pneumoniae*: CP003200, *K. quasipneumoniae* subsp. *quasipneumoniae*: AYIC00000000, *K. quasipneumoniae* subsp. *similipneumoniae*: CP084787, *K. quasivariicola*: AKYX00000000, *K. variicola* subsp. *tropica*: CP084767, *K. variicola* subsp. *variicola*: CP072130, and *K. africana*: CP084874), and *K. aerogenes* (FKIV00000000).

(0.98%, 1/102). Seventy-nine (91.86%, 79/86) carbapenemase-encoding genes of three types (bla_{KPC-2}, bla_{IMP-4} and bla_{KPC-3}) were present in 78 ST11 *K. pneumoniae* isolates, most of which were bla_{KPC-2} (96.2%, 76/79). The remaining three were bla_{IMP-4} (2.53%, 2/79) and bla_{KPC-3} (1.27%, 1/79). Apart from one ST11 *K. pneumoniae* that included both bla_{KPC-2} and bla_{IMP-4} genes, all the other 77 ST11 *K. pneumoniae* carried one carbapenem resistance gene each. The remaining bla_{KPC-2} genes ($n = 10$) were present in five different ST types, including five in ST23, two in ST15, and one in each of the three STs (ST86, ST107, and ST687). The remaining bla_{IMP-4} genes ($n = 6$) were present in six ST23 *K. pneumoniae* isolates. It is worth noting that there were five ST23 *K. pneumoniae* isolates each carrying both the bla_{KPC-2} and bla_{IMP-4} genes. bla_{NDM-1} was identified in two *K. pneumoniae* (ST685, ST6254) and three *K. variicola* subsp. *variicola* isolates (ST431, ST431, and ST429). bla_{NDM-5} was identified in two ST340 *K. pneumoniae* strains (Table 3). In addition, there

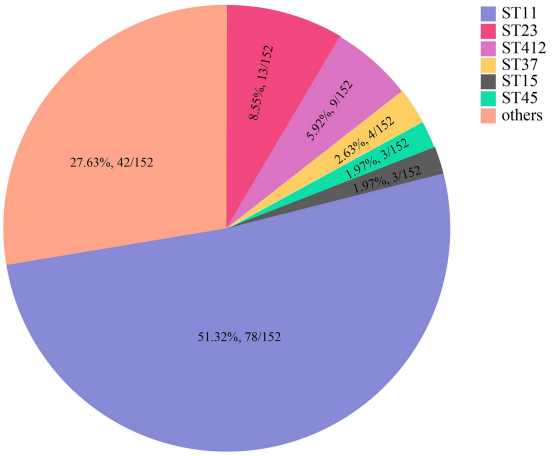

FIGURE 2
The sequence types (STs) of *K. pneumoniae*. * Only STs with at least three isolates are presented. STs with fewer isolates are included in the category "others." The detailed ST information is in Supplementary Table S3.

were 15 different aminoglycoside-modifying enzyme genes encoding APH (6), APH (3″), APH (3″), ANT (3″), ANT (2″), AAC (6), and AAC (3′). Among the aminoglycoside resistance genes, *aadA2* was the most prevalent, and 45.51% (76/167) of the isolates carried this gene (Supplementary Table S6).

Prevalence of integrons and association between the presence of Class 1 integrons and antimicrobial susceptibility

Class 1 integrase genes were found in 57.49% (96/167) of the isolates, including 93 *K. pneumoniae* and 3 *K. variicola* subsp. *variicola*. The integron-positive isolates contained 17 ST types (15 types in *K. pneumoniae* and 2 types in *K. variicola* subsp. *variicola*), and within these *K. pneumoniae* isolates, ST11 showed the highest prevalence (75.27%, 70/93; Supplementary Table S7).

The MIC results demonstrated that the 96 isolates with class 1 integrons had a higher proportion of MDR isolates (78.13%, 75/96) than the 71 isolates with no class 1 integron (43.66%, 31/71; Table 2; Supplementary Tables S4, S5; Figure 4). The relationship between integron carriage and antimicrobial resistance levels was significant for β-lactams (cefepime, ceftazidime, aztreonam and meropenem), aminoglycosides (gentamicin and amikacin), nalidixic acid, fosfomycin, tetracycline, and chloramphenicol ($p < 0.05$). The MIC_{50} values in the integron-positive group were 2-to 512-fold higher than those in the integron-negative group. The MIC_{90} values did not show any difference between the two groups except for fosfomycin and tetracycline, which were 2-and 4-fold higher in the integron-positive group than in the integron-negative group, respectively (Table 4). In contrast, the tigecycline resistance phenotype appeared to be more frequent in the integron-negative group than in the integron-positive group.

Furthermore, among the 143 types of drug resistance genes, 82 were present in both groups, and 36 and 25 were found only in the integron-positive and integron-negative groups, respectively (Supplementary Table S6). In the 82 types of resistance genes that appeared in both groups, the frequencies of a large number of resistance genes differed between them. The frequencies of 36 resistance genes were 1.01-to 13.68-fold higher in the integron-positive group than in the integron-negative group. These genes were mainly related to resistance to aminoglycosides [*aadA2*, *aadA5*, *aph (3′)-Ia*, *aac(3)-IId*, etc.], β-lactams (bla_{SHV-66}, $bla_{SHV-142}$, $bla_{CTX-M-65}$, $bla_{CTX-M-14}$, etc.), tetracycline [*tet (A)* and *tet (D)*], fluoroquinolone (*qnrS1*, *qnrB4* and *emrB*), and so on. The prevalence of *aadA2* was higher in the integron-positive group than in the integron-negative group (75.0% vs. 5.63%). Twenty-seven types of resistance genes showed almost the same frequencies in both groups. Most (19 types) of them were efflux pump genes related to aminoglycosides (*mdtC*, *baeR*, *cpxA*, *kdpE*, etc.) and fluoroquinolone (*mdtK* and *mdtH*). The frequencies of the remaining 19 resistance genes, however, were 0.07-to 0.99-fold higher in the integron-negative group than in the integron-positive group, which included efflux pump genes related to the antimicrobials aminoglycosides (*mdtB*, *mdtA* and *crcB*) and fluoroquinolone (*emrR*), as well as β-lactams (bla_{SHV-94} and bla_{SHV-33}).

The prevalence rates of 36 types of resistance genes uniquely present in the integron-positive group ranged between 1.04% (1/96) and 10.42% (10/96), and they were mainly related to the antimicrobial aminoglycosides [*aadA16*, *aac (6′)-Ib9*, *aac (6′)-Ib*, *aac (6′)-Ib10*, *ant (2″)-Ia*, etc.], fluoroquinolone (*qnrA1*, *qnrB1*, *qnrB2* and *qnrB20*), β-lactams (bla_{SHV-11}, bla_{SHV-27}, bla_{IMP-4}, etc.), rifamycin (*arr-2* and *arr-3*) and so on, with the most frequent gene being *arr-3*. The prevalence of 25 types of resistance genes uniquely in the integron-negative group was between 1.41% (1/71) and 15.49% (11/71), and they were related to the antimicrobials fosfomycin and (*fosA5* and *fosA3*), fluoroquinolone (*qnrS2* and *qnrH*), β-lactams (bla_{SHV-2}, bla_{SHV-28}, bla_{SHV-66}, etc.), among others, with the most frequent gene being *fosA5*.

Analysis of gene cassettes

A total of 169 resistance gene cassettes containing 19 types of antimicrobial resistance genes were found, among which 102 resistance gene cassettes were from 79 isolates with draft genomes and 67 were from 17 isolates with complete genomes. These resistance genes were related to antimicrobials such as aminoglycosides [*aadA2*, *aadA5*, *aadA16*, *ant (2″)-Ia*, *ant (3″)-IIa*, *aac (6′)-Ib-cr*, *aac (6′)-Ib4*, *aac (6′)-Ib9* and *aac (6′)-Ib10*], carbapenems (bla_{IPM-4}), β-lactams (bla_{OXA-1} and bla_{OXA-10}), trimethoprim (*dfrA12*, *dfrA14*, *dfrA27*), rifampin (*arr2*, *arr3*), and chloramphenicol (*cmlA5* and *catB3*). The most prevalent resistance gene was *aadA2* (24.26%, 41/169), followed by *ant (3″)-IIa* (20.12%, 34/169), *aadA5* (8.88%, 15/169) and *dfrA14* (8.88%, 15/169; Table 5).

To determine the structure and location of the class 1 integrons, the complete genomes of 17 integrase gene-positive isolates (KP16, KP122, KP127, KP165, KP167, KP169, KP20, KP307, KP357, KP389, KP431, KP443, KP446, KP494, KP537, KP598, and KP61) that showed a relatively wide resistance spectrum or high resistance levels or carried more resistance genes were obtained by PacBio sequencing. Among the 17 complete genomes, a total of 29 class 1 integrons with 12 different groups of gene cassette arrays were identified (Table 6; Supplementary Figure S2). Two isolates (KP169 and KP307) each

10.3389/fmicb.2023.985102

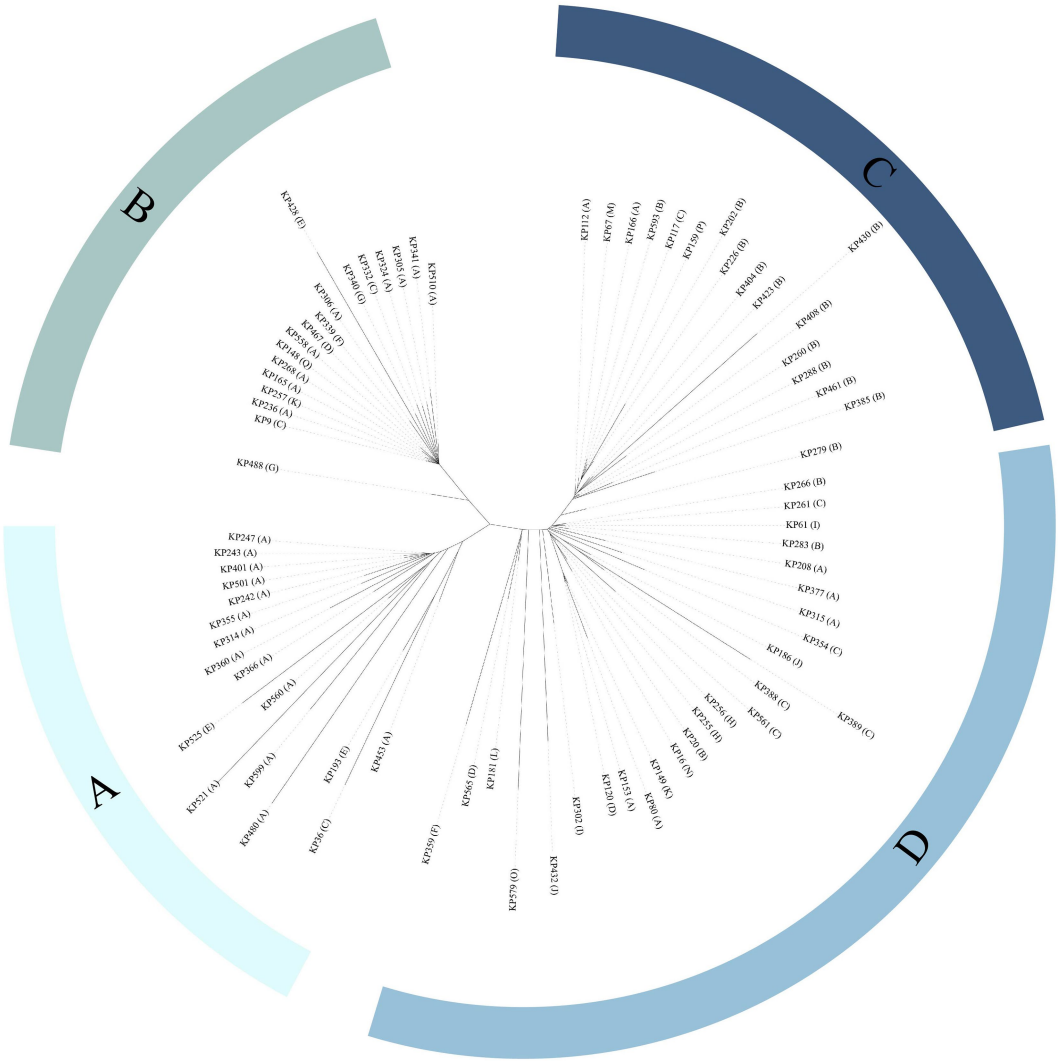

FIGURE 3
Phylogenetic tree of ST11 *Klebsiella* isolates. Department abbreviations: A, Intensive Care Unit; B, Hematology Department; C, Respiratory Medicine Department; D, Neurointerventional Department; E, Community Hospital; F, Neurosurgery Department; G, General Surgery Department; H, Urology Department; I, Geriatrics Outpatient Clinic; J, Rehabilitation Medicine Department; K, Cardiac Macrovascular Surgery Department; L, Nephrology Department; M, Neurology Department; N, Orthopedics Department; O, Department of Infection Disease; P, Department of Plastic Surgery; Q, Cardiovascular Medicine Department.

TABLE 1 The sequence types (STs) of *K. variicola* subsp. *variicola*.

Strain	ST	leuS	pgi	pgk	phoE	pyrG	rpoB	fusA
KP125	333	1	18	12	79	56	33	2
KP124	427	11	1	2	41	1	1	6
KP136	427	11	1	2	41	1	1	6
KP141	428	19	4	5	46	5	7	4
KP307	429	31	60	2	8	58	1	4
KP325	430	9	79	12	8	3	5	2
KP553	431	9	24	3	94	54	1	6
KP538	431	9	24	3	94	54	1	6
KP576	432	93	3	6	32	1	1	4
KP580	433	10	24	44	24	1	1	2
KP94	434	1	36	12	25	3	1	2

TABLE 2 Comparison of the integron-positive and integron-negative isolates in terms of the resistance rates against 11 antimicrobials.

Antimicrobial	Integron-positive isolates (*n*=96)		Integron-negative isolates (*n*=71)		Total isolates (*n*=167)		*p* value
	R no. (%)	S no. (%)	R no. (%)	S no. (%)	R no. (%)	S no. (%)	
Amikacin	44 (45.83)	50 (52.08)	16 (22.54)	54 (76.06)	60 (35.93)	104 (62.28)	0.01763
Gentamicin	52 (54.17)	42 (43.75)	21 (29.58)	50 (70.42)	73 (43.71)	92 (55.09)	0.001464
Cefepime	73 (76.04)	21 (21.87)	30 (42.25)	39 (54.92)	103 (61.68)	60 (35.93)	1.491e-05
Ceftazidime	70 (72.92)	23 (23.96)	29 (40.85)	40 (56.34)	99 (59.28)	63 (37.72)	2.165e-05
Aztreonam	71 (73.96)	23 (23.96)	30 (42.25)	41 (57.75)	101 (60.48)	64 (38.32)	2.237e-05
Meropenem	67 (69.79)	29 (30.21)	26 (36.62)	45 (63.38)	93 (55.69)	74 (44.31)	3.549e-05
Tetracycline	56 (58.33)	17 (17.71)	22 (28.17)	30 (42.25)	78 (46.71)	47 (28.14)	0.000151
Tigecycline	29 (30.21)	44 (45.83)	12 (16.90)	40 (56.34)	41 (24.55)	84 (50.30)	0.05599
Chloramphenicol	38 (39.58)	45 (46.88)	18 (25.35)	50 (70.42)	56 (33.53)	95 (56.89)	0.01786
Fosfomycin	66 (68.75)	26 (27.08)	12 (16.90)	42 (59.15)	78 (46.71)	68 (40.72)	6.747e-09
Nalidixic acid	74 (77.08)	22 (22.92)	33 (46.48)	38 (53.52)	107 (64.07)	60 (35.93)	7.638e-05

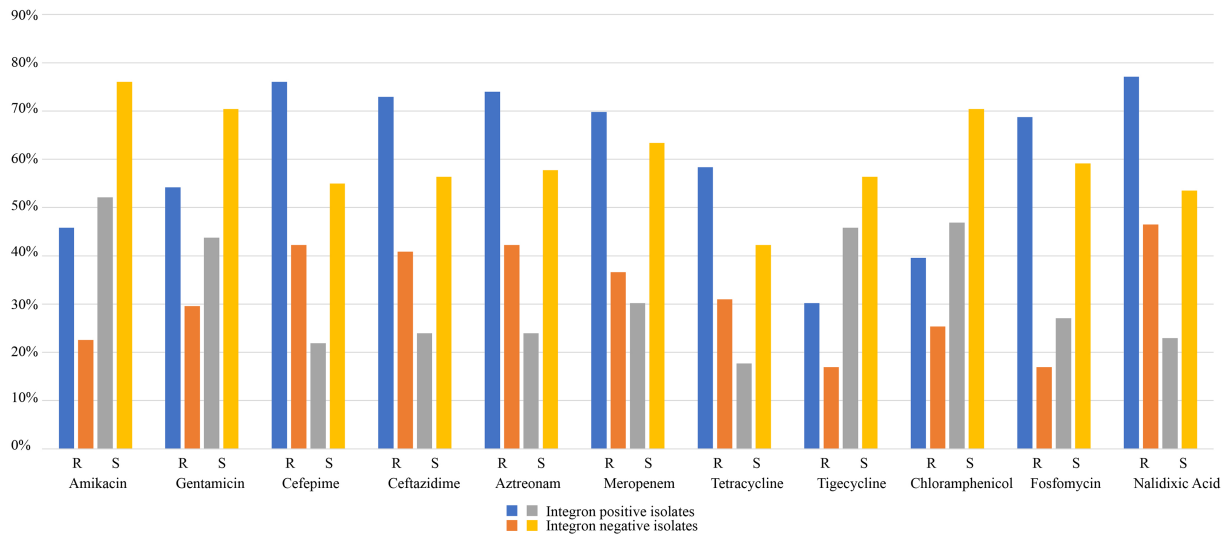

FIGURE 4
Comparison of the integron-positive and integron-negative isolates in terms of resistance rates against 11 antimicrobials.

contained three integrons with different structures, and eight isolates (KP16, KP122, KP127, KP167, KP357, KP389, KP431, and KP598) each harbored two integrons with different structures, whereas the remaining seven each harbored one integron. The most numerous arrays of gene cassettes were *int1-aadA2-qacEΔ1-sul1* (*n*=7), *int1-dfrA14* (*n*=5) and *int1-aac (6′)-Ib-cr-arr-3-dfrA27-aadA16-qacEΔ1-sul1* (*n*=5). Analysis of the location of integrons in the 17 complete genomes revealed that except for the arrays of *int1-aadA2-qacEΔ1-sul1*, which were all encoded on chromosomes, the other 11 arrays were all encoded on plasmids (Table 6). Further analysis of the amino acid sequences of the integrase proteins of the 17 complete genomes showed that they were different and could be clustered into four groups with lengths of 296, 319, 337 and 370 a (Supplementary Figure S3). Each of the integrase groups contained variants of different gene cassette arrays. Interestingly, one plasmid of KP167 (pKP167-261) carried two integrons with almost the same sequences of five gene-cassette arrays (Table 6; Supplementary Figure S2).

Genomic features of the KP167 plasmid pKP167-261

The complete genome of KP167 was composed of a 5.31-Mb chromosome and three plasmids. The pKP167-261 plasmid (CP098759) was 261,525 bp in size with a 51.8% GC content and harbored 280 coding sequences (CDSs). pKP167-261 was an IncFIB (K) plasmid encoding the plasmid replication genes *repFII* and *repFIB*. Based on ≥80% similarity with functionally characterized resistance genes, plasmid pKP167-261 harbored 20 resistance genes encoded in the two copies of the IS*26*-Int1 complex resistance region (one of 19,136 bp in length and the other of 19,135 bp) in the form of

TABLE 3 The carbapenemase-encoding genes in different sequence types (STs).

Gene	Species	ST	Frequency
bla_{KPC-2}	*K. pneumoniae*	11	76
	K. pneumoniae	23	5
	K. pneumoniae	15	2
	K. pneumoniae	86	1
	K. pneumoniae	107	1
	K. pneumoniae	687	1
bla_{IMP-4}	*K. pneumoniae*	11	2
	K. pneumoniae	23	6
bla_{NDM-1}	*K. pneumoniae*	685	1
	K. pneumoniae	6,254	1
	K. variicola subsp. *variicola*	431	2
	K. variicola subsp. *variicola*	429	1
bla_{NDM-5}	*K. pneumoniae*	340	2
bla_{KPC-3}	*K. pneumoniae*	11	1

TABLE 4 The 50 and 90% minimum inhibitory concentration (MIC_{50} and MIC_{90}) values of the 11 tested antimicrobials against integron-positive and integron-negative isolates.

Antimicrobial	Integron-positive isolates (*n*=96)		Integron-negative isolates (*n*=71)	
	MIC_{50} (mg/L)	MIC_{90} (mg/L)	MIC_{50} (mg/L)	MIC_{90} (mg/L)
Amikacin	8	128	1	128
Gentamicin	32	128	0.25	128
Aztreonam	128	128	0.25	128
Cefepime	128	128	0.25	128
Ceftazidime	64	128	0.25	128
Chloramphenicol	16	512	8	512
Fosfomycin	256	512	64	256
Meropenem	128	128	0.25	128
Nalidixic acid	512	512	16	512
Tetracycline	16	512	8	128
Tigecycline	4	16	2	16

TABLE 5 Resistance gene cassettes in 96 integron-positive isolates.

Resistance gene cassette	No. of isolates
aadA2	41
ant(3″)-IIa	34
aadA5	15
dfrA14	15
aac(6′)-Ib-cr	11
arr-3	10
aadA16	7
dfrA27	7
bla_{IMP-4}	6
aac(6′)-Ib4	4
catB3	4
arr-2	3
aac(6′)-Ib9	2
cmlA5	2
dfrA12	2
bla_{OXA-1}	2
bla_{OXA-10}	2
ant(2″)-Ia	1
aac(6′)-Ib10	1

a tandem repeat. Each copy contained a class 1 integron region with 5 resistance genes flanked by IS*26* (AR cassette 1) [IS*26*-*int1*-*arr-2*-*cmlA5*-bla_{OXA-10}-*ant (3″)-IIa*-*dfrA14*-IS*26* (7,340 bp in length), IS*26*-*int1*-*arr-2*-*cmlA5*-bla_{OXA-10}-*ant (3″)-IIa*-*dfrA14*-*orf*-ΔIS*26* (7,339 bp in length)] and a fragment (11,851 bp in length) encoding five resistance genes (AR cassette 2) [*sul2*, *aph (3″)-Ib*, *aph (6)-Id*, *tet (A)* and *floR*] (Supplementary Figure S2).

Eight plasmids showing relatively high nucleotide sequence similarity (coverage ≥80% and identity ≥90%) with pKP167-261 were retrieved from the NCBI nucleotide database (Table 7). All 8 plasmids originated from *K. pneumoniae*. Four of those plasmids, pCY814036-iucA (CP093152.1; 257,343 bp), p130411-38,618_1 (MK649826.1; 241,799 bp), pVir_115011 (CP089955.1; 257,157 bp) and pSCH6109-Vir (CP050860.1; 242,628 bp), shared 100% coverage and 100% identity, 100% coverage and 100% identity, 100% coverage and 100% identity and 100% coverage and 99.95% identity with pKP167-261, respectively. The main difference among pKP167-261, pCY814036-iucA, p130411-38,618_1, pSCH6109-Vir and pVir_115011 was the copy number of the IS*26*-Int1 complex resistance region mentioned above. pCY814036-iucA, pVir_115011, p130411-38,618_1 and pSCH6109-Vir each had only one copy, but pKP167-261 had two copies (Figures 5, 6). The IS*26*-Int1 complex resistance region in the plasmids pVir_115011, p130411-38,618_1 and pSCH6109-Vir were nearly identical, and they shared 100% coverage and 99.92 to 99.99% identity with that in this work (Figure 5; Supplementary Table S8).

Comparative genomic analysis of IS*26*-Int1 complex resistance regions

By comparing the two copies of the class 1 integron sequences, it was found that the difference between the two was that in one of them, a nucleotide was lost in the IS*26* sequence adjacent to *dfrA14*, splitting the intact IS*26* into two parts: a small *orf* and a truncated IS*26* (ΔIS*26*; Figure 6). Using one copy of the IS*26*-Int1 complex resistance region (19,136 bp) as a query to search for homologous sequences in the NCBI nucleotide database, 112 sequences sharing ≥80% coverage and ≥99% identity were found. A total of 96.4% (108/112) of the sequences came from plasmids, including the nine plasmids mentioned above (Supplementary Table S8). Most of the sequences were from Enterobacteriaceae, and only a few sequences were from bacteria outside that family, such as *Aeromonas* and *Vibrio* species. No

TABLE 6 Gene cassette arrays in 17 integron-positive isolates with complete genome sequences.

Strain	Gene cassette	Size (bp)	Location	No. of isolates
KP16, KP20, KP61, KP165, KP389, KP122, KP127	*aadA2*	780	Chromosome	7
KP16, KP431, KP169, KP307*, KP598	*dfr14*	474	Plasmid	5
KP122, KP127, KP431, KP443, KP357	*aac(6′)-Ib-cr-arr-3-dfrA27-aadA16*	2,781	Plasmid	5
KP307*, KP357	*dfrA12-aadA2*	1,697	Plasmid	2
KP446, KP494	bla_{IMP-4}*-orf-orf-aac(6′)-Ib9-catB3*	4,534	Plasmid	2
KP167	*arr2-cmlA5-*bla_{OXA-10}*-ant(3″)-IIa-dfrA14* (duplicate)	4,454	Plasmid	1
KP389	*aadA5*	780	Plasmid	1
KP169	*ant(2″)-Ia*	648	Plasmid	1
KP598	*arr2-orf-aac(6′)-Ib-cr*	2,751	Plasmid	1
KP537	*aac(6′)-Ib-cr-*bla_{OXA-1}*-catB3*	2,331	Plasmid	1
KP169	*aac(6′)-Ib-cr-*bla_{OXA-1}*-catB3-arr-3*	2,868	Plasmid	1
KP307*	*aac(6′)-Ib10-arr-3-dfrA27-aadA16*	2,781	Plasmid	1

*Only the last isolate on the list (KP307) belonged to *K. variicola* subsp. *variicola*; all the others belonged *K. pneumoniae*.

plasmid or chromosome was found to contain two copies of this IS*26*-Int1 complex resistance region.

The IS*26*-flanked class 1 integron sequences (AR cassette 1) of the IS*26*-Int1 complex resistance region identified in this work did not have 3'-CS (*qacEΔ1*/*sul1*). Analyzing the 84 sequences showing the highest similarity (coverage ≥95% and identity ≥99%) to the class 1 integron sequence [IS*26-int1-arr-2-cmlA5-*bla_{OXA-10}*-ant(3″)-IIa-dfrA14*-IS*26*] obtained from the NCBI nucleotide database revealed that only 3 sequences were chromosomally encoded, while all the others were from plasmids. All of them came from 21 different bacterial species, among which *E. coli* was the most abundant (41.67%, 35/84), followed by *K. pneumoniae* (33.33%, 28/84). Among the 84 sequences, 13 were identical to the sequence (7,340 bp in length) identified in this work, including 7 from *E. coli*, 2 from *K. pneumoniae* and 4 from 4 different bacterial species, including *E. fergusonii, A. hydrophila*, *K. grimontii* and *S. enterica* subsp. enterica serovar Derby (Table 8). The sequence in *A. hydrophila* was encoded in the chromosome, and the other 12 were from the plasmids.

When searching for homologous sequences of the five-resistance gene encoding fragments (AR cassette 2; 11,851 bp) of the IS*26*-Int1 complex resistance region, a total of 361 sequences sharing high nucleotide sequence similarity (coverage ≥95% and identity ≥99%) were obtained from the NCBI nucleotide database. All the sequences were encoded on plasmids except nine, which were on chromosomes. They were from 80 different species, with the greatest number derived from *E. coli* (32.69%, 118/361), followed by *K. pneumoniae* (13.85%, 50/361). Four sequences (MK649826.1, CP093152.1, CP089955.1, CP068973.1) were identical to those in this work, and all of them were from plasmids of *K. pneumoniae* (Table 9).

Discussion

In this study, based on ANI analysis, 167 clinically identified *K. pneumoniae* isolates were classified as four different *Klebsiella* species, including three species of the *Klebsiella pneumoniae* complex group (*K. pneumoniae*, *K. variicola* subsp. *variicola* and *K. quasipneumoniae* subsp. *similipneumoniae*) and *K. aerogenes*. *K. variicola* and *K. quasipneumoniae* are relatively common pathogens causing hospital-acquired infections, but traditional clinical laboratory methods (MALDI-TOF MS, multilocus sequence typing, or capsule genotyping) may misclassify them as *K. pneumoniae*, which would underestimate the clinical infection they cause and the actual prevalence (Wyres et al., 2020; Ohama et al., 2022).

Forty-one STs, including five novel STs, were identified in 152 *K. pneumoniae* isolates, of which ST11 (51.32%, 78/152) was identified as the dominant sequence type in the hospital, which is consistent with previous results (Kong et al., 2020). In the present study, most ST11 *K. pneumoniae* strains carried class 1 integrons (89.74%, 70/78) and were MDR (70.51%, 55/78). Additionally, more than 90% of the carbapenemase genes (mainly bla_{KPC-2}) were encoded in ST11 *K. pneumoniae*. From the phylogenetic analysis (Figure 3), clonal outbreaks of ST11 *K. pneumoniae* could be found in some departments, especially in the Intensive Care Unit (A) and Hematology Department (B). The clonal transmission events may result from more complex pathways, intermediate patients, or environmental sources; these factors need further study (Sui et al., 2018). As an important opportunistic pathogen associated with nosocomial bacterial infections, studies have revealed that ST11 *K. pneumoniae* has a high prevalence of virulence factors favoring binding, biofilm formation, colonization and escape from phagocytosis, which can allow clones of this pathogen to successfully spread worldwide (Andrade et al., 2014; Liao et al., 2020b). During the clonal spread of ST11 *K. pneumoniae* strains, the diverse genomic structures of clinical pathogens may help to adapt to the complex and strong selective pressure of the clinical environment (Barrios-Camacho et al., 2019; Rodriguez-Medina et al., 2020).

More than half (57.49%) of the clinical *Klebsiella* isolates from Zhejiang, China, carried class 1 integrons, which was slightly higher than those previously reported in other districts, such as Beijing, or elsewhere in China (Li et al., 2013; Liao et al., 2020a). Compared with the class 1 integron–negative group, the class 1 integron-positive

10.3389/fmicb.2023.985102

TABLE 7 Plasmids similar to pKP167-261 in the NCBI nucleotide collection databases.

Strain	Plasmid	Size (bp)	Coverage (%)	Identity (%)	Accession No.
Klebsiella pneumoniae CY814036	pCY814036-iucA	257,343	100	100	CP093152.1
Klebsiella pneumoniae 130,411–38,618	p130411-38,618_1	241,799	100	100	MK649826.1
Klebsiella pneumoniae WCHKP115011	pVir_115011	257,157	100	100	CP089955.1
Klebsiella pneumoniae SCH6109	pSCH6109-Vir	242,628	100	99.95	CP050860.1
Klebsiella pneumoniae N201205880	p205880-2FIIK	229,479	84	99.97	MN824002.1
Klebsiella pneumoniae Kpn47	pKpn47-FIIK	248,876	86	100	MN821369.1
Klebsiella pneumoniae R46	pR46-270	270,566	90	100	CP035776.1
Klebsiella pneumoniae	pWP2-W18-ESBL-06_1 DNA	140,912	81.61	96.77	AP021930.1

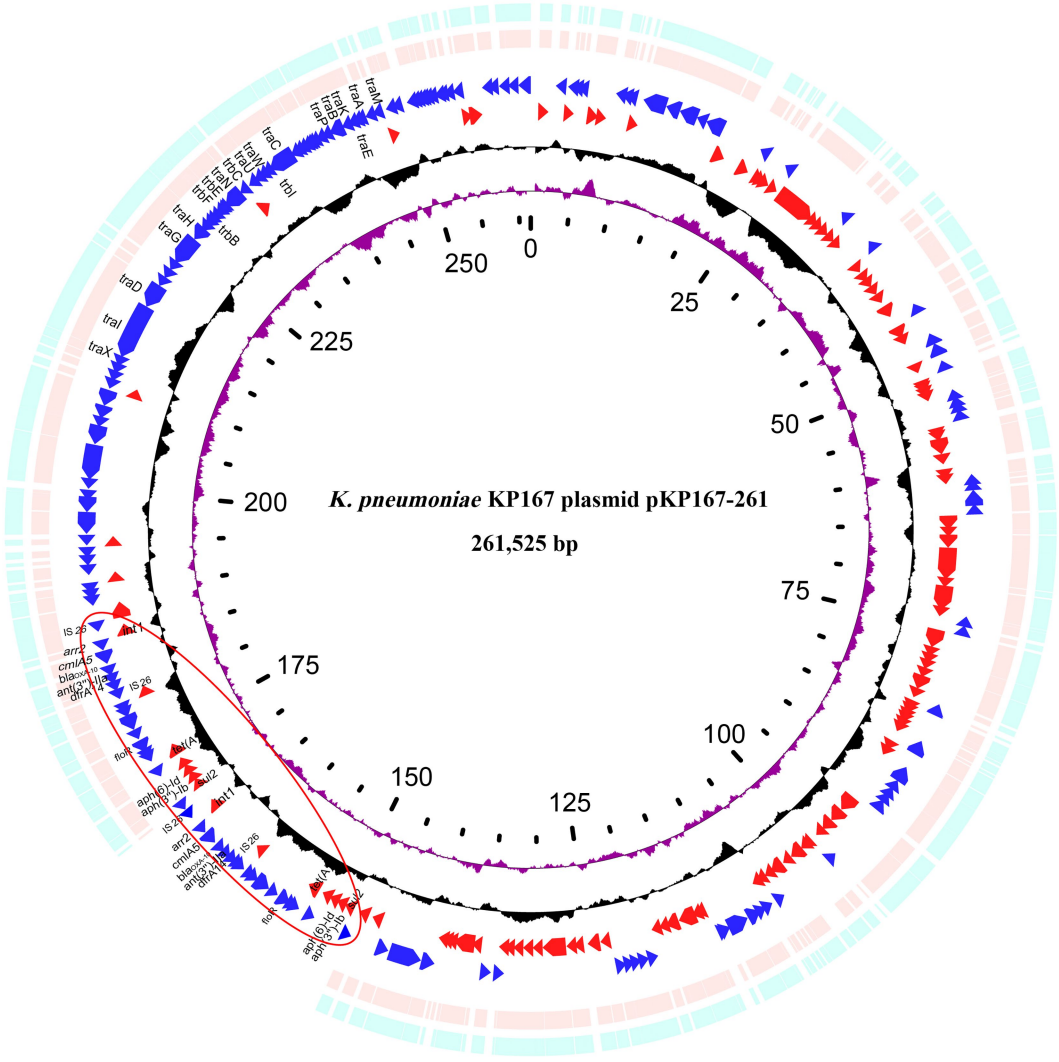

FIGURE 5
Genomic comparison of pKP167-261 with the two most similar plasmids. From outside to inside: circles 1 and 2 are homologous regions of the *K. pneumoniae* SCH6109 plasmid pSCH6109-Vir (CP050860.1) and the *K. pneumoniae* 130,411–38,618 plasmid p130411-38,618_1 (MK649826.1) compared to pKP167-261 with unmatched regions left blank; circles 3 and 4 display predicted ORFs encoded in the forward and reverse strands, respectively; circles 5 and 6 represent the GC content and GC skew, respectively; and circle 7 shows the scale in kilobases. The tandem repeat structure of IS*26*-Int1 complex resistance regions is highlighted with a red circle.

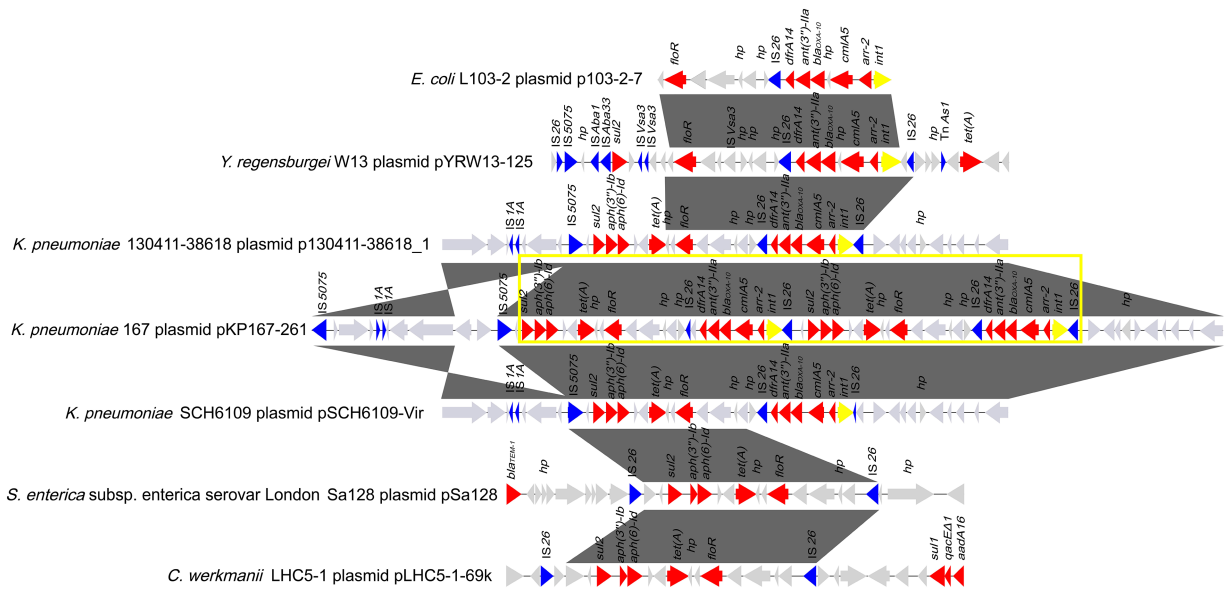

FIGURE 6
Comparative analysis of the genomic context of the two copies of the IS*26*-Int1 complex resistance region in a tandem repeat structure in pKP167-261. This structure was compared among the sequences of different sources. Genes are denoted by arrows and colored according to gene function classification. Gray shading denotes regions of homology (>90% nucleotide sequence identity). The accession numbers of the sequences are as follows: *E. coli* L103-2 plasmid p103-2-7 (CP034849.1); *Yokenella regensburgei* W13 plasmid pYRW13-125 (CP050812.1); *K. pneumoniae* 130,411–38,618 plasmid p130411-38,618_1 (MK649826.1); *K. pneumoniae* KP167 plasmid pKP167-261 (CP098759); *K. pneumoniae* SCH6109 plasmid pSCH6109-Vir (CP050860.1); *S. enterica* subsp. enterica serovar London Sa128 plasmid pSa128 (MG870194.1) and *Citrobacter werkmanii* LHC5-1 plasmid pLHC5-1-69k (CP084291.1). The tandem repeat structure consisting of two copies of the IS*26*-Int1 complex resistance region is highlighted with a yellow box.

TABLE 8 Homologous sequences of antibiotic resistance (AR) cassette 1 in IS*26*-Int1 complex resistance regions.

Bacterium	Location	Coverage (%)	Identity (%)	Accession No.
Escherichia fergusonii HNCF11W	pHNCF11W-130 kb	100	100	CP053046.1
Escherichia coli H3	A	100	100	CP010168.1
Klebsiella pneumoniae CY814036	pCY814036-iucA	100	100	CP093152.1
Klebsiella grimontii 2,481,359	p2481359-2	100	100	CP067382.1
Escherichia coli LD67-1	pLD67-1-157 kb	100	100	CP061187.1
Escherichia coli 98.1	p1	100	100	CP059954.1
Escherichia coli	pIncX1_p1	100	100	MN783746.1
Escherichia coli L103-2	p103-2-7	100	100	CP034849.1
Escherichia coli CT29	p. CT29-P4	100	100	CP032077.1
Escherichia coli RCAD0514	pRCAD0514EC-1	100	100	CP034107.1
Klebsiella pneumoniae fekpn2511	pfekpn2511-3	100	100	CP068975.1
Salmonella enterica subsp. enterica serovar Derby SA1982	Unnamed plasmid	100	100	MT513102.1
Aeromonas hydrophila ZYAH75	Chromosome	100	100	CP016990.1

group exhibited much higher resistance rates against a number of antimicrobials, such as nalidixic acid (77.08% vs. 46.48%), cefepime (76.04% vs. 42.25%), aztreonam (73.96% vs. 42.25%) and ceftazidime (72.92% vs. 40.85%), and likewise, the proportion of MDR isolates was significantly higher in the class 1 integron-positive group (78.12% vs. 46.48%). In addition, more types and numbers of resistance genes were identified in class 1 integron-positive isolates than in class 1 integron-negative isolates, such as *aadA2* (75% vs. 5.65%), bla_{KPC-2} (77.08% vs. 16.90%), and bla_{SHV-1} (73.96% vs. 12.68%), which made the class 1 integron-positive group show higher MIC levels for the corresponding antimicrobials. Similar to transposons, class 1 integrons can capture resistance genes from bacteria of various sources (Fluit and Schmitz, 2004; Firoozeh et al., 2019; Farhadi et al., 2021). It plays an important role in resistance gene spreading by means of horizontal gene transfer between bacteria of different species or genera, resulting in the increasing emergence of MDR bacteria, especially clinical pathogens (Fluit and Schmitz, 2004; Firoozeh et al., 2019; Farhadi et al., 2021).

TABLE 9 Homologous sequences of antibiotic resistance (AR) cassette 2 of IS26-Int1 complex resistance regions.

Bacterium	Location	Coverage (%)	Identity (%)	Accession No.
Klebsiella pneumoniae 130,411–38,618	p130411-38,618_1	100	100	MK649826.1
Klebsiella pneumoniae CY814036	pCY814036-iucA,	100	100	CP093152.1
Klebsiella pneumoniae WCHKP115011	pVir_115011	100	100	CP089955.1
Klebsiella pneumoniae fekpn2511	pfekpn2511-1	100	100	CP068973.1

Twelve groups of 29 complete class 1 integrons were identified in 17 isolates with complete genome sequences. Almost all class 1 integrons or resistance gene cassettes from the remaining class 1 integrase gene-positive strains (without complete genome sequences) could be mapped to one of the 12 groups of class 1 integrons. These 12 groups of integrons were all present in *Enterobacteriaceae*, especially *E. coli* and *K. pneumoniae* (Li et al., 2013; Liu et al., 2015). The class 1 integron *int1-aadA2-qacEΔ1-sul1* was found to be encoded in the chromosomes of seven *K. pneumoniae* of two ST types, ST11 and ST340. NCBI nucleotide database searching revealed that this integron was also found in 108 *K. pneumoniae* chromosomes and two plasmids, one of which came from *K. pneumoniae*, while the other came from *Enterobacter asburiae* (Supplementary Table S9). This suggested that this integron had undergone horizontal transfer between different bacterial species.

The integron carrying the carbapenemase gene bla_{IMP-4} (*bla*$_{IMP-4}$*-orf-orf-aac (6′)-Ib9-catB3*) was found in two *K. pneumoniae* isolates (KP446 and KP494). A similar integron was found in the plasmids of 9 *Klebsiella* strains and 1 *E. asburiae* strain available in public nucleotide databases (Supplementary Table S10). The carbapenem resistance gene carried by the integrons and encoded on the plasmids may lead to a broadened distribution of carbapenem resistance within and between species of different genera and may increase the severity of problems caused by MDR bacteria (van Duin and Doi, 2017).

Interestingly, one plasmid contained two copies of the IS*26*-Int1 complex resistance region, with each IS*26*-Int1 complex resistance region consisting of a class 1 integron fragment (AR cassette 1) and a multidrug resistance fragment (AR cassette 2). Previously, IS*26* has been involved in the amplification of resistance gene-related sequences (Hansen et al., 2019; Harmer et al., 2022). In the present study, comparative genomic analysis revealed that the single IS*26*-Int1 complex resistance region was found in a variety of bacterial genera in the family *Enterobacteriaceae* (*Escherichia*, *Klebsiella*, *Salmonella*) and was mostly encoded on plasmids, but none of them contained this double copy structure (Garza-Ramos et al., 2009; Ji et al., 2022). The BLASTN search of AR cassette 1 and AR cassette 2 suggested that both structures are widely present in different bacterial species and associated with IS*26*. In the present study, we did not identify this duplication in another genome, perhaps because of a high fitness cost or because strains carrying this structure have not been submitted (Adler et al., 2014; McGann et al., 2014). Additional work is necessary to determine the implications of this duplication structure.

Gene duplication/amplification constitutes an important adaptive mechanism in bacteria, and under the strength of clinal antibiotic selection pressure, amplification of antibiotic resistance genes could have a specific clinical impact, including leading to higher expression of these genes and levels of resistance to antibiotics, ultimately contributing to bacterial survival (Sandegren and Andersson, 2009). Bacteria with multiple copies of identical resistance genes or gene arrays have been frequently identified. The recombinant with four copies of bla_{GES-5} had a 2-to 4-fold increase in its MIC levels for the tested β-lactam antimicrobials compared with that carrying one copy of bla_{GES-5}, and bla_{GES-5} was expressed more abundantly in the former (by approximately twofold) than in the latter. The presence of multiple copies of the bla_{OXA-58} gene resulted in high-level resistance to carbapenems in *Acinetobacter baumannii*, and duplication of a 36.4 kb region encompassing bla_{SHV-11} in a clinical isolate of *K. pneumoniae* increased a 16-fold MIC level to amoxicillin (Bertini et al., 2007; Duvernay et al., 2011; Xu et al., 2018). Therefore, it is essential to identify and monitor the occurrence of resistance gene duplication/amplification and its possible impact on the MIC and treatment failure of relevant antibiotics.

Conclusion

In this study, based on whole genome sequencing, the species identification of 167 clinical *Klebsiella* isolates revealed three additional species: *K. variicola* subsp. *variicola*, *K. quasipneumoniae* subsp. *similipneumoniae* and *K. aerogenes* identified by ANI that were not identified by either the common clinical laboratory method or 16S rRNA gene homology analysis. Accurate identification of *Klebsiella* species contributed to the clinical monitoring of the prevalence of pathogenic bacteria and the designation and implementation of novel control strategies. A total of 169 resistance gene cassettes encoding 19 types of resistance genes were found in 96 integrase gene-positive isolates. Among the 17 complete genomes, 12 groups of class 1 integrons were identified, among which one group was encoded on chromosomes, while the others were encoded on the plasmids. One plasmid carrying two copies of the IS*26*-Int1 complex resistance region in a tandem repeat form is reported for the first time in this work. These findings indicate that continuing research on the genome structures of pathogenic bacteria, focusing on resistance-related sequence structures, is of great importance for elucidating the molecular backgrounds of pathogens and the mechanisms through which resistance emerges and spreads.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number (s) can be found in the article/Supplementary material.

Ethics statement

Individual patient data were not involved, and only anonymous residual clinical samples obtained during routine hospital laboratory procedures were used in this study. This study was approved by the ethics committee of Zhejiang Hospital, Hangzhou, Zhejiang, China.

Author contributions

QB, YH, TX, and JlL conceived and designed the experiments. LaW, MZ, CY, YZ, XH, LiW, JX, and JwL performed the experiments. LaW, MZ, and JlL performed data analysis and interpretation. LaW, MZ, QB, TX, and JlL drafted the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Zhejiang Provincial Natural Science Foundation of China (LGF19H200003); the Science & Technology Project of Wenzhou City, China (N20210001, Y2020112); the Natural Science Foundation of China (81960381); and the Science and Technology Planning Project of Zhejiang Province (LGN19C180002).

Acknowledgments

The authors would like to acknowledge all study participants and all individuals who contributed to this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.985102/full#supplementary-material

References

Adler, M., Anjum, M., Berg, O. G., Andersson, D. I., and Sandegren, L. (2014). High fitness costs and instability of gene duplications reduce rates of evolution of new genes by duplication-divergence mechanisms. *Mol. Biol. Evol.* 31, 1526–1535. doi: 10.1093/molbev/msu111

Andrade, L. N., Vitali, L., Gaspar, G. G., Bellissimo-Rodrigues, F., Martinez, R., and Darini, A. L. (2014). Expansion and evolution of a virulent, extensively drug-resistant (polymyxin B-resistant), QnrS1-, CTX-M-2-, and KPC-2-producing Klebsiella pneumoniae ST11 international high-risk clone. *J. Clin. Microbiol.* 52, 2530–2535. doi: 10.1128/JCM.00088-14

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Barrios-Camacho, H., Aguilar-Vera, A., Beltran-Rojel, M., Aguilar-Vera, E., Duran-Bedolla, J., Rodriguez-Medina, N., et al. (2019). Molecular epidemiology of Klebsiella variicola obtained from different sources. *Sci. Rep.* 9:10610. doi: 10.1038/s41598-019-46998-9

Bertini, A., Poirel, L., Bernabeu, S., Fortini, D., Villa, L., Nordmann, P., et al. (2007). Multicopy blaOXA-58 gene as a source of high-level resistance to carbapenems in *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* 51, 2324–2328. doi: 10.1128/AAC.01502-06

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Cambray, G., Guerout, A. M., and Mazel, D. (2010). Integrons. *Annu. Rev. Genet.* 44, 141–166. doi: 10.1146/annurev-genet-102209-163504

Clarridge, J. E. (2004). Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin. Microbiol. Rev.* 17, 840–862. doi: 10.1128/CMR.17.4.840-862.2004

CLSI. (2019). *Performance Standards for Antimicrobial Susceptibility Testing, CLSI Supplement M100. 29th*. Wayne, PA: Clinical and Laboratory Standards Institute.

Deng, Y., Bao, X., Ji, L., Chen, L., Liu, J., Miao, J., et al. (2015). Resistance integrons: class 1, 2 and 3 integrons. *Ann. Clin. Microbiol. Antimicrob.* 14:45. doi: 10.1186/s12941-015-0100-6

Diancourt, L., Passet, V., Verhoef, J., Grimont, P. A., and Brisse, S. (2005). Multilocus sequence typing of *Klebsiella pneumoniae* nosocomial isolates. *J. Clin. Microbiol.* 43, 4178–4182. doi: 10.1128/JCM.43.8.4178-4182.2005

Domingues, S., Da Silva, G. J., and Nielsen, K. M. (2015). Global dissemination patterns of common gene cassette arrays in class 1 integrons. *Microbiology* 161, 1313–1337. doi: 10.1099/mic.0.000099

Duvernay, C., Coulange, L., Dutilh, B., Dubois, V., Quentin, C., and Arpin, C. (2011). Duplication of the chromosomal blaSHV-11 gene in a clinical hypermutable strain of *Klebsiella pneumoniae*. *Microbiology* 157, 496–503. doi: 10.1099/mic.0.043885-0

Effah, C. Y., Sun, T., Liu, S., and Wu, Y. (2020). *Klebsiella pneumoniae*: an increasing threat to public health. *Ann. Clin. Microbiol. Antimicrob.* 19:1. doi: 10.1186/s12941-019-0343-8

Farhadi, M., Ahanjan, M., Goli, H. R., Haghshenas, M. R., and Gholami, M. (2021). High frequency of multidrug-resistant (MDR) *Klebsiella pneumoniae* harboring several beta-lactamase and integron genes collected from several hospitals in the north of Iran. *Ann. Clin. Microbiol. Antimicrob.* 20:70. doi: 10.1186/s12941-021-00476-1

Firoozeh, F., Mahluji, Z., Khorshidi, A., and Zibaei, M. (2019). Molecular characterization of class 1, 2 and 3 integrons in clinical multi-drug resistant *Klebsiella pneumoniae* isolates. *Antimicrob. Resist. Infect. Control* 8:59. doi: 10.1186/s13756-019-0509-3

Fluit, A. C., and Schmitz, F. J. (1999). Class 1 integrons, gene cassettes, mobility, and epidemiology. *Eur. J. Clin. Microbiol. Infect. Dis.* 18, 761–770. doi: 10.1007/s100960050398

Fluit, A. C., and Schmitz, F. J. (2004). Resistance integrons and super-integrons. *Clin. Microbiol. Infect.* 10, 272–288. doi: 10.1111/j.1198-743X.2004.00858.x

Garza-Ramos, U., Davila, G., Gonzalez, V., Alpuche-Aranda, C., Lopez-Collada, V. R., Alcantar-Curiel, D., et al. (2009). The blaSHV-5 gene is encoded in a compound transposon duplicated in tandem in *Enterobacter cloacae*. *Clin. Microbiol. Infect.* 15, 878–880. doi: 10.1111/j.1469-0691.2009.02790.x

Goldstein, C., Lee, M. D., Sanchez, S., Hudson, C., Phillips, B., Register, B., et al. (2001). Incidence of class 1 and 2 integrases in clinical and commensal bacteria from livestock, companion animals, and exotics. *Antimicrob. Agents Chemother.* 45, 723–726. doi: 10.1128/AAC.45.3.723-726.2001

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0

Hall, R. M., and Collis, C. M. (1995). Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. *Mol. Microbiol.* 15, 593–600. doi: 10.1111/j.1365-2958.1995.tb02368.x

Hansen, K. H., Andreasen, M. R., Pedersen, M. S., Westh, H., Jelsbak, L., and Schonning, K. (2019). Resistance to piperacillin/tazobactam in Escherichia coli resulting from extensive IS26-associated gene amplification of blaTEM-1. *J. Antimicrob. Chemother.* 74, 3179–3183. doi: 10.1093/jac/dkz349

Harmer, C. J., Lebreton, F., Stam, J., Mcgann, P. T., and Hall, R. M. (2022). Mechanisms of IS26-mediated amplification of the aphA1 gene leading to tobramycin resistance in an *Acinetobacter baumannii* isolate. *Microbiol. Spectr.* 10:e0228722. doi: 10.1128/spectrum.02287-22

Holt, K. E., Wertheim, H., Zadoks, R. N., Baker, S., Whitehouse, C. A., Dance, D., et al. (2015). Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proc. Natl. Acad. Sci. U. S. A.* 112, E3574–E3581. doi: 10.1073/pnas.1501049112

Jain, C., Rodriguez, R. L., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114. doi: 10.1038/s41467-018-07641-9

Ji, F., Liu, S., Wang, X., Zhao, J., Zhu, J., Yang, J., et al. (2022). Characteristics of the multiple replicon plasmid IncX1-X1 in multidrug-resistant Escherichia coli from Malayan pangolin (*Manis javanica*). *Integr. Zool.* 2022:12637. doi: 10.1111/1749-4877.12637

Kong, Z., Liu, X., Li, C., Cheng, S., Xu, F., and Gu, B. (2020). Clinical molecular epidemiology of carbapenem-resistant *Klebsiella pneumoniae* among pediatric patients in Jiangsu Province, China. *Infect. Drug Resist.* 13, 4627–4635. doi: 10.2147/IDR.S293206

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116

Lam, M. M. C., Wick, R. R., Watts, S. C., Cerdeira, L. T., Wyres, K. L., and Holt, K. E. (2021). A genomic surveillance framework and genotyping tool for *Klebsiella pneumoniae* and its related species complex. *Nat. Commun.* 12:4188. doi: 10.1038/s41467-021-24448-3

Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, B., Hu, Y., Wang, Q., Yi, Y., Woo, P. C., Jing, H., et al. (2013). Structural diversity of class 1 integrons and their associated gene cassettes in *Klebsiella pneumoniae* isolates from a hospital in China. *PLoS One* 8:e75805. doi: 10.1371/journal.pone.0075805

Liao, W., Li, D., Liu, F., Du, F. L., Long, D., Zhang, W., et al. (2020a). Distribution of integrons and phylogenetic groups among highly virulent serotypes of *Klebsiella pneumoniae* in a Chinese tertiary hospital. *J. Glob. Antimicrob. Resist.* 21, 278–284. doi: 10.1016/j.jgar.2019.11.016

Liao, W., Liu, Y., and Zhang, W. (2020b). Virulence evolution, molecular mechanisms of resistance and prevalence of ST11 carbapenem-resistant *Klebsiella pneumoniae* in China: a review over the last 10 years. *J. Glob. Antimicrob. Resist.* 23, 174–180. doi: 10.1016/j.jgar.2020.09.004

Lima, A. M., De Melo, M. E., Alves, L. C., Brayner, F. A., and Lopes, A. C. (2014). Investigation of class 1 integrons in *Klebsiella pneumoniae* clinical and microbiota isolates belonging to different phylogenetic groups in Recife, state of Pernambuco. *Rev. Soc. Bras. Med. Trop.* 47, 165–169. doi: 10.1590/0037-8682-0021-2014

Liu, Z., Zhang, Z., Yan, H., Li, J., and Shi, L. (2015). Isolation and molecular characterization of multidrug-resistant *Enterobacteriaceae* strains from pork and environmental samples in Xiamen, China. *J. Food Prot.* 78, 78–88. doi: 10.4315/0362-028X.JFP-14-172

Long, S. W., Linson, S. E., Ojeda Saavedra, M., Cantu, C., Davis, J. J., Brettin, T., et al. (2017). Whole-genome sequencing of human clinical *Klebsiella pneumoniae* isolates reveals misidentification and misunderstandings of *Klebsiella pneumoniae*, *Klebsiella variicola*, and *Klebsiella quasipneumoniae*. *mSphere* 2:e00290. doi: 10.1128/mSphereDirect.00290-17

Magiorakos, A. P., Srinivasan, A., Carey, R. B., Carmeli, Y., Falagas, M. E., Giske, C. G., et al. (2012). Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin. Microbiol. Infect.* 18, 268–281. doi: 10.1111/j.1469-0691.2011.03570.x

Martin, R. M., and Bachman, M. A. (2018). Colonization, infection, and the accessory genome of *Klebsiella pneumoniae*. *Front. Cell. Infect. Microbiol.* 8:4. doi: 10.3389/fcimb.2018.00004

Mazel, D. (2006). Integrons: agents of bacterial evolution. *Nat. Rev. Microbiol.* 4, 608–620. doi: 10.1038/nrmicro1462

Mcgann, P., Courvalin, P., Snesrud, E., Clifford, R. J., Yoon, E. J., Onmus-Leone, F., et al. (2014). Amplification of aminoglycoside resistance gene aphA1 in *Acinetobacter baumannii* results in tobramycin therapy failure. *MBio* 5:e00915. doi: 10.1128/mBio.00915-14

Moura, A., Soares, M., Pereira, C., Leitao, N., Henriques, I., and Correia, A. (2009). INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. *Bioinformatics* 25, 1096–1098. doi: 10.1093/bioinformatics/btp105

Navon-Venezia, S., Kondratyeva, K., and Carattoli, A. (2017). *Klebsiella pneumoniae*: a major worldwide source and shuttle for antibiotic resistance. *FEMS Microbiol. Rev.* 41, 252–275. doi: 10.1093/femsre/fux013

Ohama, Y., Nomura, Y., Mizoguchi, M., Higurashi, Y., Okamoto, K., and Harada, S. (2022). Accurate identification of *Klebsiella variicola* by MALDI-TOF mass spectrometry in clinical microbiology laboratories. *Microbiol. Spectr.* 10:e0284422. doi: 10.1128/spectrum.02844-22

Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., et al. (2016). Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 17:132. doi: 10.1186/s13059-016-0997-x

Paczosa, M. K., and Mecsas, J. (2016). *Klebsiella pneumoniae*: going on the offense with a strong defense. *Microbiol. Mol. Biol. Rev.* 80, 629–661. doi: 10.1128/MMBR.00078-15

Partridge, S. R., Kwong, S. M., Firth, N., and Jensen, S. O. (2018). Mobile genetic elements associated with antimicrobial resistance. *Clin. Microbiol. Rev.* 31:e00088. doi: 10.1128/CMR.00088-17

Rodriguez-Medina, N., Barrios-Camacho, H., Duran-Bedolla, J., and Garza-Ramos, U. (2019). Klebsiella variicola: an emerging pathogen in humans. *Emerg. Microbes Infect.* 8, 973–988. doi: 10.1080/22221751.2019.1634981

Rodriguez-Medina, N., Martinez-Romero, E., De La Cruz, M. A., Ares, M. A., Valdovinos-Torres, H., Silva-Sanchez, J., et al. (2020). A *Klebsiella variicola* plasmid confers hypermucoviscosity-like phenotype and alters capsule production and virulence. *Front. Microbiol.* 11:579612. doi: 10.3389/fmicb.2020.579612

Sandegren, L., and Andersson, D. I. (2009). Bacterial gene amplification: implications for the evolution of antibiotic resistance. *Nat. Rev. Microbiol.* 7, 578–588. doi: 10.1038/nrmicro2174

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference Centre for bacterial insertion sequences. *Nucleic Acids Res.* 34, D32–D36. doi: 10.1093/nar/gkj014

Silva, I. R., Larsen, D. M., Meyer, A. S., and Mikkelsen, J. D. (2011). Identification, expression, and characterization of a novel bacterial RGI lyase enzyme for the production of bio-functional fibers. *Enzym. Microb. Technol.* 49, 160–166. doi: 10.1016/j.enzmictec.2011.04.015

Sui, W., Zhou, H., Du, P., Wang, L., Qin, T., Wang, M., et al. (2018). Whole genome sequence revealed the fine transmission map of carbapenem-resistant *Klebsiella pneumonia* isolates within a nosocomial outbreak. *Antimicrob. Resist. Infect. Control* 7:70. doi: 10.1186/s13756-018-0363-8

Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics* 27, 1009–1010. doi: 10.1093/bioinformatics/btr039

Tao, S., Chen, H., Li, N., and Liang, W. (2022). The application of the CRISPR-Cas system in antibiotic resistance. *Infect. Drug Resist.* 15, 4155–4168. doi: 10.2147/IDR.S370869

Van Duin, D., and Doi, Y. (2017). The global epidemiology of carbapenemase-producing *Enterobacteriaceae*. *Virulence* 8, 460–469. doi: 10.1080/21505594.2016.1222343

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 13:e1005595. doi: 10.1371/journal.pcbi.1005595

Wyres, K. L., and Holt, K. E. (2018). *Klebsiella pneumoniae* as a key trafficker of drug resistance genes from environmental to clinically important bacteria. *Curr. Opin. Microbiol.* 45, 131–139. doi: 10.1016/j.mib.2018.04.004

Wyres, K. L., Lam, M. M. C., and Holt, K. E. (2020). Population genomics of *Klebsiella pneumoniae*. *Nat. Rev. Microbiol.* 18, 344–359. doi: 10.1038/s41579-019-0315-1

Xu, X., Li, X., Luo, M., Liu, P., Su, K., Qing, Y., et al. (2017). Molecular characterisations of integrons in clinical isolates of *Klebsiella pneumoniae* in a Chinese tertiary hospital. *Microb. Pathog.* 104, 164–170. doi: 10.1016/j.micpath.2017.01.035

Xu, T., Wang, J., Ying, J., Zhu, T., Liu, Y., Xu, L., et al. (2018). Characterisation of a class 1 integron associated with the formation of quadruple Bla(GES-5) cassettes from an IncP-1beta group plasmid in Pseudomonas aeruginosa. *Int. J. Antimicrob. Agents* 52, 485–491. doi: 10.1016/j.ijantimicag.2018.07.002

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 20 April 2023
DOI 10.3389/fmicb.2023.1106157

OPEN ACCESS

EDITED BY
Hemda Garelick,
Middlesex University, United Kingdom

REVIEWED BY
Ilunga Kamika,
University of South Africa, South Africa
Dong Yang,
Tianjin Institute of Environmental and Operational Medicine, China
Yolanda Moreno,
Universitat Politècnica de València, Spain

*CORRESPONDENCE
Yong Xia
yxia@stu.edu.cn

†These authors share first authorship

SPECIALTY SECTION
This article was submitted to Antimicrobials, Resistance and Chemotherapy, a section of the journal Frontiers in Microbiology

RECEIVED 07 December 2022
ACCEPTED 27 March 2023
PUBLISHED 20 April 2023

CITATION
Jiao X, Guo W, Li X, Yao F, Zeng M, Yuan Y, Guo X, Wang M, Xie QD, Cai L, Yu F, Yu P and Xia Y (2023) New insight into the microbiome, resistome, and mobilome on the dental waste water in the context of heavy metal environment.
Front. Microbiol. 14:1106157.
doi: 10.3389/fmicb.2023.1106157

New insight into the microbiome, resistome, and mobilome on the dental waste water in the context of heavy metal environment

Xiaoyang Jiao[1†], Wenyan Guo[2†], Xin Li[1], Fen Yao[3], Mi Zeng[1], Yumeng Yuan[1], Xiaoling Guo[1], Meimei Wang[1], Qing Dong Xie[1], Leshan Cai[2], Feiyuan Yu[1], Pen Yu[2] and Yong Xia[2*]

[1]College of Medicine, Shantou University, Shantou, China, [2]Department of Clinical Laboratory, First Affiliated Hospital of Shantou University Medical College, Shantou, China, [3]Department of Pharmacology, College of Medicine, Shantou University, Shantou, China

Object: Hospital sewage have been associated with incorporation of antibiotic resistance genes (ARGs) and mobile genetic elements (MGEs) into microbes, which is considered as a key indicator for the spread of antimicrobial resistance (AMR). The compositions of dental waste water (DWW) contain heavy metals, the evolution of AMR and its effects on the water environment in the context of heavy metal environment have not been seriously investigated. Thus, our major aims were to elucidate the evolution of AMR in DWW.

Methods: DWW samples were collected from a major dental department. The presence of microbial communities, ARGs, and MGEs in untreated and treated (by filter membrane and ozone) samples were analyzed using metagenomics and bioinformatic methods.

Results: DWW-associated resistomes included 1,208 types of ARGs, belonging to 29 antibiotic types/subtypes. The most abundant types/subtypes were ARGs of multidrug resistance and of antibiotics that were frequently used in the clinical practice. *Pseudomonas putida*, *Pseudomonas aeruginosa*, *Chryseobacterium indologenes*, *Sphingomonas laterariae* were the main bacteria which hosted these ARGs. Mobilomes in DWW consisted of 93 MGE subtypes which belonged to 8 MGE types. Transposases were the most frequently detected MGEs which formed networks of communications. For example, ISCrsp1 and tnpA.5/4/11 were the main transposases located in the central hubs of a network. These significant associations between ARGs and MGEs revealed the strong potential of ARGs transmission towards development of antimicrobial-resistant (AMR) bacteria. On the other hand, treatment of DWW using membranes and ozone was only effective in removing minor species of bacteria and types of ARGs and MGEs.

Conclusion: DWW contained abundant ARGs, and MGEs, which contributed to the occurrence and spread of AMR bacteria. Consequently, DWW would seriously increase environmental health concerns which may be different but have been well-documented from hospital waste waters.

KEYWORDS

dental wastewater, resistome, mobilome, antibiotic resistance genes, antimicrobial resistance

Introduction

Hospital wastewater is a major "breeding" ground for various pathogens, antibiotic resistance bacteria (ARB) and antibiotic resistance genes (ARGs), and has generated continued environmental health concerns (Rizzo et al., 2013). The major reasons for the concerns are that the wastewater facilitated ARG-exchange events among bacteria and generation of multi-drug resistant (MDR) bacteria (Bondarczuk et al., 2016). However, similar concerns for dental waste water (DWW) have not been specifically investigated.

The DWW has some specific differences from that of hospital sewage. For example, certain oral bacteria (i.e., *Pseudomonas. gingivalis*) were associated with development of oral and gastrointestinal cancers (Ahn et al., 2012; Olsen and Yilmaz, 2019). DWW contains non-infectious toxic wastes that include acrylic resin scraps, metal alloys, porcelain, gypsum and dental amalgam, as well as abundant heavy metals, such as mercury, silver, tin, zinc, and copper which have toxic properties (Clarkson et al., 2003; Jones, 2004; Kao et al., 2004; Vandeven and McGinnis, 2004). Consequently, most investigations on health hazards from DWW have been focused onto amalgam and other metals (Al-Khatib and Darwish, 2004; Muhamedagic et al., 2009), and acrylic resin filling materials (Binner et al., 2022). However, investigations using holistic and more sophisticated technologies have not been reported yet.

Metals and biocides may co-select for antimicrobial resistance (AMR; Gelalcha et al., 2017; Pal et al., 2017). Metal contaminations have been reported to significantly influence the diversity, abundance and mobility potential of a broad spectrum of ARGs in urban soils (Song et al., 2017; Zhao et al., 2019). In addition, co-selections of antibiotic-and metal-resistance have been associated with arsenic (As), cadmium (Cd), cobalt (Co), chromium (Cr), copper (Cu), mercury (Hg), nickel (Ni), lead (Pb), and zinc (Zn) (Pal et al., 2015; Song et al., 2017; Zhao et al., 2019). Another study showed that metal contamination in soil increased the potential for horizontal gene transfer (HGT) of ARGs *via* co-selection of ARGs and MGEs, thereby generating a pool of high-risk mobile ARGs (Martínez et al., 2015). With the presence of heavy metals in DWW as opposed to hospital waste water, DWW may involve novel mechanisms for ARGs evolution, HGT development and transmission of AMR. Unfortunately, there has not been a report on such investigation, especially using resistome and mobilome.

Standard handling and disposal of potentially infectious and toxic DWW has been implemented. Many dental clinics have chair-side primary and secondary filter traps which remove approximately 60% of large particles from discharges (Westman and Tuominen, 2000; Johnson and Pichay, 2001; Adegbembo et al., 2002). In addition, membrane bioreactors (MBR) in combination with biological degradation and membrane separation, have been used to remove infectious and non-infectious agents from effluents (Diehl and LaPara, 2010; Ju et al., 2016). To our knowledge, there has been no reports simultaneously identifying the bacterial communities, resistome and mobilome in DWW. Thus, the overall aim of our study was to investigate the abundance and components of bacterial communities, ARGs and MGEs in treated and untreated DWW from a single dental department. The investigation utilized advanced metagenomic and bioinformatic methods to provide in-depth characterizations of the DWW. Our investigation provides novel information on AMR evolution under high metal pressure and on environmental health concerns.

Methods

Dental waste water treatment and sample collection

The DWW samples from each washbasin or dental chair in the department were discharged *via* pipes with filters, to remove large particles, and then into a regulating pool in a tank outside the department. In the tank, the discharged water was homogenized and when the accumulated DWW reached a certain level, it triggered a high voltage discharge which produced ozone and activation of a lift pump which circulate the waste water. After the lift pump stopped working, ozone disinfection continued for another 20 min. In addition, the tank was regularly disinfected once a week by adding chlorine dioxide tablets 5–10 tablets/time (chlorine content 10%) for 30 min.

DWW samples (untreated and treated) of 1 liter each were collected from the specific discharge from the dental department (without mixing with discharge from other sources), weekly from June to July 2021. A total of nine samples were collected in sterile bottles and delivered on ice to the diagnostic microbiology laboratory within 1 h. In the laboratory, each sample was centrifuged at the speed of 10,000 rpm for 5 min at 4°C. The sediments were stored at −80°C until further analysis.

Metagenome sequencing (DNA extraction and identification)

Microbial DNAs from the sewage sediments were extracted using the E.Z.N.A.® soil DNA kit (Omega Bio-Tek, Norcross, GA, United States). DNA concentrations were measured by using the Qubit® dsDNA Assay KitinQubit® 2.0 Fluorometer (Life Technologies, CA, United States), and about 1 μg of DNA (OD: 1.8–2.0) from each sample was used to construct a library. Sequencing libraries were generated using NEB Next®Ultra™ DNA Library Prep Kit for the Illumina (NEB, United States) analysis, and libraries were analyzed using the Agilent 2,100 Bioanalyzer and quantified using PCR. The thermal cycling conditions consisted of initial denaturation at 98°C for 30 s, 12 cycles of 98°C for 10 s, 65°C for 75 s; and a final extension of 5 min at 65°C. Clustering of the index-coded samples were performed on a cBot Cluster Generation System. After the cluster generation, the library preparations were sequenced on an Illumina platform, and paired-end reads were generated. The bacterial genomic sequences were deposited in the NCBI Sequence Read Archive with an accession number (PRJNA869027) which can be shared with readers.

Raw sequence pre-processing

The raw data obtained by sequencing using the Illumina sequencing platform has a certain percentage of low-quality data, and in order to ensure accurate and reliable results for subsequent analysis,

the raw sequencing data need to be preprocessed, including quality control [Trimmomatic (v 0.39; Sewe et al., 2022) parameter: ILLUMINACLIP: adapters _path:2:30:10 SLIDINGWINDOW:4:20 MINLEN:50], and de-hosting sequences (Bowtie2 parameter: --very-sensitive) to obtain clean data for subsequent analysis. The key parameters are explained below: removal of splice sequences (parameter ILLUMINACLIP: adapters_path:2:30:10); scanning sequences (4 bp sliding window size) and excising subsequent sequences if the average quality score is below 20 (99% correct; parameter SLIDINGWINDOW:4:20); and removing sequences with a final length of less than 50 bp (parameter MINLEN:50).

Bioinformatics analyses

Short-read sequencing data were used to identify MGEs and ARGs by the Comprehensive Antibiotic Resistance Database protein homolog model version 1.1.2 (CARD; McArthur et al., 2013) and the ResFinder version 2.1. The MGE database is available from https://github.com/KatariinaParnanen/Mobile Genetic Element Database. Once ARGs and MGEs were identified within assembled contigs, the next step involved identifying which contigs contained both ARGs and MGEs. Co-occurring placements within a single contig were considered as evidence for putative genomic colocalization (Paetzold et al., 2019). Reads were assembled individually into contigs by using MEGAHIT (v 1.1.1), with the following parameters: -k-list 39, 49, …, 129, 141 -mincontig-len 1,000. The qualities of assemblies were evaluated by using QUAST (v 5.0.2; Gurevich et al., 2013). The ORFs on ACCs were annotated or retrieved in the CARD database by using Bowtie (2-2.2.9). According to the result of CARD annotation, MGEs which were located on ACCs were identified in the MGE database by using Bowtie (2-2.2.9). Annotations were categorized as MGEs based on string matches to one of the following keywords (Langmead and Salzberg, 2012).

Network analyses

Network analyses were performed with R using the Vegan and Hmisc packages, and visualizations were conducted on the interactive platform of Gephi 0.9.2. Ggplot2 and pheatmap packages were used to draw a clustering heatmap of ARGs abundance in the samples, and the Hmisc package was used to calculate the correlation matrix for making the network map (Feng et al., 2015). Spearman's rank correlations were used to construct the co-occurrence networks between ARGs and MGEs, ARGs subtypes and microbial communities that occurred in at least 80% of all samples (Karaolia et al., 2021). A correlation between any two items was considered statistically significant if Spearman's correlation coefficient (ρ) was ≥0.7 and the value of p was <0.001.

Results

Diversity of bacterial community, ARG, and MGEs in the DWW

Characteristics of bacterial communities in both the untreated and treated groups of DWW were determined. Alpha diversity including Shannon index/diversity, Simpson index/diversity, richness index and evenness index showed a similar trend between the two groups of DWW. Thus, the Shannon diversity was selected as representative of the alpha diversity (Leiviska and Risteela, 2022) and there was no significant difference between the two groups of DWW samples ($p > 0.05$). For example, the diversity of bacteria and ARGs was insignificantly higher in the treated sewage than in the untreated group, while the diversity of MGEs was insignificantly lower in the treated than that in untreated sewages (Figures 1A–C).

Beta diversity was used to reveal differences in species complexity. The principal coordinate analysis (PCoA) based on Bray-Cutis distance was used to analyze the Beta diversity of OTUs, ARGs and MGEs in both DWW samples, and the PERMANOVA analysis to check whether there was a significant difference in community composition structures between the two groups. The results show that there was no significant difference in bacteria, ARGs and MGEs composition between the two groups. The PCA analyses show that the standard treatment of DWW did not appear to have significant impact on the microbial communities in the waste water (Figures 1D–F).

Microbiome in the DWW

In total, 1, 574 microbial species were identified in the DWW. Among them, there were 1,514 types of bacteria (99.89%), 37 types of fungi (0.093%), 11 types of phages (0.003%), 7 types of Archaea (0.012%), and 5 types of viruses (0.002%). Then, we focus on the bacteria as it is the most abundant component. A total of 4 bacterial phyla with relative abundance of over 1% were identified. The most abundant phyla were Proteobacteria (62.27%), followed by Bacteroidetes (26.66%), Actinobacteria (4.93%), Firmicutes (4.62%), and other phyla (1.52%). In the general level, the most abundant genus was *Pseudomonas* (25.67%), followed by *Chryseobacterium* (19.43%), *Comamonas* (7.98%), *Stenotrophomonas* (3.33%), *Delftia* (3.07%), *Sphingobium* (2.80%), *Morganella* (2.77%), *Afipia* (2.66%), *Prevotella* (2.41%), *Azospira* (2.03%), *Cupriavidus* (1.91%), *Streptococcus* (1.48%), *Aeromonas* (1.42%), *Elizabethkingia* (1.42%), *Neisseria* (1.30%), and *Actinomyces* (1.17%). Among bacterial species, the most abundant species was *Chryseobacterium indologenes* (19.32%), followed by *Pseudomonas putida* (10.32%), *Pseudomonas* sp. *LTGT-11-2Z* (7.35%), *Pseudomonas aeruginosa* (3.88%), *Comamonas terrigena* (3.07%), *Morganella morganii* (2.77%), *Sphingobium yanoikuyae* (2.55%), *Delftia tsuruhatensis* (2.54%), *Afipia broomeae* (2.52%), *Comamonas testosterone* (2.50%), *Stenotrophomonas maltophilia* (2.32%), *Comamonas thiooxydans* (2.18%), *Azospira oryzae* (2.03%), *Cupriavidus metallidurans* (1.85%), *Pseudomonas* sp. *VLB120* (1.73%; only presented bacteria with relative abundance over 1%; Figure 2A).

Variations and relative abundances of ARGs and MGEs types/subtypes in the DWW

The abundance and structure of ARGs were measured. In total, 1,208 types of ARGs were found, belonging to 29 antibiotic types/subtypes. Among them, the most abundant types or subtypes were multi-drug resistant (523), Aminoglycoside (96), Cephalosporin (95), Fluoroquinolone (86), Tetracycline (46), Peptide (37), Cephamycin

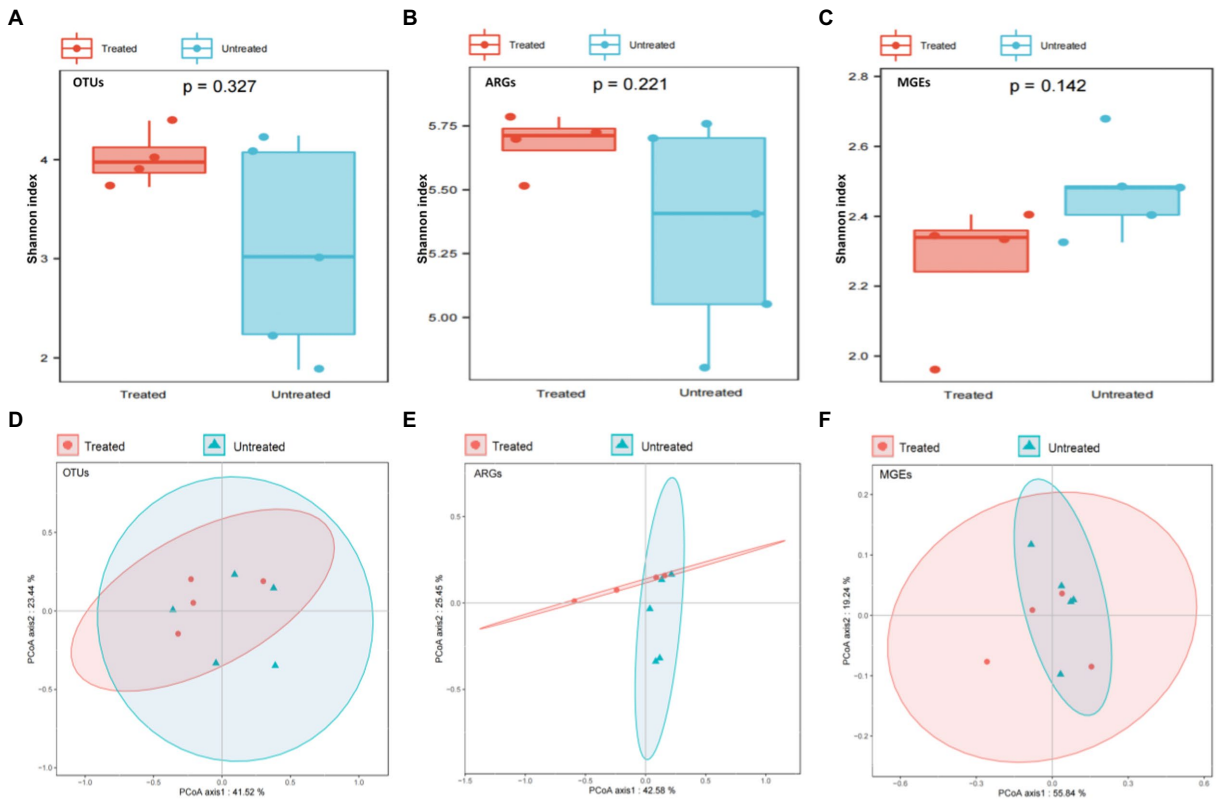

FIGURE 1
Comparison of diversities among OTUs, ARGs and MGEs in the untreated and treated dental waste waters. **(A)** alpha-diversity of OTUs; **(B)** alpha-diversity of ARGs; **(C)** alpha-diversity of MGEs; **(D)** Beta diversity of OTUs; **(E)** Beta diversity of ARGS; **(F)** Beta diversity of MGEs.

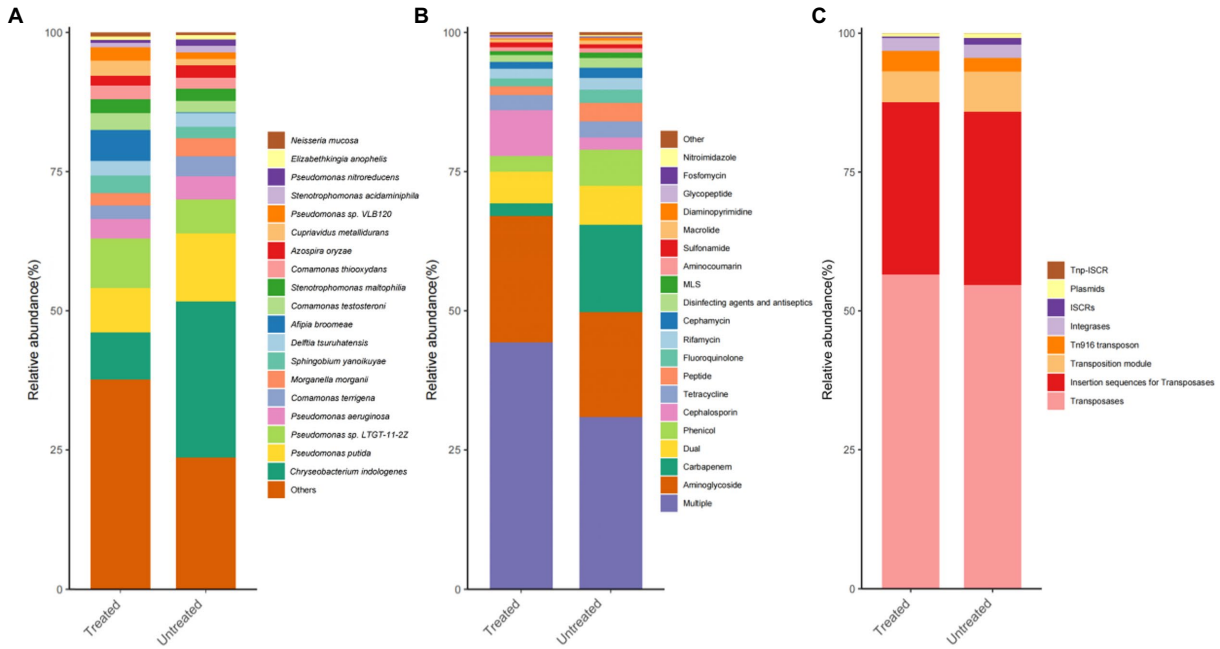

FIGURE 2
Histograms of the relative abundance distributions of the TOP 20 dominant species, ARGs and MGEs in dental waste waters. The abscissa represents the sample groups, and the ordinate corresponds to the proportion of dominant species. **(A)** Bacteria; **(B)** ARGs; and **(C)** MGEs. The color order from top to bottom of the histogram corresponds to the color order of the legend on the right.

(36), Phenicol (35), Glycopeptide (32), Carbapenem (20), Diaminopyrimidine (17), Rifamycin (14), Macrolide (12), Penam (12), MLS (12), Fosfomycin (8), Disinfecting agents and antiseptics (6), Aminocoumarin (5), Lincosamide (5), Sulfonamide (4), Streptogramin (3), Mupirocin (3), Antibacterial free fatty acids (2), Bicyclomycin (1), Elfamycin (1), Nitroimidazole (1), Pleuromutilin (1), and Others (6). Our results show that abundant ARGs persistent in sewage, most of which belong to antibiotics commonly used in clinical practice (Figure 2B).

A total of 93 MGE subtypes belonging to 8 MGE types were found in the DWW. Among them, Transposases (25) was the most frequently detected MGE, followed by Plasmids (28), Insertion sequences for Transposases (23), Tn916 transposon (19), Integrases (4), ISCRs (3), Transposition Module (2), and TNP-ISCR (2; in subtypes; Figure 2C).

Removal efficiency of bacteria, ARGs and MGEs by treatment

Our data indicate that, on average, the relative abundance of nine bacteria was reduced by the treatment (see Methods section) of DWW. For example, *Chryseobacterium Indologenes* were reduced from 28.04 to 8.43% and *Pseudomonas putida* from 12.2 to 7.98%. Other reduced bacteria included *Pseudomonas aeruginosa*, *Comamonas terrigena*, *Morganella morganii*, *Azospira oryzae*, *Stenotrophomonas acidaminiphila*, *Pseudomonas nitroreducens*, *Elizabethkingia anopheles*, *Aeromonas* sp. ASNIH1, and *Veillonella parvula*. On the contrary, the average relative abundance of 9 bacteria was slightly increased after treatment, including *Pseudomonas* sp. Ltgt-11-2z, *Cupriavidus metallidurans*, *Pseudomonas* sp. VLB120, *Sphingobium yanoikuyae*, *Delftia tsuruhatensis*, *Afipia broomeae*, *Comamonas testosterone*, *Stenotrophomonas maltophilia*, *Comamonas thiooxydans*, and *Neisseria mucosa*. However, the differences between the treated and untreated waste water samples were not statistically significant (Figure 3A; Supplementary Table 1).

To provide more accurate determination of changes in bacteria between the untreated and treated groups of DWW, LEfSe analysis was used to identify taxa with differential abundance based on bacteria with LDA threshold >2. Our results revealed 29 taxa with significant differences in both groups: 28 were in the untreated samples, mainly *Alphaproteobacteria*, *Gammaproteobacteria*, and *Actinobacteria*, while only *Betaproteobacteria Acidovoraxavenae* in the class *β-Proteobacteria* among the treated samples.

The relative abundance of 20 ARGs were compared in untreated and treated samples: ARGs of Carbapenem and Phenicol were reduced, while ARGs of Cephalosporin were increased by treatment. The ranges of change were larger than that of other types of ARGs although these changes were not significant. Specifically, subtypes of IND, CGB-1, Paer-catB6, and catB8 had higher clearance rate (>70%) through treatment. On the contrary, subtypes of APH(6)-Id, APH (3″)-Ib, AAC(3)-IIa, and AAC(3)-IIc were slightly increased in abundance after treatment (Figure 3B; Supplementary Tables 2–5).

The differential ARGs and MGEs were evaluated based on LEfSe analysis. The LDA histograms of ARGs and MGEs were presented in Figures 4, 5. The lengths of the bars represent the contribution from different species (LDA Score). The featured ARGs (LDA > 2) were mainly TEM (84 subtypes), tet 39, tet 41, AAC-3 (2 subtypes), AAC-2, VanB, and dfrC in the untreated samples while only dfrA12, dfrA13, and OXA-209 were detected in the treated samples. Among MGEs, only tnpAa (LDA > 2) showed the biggest difference between the treated and untreated samples. The ARG classifications before and after treatment were shown in Figure 5.

Correlations among bacterial communities and ARGs

To further evaluate correlations between ARGs and the more dominant genera, the top abundant ARGs (100 subtypes) and the top 30 bacterial species were selected for the Spearman correlation coefficients analysis (Langmead and Salzberg, 2012). From the analysis, the positive-strong correlations ($r > 0.8$, $p < 0.01$) were selected for building a network of co-occurrences. A co-occurrence network contained 88 nodes (27 bacteria, 95 antibiotic subtypes) and 101 edges. Among all the bacteria, *Pseudomonas putida*, *Pseudomonas aeruginosa*, *Chryseobacterium indologenes*, *Sphingomonas laterariae* were located in the central hub. In particular, *Pseudomonas putida* correlated with Mex (11 subtypes), OprN/J/M, TriC, OpmH, mdtB/F, AxyY, acrB/D, MuxB/C, AcrF, Paer-CpxR, and amrB; *Pseudomonas aeruginosa* with OprM/N, Mex (10 subtypes), MuxB/C, mdtF, amrB; *Sphingomonas laterariae* with sul1, ANT3Ii-AAC6-IID, AAC-6-IB-Su/-HZ, AAC(6′; 8 subtypes), and AAC-3Ib-AAC-6Ib; *Chryseobacterium indologe* with IND (15 subtypes), and CGB-1; *Sphingomonas laterariae* with sul1, ANT3Ii-AAC6-IID, AAC-6 (11 subtypes), AAC-3Ib and AAC-6Ib; *Cupriavidus metallidurans* with aadA/A8, macB, mtrD; *Morganella morganii* with CRP、aadA (7 subtypes); *Neisseria mucosa* with macB; *Pseudomonas nitroreducens* with aadA; *Pseudomonas* sp. *LTGT.11.2Z* with APH(3″)-Ib, APH(6)-Id, and AAC(3)-IIc; *Pseudomonas sp. VLB120* with sul1, APH(3″)-Ib, and APH(6)-Id; *Sphingobium yanoikuyae* with APH(3″)-Ib and APH(6)-Id; *Stenotrophomonas maltophilia* with AAC(3)-IIb/-IIa/-IIc, APH(3″)-Ib, and APH(6)-Id; *Tannerella forsythia* with ErmF; and *Veillonella parvula* with tetM (Figure 6A).

Correlations among ARGs and MGEs

Spearman correlation coefficients were used to evaluate correlations between ARGs and MGEs, using the top 100 ARG subtypes and 93 MGE subtypes. The positive-strong correlations ($r > 0.8$, $p < 0.01$) were selected for a network co-occurrence analysis. The co-occurrence network consisted of 109 nodes and 175 edges. Specifically, ISCrsp1 and tnpA.5/4/11 were located in the central hubs of the network, with the largest number of ARGs connected to them. In addition, ISCrsp1 was mainly correlated with IND (15 subtypes); delta.tnpA with AAC(3; 3 subtypes), aadA (6 subtypes), APH(3″)-Ib, and APH(6)-Id; tnpA.5 with oqxB, Mex (10 subtypes), Paer-CpxR, acrB/D, mdtB/F, AxyY, TriC, adeF, ceoB, OpmH, OprN/J, AcrF, and amrB; tnpA5 with ANT(2″)-Ia, aadA (9 subtypes), and AAC(6′)-IIa; tnpA4 with Mex (12 subtypes), MuxB/C, AcrF, mdtB/F, AxyY, TriC, OpmH, OprM/J/N, acrB/D, amrB, CRP, oqxB, and smeE/B; tnpA11with Mex (9 subtypes), acrB/D/F, MuxB/C, mdtB/F, AxyY, amrB, and OprM; IncP.6. with adeF, ceoB, CRP, oqxB, and mdsB; Tn916 with orf (9 subtypes), and tetM; tnpA10 with APH(3″)-Ib, APH(6)-Id, AAC(3)-IIc/-IIa/-IIb, and sul1(Figure 6B).

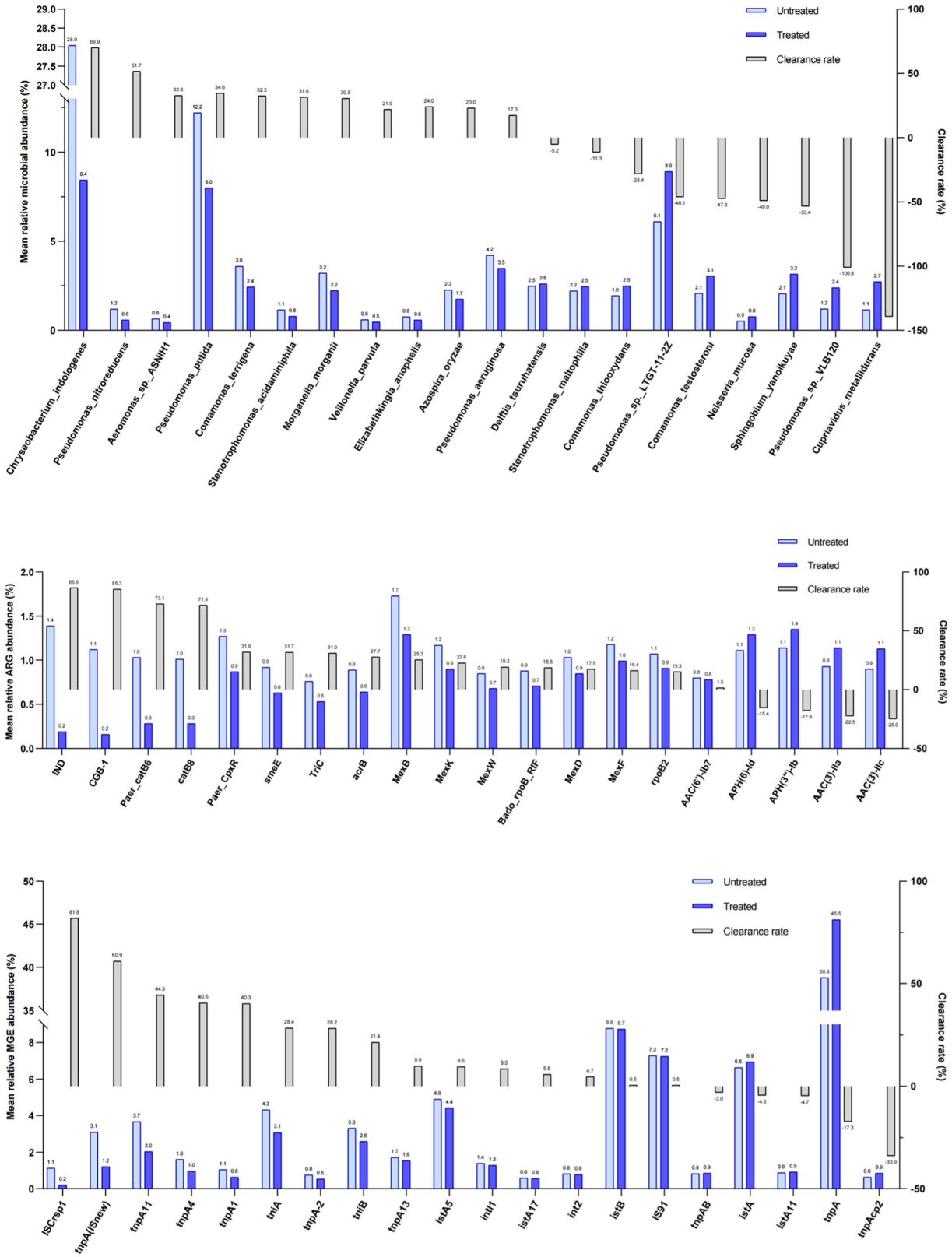

FIGURE 3
Changes in relative abundance of bacteria, ARGs and MGEs in the untreated and treated dental waste waters and their clearance rate. Light blue and dark blue columns show the mean relative abundance of bacteria **(A)**, ARGs **(B)**, and MGEs**(C)** in the two groups. The gray columns show the corresponding clearance rates.

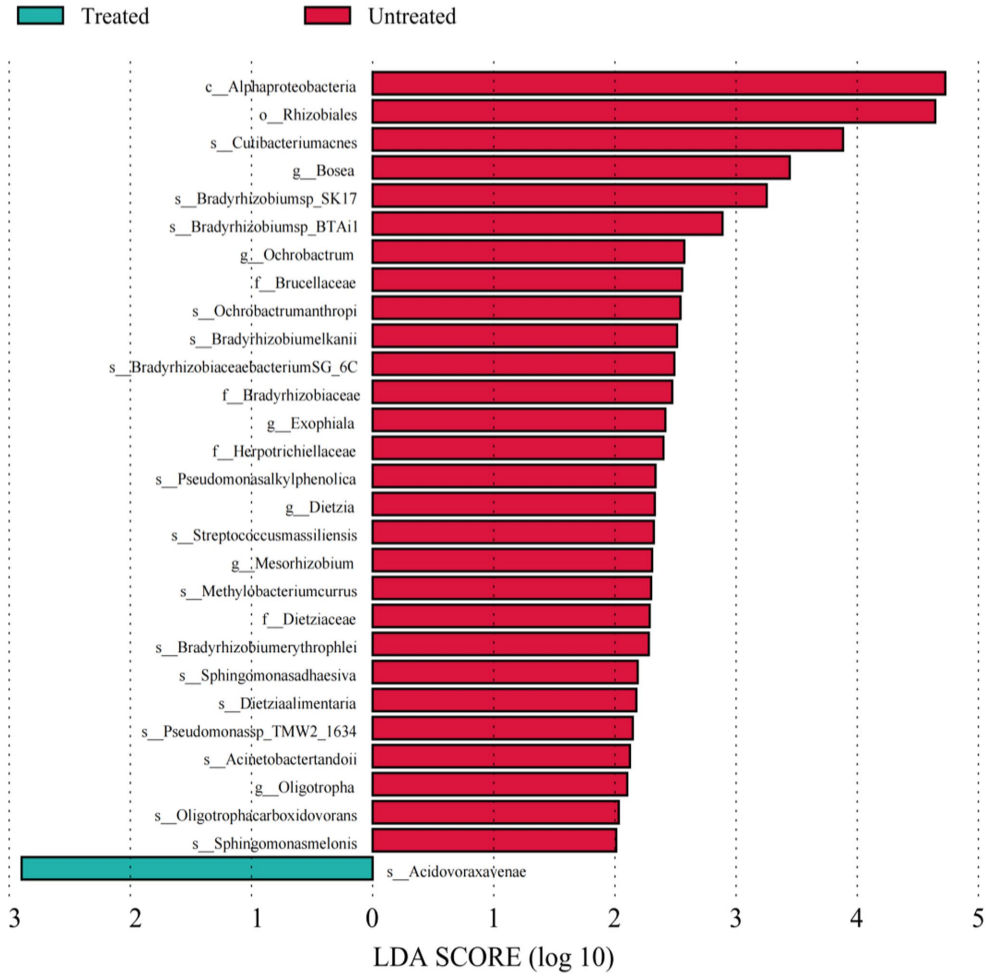

FIGURE 4
Linear discriminant analyses of bacterial species. Each column represents a bacterium, and the length of the column corresponds to the LDA value. The larger the LDA value, the larger the difference. The color of the bar corresponds to the grouping of characteristic bacteria.

Discussion

Studies on hospital waste waters have shown strong associations between their contaminants (ARG, MGE, and antibiotic resistant microbes) and environmental health problems (Su et al., 2017; Quintela-Baluja et al., 2019; Cai et al., 2021). DWW may contain the similar types of contaminants as found in hospital waste but also a substantial amount of heavy metals which may influence interactions among ARG, MGE and microbes. Since metals increase the potential for ARGs spread *via* co-selection of ARGs and MGEs, co-existence of the metals and ARGs would make the DWW to be a novel niche for studying AMR emergence and environmental health concerns. However, there have been very limited reports on environmental health hazards with DWW, especially with new technology such as resistome and mobilome in our investigation. By using a metagenomics approach, DWW samples from one major dental department were found to have resistome which included 1,208 types of ARGs belonging to 29 antibiotic types/subtypes. The most abundant ones were ARGs of multidrug resistance, followed by ARGs of Aminoglycoside, Cephalosporin, Fluoroquinolone, Tetracycline, Peptide, Cephamycin, Phenicol, Glycopeptide, Carbapenem, Diaminopyrimidine, Rifamycin, Macrolide, Penam, MLS, and Fosfomycin. Importantly, all of the mentioned resistance was to antibiotics which were commonly used in clinical practice in the hospital but were less frequently used in the dental department where the waste water samples were collected. Our results are intriguing as well as meaningful because DWW was thought to be rarely involved in the transmission of AMR. Furthermore, a wide variety of ARGs were unexpectedly found in DWW which might have been influenced by the abundant metals. These unique features need to be further investigation in order to better understand mechanisms and to develop more effective prevention strategies.

Microbiomes have been considered as an important driver for ARG disseminations in the environment (Baym et al., 2016; Jia et al., 2017; Chen et al., 2019; Yu et al., 2020). The source of ARGs in the DWW may come from oral microbiome. Indeed, our collected DWW samples included 1,514 types of bacteria, 37 types of fungi, 11 types of phages, 7 types of Archaea, and 5 types of viruses. Among them, bacteria were the majority while the most abundant phyla and genus were Proteobacteria (76.4%), and Pseudomonas (25.67%), particularly *Pseudomonas putida*, *Pseudomonas* sp. *LTGT-11-2Z*, and *Pseudomonas aeruginosa*. Importantly, they also belonged to the important

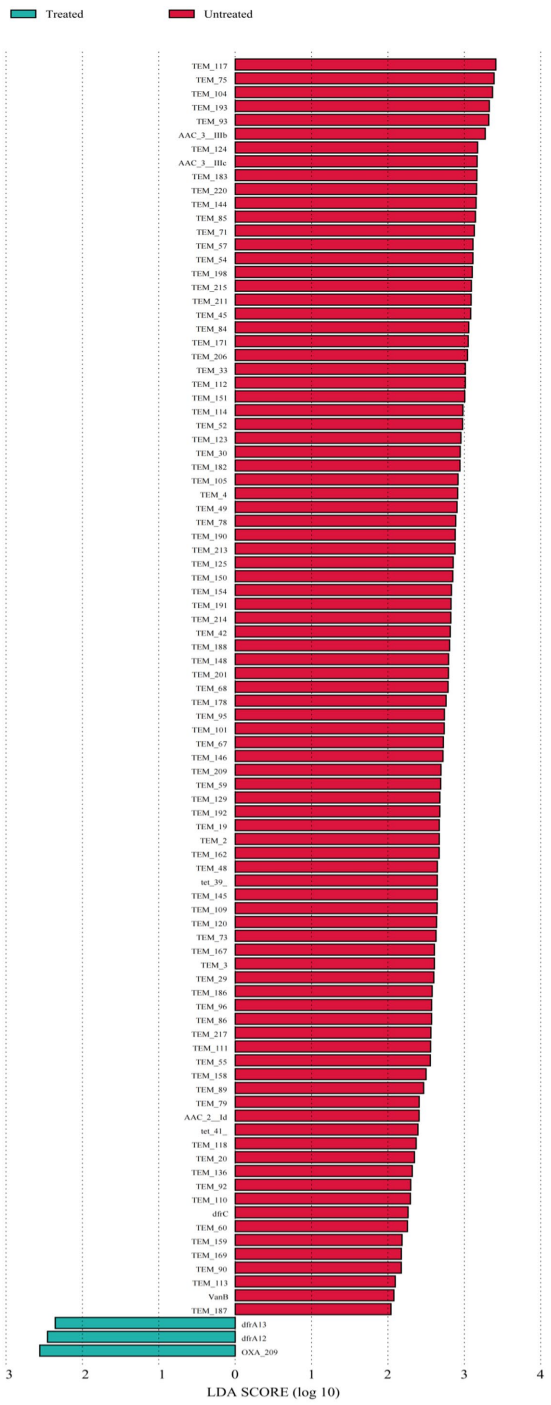

FIGURE 5
ARGs linear discriminant analyses. Each horizontal column represents an ARG, and the length of the column corresponds to the LDA value. The larger the LDA value, the larger the difference. The color of this bar corresponds to the grouping of feature ARGs.

pathogens found in dental clinic. A previous study revealed that composition of the microbial community in waste water was associated with 68.2% of the variations in ARGs (Zhang et al., 2016). Using the association data from our metagenomic analyses, network and binning analyses were conducted as shown in other reports (Guo et al., 2017; Liu et al., 2019; Sun et al., 2021). Our analyses revealed a complicated co-occurrence network which contained 88 nodes and 101 edges, and which involved 27 bacteria and 95 ARGs subtypes. Specifically, *Pseudomonas putida*, *Pseudomonas aeruginosa*, *Chryseobacterium indologenes*, and *Sphingomonas laterariae* were located in the central hubs of the network and involved with abundant ARGs referring to various antibiotics. Some of the ARBs which were associated with resistance to multiple drugs have been reported to contribute to increased morbidity and mortality among patients (Morrison and Zembower, 2020). Therefore, understanding existence of networks for ARGs and bacteria would provide valuable information in predicting novel ARB and in designing prevention protocols against emerging AMR (Chng et al., 2020).

Mobilome is defined as all detectable HGT elements within a given metagenomic dataset and these elements included plasmids, integrative conjugative elements (ICE), transposons, and insertional repeat sequences (Pal et al., 2015; Ju et al., 2019; Yun et al., 2021). In this study, mobilome of the DWW was composed of 93 MGE subtypes belonging to 8 MGE types. Among them, transposases were the most frequently detected MGEs. With our correlation analyses, most of ARGs showed significant correlations with total abundance of MGEs. The co-occurrence network consisted of 109 nodes and 175 edges. Among all MGEs, ISCrsp1, and tnpA.5/4/11 were located in the central hubs of the network and might serve as links to different ARG types. With the large number of ARGs connected to them, they took on the active role of ARG dissemination. Previous studies indicate that most co-occurring ARGs with metals also co-occurred with MGEs (Song et al., 2017; Zhao et al., 2019).

Heavy metals can promote resistance to antibiotics either *via* cross-resistance (a single genetic unit conferred resistant to both metals and antibiotics), co-resistance (both metal resistant genes (MRGs) and ARGs are associated with same MGEs), or co-regulation (both metal and antibiotic resistance shared their regulatory systems; Baker-Austin et al., 2006; Imran et al., 2019). Moreover, the relative MRGs and ARGs abundances would increase with increasing metals concentration (Hu et al., 2017). The metal-driven selection of AMR is markedly greater when both MRGs and ARGs are situated on the same MGEs (e.g., plasmids, transposons, and integrons; Di Cesare et al., 2016; Hu et al., 2017). For example, int1 has been closely associated with MRG czcA, coding for cobalt (Co), zinc (Zn), and cadmium (Cd) resistance, and beta-lactamase resistance (Stokes et al., 2006; Gillings et al., 2008), indicating that MRGs and ARGs may be transferred simultaneously to host bacteria *via* int1 in the HGT process (Gupta et al., 2022). Transformation is the main pathway of HGT, which can take place in more than 80 naturally competent bacterial species with distant phylogenetical backgrounds, even consisting of human pathogens (Thomas and Nielsen, 2005; Maeusli et al., 2020). Due to the prevalence of extracellular ARGs, antibiotics and naturally competent bacteria, the environmental transformation of ARGs is estimated to be quite frequent and is one of the predominant pathways to spread AMR (de Aldecoa et al., 2017). Ag, CuO and ZnO-based NPs/ions could promote the natural transformation of plasmids harboring ARGs (Zhang et al., 2022), and the promoting effect can occur at clinically relevant concentrations (Hernandez-Sierra et al., 2008; Mohler et al., 2018) or realistic concentrations within aquatic environments (Brunetti et al., 2015). On the other hand, heavy metals pollution has altered bacterial diversity and abundance, as the bacterial population is sensitive to heavy metals (Chen et al., 2018), and the long-term presence of high concentrations of metals in polluted water may increase heavy metal resistance in a

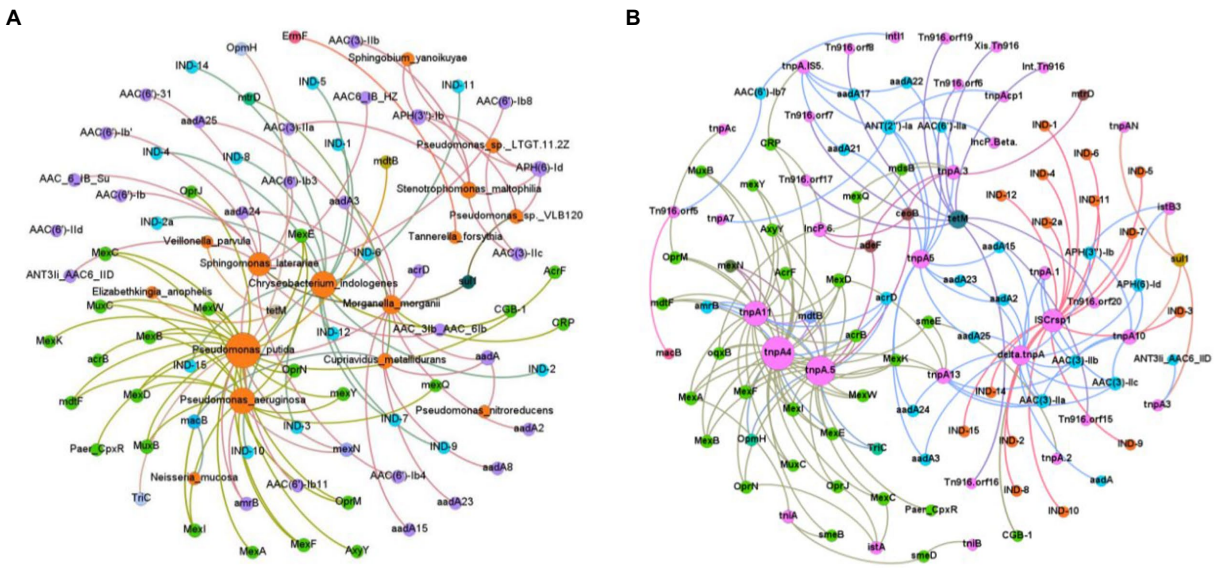

FIGURE 6
Network analysis of co-occurrence patterns. Each node connection indicates strong (Spearman correlation coefficient ρ>0.8) and significant (p<0.01) correlation. The size of each node is proportional to the number of connections. **(A)** Co-occurrence network of bacteria and ARG. **(B)** Co-occurrence network of ARGs and MGEs.

variety of bacteria (Gupta et al., 2022). These observations suggest an underlying metal-driven co-selection process which was linked with existence of cross-resistance (Li et al., 2017; Zhao et al., 2019). Furthermore, MGEs are actively involved in HGT of ARGs in neighboring microbial communities (Gupta et al., 2018). Consequently, when DWW are released into the environment, it becomes very difficult to efficiently eliminate the generation of ARB (Rakita et al., 2020).

In our study, the main limitation is that metal concentrations were not determined due to our limitations of analytical techniques. Theoretically, a large amount of heavy metals discharged from clinical practice in every day will inevitably lead to a large amount of heavy metals in DWW. On the other hand, previous reports indicate the presence of high levels of Cu, Zn, Hg and MeHg in DWW (Rani et al., 2015). Thus, co-selection and cross-resistance would occur in DWW. If ARGs and MRG is found on the same MGEs, and this physical linkage results in co-resistance. Cross-resistance is another co-selection mechanism which occurred when single genes encoded resistance to both antibiotics and metals (Li et al., 2017; Zhao et al., 2019). Better understanding of how metals influence formation of ARGs and MGEs would provide insights into novel mechanisms of HGT and emergence of ARB in the future.

The generation and mobility of clinically relevant ARGs in waste waters post significant risk to human health (Carvalho and Santos, 2016; Liu et al., 2018b; Slizovskiy et al., 2020). Our data clearly show the abundance of ARGs and MGEs in the DWW, therefore more effective treatment of the waste water is of great importance. Unfortunately, few existing processes have been designed to remove ARGs, and our data as well as others indicate that such processes were not highly effective (Mao et al., 2015; Bengtsson-Palme et al., 2016; Di Cesare et al., 2016; Pazda et al., 2019; Vinzelj et al., 2020; Li et al., 2022). The treatment process in our dental department utilized ozone, which reacts directly or indirectly *via* a hydroxyl radical mechanism to reduce organic and inorganic materials to become more biodegradable, and which efficiently inactivate a wide range of microorganisms (Tripathi and Tripathi, 2011). Our results show that only a few bacteria with clearance rate higher than 30% were observed. On the contrary, some others were increased. Considering that metagenomics only detects bacterial DNA, the data do not represent activity and integrity of bacteria. Thus, clearance or abundance of bacteria should be reconfirmed *via* bacterial isolates.

As to resistome and mobilome, our data show that ozone treatment had no obvious effects in changing the abundance of resistome and mobilome, although a previous study revealed that antibiotic-resistant hosts and resistant genes were significantly inactivated by ozone treatments (Pei et al., 2016), as well as a significant portion of only MLS and tetracycline genes (Raza et al., 2022). On the other hand, beta lactam ARGs were increased by UV-, chlorine-, and ozone-based treatment strategies (Guo et al., 2013; Alexander et al., 2016; Ferro et al., 2017; Liu et al., 2018a). These different observations are likely due to the use of different experimental designs, sample sizes and technologies. More systematic studies are needed to identify efficacy of waste water treatment protocols.

ARGs and MGEs have been listed as serious and emerging environmental pollutants and health problems from hospital waste waters (Gillings et al., 2008; Xu et al., 2016; Chen et al., 2019). Our data strengthened the addition of DWW to the list of concerns. Our study reveals that DWW harbored a significant and diversity of microbes, ARGs, and MGEs, providing a persistent selection pressure (in the presence of heavy metals) and possibly resulting in the occurrence or emergence of novel antimicrobial determinants. Our observations provide evidence which underscore the need for improved disinfection methods, and for monitoring the waste prior to disposal. Effective effort would lead to reducing the spread of drug resistant bacteria into hospitals and communities.

Author's note

The proposal was approved by the institutional review board of the First Affiliated Hospital of Shantou University Medical College.

Data availability statement

The dataset presented in this study are deposited and can be found in an online repository. The name of the repository and accession number(s) can be found below: NCBI repository - https://www.ncbi.nlm.nih.gov/, PRJNA869027.

Ethics statement

The studies involving human participants were reviewed and approved by First Affiliated Hospital of Shantou University Medical College. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

XJ and WG contributed equally to this manuscript. YX and WG carried out the sample collection, experimental studies, and drafted the manuscript. FYa, QX, LC, FYu, and PY participated in the innovative design of the study. XL, MZ, YY, XG, and MW, performed the statistical analysis. QZ and XJ conceived the study, participated in its design and coordination, and helped to revise the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1106157/full#supplementary-material

References

Adegbembo, A. O., Watson, P. A., and Lugowski, S. J. (2002). The weight of wastes generated by removal of dental amalgam restorations and the concentration of mercury in dental wastewater. *J. Can. Dent. Assoc.* 68, 553–558.

Ahn, J., Chen, C. Y., and Hayes, R. B. (2012). Oral microbiome and oral and gastrointestinal cancer risk. *Cancer Causes Control* 23, 399–404. doi: 10.1007/s10552-011-9892-7

Alexander, J., Knopp, G., Dotsch, A., Wieland, A., and Schwartz, T. (2016). Ozone treatment of conditioned wastewater selects antibiotic resistance genes, opportunistic bacteria, and induce strong population shifts. *Sci. Total Environ.* 559, 103–112. doi: 10.1016/j.scitotenv.2016.03.154

Al-Khatib, I. A., and Darwish, R. (2004). Assessment of waste amalgam management in dental clinics in Ramallah and al-Bireh cities in Palestine. *Int. J. Environ. Health Res.* 14, 179–183. doi: 10.1080/09603120420002000218598

Baker-Austin, C., Wright, M. S., Stepanauskas, R., and McArthur, J. V. (2006). Co-selection of antibiotic and metal resistance. *Trends Microbiol.* 14, 176–182. doi: 10.1016/j.tim.2006.02.006

Baym, M., Lieberman, T. D., Kelsic, E. D., Chait, R., Gross, R., Yelin, I., et al. (2016). Spatiotemporal microbial evolution on antibiotic landscapes. *Science* 353, 1147–1151. doi: 10.1126/science.aag0822

Bengtsson-Palme, J., Hammaren, R., Pal, C., Ostman, M., Bjorlenius, B., Flach, C. F., et al. (2016). Elucidating selection processes for antibiotic resistance in sewage treatment plants using metagenomics. *Sci. Total Environ.* 572, 697–712. doi: 10.1016/j.scitotenv.2016.06.228

Binner, H., Kamali, N., Harding, M., and Sullivan, T. (2022). Characteristics of wastewater originating from dental practices using predominantly mercury-free dental materials. *Sci. Total Environ.* 814:152632. doi: 10.1016/j.scitotenv.2021.152632

Bondarczuk, K., Markowicz, A., and Piotrowska-Seget, Z. (2016). The urgent need for risk assessment on the antibiotic resistance spread via sewage sludge land application. *Environ. Int.* 87, 49–55. doi: 10.1016/j.envint.2015.11.011

Brunetti, G., Donner, E., Laera, G., Sekine, R., Scheckel, K. G., Khaksar, M., et al. (2015). Fate of zinc and silver engineered nanoparticles in sewerage networks. *Water Res.* 77, 72–84. doi: 10.1016/j.watres.2015.03.003

Cai, L., Sun, J., Yao, F., Yuan, Y., Zeng, M., Zhang, Q., et al. (2021). Antimicrobial resistance bacteria and genes detected in hospital sewage provide valuable information in predicting clinical antimicrobial resistance. *Sci. Total Environ.* 795:148815. doi: 10.1016/j.scitotenv.2021.148815

Carvalho, I. T., and Santos, L. (2016). Antibiotics in the aquatic environments: a review of the European scenario. *Environ. Int.* 94, 736–757. doi: 10.1016/j.envint.2016.06.025

Chen, Y., Jiang, Y. M., Huang, H. Y., Mou, L. C., Ru, J. L., Zhao, J. H., et al. (2018). Long-term and high-concentration heavy-metal contamination strongly influences the microbiome and functional genes in Yellow River sediments. *Sci. Total Environ.* 637-638, 1400–1412. doi: 10.1016/j.scitotenv.2018.05.109

Chen, Y., Su, J. Q., Zhang, J., Li, P., Chen, H., Zhang, B., et al. (2019). High-throughput profiling of antibiotic resistance gene dynamic in a drinking water river-reservoir system. *Water Res.* 149, 179–189. doi: 10.1016/j.watres.2018.11.007

Chng, K. R., Li, C., Bertrand, D., Ng, A. H. Q., Kwah, J. S., Low, H. M., et al. (2020). Cartography of opportunistic pathogens and antibiotic resistance genes in a tertiary hospital environment. *Nat. Med.* 26, 941–951. doi: 10.1038/s41591-020-0894-4

Clarkson, T. W., Magos, L., and Myers, G. J. (2003). The toxicology of mercury — current exposures and clinical manifestations. *N. Engl. J. Med.* 349, 1731–1737. doi: 10.1056/NEJMra022471

de Aldecoa, A. L. I., Zafra, O., and Gonzalez-Pastor, J. E. (2017). Mechanisms and regulation of extracellular DNA release and its biological roles in microbial communities. *Front. Microbiol.* 8:19. doi: 10.3389/fmicb.2017.01390

Di Cesare, A., Eckert, E. M., D'Urso, S., Bertoni, R., Gillan, D. C., Wattiez, R., et al. (2016). Co-occurrence of integrase 1, antibiotic and heavy metal resistance genes in municipal wastewater treatment plants. *Water Res.* 94, 208–214. doi: 10.1016/j.watres.2016.02.049

Diehl, D. L., and LaPara, T. M. (2010). Effect of temperature on the fate of genes encoding tetracycline resistance and the integrase of class 1 integrons within anaerobic and aerobic digesters treating municipal wastewater solids. *Environ. Sci. Technol.* 44, 9128–9133. doi: 10.1021/es102765a

Feng, Q., Liang, S., Jia, H., Stadlmayr, A., Tang, L., Lan, Z., et al. (2015). Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat. Commun.* 6:6528. doi: 10.1038/ncomms7528

Ferro, G., Guarino, F., Cicatelli, A., and Rizzo, L. (2017). Beta-lactams resistance gene quantification in an antibiotic resistant Escherichia coli water suspension treated by advanced oxidation with UV/H2O2. *J. Hazard. Mater.* 323, 426–433. doi: 10.1016/j.jhazmat.2016.03.014

Gelalcha, A. G., Kebede, A., and Mamo, H. (2017). Light-emitting diode fluorescent microscopy and Xpert MTB/RIF(R) assay for diagnosis of pulmonary tuberculosis

among patients attending Ambo hospital, West-Central Ethiopia. *BMC Infect. Dis.* 17:613. doi: 10.1186/s12879-017-2701-5

Gillings, M., Boucher, Y., Labbate, M., Holmes, A., Krishnan, S., Holley, M., et al. (2008). The evolution of class 1 integrons and the rise of antibiotic resistance. *J. Bacteriol.* 190, 5095–5100. doi: 10.1128/JB.00152-08

Guo, J., Li, J., Chen, H., Bond, P. L., and Yuan, Z. (2017). Metagenomic analysis reveals wastewater treatment plants as hotspots of antibiotic resistance genes and mobile genetic elements. *Water Res.* 123, 468–478. doi: 10.1016/j.watres.2017.07.002

Guo, M. T., Yuan, Q. B., and Yang, J. (2013). Microbial selectivity of UV treatment on antibiotic-resistant heterotrophic bacteria in secondary effluents of a municipal wastewater treatment plant. *Water Res.* 47, 6388–6394. doi: 10.1016/j.watres.2013.08.012

Gupta, S., Graham, D. W., Sreekrishnan, T. R., and Ahammad, S. Z. (2022). Effects of heavy metals pollution on the co-selection of metal and antibiotic resistance in urban rivers in UK and India. *Environ. Pollut.* 306:11. doi: 10.1016/j.envpol.2022.119326

Gupta, S. K., Shin, H., Han, D., Hur, H. G., and Unno, T. (2018). Metagenomic analysis reveals the prevalence and persistence of antibiotic-and heavy metal-resistance genes in wastewater treatment plant. *J. Microbiol.* 56, 408–415. doi: 10.1007/s12275-018-8195-z

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Hernandez-Sierra, J. F., Ruiz, F., Pena, D. C. C., Martinez-Gutierrez, F., Martinez, A. E., Guillen, A. D. P., et al. (2008). The antimicrobial sensitivity of Streptococcus mutans to nanoparticles of silver, zinc oxide, and gold. *Nanomed. Nanotechnol. Biol. Med.* 4, 237–240. doi: 10.1016/j.nano.2008.04.005

Hu, H. W., Wang, J. T., Li, J., Shi, X. Z., Ma, Y. B., Chen, D. L., et al. (2017). Long-term nickel contamination increases the occurrence of antibiotic resistance genes in agricultural soils. *Environ. Sci. Technol.* 51, 790–800. doi: 10.1021/acs.est.6b03383

Imran, M., Das, K. R., and Naik, M. M. (2019). Co-selection of multi-antibiotic resistance in bacterial pathogens in metal and microplastic contaminated environments: An emerging health threat. *Chemosphere* 215, 846–857. doi: 10.1016/j.chemosphere.2018.10.114

Jia, S., Zhang, X. X., Miao, Y., Zhao, Y., Ye, L., Li, B., et al. (2017). Fate of antibiotic resistance genes and their associations with bacterial community in livestock breeding wastewater and its receiving river water. *Water Res.* 124, 259–268. doi: 10.1016/j.watres.2017.07.061

Johnson, W. J., and Pichay, T. J. (2001). Dentistry, amalgam, and pollution prevention. *J. Calif. Dent. Assoc.* 29, 509–516. doi: 10.1080/19424396.2001.12223199

Jones, D. W. (2004). Putting dental mercury pollution into perspective. *Br. Dent. J.* 197, 175–177. doi: 10.1038/sj.bdj.4811564

Ju, F., Beck, K., Yin, X., Maccagnan, A., McArdell, C. S., Singer, H. P., et al. (2019). Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. *ISME J.* 13, 346–360. doi: 10.1038/s41396-018-0277-8

Ju, F., Li, B., Ma, L., Wang, Y., Huang, D., and Zhang, T. (2016). Antibiotic resistance genes and human bacterial pathogens: co-occurrence, removal, and enrichment in municipal sewage sludge digesters. *Water Res.* 91, 1–10. doi: 10.1016/j.watres.2015.11.071

Kao, R. T., Dault, S., and Pichay, T. (2004). Understanding the mercury reduction issue: the impact of mercury on the environment and human health. *J. Calif. Dent. Assoc.* 32, 574–579. doi: 10.1080/19424396.2004.12224004

Karaolia, P., Vasileiadis, S., Michael, S. G., Karpouzas, D. G., and Fatta-Kassinos, D. (2021). Shotgun metagenomics assessment of the resistome, mobilome, pathogen dynamics and their ecological control modes in full-scale urban wastewater treatment plants. *J. Hazard. Mater.* 418:126387. doi: 10.1016/j.jhazmat.2021.126387

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Leiviska, T., and Risteela, S. (2022). Analysis of pharmaceuticals, hormones and bacterial communities in a municipal wastewater treatment plant - comparison of parallel full-scale membrane bioreactor and activated sludge systems. *Environ. Pollut.* 292:118433. doi: 10.1016/j.envpol.2021.118433

Li, L.-G., Xia, Y., and Zhang, T. (2017). Co-occurrence of antibiotic and metal resistance genes revealed in complete genome collection. *ISME J.* 11, 651–662. doi: 10.1038/ismej.2016.155

Li, Y., Zhang, X., Morgan, V. L., Lohman, H. A. C., Rowles, L. S., Mittal, S., et al. (2022). QSDsan: An integrated platform for quantitative sustainable design of sanitation and resource recovery systems. *Environ. Sci. Water Res. Technol.* 8, 2289–2303.

Liu, Z., Klumper, U., Liu, Y., Yang, Y., Wei, Q., Lin, J. G., et al. (2019). Metagenomic and metatranscriptomic analyses reveal activity and hosts of antibiotic resistance genes in activated sludge. *Environ. Int.* 129, 208–220. doi: 10.1016/j.envint.2019.05.036

Liu, X., Lu, S., Guo, W., Xi, B., and Wang, W. (2018b). Antibiotics in the aquatic environments: a review of lakes, China. *Sci. Total Environ.* 627, 1195–1208. doi: 10.1016/j.scitotenv.2018.01.271

Liu, S. S., Qu, H. M., Yang, D., Hu, H., Liu, W. L., Qiu, Z. G., et al. (2018a). Chlorine disinfection increases both intracellular and extracellular antibiotic resistance genes in a full-scale wastewater treatment plant. *Water Res.* 136, 131–136. doi: 10.1016/j.watres.2018.02.036

Maeusli, M., Lee, B., Miller, S., Reyna, Z., Lu, P., Yan, J., et al. (2020). Horizontal gene transfer of antibiotic resistance from *Acinetobacter baylyi* to *Escherichia coli* on lettuce and subsequent antibiotic resistance transmission to the gut microbiome. *mSphere* 5:7. doi: 10.1128/mSphere.00329-20

Mao, D., Yu, S., Rysz, M., Luo, Y., Yang, F., Li, F., et al. (2015). Prevalence and proliferation of antibiotic resistance genes in two municipal wastewater treatment plants. *Water Res.* 85, 458–466. doi: 10.1016/j.watres.2015.09.010

Martínez, J. L., Coque, T. M., and Baquero, F. (2015). What is a resistance gene? Ranking risk in resistomes. *Nat. Rev. Microbiol.* 13, 116–123. doi: 10.1038/nrmicro3399

McArthur, A. G., Waglechner, N., Nizam, F., Yan, A., Azad, M. A., Baylay, A. J., et al. (2013). The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* 57, 3348–3357. doi: 10.1128/AAC.00419-13

Mohler, J. S., Sim, W., Blaskovich, M. A. T., Cooper, M. A., and Ziora, Z. M. (2018). Silver bullets: a new lustre on an old antimicrobial agent. *Biotechnol. Adv.* 36, 1391–1411. doi: 10.1016/j.biotechadv.2018.05.004

Morrison, L., and Zembower, T. R. (2020). Antimicrobial Resistance. *Gastrointest. Endosc. Clin. N. Am.* 30, 619–635. doi: 10.1016/j.giec.2020.06.004

Muhamedagic, B., Muhamedagic, L., and Masic, I. (2009). Dental office waste - public health and ecological risk. *Mater. Soc.* 21, 35–38. doi: 10.5455/aim.2009.21.35-39

Olsen, I., and Yilmaz, O. (2019). Possible role of *Porphyromonas gingivalis* in orodigestive cancers. *J. Oral Microbiol.* 11:1563410. doi: 10.1080/20002297.2018.1563410

Paetzold, B., Willis, J. R., Pereira de Lima, J., Knodlseder, N., Bruggemann, H., Quist, S. R., et al. (2019). Skin microbiome modulation induced by probiotic solutions. *Microbiome* 7:95. doi: 10.1186/s40168-019-0709-3

Pal, C., Asiani, K., Arya, S., Rensing, C., Stekel, D. J., Larsson, D. G. J., et al. (2017). Metal resistance and its association with antibiotic resistance. *Adv. Microb. Physiol.* 70, 261–313. doi: 10.1016/bs.ampbs.2017.02.001

Pal, C., Bengtsson-Palme, J., Kristiansson, E., and Larsson, D. G. (2015). Co-occurrence of resistance genes to antibiotics, biocides and metals reveals novel insights into their co-selection potential. *BMC Genomics* 16:964. doi: 10.1186/s12864-015-2153-5

Pazda, M., Kumirska, J., Stepnowski, P., and Mulkiewicz, E. (2019). Antibiotic resistance genes identified in wastewater treatment plant systems - a review. *Sci. Total Environ.* 697:134023. doi: 10.1016/j.scitotenv.2019.134023

Pei, J., Yao, H., Wang, H., Ren, J., and Yu, X. (2016). Comparison of ozone and thermal hydrolysis combined with anaerobic digestion for municipal and pharmaceutical waste sludge with tetracycline resistance genes. *Water Res.* 99, 122–128. doi: 10.1016/j.watres.2016.04.058

Quintela-Baluja, M., Abouelnaga, M., Romalde, J., Su, J. Q., Yu, Y., Gomez-Lopez, M., et al. (2019). Spatial ecology of a wastewater network defines the antibiotic resistance genes in downstream receiving waters. *Water Res.* 162, 347–357. doi: 10.1016/j.watres.2019.06.075

Rakita, A., Nikolic, N., Mildner, M., Matiasek, J., and Elbe-Burger, A. (2020). Re-epithelialization and immune cell behaviour in an ex vivo human skin model. *Sci. Rep.* 10:1. doi: 10.1038/s41598-019-56847-4

Rani, A., Rockne, K. J., Drummond, J., Al-Hinai, M., and Ranjan, R. (2015). Geochemical influences and mercury methylation of a dental wastewater microbiome. *Sci. Rep.* 5:12872. doi: 10.1038/srep12872

Raza, S., Shin, H., Hur, H. G., and Unno, T. (2022). Higher abundance of core antimicrobial resistant genes in effluent from wastewater treatment plants. *Water Res.* 208:117882. doi: 10.1016/j.watres.2021.117882

Rizzo, L., Manaia, C., Merlin, C., Schwartz, T., Dagot, C., Ploy, M. C., et al. (2013). Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: a review. *Sci. Total Environ.* 447, 345–360. doi: 10.1016/j.scitotenv.2013.01.032

Sewe, S. O., Silva, G., Sicat, P., Seal, S. E., and Visendi, P. (2022). Trimming and validation of Illumina short reads using trimmomatic, trinity assembly, and assessment of RNA-Seq data. *Methods Mol. Biol.* 2443, 211–232. doi: 10.1007/978-1-0716-2067-0_11

Slizovskiy, I. B., Mukherjee, K., Dean, C. J., Boucher, C., and Noyes, N. R. (2020). Mobilization of antibiotic resistance: are current approaches for colocalizing resistomes and mobilomes useful? *Front. Microbiol.* 11:1376. doi: 10.3389/fmicb.2020.01376

Song, J., Rensing, C., Holm, P. E., Virta, M., and Brandt, K. K. (2017). Comparison of metals and tetracycline as selective agents for development of tetracycline resistant bacterial communities in agricultural soil. *Environ. Sci. Technol.* 51, 3040–3047. doi: 10.1021/acs.est.6b05342

Stokes, H. W., Nesbo, C. L., Holley, M., Bahl, M. I., Gillings, M. R., and Boucher, Y. (2006). Class 1 integrons potentially predating the association with Tn402-like transposition genes are present in a sediment microbial community. *J. Bacteriol.* 188, 5722–5730. doi: 10.1128/JB.01950-05

Su, J.-Q., An, X.-L., Li, B., Chen, Q.-L., Gillings, M. R., Chen, H., et al. (2017). Metagenomics of urban sewage identifies an extensively shared antibiotic resistome in China. *Microbiome* 5:84. doi: 10.1186/s40168-017-0298-y

Sun, Y., Clarke, B., Clarke, J., and Li, X. (2021). Predicting antibiotic resistance gene abundance in activated sludge using shotgun metagenomics and machine learning. *Water Res.* 202:117384. doi: 10.1016/j.watres.2021.117384

Thomas, C. M., and Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* 3, 711–721. doi: 10.1038/nrmicro1234

Tripathi, S., and Tripathi, B. D. (2011). Efficiency of combined process of ozone and bio-filtration in the treatment of secondary effluent. *Bioresour. Technol.* 102, 6850–6856. doi: 10.1016/j.biortech.2011.04.035

Vandeven, J. A., and McGinnis, S. L. (2004). Cost-effectiveness of removing amalgam from dental wastewater. *J. Calif. Dent. Assoc.* 32, 564–573. doi: 10.1080/19424396.2004.12224003

Vinzelj, J., Joshi, A., Insam, H., and Podmirseg, S. M. (2020). Employing anaerobic fungi in biogas production: challenges & opportunities. *Bioresour. Technol.* 300:122687. doi: 10.1016/j.biortech.2019.122687

Westman, J. F., and Tuominen, T. (2000). Amalgam waste management. Issues & answers. *N. Y. State Dent. J.* 66, 20–24.

Xu, L., Ouyang, W., Qian, Y., Su, C., Su, J., and Chen, H. (2016). High-throughput profiling of antibiotic resistance genes in drinking water treatment plants and distribution systems. *Environ. Pollut.* 213, 119–126. doi: 10.1016/j.envpol.2016.02.013

Yu, K., Li, P., Chen, Y., Zhang, B., Huang, Y., Huang, F. Y., et al. (2020). Antibiotic resistome associated with microbial communities in an integrated wastewater reclamation system. *Water Res.* 173:115541. doi: 10.1016/j.watres.2020.115541

Yun, H., Liang, B., Ding, Y., Li, S., Wang, Z., Khan, A., et al. (2021). Fate of antibiotic resistance genes during temperature-changed psychrophilic anaerobic digestion of municipal sludge. *Water Res.* 194:116926. doi: 10.1016/j.watres.2021.116926

Zhang, J., Chen, M., Sui, Q., Tong, J., Jiang, C., Lu, X., et al. (2016). Impacts of addition of natural zeolite or a nitrification inhibitor on antibiotic resistance genes during sludge composting. *Water Res.* 91, 339–349. doi: 10.1016/j.watres.2016.01.010

Zhang, S., Lu, J., Wang, Y., Verstraete, W., Yuan, Z. G., and Guo, J. H. (2022). Insights of metallic nanoparticles and ions in accelerating the bacterial uptake of antibiotic resistance genes. *J. Hazard. Mater.* 421:11. doi: 10.1016/j.jhazmat.2021.126728

Zhao, Y., Cocerva, T., Cox, S., Tardif, S., Su, J. Q., Zhu, Y. G., et al. (2019). Evidence for co-selection of antibiotic resistance genes and mobile genetic elements in metal polluted urban soils. *Sci. Total Environ.* 656, 512–520. doi: 10.1016/j.scitotenv.2018.11.372

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 07 September 2023
DOI 10.3389/fmicb.2023.1242369

OPEN ACCESS

EDITED BY
Biao Tang,
Zhejiang Academy of Agricultural Sciences, China

REVIEWED BY
Chang-Wei Lei,
Sichuan University, China
Lili Zhang,
Jiangsu Academy of Agricultural Sciences (JAAS), China
Bao-Tao Liu,
Qingdao Agricultural University, China

*CORRESPONDENCE
Ruichao Li
✉ rchl88@yzu.edu.cn
Xiaorong Yang
✉ yangyangxr@163.com

†These authors have contributed equally to this work

RECEIVED 19 June 2023
ACCEPTED 28 August 2023
PUBLISHED 07 September 2023

CITATION
Peng K, Deng J, Zou N, Sun X, Huang W, Li R and Yang X (2023) Emergence of the fourth mobile sulfonamide resistance gene *sul4* in clinical *Salmonella enterica*. *Front. Microbiol.* 14:1242369. doi: 10.3389/fmicb.2023.1242369

Emergence of the fourth mobile sulfonamide resistance gene *sul4* in clinical *Salmonella enterica*

Kai Peng[1,2†], Jianping Deng[3†], Nianli Zou[3†], Xinran Sun[1,2], Weifeng Huang[4], Ruichao Li[1,2*] and Xiaorong Yang[4*]

[1]Jiangsu Co-Innovation Center for Prevention and Control of Important Animal Infectious Diseases and Zoonoses, College of Veterinary Medicine, Yangzhou University, Yangzhou, Jiangsu, China, [2]Institute of Comparative Medicine, Yangzhou University, Yangzhou, Jiangsu, China, [3]Zigong Center for Disease Control and Prevention, Zigong, Sichuan, China, [4]Center for Disease Control and Prevention of Sichuan Province, Chengdu, Sichuan, China

The fourth mobile sulfonamide resistance gene *sul4* has been discovered in many metagenomic datasets. However, there is no reports of it in cultured bacteria. In this study, a *sul4* positive clinical *Salmonella enterica* SC2020597 was obtained by conventional *Salmonella* isolation methods and characterized by species identification and antimicrobial susceptibility testing. Meanwhile, the genomic DNA was sequenced using both long-read and short-read methods. Following that, the complete genome was analyzed by bioinformatic methods. The *sul4* gene in *S. enterica* SC2020597 differed from the *sul4* identified in metagenomic data by one amino acid and could confer full resistance to sulfamethoxazole. Genetic location analysis showed that the *sul4* in SC2020597 was carried by a complex chromosomally integrated hybrid plasmid. IS*CR20*-like was strongly associated with the mobilization of *sul4* by core genetic context analysis. To the best of our knowledge, this is the first report of the emergence of *sul4* in clinically cultured *S. enterica*. More important, the *sul4* has the potential to spread to other bacteria with the help of mobile elements.

KEYWORDS

sulfonamide resistance, *sul4*, chromosomally integrated plasmid, *Salmonella enterica*, clinical

Introduction

Sulfonamides are bacteriostatic antimicrobials that inhibit bacterial cellular activity without directly killing the bacteria. They work by interfering with the synthesis of folic acid in bacteria, which is required for the formation of nucleic acids (Skold, 2000; Ovung and Bhattacharyya, 2021). In 1930's, sulfonamides were firstly introduced for the treatment of human bacterial infections (Skold, 2000; Fernandez-Villa et al., 2019; Nunes et al., 2020). In the years that followed, over 150 sulfonamides and its derivatives were applied in human and veterinary medicine as antibacterial drugs (Baran et al., 2011). Given their extensive utilization, sulfonamide-subsisting bacteria were identified in 2008 for the first time (Dantas et al., 2008), and then many species of bacteria showing sulfonamides resistant were discovered (Deng et al., 2018; Ma et al., 2022). Currently, the most common mechanism of sulfonamide resistance in the majority of bacteria was plasmid-borne, highly mobilized *sul1*, *sul2*, and *sul3*, which encode dihydropteroate synthase (Wang et al., 2014; Nunes et al., 2020; Tang et al., 2022a; Venkatesan et al., 2023). In 2017, the fourth mobile sulfonamide resistance gene *sul4* was identified in river sediment with amplicon metagenomic sequencing for the first time (Razavi et al., 2017). To date,

many metagenomic analysis have revealed that *sul4* is already present in a wide range of environmental samples from around the world (Razavi et al., 2017; Marathe et al., 2019; Hutinel et al., 2022). However, it has not yet been identified in cultured bacteria.

Salmonellosis caused by *Salmonella enterica* is a common foodborne diseases frequently occurred around the world (Newell et al., 2010; Jajere, 2019; Tang et al., 2022b). It is estimated that salmonellosis could cause 93.8 million foodborne illnesses and 155 thousand people deaths per year (Majowicz et al., 2010; Eng et al., 2015). The World Health Organization (WHO) has listed it as one of the global health concerns. Furthermore, an antibiotic resistance surveillance of foodborne pathogenic bacteria revealed that the prevalence of antibiotic resistance genes (ARGs) in *Salmonella* was serious, second only to *Escherichia coli* (Jajere, 2019). Many clinically critical ARGs, such as *mcr-1* (Li et al., 2022), bla_{NDM-5} (Wang et al., 2020) and *tet*(X4) (Wang et al., 2021), have been identified in *S. enterica* in recently years. Therefore, the public health risk posed by antibiotics resistant *S. enterica* has increasingly arisen. In this study, we identified a fourth mobile sulfonamide resistance gene *sul4* in clinical *S. enterica* for the first time, implying that *Salmonella* could be an important carrier of emerging ARGs.

Materials and methods

Bacterial isolate and antimicrobial susceptibility testing

According to previous method (Xia et al., 2009), the *S. enterica* SC2020597 was isolated from a hospital in Guangyuan, Sichuan province, China in 2020. Then, pure cultured SC2020597 was identified as *S. enterica* using matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) (Bruker, Bremen, Germany). Meanwhile, the species of isolated *Salmonella* was further confirmed by an online rMLST analysis (Jolley et al., 2018). Subsequently, the minimum inhibitory concentrations (MICs) of SC2020597 was tested by broth microdilution according to Clinical and Laboratory Standards Institute (CLSI) guidelines (https://clsi.org/). *E. coli* ATCC25922 was used for the quality control.

Conjugation assay and electroporation experiment

In order to investigate the transfer ability of *sul4*, both conjugation assays and electroporation experiments were performed. For the conjugation assay, we used SC2020597 as donor strains and *E. coli* C600 as recipients. The donor and recipient strains were cultured into the logarithmic growth phase with an OD600 value of 0.4 in LB broth, then mixed at a ratio of 1:1 and cultured overnight on LB agar plates. The transconjugants were screened on LB agar plates containing rifampin (300 mg/L) and trimethoprim/sulfamethoxazole (4/76 mg/L). For the electroporation experiment, the genomic DNA of SC2020597 was used as donor DNA, and electrocompetent cells of *S. enterica* ATCC13076 were used as recipients. Electroporation conditions were 200 Ω, 1.8 kV and 25 uF. The transconjugants were screened on LB agar plates containing trimethoprim/sulfamethoxazole (4/76 mg/L). Then, all transconjugants were confirmed by PCR methods targeted at *sul4* and 16S rDNA genes.

Genomic DNA extraction, sequencing, and cyclic plasmid detection

The genomic DNA of SC2020597 was extracted using FastPure Bacteria DNA Isolation Mini Kit (Vazyme™, China) following the protocol descripting in the manufacturer. The purity and quality of extracted genomic DNA were evaluated using NanoDrop (Thermo ScientificTM) and using a dsDNA High Sensitivity (HS) Assay kit on the Qubit 4 Fluorometer, respectively. Then, 200 μg genomic DNA was sent to GENEWIZ (Suzhou China) to subject short-read sequencing with PE150 strategy on Illumina Hiseq 2,500 platform. Meanwhile, long-read genomic sequencing of SC2020597 was conducted at Oxford Nanopore Technologies MinION platform in our laboratory. Briefly, the long-read sequencing library was prepared using the SQK-RBK109 1D Rapid Barcoding genomic DNA kit according to the user handbook. Then, the prepared library was sequenced with R9.4 flow cells on MinION and the sequencing process was managed with MinKNOW.

The cyclic plasmid form of the chromosomally integrated plasmid was detected using the inverse PCR method with primers cir_F: TTCAGACGGACTGGACATCG and cir_R: GCGGAATTGTTC AGGGGGTA. The genomic DNA of SC2020597 was used as template DNA. The master mix for application in long fragments was used for PCR amplification reaction system.

Data analysis

The short-read raw reads were filtered to remove low-quality base and adapters using fastp with default parameters (Chen et al., 2018). Then, the complete genome of SC2020597 was generated with a hybrid assembly strategy combining of clean short-read data and long-read data using Unicycler (Wick et al., 2017). Functional annotation of the complete genome was performed by a web-based RAST annotation engine (Aziz et al., 2008; Overbeek et al., 2014; Brettin et al., 2015). Antibiotic resistance genes (ARGs), plasmid replicon genes and insertion sequences (ISs) were identified using abricate tool (https://github.com/tseemann/abricate) based on AMRFinderPlus (Feldgarden et al., 2019), PlasmidFinder (Carattoli et al., 2014) and ISFinder (Siguier et al., 2006) databases, respectively. Multilocus sequence typing (MLST) of the complete bacterial genome was performed using mlst tool (https://github.com/tseemann/mlst). Using *Salmonella* genomes as input data, the web-based application SISTR was used to identify the *Salmonella* serovar (Yoshida et al., 2016). Plasmid comparisons and genetic context comparisons visualization were performed with BRIG (Alikhan et al., 2011) and Easyfig (Sullivan et al., 2011) tools.

Functional confirmation of *sul4*

To confirm the resistance function of the mutated *sul4*, TA-cloning was performed using a 5 min TA/Blunt-Zero Cloning Kit developed by Vazyme (Vazyme, China). Briefly, the *sul4* gene and its predicted promoter were amplified by PCR using primers sul4_F: TGCCTGCAGGTCGACTCTAGAACCCAAAAGTCTGTAGCCCA AA, sul4_R: ACGGCCAGTGAATTGAGCTCTGGTCTAGTIC AAAATCGATCATGT, and then cloned into pUC19 vector. Meanwhile, in order to verify the effect of the base mutation on the function of *sul4*, an unmutated *sul4* recombinant expression plasmid was constructed. Subsequently, the recombinant plasmids were

introduced chemically into *E. coli* DH5α. At last, we tested the resistance phenotype of the transconjugants using broth microdilution.

Data availability

The genome sequences of SC2020597 were deposited into the National Center for Biotechnology information (NCBI) under BioProject PRJNA946266.

Results and discussion

Characteristic of the *Salmonella enterica* isolate SC2020597

The isolate SC2020597 was recovered from a clinical patient. We identified it as *S. enterica* using MALDI-TOF-MS and confirmed by Ribosomal Multilocus Sequence Typing (rMLST) analysis. MLST analysis showed that isolate SC2020597 belonged to ST26 *S. enterica*. Serovar analysis classified the isolate as *S. enterica* subsp. *enterica* serovar Thompson, which is one of the most frequent *Salmonella* serovars involved in human infection (Eun et al., 2019). According to previous investigations, the prevalence of *S.* Thompson in clinical patients and food in China was 3.9 and 5.4%, respectively (Wang et al., 2017; Fan et al., 2020). The serovar has the potential to cause outbreaks of *Salmonella* infection. Antimicrobial susceptibility testing showed that the isolate was resistant to multiple antibiotics including kanamycin, ampicillin, tetracycline, sulfamethoxazole and trimethoprim/sulfomethoxazole, but sensitive to aztreonam, meropenem, ciprofloxacin, colistin and enrofloxacin (Supplementary Table S1).

Functional analysis of *sul4*

The draft genome of SC2020597 was obtained by short-read genome assembly. Many ARGs, plasmid replicon genes and ISs were identified in the draft genome. Of note, we found a *sul4* gene in isolate SC2020597, which had previously only been found in metagenomes (Razavi et al., 2017; Marathe et al., 2019; Hutinel et al., 2022). The *sul4* gene in SC2020597 showed 100% coverage and 99.88% nucleic acid identity to *sul4* (NG_056174). The one nucleic acid substitution of *sul4* causes one amino acid change (W120R). To verify the function of the novel *sul4*, the intact *sul4* gene and its promoter were cloned into pUC19 vector and introduced into *E. coli* DH5α. The *sul4* positive transconjugants were resistant to trimethoprim/sulfomethoxazole and had a MIC for trimethoprim/sulfomethoxazole that was more than 16-fold higher than *E. coli* DH5 with an empty vector (≤ 1/19 mg/L to >32/608 mg/L). Meanwhile, unmutated *sul4*-bearing *E. coli* DH5 also showed full resistance to trimethoprim/sulfomethoxazole (>32/608 mg/L). This demonstrated that W120R substitution had no effect to the function of *sul4*.

Genomic feature of SC2020597

To decipher the genomic structure feature of SC2020597, long-read sequencing was performed. Then, the complete genome of SC2020597 was generated by a hybrid assembly strategy using short-read and long-read data. The isolate harbored one chromosome with a length of 5, 035, 375 bp and two plasmids, pSC2020597_48k with a length of 48, 530 bp, and pSC2020597_5k with a length of 5, 754 bp. Plasmid replicon analysis showed that pSC2020597_48k was an IncFII type plasmid and pSC2020597_5k was a Col type small plasmid. Of note, no ARG was found in the two plasmids. Online blastn analysis showed that many plasmids from *Salmonella* were similar to pSC2020597_48k, indicating that such plasmids were common in *Salmonella*. Plasmid pSC2020597_5k belonged to a group of small plasmids with a broad host range that were widely distributed in Enterobacteriaceae.

A total of 22 ARGs, including *sul4*, were detected in SC2020597, and all of them were located on chromosome (Supplementary Table S2). In addition, three plasmid replicon genes were discovered on the chromosome of SC2020597. Further analysis found that a multi-drugs resistant plasmid was integrated into chromosome of SC2020597. The chromosomally integrated plasmid was 330, 232 bp in length and was designated as PSC2020597-sul4-330 k-c (Figure 1A). The phenomenon that plasmid was integrated into chromosome have been frequently reported (Muthuirulandi Sethuvel et al., 2021; Shigemura et al., 2021; Chang et al., 2022), especially in *Salmonella*, and is commonly mediated by homologous recombination of ISs. We found that PSC2020597-sul4-330 k-c was flanked by IS*1F* on chromosome of SC2020597. Meanwhile, a 7 bp direct repeat sequence was found around PSC2020597-sul4-330 k-c (Figure 1B). These findings provided strong evidences that PSC2020597-sul4-330 k-c was integrated into chromosome via the homologous recombination of IS*1R*. The precursor of PSC2020597-sul4-330 k-c was a hybrid plasmid with complex structure. Half of PSC2020597-sul4-330 k-c was composed by a typical IncHI2/HI2A plasmid. Another half was made up of flexible genetic arrays. Blastn analysis with NCBI nr database showed that PSC2020597-sul4-330 k-c was most closely related to pSIn_quan12 (GenBank: ON960352.1), which had 66% coverage and 99.98% identity to PSC2020597-sul4-330 k-c (Figure 1A). Meanwhile, many other plasmids similar to PSC2020597-sul4-330 k-c were also found in nr database, and the complete structure of PSC2020597-sul4-330 k-c could almost be covered by the genetic array of these plasmids (Figure 1A). Hence, PSC2020597-sul4-330 k-c was most likely formed through the recombination of diversity of plasmids and then integrated into chromosome. The discovery of *sul4* in a complex chromosomally integrated plasmid implied that *sul4* gene had already appeared and spread in plasmids. We should take notice to monitor the spread of *sul4* by plasmids or other mobile genetic elements.

The transfer ability of *sul4*

By the genetic analysis, we found that *sul4* was located on a chromosomally integrated plasmid. However, a previous study demonstrated that chromosomally integrated plasmids could also be transferred to other recipients by conjugation assay (Chang et al., 2022). Here, we verified that there is a circle form of chromosomally integrated plasmid PSC2020597-sul4-330 k-c in some clones of SC2020597 using the inverse PCR method and confirmed by Sanger sequencing (Supplementary Figure S1). This phenomenon implied that the chromosomally integrated plasmid PSC2020597-sul4-330 k-c had potential to horizontal transfer by conjugation. Subsequently, we used *E. coli* C600 as recipients to test the transfer of *sul4* by conjugation assay. The *sul4* gene could not be transferred into *E. coli* C600 after multiple tries. Then, we attempted to verify the transfer

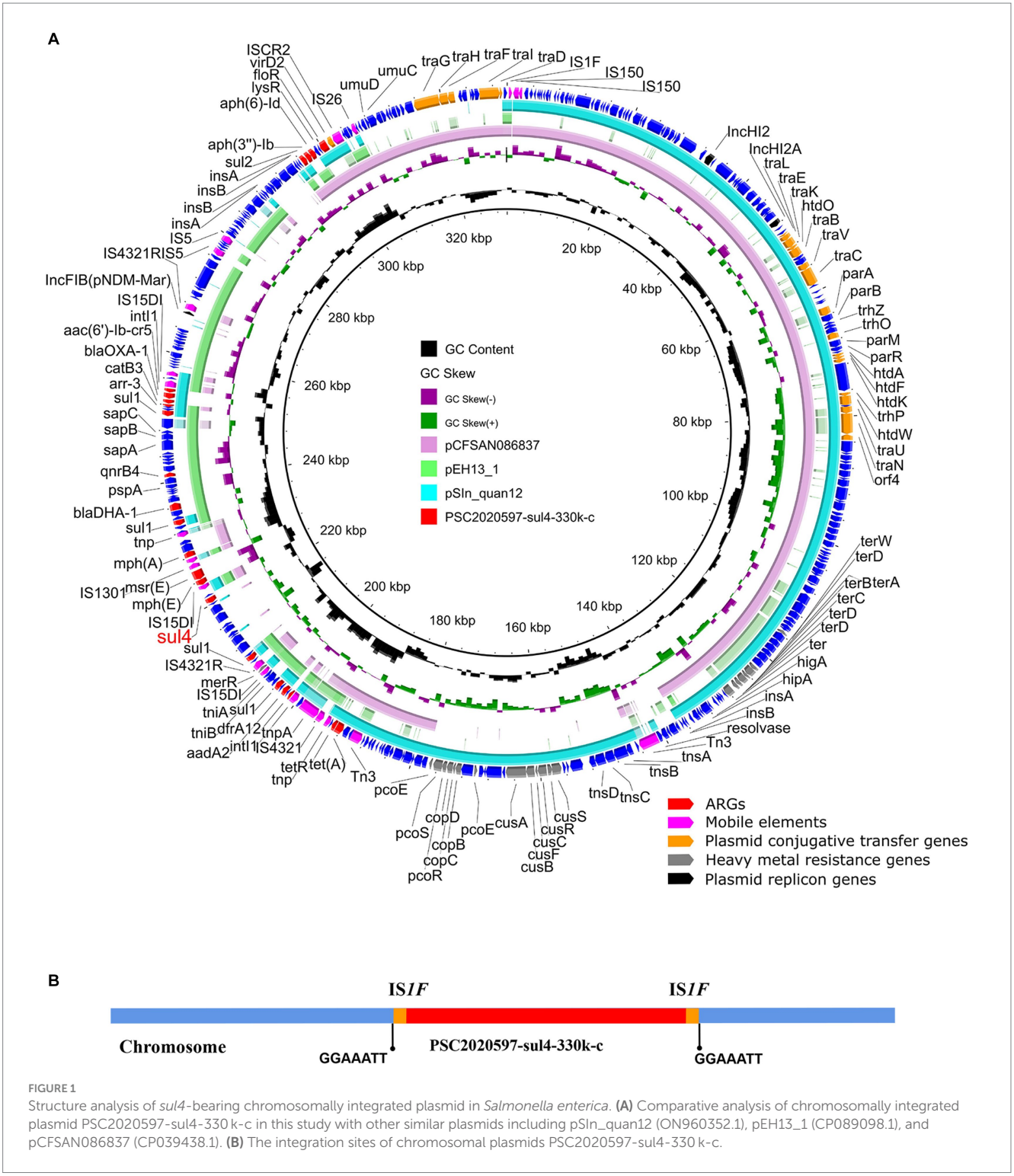

FIGURE 1
Structure analysis of *sul4*-bearing chromosomally integrated plasmid in *Salmonella enterica*. **(A)** Comparative analysis of chromosomally integrated plasmid PSC2020597-sul4-330 k-c in this study with other similar plasmids including pSln_quan12 (ON960352.1), pEH13_1 (CP089098.1), and pCFSAN086837 (CP039438.1). **(B)** The integration sites of chromosomal plasmids PSC2020597-sul4-330 k-c.

ability of *sul4* using electroporation experiments with *S. enterica* ATCC13076 as recipients. The *sul4*-positive *S. enterica* ATCC13076 was successfully screened and showed resistance to trimethoprim/sulfomethoxazole.

The core genetic structure of *sul4*

Mobile elements played a key role in the dissemination of ARGs (Partridge et al., 2018). Previous research discovered an IS*CR* family transposase IS*CR20*-like (GneBank: MG649402.1) downstream of *sul4* that could be mobilized along with their adjacent genes via rolling-circle transposition without the assistance of any other transposase protein (Razavi et al., 2017). Therefore, *sul4*-IS*CR20*-like was considered to be mobilizable as an entire integrin (Razavi et al., 2017). Subsequently, we investigated the core genetic structure of *sul4* in PSC2020597-sul4-330 k-c. We also found an IS*CR20*-like in the downstream of *sul4*, which was consistent with previous structure of *sul4* in metagenome (GneBank: MG649402.1) (Figure 2). In addition, other *sul4*-bearing genetic contexts associated with IS*CR20*-like or truncated

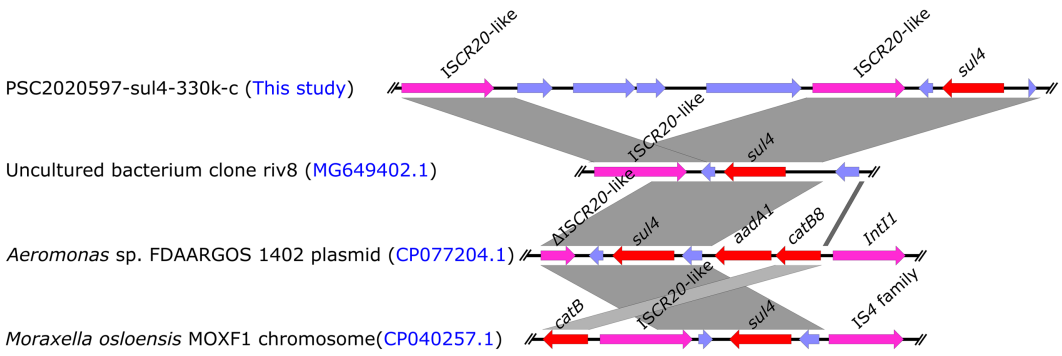

FIGURE 2
Linear comparison of the core genetic structures of *sul4* gene. Comparative analysis of the genetic context of *sul4* in PSC2020597-sul4-330 k-c with the genetic contexts of *sul4* in uncultured riv8 (MG649402.1), isolates FDAARGOS 1402 (CP077204.1), and MOXF1 (CP040257.1).

IS*CR20*-like were also found in other bacterial chromosome or plasmid in NCBI nr database (Figure 2). These findings demonstrated that the dissemination of *sul4* was probably driven by IS*CR20*-like. Although the transfer of *sul4*-IS*CR20*-like has not been verified by experiment, other similar transfer events have been confirmed, such as *tet*(X4)-IS*CR2* (He et al., 2019). Hence, the mobilization of *sul4*-IS*CR20*-like was likely to happen and promoted the dissemination of *sul4*. Of note, *sul4* was found to be integrated into a class 1 integron in the plasmid of *Aeromonas* sp. FDAARGOS 1402 (Figure 2). It means that novel mobile genetic structure harboring *sul4* is emerging, which may accelerate the propagation of *sul4*. Therefore, we should take more attention to the surveillance of *sul4* in order to slow its spread.

Conclusion

To our best knowledge, this is the first report of the emergence of *sul4* in clinically cultured *S. enterica*. The genetic location analysis showed that *sul4* had already emerged in common and continually evolving hybrid plasmids of Enterobacteriaceae. Meanwhile, we found that IS*CR20*-like played an important role in the spread of *sul4*. These findings demonstrated that *sul4* has the potential to be prevalent in various bacteria, and further reduce the effect of sulfonamide.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/, PRJNA946266.

Author contributions

KP, JD, and NZ: conception and design. KP, JD, NZ, and XS: methodology. WH: collection and assembly of data. KP and WH: data analysis and interpretation. KP: writing—original draft. RL and XY: writing—reviewing and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Sichuan Science and Technology Program (2022ZDZX0017), Postgraduate Research & Practice Innovation Program of Jiangsu Province (SJCX21_1631) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1242369/full#supplementary-material

References

Alikhan, N. F., Petty, N. K., Ben Zakour, N. L., and Beatson, S. A. (2011). BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. doi: 10.1186/1471-2164-12-402

Aziz, R. K., Bartels, D., Best, A. A., Dejongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST server: rapid annotations using subsystems technology. *BMC Genom.* 9:75. doi: 10.1186/1471-2164-9-75

Baran, W., Adamek, E., Ziemianska, J., and Sobczak, A. (2011). Effects of the presence of sulfonamides in the environment and their influence on human health. *J. Hazard. Mater.* 196, 1–15. doi: 10.1016/j.jhazmat.2011.08.082

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., et al. (2015). RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* 5:8365. doi: 10.1038/srep08365

Carattoli, A., Zankari, E., Garcia-Fernandez, A., Larsen, M. V., Lund, O., Villa, L., et al. (2014). In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* 58, 3895–3903. doi: 10.1128/AAC.02412-14

Chang, M. X., Zhang, J., Zhang, J. F., Ding, X. M., Lu, Y., Zhang, J., et al. (2022). Formation, transmission, and dynamic evolution of a multidrug-resistant chromosomally integrated plasmid in Salmonella Spp. *Front. Microbiol.* 13:846954. doi: 10.3389/fmicb.2022.846954

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Dantas, G., Sommer, M. O., Oluwasegun, R. D., and Church, G. M. (2008). Bacteria subsisting on antibiotics. *Science* 320, 100–103. doi: 10.1126/science.1155157

Deng, Y., Li, B., and Zhang, T. (2018). Bacteria that make a meal of sulfonamide antibiotics: blind spots and emerging opportunities. *Environ. Sci. Technol.* 52, 3854–3868. doi: 10.1021/acs.est.7b06026

Eng, S. K., Pusparajah, P., Ab Mutalib, N. S., Ser, H. L., Chan, K. G., and Lee, L. H. (2015). Salmonella: A review on pathogenesis, epidemiology and antibiotic resistance. *Front. Life Sci.* 8, 284–293. doi: 10.1080/21553769.2015.1051243

Eun, Y., Jeong, H., Kim, S., Park, W., Ahn, B., Kim, D., et al. (2019). A large outbreak of Salmonella enterica serovar Thompson infections associated with chocolate cake in Busan Korea. *Epidemiol. Health* 41:e2019002. doi: 10.4178/epih.e2019002

Fan, J., Zhang, L., He, J., Zhao, M., Loh, B., Leptihn, S., et al. (2020). Plasmid dynamics of mcr-1-positive Salmonella spp. in a general Hospital in China. *Front. Microbiol.* 11:604710. doi: 10.3389/fmicb.2020.604710

Feldgarden, M., Brover, V., Haft, D. H., Prasad, A. B., Slotta, D. J., Tolstoy, I., et al. (2019). Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob. Agents Chemother.* 63:e00483-19. doi: 10.1128/AAC.00483-19

Fernandez-Villa, D., Aguilar, M. R., and Rojo, L. (2019). Folic acid antagonists: antimicrobial and Immunomodulating mechanisms and applications. *Int. J. Mol. Sci.* 20:4996. doi: 10.3390/ijms20204996

He, T., Wang, R., Liu, D., Walsh, T. R., Zhang, R., Lv, Y., et al. (2019). Emergence of plasmid-mediated high-level tigecycline resistance genes in animals and humans. *Nat. Microbiol.* 4, 1450–1456. doi: 10.1038/s41564-019-0445-2

Hutinel, M., Larsson, D. G. J., and Flach, C. F. (2022). Antibiotic resistance genes of emerging concern in municipal and hospital wastewater from a major Swedish city. *Sci. Total Environ.* 812:151433. doi: 10.1016/j.scitotenv.2021.151433

Jajere, S. M. (2019). A review of Salmonella enterica with particular focus on the pathogenicity and virulence factors, host specificity and antimicrobial resistance including multidrug resistance. *Vet World* 12, 504–521. doi: 10.14202/vetworld.2019.504-521

Jolley, K. A., Bray, J. E., and Maiden, M. C. J. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* 3:124. doi: 10.12688/wellcomeopenres.14826.1

Li, R., Peng, K., Huang, W., Sun, X., Huang, Y., Lei, G., et al. (2022). The genomic epidemiology of mcr-positive Salmonella enterica in clinical patients from 2014 to 2017 in Sichuan, China and global epidemiological features. *J. Infect.* 85, 702–769. doi: 10.1016/j.jinf.2022.08.042

Ma, J., Zhou, W., Wu, J., Liu, X., Lin, J., Ji, X., et al. (2022). Large-scale studies on antimicrobial resistance and molecular characterization of Escherichia coli from food animals in developed areas of eastern China. *Microbiol. Spectr.* 10:e0201522. doi: 10.1128/spectrum.02015-22

Majowicz, S. E., Musto, J., Scallan, E., Angulo, F. J., Kirk, M., O'brien, S. J., et al. (2010). The global burden of Nontyphoidal Salmonella gastroenteritis. *Clin. Infect. Dis.* 50, 882–889. doi: 10.1086/650733

Marathe, N. P., Berglund, F., Razavi, M., Pal, C., Droge, J., Samant, S., et al. (2019). Sewage effluent from an Indian hospital harbors novel carbapenemases and integron-borne antibiotic resistance genes. *Microbiome* 7:97. doi: 10.1186/s40168-019-0710-x

Muthuirulandi Sethuvel, D. P., Anandan, S., Murugan, D., Asokan, K., Vasudevan, K., Jacob, J. J., et al. (2021). Hybrid genome assembly of Shigella sonnei reveals the novel finding of chromosomal integration of an IncFII plasmid carrying a mphA gene. *Access Microbiol.* 3:000189

Newell, D. G., Koopmans, M., Verhoef, L., Duizer, E., Aidara-Kane, A., Sprong, H., et al. (2010). Food-borne diseases - the challenges of 20 years ago still persist while new ones continue to emerge. *Int. J. Food Microbiol.* 139, S3–S15. doi: 10.1016/j.ijfoodmicro.2010.01.021

Nunes, O. C., Manaia, C. M., Kolvenbach, B. A., and Corvini, P. F. (2020). Living with sulfonamides: a diverse range of mechanisms observed in bacteria. *Appl. Microbiol. Biotechnol.* 104, 10389–10408. doi: 10.1007/s00253-020-10982-5

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* 42, D206–D214. doi: 10.1093/nar/gkt1226

Ovung, A., and Bhattacharyya, J. (2021). Sulfonamide drugs: structure, antibacterial property, toxicity, and biophysical interactions. *Biophys. Rev.* 13, 259–272. doi: 10.1007/s12551-021-00795-9

Partridge, S. R., Kwong, S. M., Firth, N., and Jensen, S. O. (2018). Mobile genetic elements associated with antimicrobial resistance. *Clin. Microbiol. Rev.* 31:e00088-17. doi: 10.1128/CMR.00088-17

Razavi, M., Marathe, N. P., Gillings, M. R., Flach, C. F., Kristiansson, E., and Joakim Larsson, D. G. (2017). Discovery of the fourth mobile sulfonamide resistance gene. *Microbiome* 5:160. doi: 10.1186/s40168-017-0379-y

Shigemura, H., Maeda, T., Nakayama, S., Ohishi, A., Carle, Y., Ookuma, E., et al. (2021). Transmission of extended-spectrum cephalosporin-resistant Salmonella harboring a blaCMY-2-carrying IncA/C2 plasmid chromosomally integrated by ISEcp1 or IS26 in layer breeding chains in Japan. *J. Vet. Med. Sci.* 83, 1345–1355. doi: 10.1292/jvms.21-0085

Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference Centre for bacterial insertion sequences. *Nucleic Acids Res.* 34, D32–D36. doi: 10.1093/nar/gkj014

Skold, O. (2000). Sulfonamide resistance: mechanisms and trends. *Drug Resist. Updat.* 3, 155–160. doi: 10.1054/drup.2000.0146

Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics* 27, 1009–1010. doi: 10.1093/bioinformatics/btr039

Tang, B., Chang, J., Chen, Y., Lin, J., Xiao, X., Xia, X., et al. (2022a). *Escherichia fergusonii*, an underrated repository for antimicrobial resistance in food animals. *Microbiol. Spectr.* 10:e0161721. doi: 10.1128/spectrum.01617-21

Tang, B., Elbediwi, M., Nambiar, R. B., Yang, H., Lin, J., and Yue, M. (2022b). Genomic characterization of antimicrobial-resistant Salmonella enterica in duck, chicken, and pig farms and retail Markets in Eastern China. *Microbiol Spectr* 10:e0125722. doi: 10.1128/spectrum.01257-22

Venkatesan, M., Fruci, M., Verellen, L. A., Skarina, T., Mesa, N., Flick, R., et al. (2023). Molecular mechanism of plasmid-borne resistance to sulfonamide antibiotics. *Nat. Commun.* 14:4031. doi: 10.1038/s41467-023-39778-7

Wang, Y., Cao, C., Alali, W. Q., Cui, S., Li, F., Zhu, J., et al. (2017). Distribution and antimicrobial susceptibility of foodborne Salmonella Serovars in eight provinces in China from 2007 to 2012 (except 2009). *Foodborne Pathog. Dis.* 14, 393–399. doi: 10.1089/fpd.2016.2237

Wang, Z., He, J., Li, Q., Tang, Y., Wang, J., Pan, Z., et al. (2020). First detection of NDM-5-positive Salmonella enterica Serovar typhimurium isolated from retail pork in China. *Microb. Drug Resist.* 26, 434–437. doi: 10.1089/mdr.2019.0323

Wang, Y. N., Liu, F., Xu, X. B., Huang, H., Lyu, N., Ma, S. F., et al. (2021). Detection of plasmid-mediated Tigecycline resistance gene tet(X4) in a Salmonella enterica Serovar Llandoff isolate. *Infect. Microb. Dis.* 3, 198–204. doi: 10.1097/IM9.0000000000000077

Wang, N., Yang, X., Jiao, S., Zhang, J., Ye, B., and Gao, S. (2014). Sulfonamide-resistant bacteria and their resistance genes in soils fertilized with manures from Jiangsu Province Southeastern China. *PLoS One* 9:e112626. doi: 10.1371/journal.pone.0112626

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 13:e1005595. doi: 10.1371/journal.pcbi.1005595

Xia, S., Hendriksen, R. S., Xie, Z., Huang, L., Zhang, J., Guo, W., et al. (2009). Molecular characterization and antimicrobial susceptibility of Salmonella isolates from infections in humans in Henan Province, China. *J. Clin. Microbiol.* 47, 401–409. doi: 10.1128/JCM.01099-08

Yoshida, C. E., Kruczkiewicz, P., Laing, C. R., Lingohr, E. J., Gannon, V. P., Nash, J. H., et al. (2016). The Salmonella in silico typing resource (SISTR): an open web-accessible tool for rapidly typing and subtyping draft Salmonella genome assemblies. *PLoS One* 11:e0147101. doi: 10.1371/journal.pone.0147101

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 15 November 2023
DOI 10.3389/fmicb.2023.1284929

OPEN ACCESS

EDITED BY
Xin Wang,
National Institute for Communicable Disease Control and Prevention (China CDC), China

REVIEWED BY
Meiying Yan,
National Institute for Communicable Disease Control and Prevention (China CDC), China
Chang-Wei Lei,
Sichuan University, China

*CORRESPONDENCE
Jennie Fischer
✉ jennie.fischer@bfr.bund.de

RECEIVED 29 August 2023
ACCEPTED 25 October 2023
PUBLISHED 15 November 2023

CITATION
Bartsch LJ, Borowiak M, Deneke C, Gruetzke J, Hammerl J-A, Malorny B, Szabo I, Alter T, Nguyen KK and Fischer J (2023) Genetic characterization of a multidrug-resistant *Salmonella enterica* serovar Agona isolated from a dietary supplement in Germany. *Front. Microbiol.* 14:1284929. doi: 10.3389/fmicb.2023.1284929

Genetic characterization of a multidrug-resistant *Salmonella enterica* serovar Agona isolated from a dietary supplement in Germany

Lee Julia Bartsch[1], Maria Borowiak[1], Carlus Deneke[1], Josephine Gruetzke[1], Jens-Andre Hammerl[1], Burkhard Malorny[1], Istvan Szabo[1], Thomas Alter[2], Kim Katherine Nguyen[3] and Jennie Fischer[1]*

[1]Department Biological Safety, German Federal Institute for Risk Assessment, Berlin, Germany, [2]Institute of Food Safety and Food Hygiene, School of Veterinary Medicine, Freie Universität Berlin, Berlin, Germany, [3]Bavarian Health and Food Safety Authority, Erlangen, Germany

Salmonella enterica subsp. *enterica* serovar Agona has a history of causing food-borne outbreaks and any emergence of multidrug-resistant (MDR) isolates in novel food products is of concern. Particularly, in food products frequently consumed without sufficient heating prior to consumption. Here, we report about the MDR isolate, 18-SA00377, which had been isolated from a dietary supplement in Germany in 2018 and submitted to the German National Reference Laboratory for *Salmonella*. WGS-based comparative genetic analyses were conducted to find a potential reservoir of the isolate itself or mobile genetic elements associated with MDR. As a phylogenetic analysis did not yield any closely related *S.* Agona isolates, either globally or from Germany, a detailed analysis of the largest plasmid (295,499 bp) was performed as it is the main carrier of resistances. A combined approach of long-read and short-read sequencing enabled the assembly of the isolate's chromosome and its four plasmids. Their characterization revealed the presence of 23 different antibiotic resistance genes (ARGs), conferring resistance to 12 different antibiotic drug classes, as well as genes conferring resistance to six different heavy metals. The largest plasmid, pSE18-SA00377-1, belongs to the IncHI2 plasmid family and carries 16 ARGs, that are organized as two distinct clusters, with each ARG associated with putative composite transposons. Through a two-pronged approach, highly similar plasmids to pSE18-SA00377-1 were identified in the NCBI database and a search for *Salmonella* isolates with a highly similar ARG resistance profile was conducted. Mapping and structural comparisons between pSE18-SA00377-1 and these plasmids and *Salmonella* isolates showed that both the plasmid backbone and identical or similar ARG clusters can be found not only in *Salmonella* isolates, originating mostly from a wide variety of livestock, but also in a diverse range of bacterial genera of varying geographical origins and isolation sources. Thus, it can be speculated that the host range of pSE18-SA00377-1 is not restricted to *Salmonella* and its spread already occurred in different bacterial populations. Overall, this hints at a complex history for pSE18-SA00377-1 and highlights the importance of surveilling multidrug-resistant *S. enterica* isolates, especially in novel food items that are not yet heavily regulated.

KEYWORDS

Salmonella Agona, dietary supplement, whole genome sequencing, antimicrobial resistance, plasmid

1. Introduction

The bacterial genus *Salmonella* spans two species and more than 2,500 serovars (Grimont and Weill, 2007). Non-typhoidal serovars of one of its constituent subspecies, *Salmonella enterica* subsp. *enterica*, are the principal cause of food-borne illness worldwide manifesting as enteric infections (salmonellosis) with symptoms such as diarrhea, fever, abdominal pain and vomiting. Salmonellosis is a worldwide public health issue. In 2021, it was the second most reported zoonosis with over 60,000 reported cases in humans in the European Union (EU), including 71 deaths (European Food Safety Authority and European Centre for Disease Prevention and Control, 2022b). Additionally, the World Health Organization (WHO) estimates that salmonellosis accounts for more than 150,000 deaths worldwide annually (Ao et al., 2015). Despite the large number of serovars, fewer than 100 serovars are the main causative agents of human salmonellosis.

Starting in the 1970s, the serovar *Salmonella enterica* subsp. *enterica* serovar (*S.*) Agona has emerged on the radar of public health agencies, when a major *S.* Agona outbreak in five different countries was traced to animal feed based on contaminated fish meal from Peru (Clark et al., 1973). *S.* Agona now ranks among the top 20 most frequent serovars both in the European Economic Area (EEA) and European Union, having caused 784 reported cases of human salmonellosis from 2019 to 2021 (European Food Safety Authority and European Centre for Disease Prevention and Control, 2022a). In the past decades, numerous *S.* Agona outbreaks have occurred, ranging from smaller, localized outbreaks to larger, transnational and even transcontinental outbreaks. The first recognized transcontinental *S.* Agona outbreak occurred in 1994/95 in England, Wales, the United States (Killalea et al., 1996), and Israel. Here, the origin was traced to contaminated kosher savory snacks, frequently consumed by children (Shohat et al., 1996). A further outbreak in the United States occurred in a similar isolation source, dry rice and wheat cereal products, in 1998 and re-occurred 10 years later in 2008 (Russo et al., 2013). A similar re-emergence of *S.* Agona contamination in the same food matrix and linked to the same production facility, but 12 years apart, has also been stipulated with the 2005 (Brouard et al., 2007) and 2017 (Jourdan-da Silva et al., 2018) outbreaks in infant milk formula. Other past *S.* Agona outbreaks could be traced to air-dried beef products (Taylor et al., 1998), pre-cooked meat products (Nicolay et al., 2011), fresh papaya (Mba-Jonas et al., 2018) and aniseed tea (Koch et al., 2005). Notably, these outbreaks occurred in a diverse range of mostly dried food products that are frequently consumed by children and infants without sufficient prior cooking.

Since salmonellosis symptoms are ordinarily mild, treatments usually do not include antibiotics, except for serious infections or in infections of vulnerable populations such as infants, the elderly and immunocompromised. However, excessive and improper use of antibiotics in the treatment of food-producing animals has led to the emergence of antimicrobial resistance (AMR) as well as multidrug-resistance (MDR) in *Salmonella* species. Particularly alarming is the occurrence of isolates with increased resistance to ciprofloxacin or combined resistance to fluoroquinolones and third generation cephalosporins (European Food Safety Authority and European Centre for Disease Prevention and Control, 2023). This trend poses a major threat to public health through contamination in the food chain. *Salmonella* has been shown to harbor antimicrobial resistances genes (ARGs) including those, causing resistance against last-resort antibiotics such as colistin (Borowiak et al., 2017; Fernández et al., 2018) encoded on various plasmid families (Carattoli, 2003; Rozwandowicz et al., 2018). Therefore, monitoring and surveillance of antimicrobial resistance in *Salmonella* is crucial to prevent further spread and infection. Particular attention has to be given to novel foods such as insect-derived foods or dietary supplements as they are often produced using new methods, with novel ingredients from unfamiliar sources. Especially, the recent increase in the consumption of dietary supplements in the EU (Lordan, 2021) is worrying as similar matrices have caused *Salmonella* outbreaks before – including *S.* Agona outbreaks. None of these matrices – dietary supplements, cereals, etc. – are heated or are sufficiently heated immediately prior to consumption. Thus, potential contaminations with *Salmonella* spp. pose a serious health risk to the consumer, especially since infants and vulnerable or ill populations frequently consume these products. This is evident in the 2021 outbreak of salmonellosis by *S.* Typhimurium in Denmark, which was linked to contaminated herbal supplement capsules containing psyllium husk (Technical University of Denmark, 2022).

Here, we have investigated a multidrug-resistant *S.* Agona, isolated from dietary supplements in 2018. We also sequenced and characterized its very large constituent plasmid harboring many antimicrobial and heavy metal resistance genes on mobile genetic elements and tried to reconstruct its origin. Structural comparisons revealed a composite plasmid structure, found in isolates from numerous other genera, geographic origins and isolation matrices highlighting the enormous potential for shuffling resistance determinants within an epidemiologically widely disseminated plasmid backbone.

2. Materials and methods

2.1. Strain collection and isolation

On average, the National Reference Laboratory (NRL) for *Salmonella* in Germany receives around 3,000–4,000 *Salmonella* isolates annually from different sources. Since 2018, whole-genome short-read sequencing is completed on all *S.* Enteritidis isolates, as well as isolates from national zoonosis monitoring and control programs. Since 2019, these isolates are further supplemented with sequencing of project- and outbreak-specific isolates, rough *Salmonella* isolates, and since 2020 isolates arising from Phase I of GenoSalmSurv (Uelze et al., 2021). Since 2021, foodborne isolates,

isolates from Phase II of GenoSalmSurv and isolates arising from antimicrobial resistance monitoring according to Commission Implementing Decision (EU) 2020/1729 are also sequenced.

Isolate 18-SA00377 was submitted to the NRL for *Salmonella* in 2018. The sample was isolated from a dietary supplement that had been submitted to the Bavarian Health and Food Safety Authority in 2018, after having been recalled by the distributor in late 2017. According to submission metadata supplied by the Bavarian authority, the dietary supplement consisted of gelatin capsules filled with plant material enriched with vitamins and magnesium, intended for twice daily consumption.

2.2. Serotyping of isolates

The *S.* Agona isolates of this study underwent classical serotyping according to the White-Kauffmann-Le Minor scheme (Grimont and Weill, 2007) by slide agglutination with the O- and H-antigen specific sera (Sifin Diagnostics, Berlin, Germany) as part of the routine diagnostics of the NRL for *Salmonella*.

2.3. Antibiotic susceptibility testing

Antibiotic susceptibility testing was performed by the NRL for Antibiotic Resistance at the BfR, using broth microdilutions or disk diffusion. Following CLSI guidelines (version A7-M11) for broth microdilution procedures, minimum inhibitory concentrations (MIC) were obtained for 37 different antibiotics. These were interpreted following the epidemiological cutoff values provided by EUCAST (v.10.0) (The European Committee on Antimicrobial Susceptibility Testing, 2021).

2.4. S1-pulsed-field gel electrophoresis (S1-PFGE)

Preparation of agarose plug molds and subsequent digestion by S1 nuclease (Thermo Fisher Scientific) were performed using a previously described protocol (Rodríguez et al., 2009). A plasmid pattern by PFGE was generated using the CHEF-DR III system (Bio-Rad Laboratories, Madrid, Spain) under standardized run conditions described by PulseNet standardized protocol (available at[1]). The *Salmonella* Braenderup strain H9812 (digested with the restriction endonuclease "XbaI" (Thermo Fisher Scientific, Darmstadt, Germany)) was used for size comparisons.

2.5. Whole genome sequencing (WGS) of *Salmonella* Agona isolate 18-SA00377

For sequencing of 18-SA00377, genomic DNA was extracted using the PureLink Genomic DNA Mini Kit (Thermo Fisher Scientific, Waltham, MA, USA) and sequenced using MiSeq (Illumina, San Diego, CA, USA) and Oxford Nanopore Technologies (ONT, Oxford, UK) MinION devices.

A short-read sequencing library was prepared using the sparQ DNA Frag & Library Prep Kit (Quantabio Beverly, MA, USA). Paired-end sequencing was performed in 2 × 151 bp cycles on an Illumina MiSeq instrument using the MiSeq Reagent Kit v3 (600 cycle). Trimming of short-reads using fastp v0.19.5 (Chen et al., 2018) resulted in 1.9 million high quality reads (≥ 89% Q30).

An Oxford Nanopore (ONT) sequencing library was prepared according to manufacturer's protocol using the Rapid Barcoding Kit (SQK-RBK004) and sequenced on an ONT MinION sequencer MK1C using a FLO-MIN106 R9.4.1 flow cell. Basecalling was performed with Albacore v2.3.1 (available at[2]). Obtained reads were trimmed using Porechop v0.2.3 (available at[3]), filtered, and quality checked using NanoStat v1.4.0 and NanoFilt v2.7.1 (De Coster et al., 2018), respectively. In total, 123,777 reads with a read length N50 value of 11,288 bp and a mean read quality score of 10.3 were available.

Both data sets were assembled and circularized using Unicycler v0.4.8 (Wick et al., 2017) including Pilon (Walker et al., 2014). Default parameters were used for all software unless otherwise noted.

2.6. Genotypic characterization and visualization

Characterization of the *S.* Agona sequences from the NRL for *Salmonella*, including 18-SA00377 and its four plasmids, was completed using the BakCharak pipeline version 3.0.4 (including -species *Salmonella* and -bakta options) (available at[4]). In short, the pipeline relies on the following tools: NCBI AMRFinderPlus (Feldgarden et al., 2021) to find antimicrobial resistance genes, ABRicate (available at[5]) to classify plasmids using CGE PlasmidFinder (Carattoli et al., 2014) and to find virulence factors using VFDB (Chen et al., 2015), Mash (Ondov et al., 2016) to find a nearest reference and plasmid reference using NCBI RefSeq (O'Leary et al., 2015) and NCBI plasmid database, respectively, Blastn (Camacho et al., 2009) to blast plasmids against the NCBI plasmid database, Bakta (Schwengers et al., 2021) for annotations using the Bakta database (Schwengers, 2021), and lastly, Platon (Schwengers et al., 2020) to predict plasmid contigs utilizing the Platon database (Schwengers, 2020). Further, *Salmonella*-specific characterization occurred using the SISTR tool with the sistr database and sistr 330 locus scheme for serotyping and cgMLST, respectively (Yoshida et al., 2016).

The annotations were supplemented with NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (Tatusova et al., 2016) and the following tools: ISfinder v. 2–2022-04-06 (Siguier et al., 2006), *PhiSpy* v.4.2.21 (Akhter et al., 2012), MobileElementFinder v1.0.3 (Johansson et al., 2020) and SPIFinder2 (Roer et al., 2016). For SPIFinder2,

1 https://www.cdc.gov/pulsenet/pdf/ecoli-shigella-salmonella-pfge-protocol-508c.pdf

2 https://community.nanoporetech.com

3 https://github.com/rrwick/Porechop

4 https://gitlab.com/bfr_bioinformatics/bakcharak

5 https://github.com/tseemann/abricate

thresholds of 80 and 90% of minimum sequence coverage and identity, respectively, were used to identify *Salmonella* Pathogenicity Islands (SPIs). Visualization occurred with circos v.0.69–6 (Krzywinski et al., 2009).

2.7. Short-read sequencing of related *Salmonella* Agona isolates

For a representative selection of the *S.* Agona isolates received by the strain collection of the German NRL for *Salmonella*, isolates of different isolation sources, isolation years and geographic origins were sequenced (Supplementary File 6). Genomic DNA was prepared as described above and short-read libraries were prepared using the Nextera DNA Flex Library Prep Kit (Illumina, San Diego, CA, USA Illumina). Sequencing occurred on either a NextSeq 500 or a MiSeq device (Illumina, San Diego, CA, USA). MiSeq sequencing was performed as described for isolate 18-SA00377-0. For the NextSeq, paired-end sequencing was performed in 2 × 151 bp using the NextSeq 500/550 Mid Output Kit v2.5 (300 Cycles).

2.8. Sequence data, allele calling and visualization

Short-read data of 18-SA00377 is available in BioProject PRJNA706607 and was compared to genetically similar isolates on NCBI Pathogen Detection database (accessed on 12.12.2022 at[6]) based on Single Nucleotide Polymorphisms (SNPs). Within the SNP cluster, three closely related isolates were identified and included in the phylogenetic analysis.

Additionally, all available *S.* Agona sequences (157 isolates, excluding duplicates) from the NRL for *Salmonella* (BioProject PRJEB31846 and BioProject PRJNA742494) were included for phylogenetic comparisons as well as the following six isolates: one MDR *S.* Agona sequence isolated from a silver gull and five isolates from recent European *S.* Agona outbreaks (Supplementary File 6). Where required, read-based sequencing data was assembled with the Assembly-based QUality Assessment for Microbial Isolate Sequencing (AQUAMIS) pipeline (Deneke et al., 2021a). A *S.* Agona isolate (GCA_011632245.1) with a highly similar resistance profile to 18-SA00377 (more than 80% overlap of ARGs) was also included in the phylogenetic comparison. Thus, a total of 167 isolates were included in the phylogenetic analysis using the cgMLST workflow ChewieSnake (Deneke et al., 2021b). This workflow implements the allele calling software chewBBACA and generates an allele distance matrix, cluster membership, and phylogeny. The resulting allele distance matrix was visualized as a minimum spanning tree using iTOL (Letunic and Bork, 2021). Isolates were classified based on the number of different antimicrobial resistance classes conferred by their ARGs according to the comprehensive antibiotic resistance database (Alcock et al., 2019).

2.9. Average nucleotide identity (ANI)-based plasmid comparisons

The plasmid pSE18-SA00377-1 was analyzed using two tools in order to find closed plasmid sequences with high levels of similarity: the MOB-cluster tool from the MOB-suite (Robertson and Nash, 2018), which utilizes fast genomic distance estimation using Mash, and the web service COPLA (available at[7]), a plasmid classifying tool. The resulting plasmids deemed similar to pSE18-SA00377-1 by these two tools were characterized using BakCharak version 3.0.4 and their Average Nucleotide Identity (ANI) scores were calculated using FastANI (Jain et al., 2018). Plasmids with the highest ANI were mapped against pSE18-SA00377-1 using minimap2 version 2.22-r1101 (−asm5 option). The 11 plasmids with the highest mapping coverages against pSE18-SA00377-1 were visualized using circos v. 0.69–6 6 (Krzywinski et al., 2009). Easyfig v. 2.2 (Sullivan et al., 2011) was utilized to compare the structural organization of two subregional clusters of pSE18-SA00377-1 with four of these plasmids. Easyfig visualized the BLAST results from comparing the two clusters (100–146 kb and 220–260 kb) with the four plasmids with the highest mapping coverage (NC_012555, CP011601, CP042552, and NC_012556).

2.10. Querying of NCBI pathogen detection databases for high similarity ARG resistance profiles to *Salmonella* Agona isolate 18-SA00377

The AMR resistance profile of the 18-SA00377 isolate available on NCBI was compared to all available isolates of *S. enterica*, *E. coli* and *Shigella* sp., *Klebsiella pneumoniae*, *Enterobacter* sp., *Acinetobacter baumannii* on the respective NCBI Pathogen Detection databases (available at https://www.ncbi.nlm.nih.gov/pathogens)[8] using a custom R script (Supplementary File 8) (databases download on 31.10.2022). The aim was to identify further *Salmonella* strains harboring plasmids that are similar to pSE18-SA00377-1, but for which no closed plasmid sequences were available. The draft genomes of all *Salmonella* genomes with a resistance profile that shared a minimum of 80% of the ARGs of 18-SA00377 were downloaded (Table 1), and mappings to pSE18-SA00377-1 with minimap2 version 2.22-r1101 (−asm5 option) (Li, 2018) were visualized using circos v. 0.69–6 (Krzywinski et al., 2009).

2.11. Filter mating conjugation experiments

To analyze transferability of pSE18-SA00377-1, filter mating conjugation studies using sodium azide-resistant *E. coli* K-12 J53 recipient cells were conducted. Selective marker for p18-SA00377-1 was the *tetD* gene, located on the plasmid (see

6 https://www.ncbi.nlm.nih.gov/pathogens

7 https://castillo.dicom.unican.es/copla

8 https://www.ncbi.nlm.nih.gov/pathogens/

TABLE 1 Available metadata details of the nine *Salmonella* isolates mapped against pSE18-SA00377-1, covering their size in bp, date of collection, country of origin, isolation source, mapping coverage (%) to pSE-18-SA0037-1, assembly level and their accession numbers.

Color in Figure 6	Isolate	Collection year	Country of origin	Isolation source	Mapped coverage (%)	Assembly level	Accession number
	Salmonella enterica subsp. *enterica* serovar Agona	2018	Germany	Human (clinical)	94	Contig	GCA_020159705.1
	Salmonella enterica subsp. *enterica* serovar 4,[5],12:i:-	2015	USA	Animal (*Sus scrofa domesticus*)	93	Contig	GCA_007760705.1
	Salmonella enterica	NA	USA	Clinical	92	Contig	GCA_006389875.1
	Salmonella enterica subsp. *enterica* serovar 4,[5],12:i:-	2014	USA	Clinical (human)	92	Contig	GCA_006389855.1
	Salmonella enterica	2015	USA	Food (porcine)	92	Contig	GCA_005610165.1
	Salmonella enterica subsp. *enterica* serovar Agona	2019	USA	Animal (*Sus scrofa domesticus*)	92	Contig	GCA_011632245.1
	Salmonella enterica subsp. *enterica* serovar 4,12:i:-	2015	USA	Animal (*Sus scrofa*)	92	Contig	GCA_007756295.1
	Salmonella enterica	NA	NA	NA	89	Contig	GCA_011585645.1
	Salmonella enterica	2015	USA	Animal (*Bos taurus*)	88	Contig	GCA_005899425.1

Table 2). Filter mating conjugation was performed as previously described (Hadziabdic et al., 2018). The reaction mixtures were plated on transconjugant selective LBA plates containing 12.5 mg/liter tetracycline (tetracycline, Sigma-Aldrich, Steinheim, Germany) and 100 mg/liter sodium azide (NaN_3, Sigma-Aldrich, Darmstadt, Germany) and incubated at 37°C for approximately 42 h. A selection of potential transconjugants were picked and singulated on new selective tet/NaN_3 plates. From these plates, single colonies were inoculated in LBL and incubated at 37°C under shaking condition (250 × rpm) for 16 h. Thermal cell lysis preparations were produced as previously described (Borowiak et al., 2017). Transconjugants were confirmed by *tetD* and J53 K12 screening PCR. The following primer pairs (with primer sequences in 5′-3′ direction and product length in parentheses) were used for the screening PCRs: For the J53 K12 screening K12R (ATCCTGCGCACCAATCAACAA; 1687 bp) and K12L (TTCCCACGGACATGAAGACTACA; 1687 bp) (Bauer et al., 2007), and for the *tet(D)* screening *tet(D)*-1 (AAACCATTACGGCATTCTGC; 787 bp) and *tet(D)*-2 (GACCGGATACACCATCCATC; 787 bp) (Ng et al., 2001).

The PCR reactions were prepared in 25 µL including 12.5 µL 2x DreamTaq Green PCR Master Mix (Thermo Scientific, Vilnius, Lithuania), 2.5 µL of each 10 µM primer dilution (s. above for details), 5.5 µL PCR grade water and 2 µL of thermal cell lysis preparation as template DNA and carried out as follows: initial denaturation for 5 min at 95°C, 30 cycles denaturation for 30 s at 95°C, primer annealing for 30 s at 54°C and elongation for 1 min [*tet(D)*] or 1:40 min (J53 K12) at 72°C followed by a final elongation step for 10 min at 72°C.

3. Results and discussion

3.1. Features of the 18-SA00377 chromosome and its plasmids

Illumina short-read and ONT long-read sequencing resulted in a hybrid assembly of one *Salmonella* chromosome with 4,856,956 bp and four plasmids: one IncHI2 plasmid (pSE18-SA00377-1) of 295,499 bp, one p0111 plasmid (pSE18-SA00377-2) of 94,574 bp, one IncX3 plasmid (pSE18-SA00377-3) of 50,931 bp and one Col (Ye4449) plasmid (pSE18-SA00377-4) of 5,284 bp. These five constituent assemblies of 18-SA00377 are available online on NCBI, grouped under the GenBank Accession number GCA_021497565.1 and BioProject PRJNA706607 with their individual Genbank Accession numbers as follows: CP071388.1 (chromosome), CP071389.1 (pSE18-SA00377-1), CP071390.1 (pSE18-SA00377-2), CP071391.1 (pSE18-SA00377-3), and CP071392.1 (pSE18-SA00377-4).

The presence of at least three plasmids was confirmed by PFGE using the S1 nuclease (Supplementary File 1). The sizes of these three plasmids were determined to be around 45, 80 and 320 kb. While varying slightly from the whole genome sequencing results this can be explained by the fact that S1-PFGE is more reliable at ascertaining plasmid sizes above 100 kb (Barton et al., 1995; Li et al., 2022) and can be unreliable for smaller plasmid sizes (Zhang et al., 2020; Juraschek et al., 2021).

Following genotypic characterization using the BakCharak pipeline and NCBI PGAP, the isolate 18-SA00377 was found to exhibit one DNA gyrase amino acid substitution at codon 83 (*gyrA_S83F*) and 23 different antibiotic resistance genes (ARGs). In total, the isolate

TABLE 2 Breakdown of SPIs, plasmid markers, antibiotic resistance genes (ARGs), heavy metal resistance genes, and results from antibiotic susceptibility testing of the 18-SA00377 isolate including its four constituent plasmids.

Isolate	SPIs [a] or Plasmid markers [b]	ARGs [c] [d] and amino acid substitution	Antibiotic susceptibility [e]	Heavy metal resistance gene [c]
SE18-SA00377-chromosome	SPI-1, SPI-2, SPI-4, SPI-8, SPI-9, and SPI-16	**Efflux transporter:** *mdsA* and *mdsB* **Fluoroquinolone:** *gyrA*_S83F **Fosfomycin:** *fosA7.6*	**Fluoroquinolone:** CIP =4 (S) and NAL >128 (R) **Fosfomycin:** FOS >4 **Aminoglycoside:** GEN >132 (R), KAN >64, and STR >32 **Beta-lactam:** AMP >64 (R), ETP <=0.015 (S), FEP =16 (R), FOT >4 (R), IMI =0.025, FOX =8 (S), PEN >2, MERO <=0.03 (S) and TAZ >8 (R) **Diaminopyrimidine:** TMP >32 (R) **Macrolide:** AZI =16 (S) **Peptide:** VAN >16 **Phenicol:** CHL >128 (R) **Polymyxins:** COL <= 1 (S) **Rifamycin:** RIF >0.5 **Sulfonamide** SMX >1,024 (R) **Tetracycline** TET >64 (R) and TGC =1 (S)	**Gold resistance:** *golS* and *golT*
pSE18-SA00377-1	RepA_1_pKPC-CAV1321, IncHI2A and IncHI2	**Aminoglycosides:** *aac(3)-Iig*, *aac(6′)-Iic*, *aadA2, aph(3′)-Ia*, *aph(3″)-Ib* and *aph(6)-Id* **Beta-lactam:** bla_{SHV-12} **Diaminopyrimidine:** *dfrA19* **Efflux transporter:** *qacEΔ1* * **Macrolide:** *ere(A)* * **Peptide antibiotic:** *mcr-9.1* **Phenicol:** *catA2* * **Rifamycin:** *arr* **Sulfonamide:** *sul1* * and *sul2* **Tetracycline:** *tet(D)*		**Arsenic resistance:** *arsB*, *arsC*, and *arsH* **Copper resistance:** *pcoE*, *pcoR*, and *pcoS* **Mercury resistance:** *merA*, *merD*, *merE*, and *merT* **Nickel/Cobalt resistance:** *rcnA* and *rcnR* **Tellurium resistance:** *terD*, *terW*, and *terZ*
pSE18-SA00377-2	p0111			
pSE18-SA00377-3	IncX3	**Aminoglycoside:** *aac(3)-IIe* **Beta-lactam:** bla_{TEM-1} **Fluoroquinolone:** *qnrS1* **Phenicol:** *floR*		
pSE18-SA00377-4	Col(Ye4449) and Col(MGD2)			

[a] SPIFinder2 (Roer et al., 2016) with minimum sequence coverage of 90% and minimum sequence identity of 80%, [b] ABRicate (available at https://github.com/tseemann/abricate) and CGE PlasmidFinder (Carattoli et al., 2014), [c] NCBI AMRFinder (Feldgarden et al., 2021) and [d] Comprehensive Antibiotic Resistance Database (Alcock et al., 2019). [e] For MIC testing the CLSI guidelines (version A7-M11) were followed and for interpretation the epidemiological cutoff values provided by EUCAST. Duplicates are marked with *.

18-SA00377 carries 27 ARGs as four ARGs [*catA2*, *qacEΔ1*, *sul1* and *ere(A)*] are in duplicate and they confer resistance to 12 different classes of antibiotics resistance (Table 2). This was supported by antimicrobial susceptibility testing against antimicrobials of nearly every drug class, excluding efflux transporters. Moreover, the isolate harbors resistance genes against six heavy metals (gold, tellurium, arsenic, mercury, copper, and nickel/cobalt) as well as containing 131 chromosomal virulence factors belonging to seven classes (adherence, antimicrobial activity/competitive advantage, effector delivery system, immune modulation, invasion, nutritional/metabolic factor, and regulation) (data not shown). The isolate's chromosome also harbors six *Salmonella* Pathogenicity Islands (SPIs) – SPI-1, SPI-2, SPI-4, SPI-8, SPI-9, and SPI-16 – with a sequence coverage of minimum 90% and sequence identity of above 80%. The following nine further SPIs were identified having a sequence coverage between 30 and 80%, but sequence identities of above 90%: SPI-3, SPI-12 (two copies), SPI-11, SPI-5, SPI-19, SPI-16 (two copies). Three of these SPIs (SPI-1, SPI-2, and SPI-4) have important roles in the virulence of *Salmonella* infections and can be found in all serovars of *Salmonella enterica*. Both SPI-1 and SPI-2 encode a distinct of type III protein secretion system (Jennings et al., 2017; Lou et al., 2019), which are, *inter alia*, important for the penetration and invasion of epithelial intestinal cells. On the other hand, the role of SPI-4 encodes a type I secretion system, which is crucial for adhesion during *Salmonella* infections (Gerlach et al., 2007). Two other SPIs – SPI-8 and SPI-9 – have been characterized based on the complete genome sequence of a *S.* Typhi CT18 strain (Parkhill et al., 2001). The function of SPI-8 is not well understood, but it has been shown to conferring resistance to bacteriocins (van Asten and van Dijk, 2005), while SPI-9, similar to SPI-4, encodes a type I secretion system as well as a RTX toxin-like protein (Velásquez et al., 2016). Lastly, SPI-16 has a role in immune evasion by carrying genes involved in O-antigen variation (Ilyas et al., 2017). Of the nine other SPIs present, the SPI-5 and SPI-3 are of particular importance as they are important mediators in host colonization and intracellular survival (Blanc-Potard et al., 1999), as well as the enteric stage of a *Salmonella* infections (Marcus et al., 2000), respectively. Overall, the presence of such a wide range of SPIs in a single isolate underlines the pathogenic potential of 18-SA00377. Thus, as previous outbreaks of *S.* Agona have shown, this serovar can cause human harm and the presence of both multidrug resistance and heavy metal resistances, highlight the importance of finding closely related isolates and establishing phylogeny.

3.2. Phylogenetic analysis of 18-SA00377

NCBI Pathogen Detection database based on Single Nucleotide Polymorphism (SNP) was searched for closely related isolates. This showed that the isolate 18-SA00377 is, with a minimum SNP distance of 31, only distantly related to three other *S.* Agona isolates: GCA_011585645.1, GCA_006296875, and GCA_020159705.1) (Supplementary File 6).

Next, all available sequences of *S.* Agona isolates from the German NRL for *Salmonella* was supplemented with *S.* Agona sequences from four other sources. Firstly, from three foodborne outbreaks: the 2017/18 infant formula outbreak in France (Jourdan-da Silva et al., 2018), the 2002/03 herbal tea outbreak (Koch et al., 2005), and the

outbreak caused by a Bavarian feed product (Dangel et al., 2019). Secondly, three isolates from the aforementioned NCBI Pathogen Detection database (GCA_011585645.1, GCA_006296875, and GCA_020159705.1) as well as the sequence of a extensively drug-resistant *S.* Agona isolated from a silver gull (GCA_012169275.1) (Cummins et al., 2020). Lastly, a *S.* Agona isolate (GCA_011632245.1) with a high similarity ARG resistance profile was also included. A cgMLST analysis of these 167 *S.* Agona sequences (Supplementary File 6) was completed and visualized as a minimum spanning tree (Figure 1) (Letunic and Bork, 2021).

As shown in Figure 1, isolate 18-SA00377 is located within a clade, here named cluster A, with three other isolates: GCA_020159705.1, GCA_011585645.1 and GCA_006296875.1 with minimum allelic distances of 60, 19 and 27, respectively. These three isolates are not isolates from the German NRL for *Salmonella*, but were found via the NCBI Pathogen Detection database search. Two isolates, GCA_020159705.1 and GCA_006296875.1, were isolated from human sources in Germany in 2018 and the United Kingdom in 2014, respectively, while for GCA_011585645.1 metadata was not available. The closest *S.* Agona isolate to 18-SA00377 from the German NRL *Salmonella* has a minimum allelic distance of 74. The closest outbreak-associated *S.* Agona isolate is SRR7253423 from the Bavarian prevalence study, isolated from animal feed in Germany in 2017 (Dangel et al., 2019). As the allelic distances both between isolates within cluster A and to the closest cluster of isolates from the NRL for *Salmonella* exceed cut-offs previously used for defining distinct *Salmonella* outbreak clusters (Simon et al., 2018; Meinen et al., 2019), a recent common ancestor cannot be pinpointed by the cgMLST of 167 *S.* Agona isolates.

The total number of ARG drug classes across the 167 *S.* Agona isolates varies considerably and is not uniformly distributed. The majority of *S.* Agona isolates, including all the available *S.* Agona sequences in the NRL for *Salmonella*, harbor ARGs against only two classes, fosfomycin (*fosA7.2*) and efflux transporter subunits (*mdsA* and *mdsB*). Only a minority of *S.* Agona isolates ($n = 8$) carry resistance genes encoding for nine or more drug classes. Nevertheless, the four isolates in cluster A all carry ARGs against a minimum of nine antibiotic classes. Isolate 18-SA00377 is one of the most resistant isolates with 23 distinct ARGs conferring resistance to 12 different classes. There is an overlap between the antibiotic classes, with all four isolates in cluster A sharing resistances against the aforementioned fosfomycin and efflux transporter subunits, as well as the peptide antibiotic colistin (*mcr-9.1*), beta-lactams, fluoroquinolones, aminoglycosides, sulfonamides, tetracyclines, and diaminopyrimidine (see Supplementary Files 6, 7). Since the cgMLST analysis did not reveal a recent common ancestor for 18-SA00377, the antibiotic resistance profiles within cluster A and the unique resistance profile of 18-SA00377 highlight the importance of finding another potential source of the resistance properties of 18-SA00377, namely associated mobile genetic elements.

3.3. Plasmid descriptions and comparisons

Next, we focused to finding isolates with similar antibacterial resistance profiles in a wider set of genera of *Enterobacteriaceae*. As the majority of its unique MDR resistance profile is due to the ARGs carried on the isolate's largest plasmid, pSE18-SA00377-1, the plasmid was characterized. Assembly of the plasmids was possible by combining long-read and short-read sequencing data. Annotation of the assembly revealed that pSE18-SA00377-1 harbors a total of 20 ARGs (Table 2), including two copies of each of *catA2*, *qacEΔ1, sul1*, and *ere(A)*. A further four ARGs (*qnrS1, bla*$_{\text{TEM-1}}$, *aac(3)-IIe*, and *floR*) are located on pSE18-SA00377-3 and each associated with a putative composite transposon of the IS*6* family (Supplementary File 4). The plasmids, pSE18-SA00377-2 and pSE18-SA00377-4, do not carry any antibiotic resistance genes, but pSE18-SA00377-4 carries two replicons, Col(MGD2) and Col(Ye4449) as well as mobilization genes (*mobC*, *mbeD*, *mbeB*, *mbA*) (Supplementary File 5).

Similar to the ARG distribution, the distribution of genes conferring heavy metal resistances is skewed toward pSE18-SA00377-1, where genes encoding against six different heavy metal resistances are located (Table 2). While the genome only carries *golS* and *golT* genes, encoding for gold resistance, the resistance genes against arsenic (*arsB, arsC*, and *arsH*), copper (*pcoE, pcoR, pcoS)*, mercury (*merA, merD, merE*), nickel/cobalt (*rcnA* and *rcnR)*, and tellurium (*terD, terW, terZ*) all are located on pSE18-SA00377-1.

The makeup of pSE18-SA00377-1 is not only limited to a large number of resistance genes, it also carries three plasmid markers: RepA_1_pKPC-CAV1321 as well as IncHI2A and IncHI2. Moreover, it also carries genes for the purpose of conjugational transfer by encoding for numerous conjugal transfer proteins (e.g., *traK, traB, traV,* etc.) as well as an *oriT and repB* replication initiator. The presence of these genes supports the assumption that the pSE18-SA00377-1 plasmid is transferrable *in vivo*.

Transmissibility of pSE18-SA00377-1 by conjugation was confirmed experimentally by *in vitro* filter mating experiments. Successful transfer of the plasmid to the recipient, *E. coli* J53 K12, occurred, supporting the bioinformatic analysis that pSE18-SA00377-1 is conjugative. Furthermore, as the recipient *E. coli* K12 J53 belongs to a different genus of the family *Enterobacteriaceae*, this supports the subsequent bioinformatic analyses that indicates that large parts of the pSE18-SA00377-1 plasmid backbone can be found in other bacterial genera and that the host range of the pSE18-SA00377-1 plasmid is not limited to *Salmonella*.

Arrangement of ARGs, heavy metal resistances, conjugation machinery and transposable elements on pSE18-SA00377-1 is concentrated within two distinct clusters, as evident in the visualization of the plasmid's annotations (Figure 2). The annotations of the remaining plasmids, pSE18-SA00377-2, pSE18-SA00377-3, and pSE18-SA00377-4, were not visualized, but their consensus annotations are available as Supplementary Files 3–5, respectively. The first cluster on pSE18-SA00377-1, ~ 95–140 kb, contains nine ARGs interspersed with transposases. Four ARGs – *aadA2*, *qacEΔ1*, *sul1*, and *dfrA19* – are part of a class 1 integron cassette as they are flanked on both sides by class 1 integron integrases (*intI1*) and have Pc and PintI1 promoters in their vicinity as well as two recombination crossover points (*attC* and *attI*). Moreover, all nine ARGs in this region are associated with putative composite transposons. Similarly, six ARGs – *aadA2*, *qacEΔ1*, *sul1, dfrA19, aph(3″)-Ib* and *aph(6)-Id* – are associated with a single putative composite transposon,

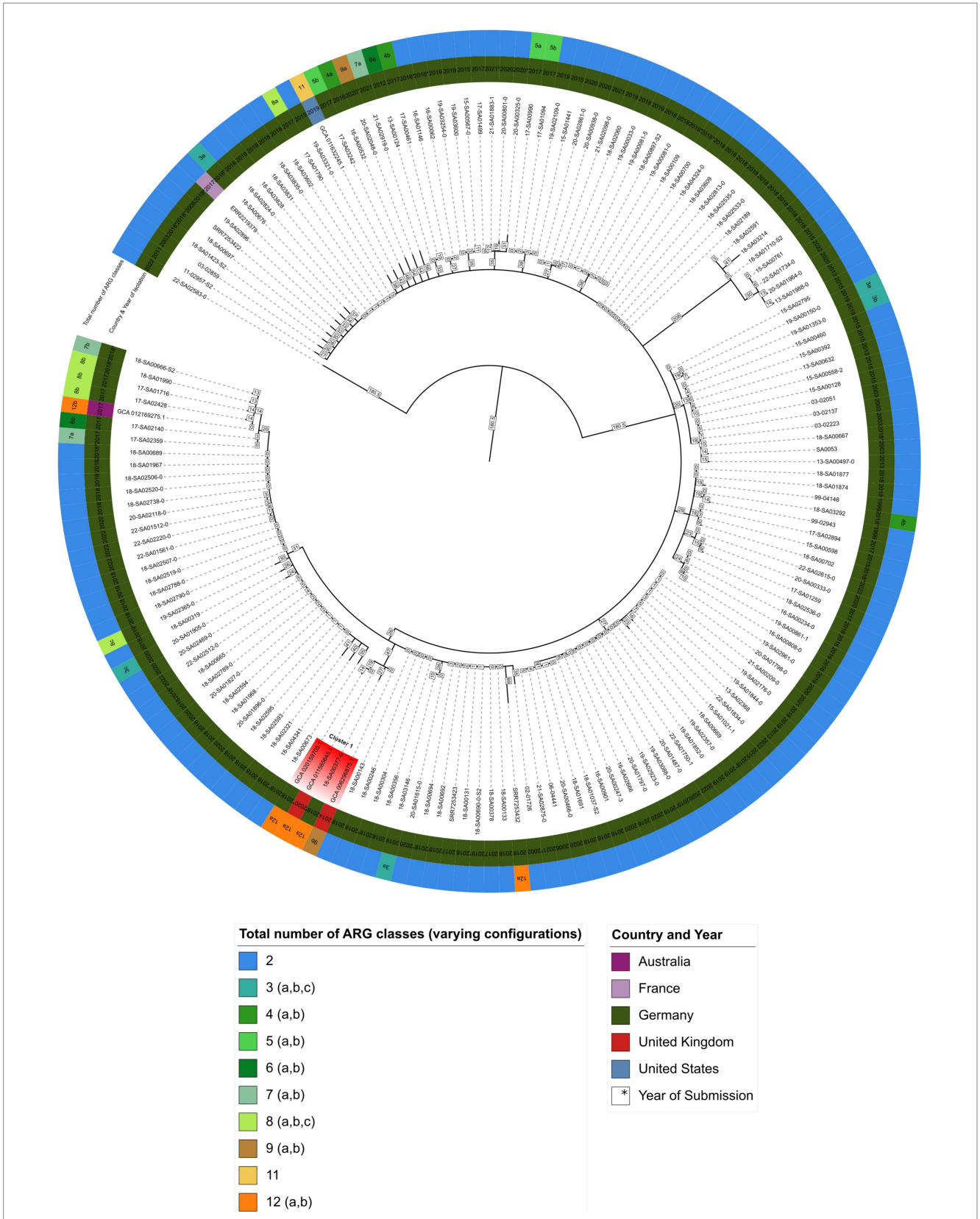

FIGURE 1
Minimum spanning tree showing hierarchical clustering between 167 *S.* Agona isolates. Inner circle indicates the country and isolation year, outer circle represents the number of ARG classes for each isolate, respectively (Supplementary File 6). Branch lengths in square boxes indicate minimum allelic distance between isolates. Varying configurations in the total number of different antibiotic resistance drug classes are indicated with a, b, and c – for a detailed breakdown see Supplementary File 7.

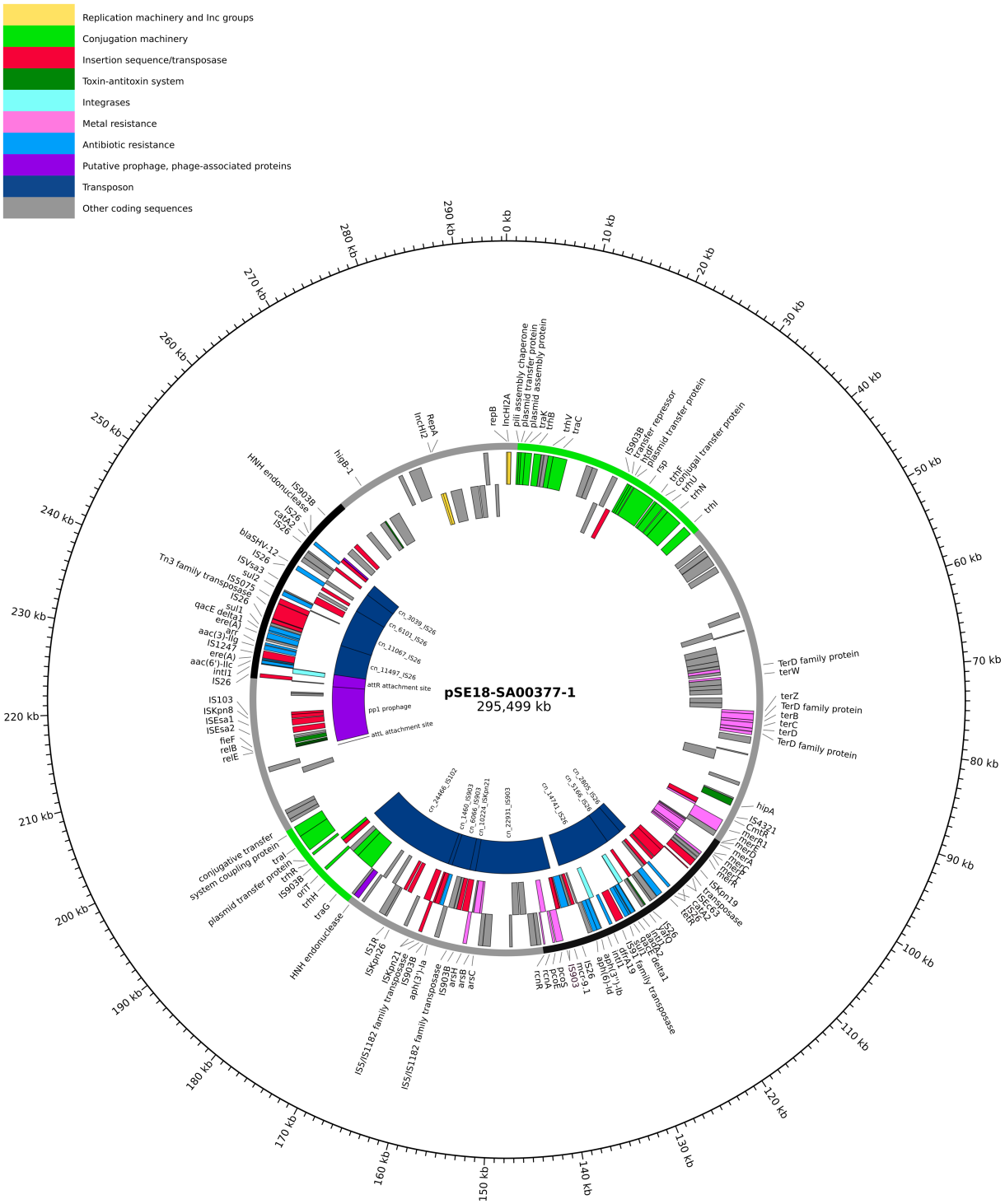

FIGURE 2
Visualization of the pSE18-SA00377-1 plasmid. Annotations are colored according to nine categories: replication machinery and incompatibility group in yellow, conjugation machinery in light green, insertion sequences and transposases in red, toxin-antitoxin system in dark green, integrases in turquoise, heavy metal resistances in pink, and antibiotic resistances in light blue. On the innermost track, the putative prophage and prophage-associated proteins are shown in purple, while putative transposons are colored dark blue. Coding sequences for other gene products are colored in gray. The outermost ring shows the plasmid backbone in gray and highlights regions of interest: cluster 1 (95–140kb) and cluster 2 (230–250kb) in black and regions with conjugation machinery in green (0–40kb and 180–200kb). For a complete list of available annotations, see Supplementary File 3.

cn_14741_IS26, which is flanked by two IS*26* elements. A similar picture emerges for heavy metal resistances, with genes encoding for copper, nickel/cobalt, and nickel resistances associated with a putative composite transposon, cn_22931_IS903.

A second cluster with a high concentration of ARGs spans the range 230–250 kb and contains further 10 ARGs. The *aac(6')-IIc* gene is associated with an incomplete class 1 integron as a class 1 integron integrase (*intI1*), Pc and PintI1 promoters and recombination

TABLE 3 Metadata details of the 11 plasmids mapped against pSE18-SA00377-1, covering their size in bp, date of collection or NCBI submission, country of origin, category of isolation source, percentage of coverage when mapped to pSE-18-SA0037-1, and their accession.

Color in Figure 2	Isolate	Size (bp)	Collection year	Country of origin	Isolation source	Mapped coverage (%)	Accession number
	Enterobacter cloacae plasmid pEC-IMP	318,782	2008 [a)]	Taiwan	Clinical	92	NC_012555
	Phytobacter ursingii strain CAV1151 plasmid pCAV1151-296	295,619	2009	USA	Clinical (human)	89	CP011601.1
	Enterobacter hormaechei strain C45 plasmid pC45_001	288,659	2013	Australia	Clinical (human)	85	CP042552
	Enterobacter cloacae plasmid pEC-IMPQ	324,503	2008 [a)]	Taiwan	Clinical	85	NC_012556
	Enterobacter asburiae strain AMA 497 plasmid pOXA436	314,137	2014	Denmark	Clinical (human urine)	85	KY863418
	Enterobacter hormaechei strain C15117 plasmid pSPRC-Echo1	339,920	2007	Australia	Clinical (Burns Unit Surveillance)	85	CP032842
	S. Typhimurium strain MU1 plasmid pIMP4-SEM1	339,962	2016	Australia	Animal (cat)	79	KX810825
	Enterobacter hormaechei strain MS7884A plasmid pMS7884A	330,060	2015	Australia	Clinical (human)	71	CP022533.1
	S. Heidelberg strain 09–036813-1A plasmid p09-036813-1A_261	261,310	2009	Canada	Animal (horse)	57	CP016526.1
	Enterobacter hormaechei subsp. *steigerwaltii* strain 34,977 plasmid p34977-263.138 kb	263,138	2009	USA	Clinical (human)	55	CP012170.1
	Citrobacter farmeri strain AUSMDU00008141 plasmid pAUSMDU8141-1	328,945	2015	Australia	Clinical (human)	51	CP022696.1

[a)] Year of submission to NCBI.

crossover points (*attC* and *attI*) are in its vicinity. Similar to the first cluster, all ARGs are located on putative composite transposons, with the first seven ARGs being associated with the cn_11497_IS26 putative composite transposon, while each of the remaining three ARGs are associated with a composite transposons of the IS*6* family. Between these two distinct regions of the plasmid, pSE18-SA00377-1 also encodes components for a type II toxin-antitoxin system (*relE*, *relB*) and an iron efflux transporter (*fieF*).

Closer analysis of the plasmid's genetic makeup revealed that this large plasmid has two distinct regions of densely clustered ARGs, which are associated with putative composite transposons. This, in conjunction with the presence of numerous copies of highly active IS*26* elements, that frequently mediate recombination in *Salmonella* spp. (Doublet et al., 2009), suggests that this plasmid has a complex evolution with frequent insertions of ARGs.

3.4. Comparative analysis of the pSE18-SA00377-1 plasmid

In order to find closely related plasmids of pSE18-SA00377-1, the outputs of two tools, the MOB-cluster tool and the COPLA web tool, were further analyzed by calculating their ANI scores. Consequent ranking by ANI score and matching antibiotic resistance profiles yielded 11 high-similarity plasmids (Table 3), which were mapped to pSE18-SA00377-1 (Figure 3).

When mapped against pSE18-SA00377-1, these 11 plasmids exhibit a nucleotide coverage exceeding 50%. They were isolated from a wide range of bacterial isolates and their geographical origins span Australia (CP022696.1, CP022533.1, CP042552, CP032842, and KX810825), Taiwan (NC_012555 and NC_012556), the United States (CP012170.1 and CP011601.1), Canada (CP016526.1), and Denmark (KY863418). Furthermore, these plasmids were mostly isolated from *Enterobacter* species in a clinical context, but were also found in *Salmonella*, *Citrobacter*, and *Phytobacter*. The *Enterobacter* and *Citrobacter* plasmids were found exclusively in clinical isolates, while the *Salmonella* plasmids were isolated from animals. Particularly, the occurrence of the *Salmonella* plasmid, KX810825, is of concern as it was found in a companion animal (Abraham et al., 2016).

Figure 3 shows that the coverages against pSE18-SA00377-1 (*E. cloacae* pEC-IMP, *Phytobacter ursingii*, *Enterobacter hormaechei*, and *E. cloacae* pEC-IMPQ) all map in the region of the first cluster of ARGs. However, for the second cluster of ARGs at 230–250 kb only the *E. cloacae* pEC-IMP plasmid (in dark purple) maps against large parts of this cluster, while the other six plasmids do not. All eleven plasmids wholly map against the majority of the pSE18-SA00377-1 plasmid backbone, including the conjugation and replication machinery.

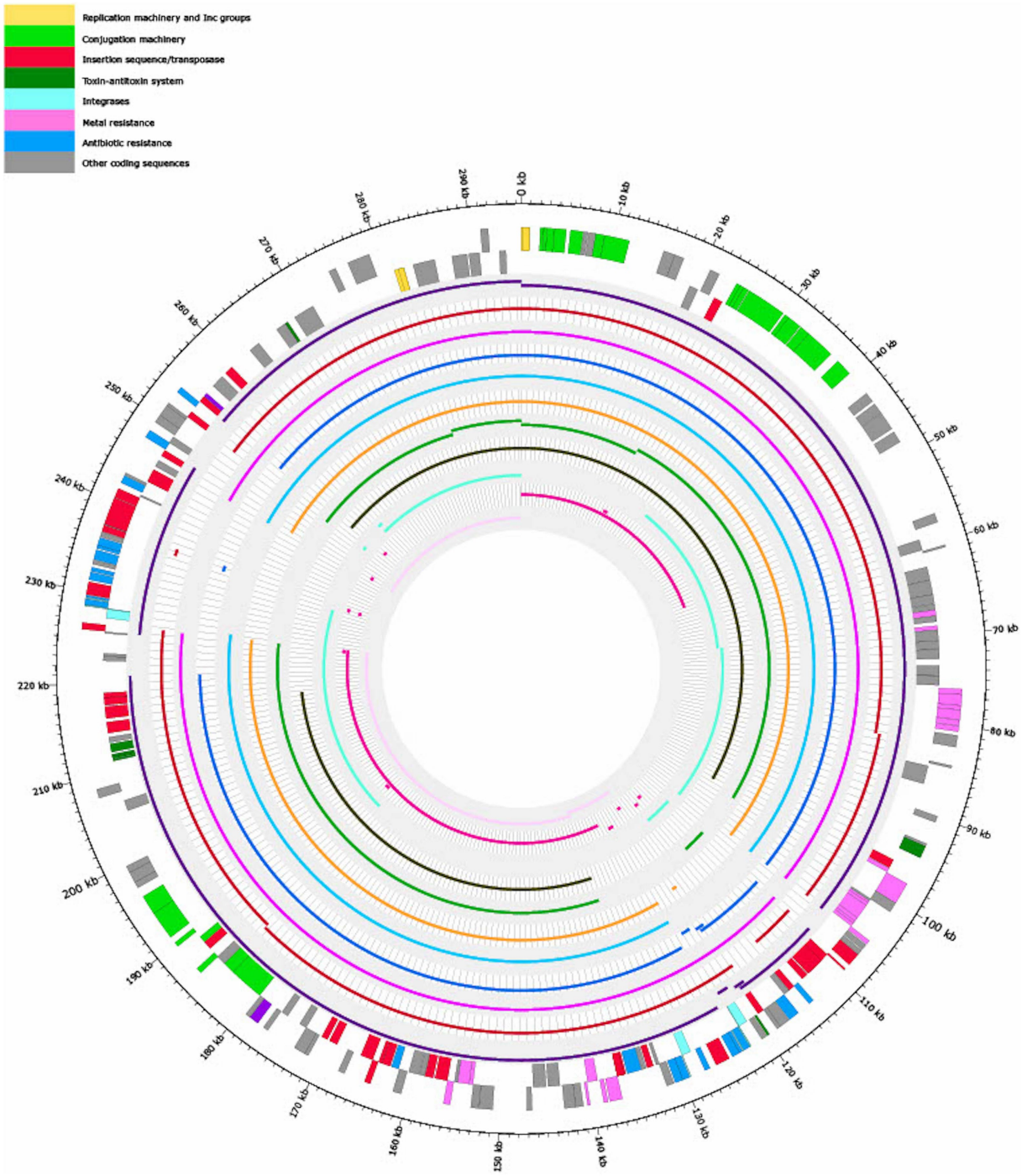

FIGURE 3
Visualization of mapping results of 11 plasmids against pSE18-SA00377-1 (outermost two tracks), in order of decreasing mapping coverage (%), with lowest coverage of the *Citrobacter farmeri* plasmid (light pink) and highest coverage of the *Enterobacter cloacae* plasmid pEC-IMP (dark purple) (see Table 3). Regions of the 11 plasmids that did not map to pSE18-SA00377-1 are not shown. The outermost black circle designates the base positions around the plasmid. Farthest two tracks on the outside represents the pSE18-SA00377-1 plasmid annotations, with replication machinery and incompatability groups colored in yellow, conjugation machinery in light green, insertion sequences and transposases in red, toxin-antitoxin system in dark green, integrases in turquoise, heavy metal resistances in pink, and antibiotic resistances in light blue. Coding sequences for other gene products are colored in gray. For labeling of annotations refer to Figure 2.

3.5. Structural representation of two pSE18-SA00377-1 subregions

For closer comparison of the organization of the ARGs and other features within the two aforementioned ARG clusters of pSE18-SA00377-1, these regions were compared to the 11 high-similarity plasmids using BLAST and visualized using easyfig. This closer inspection revealed that the organization of the first cluster of pSE18-SA00377-1, 100–146 kb, is conserved across four plasmids NC_012555, CP011601, CP042552, and NC_012556 (Figure 4).

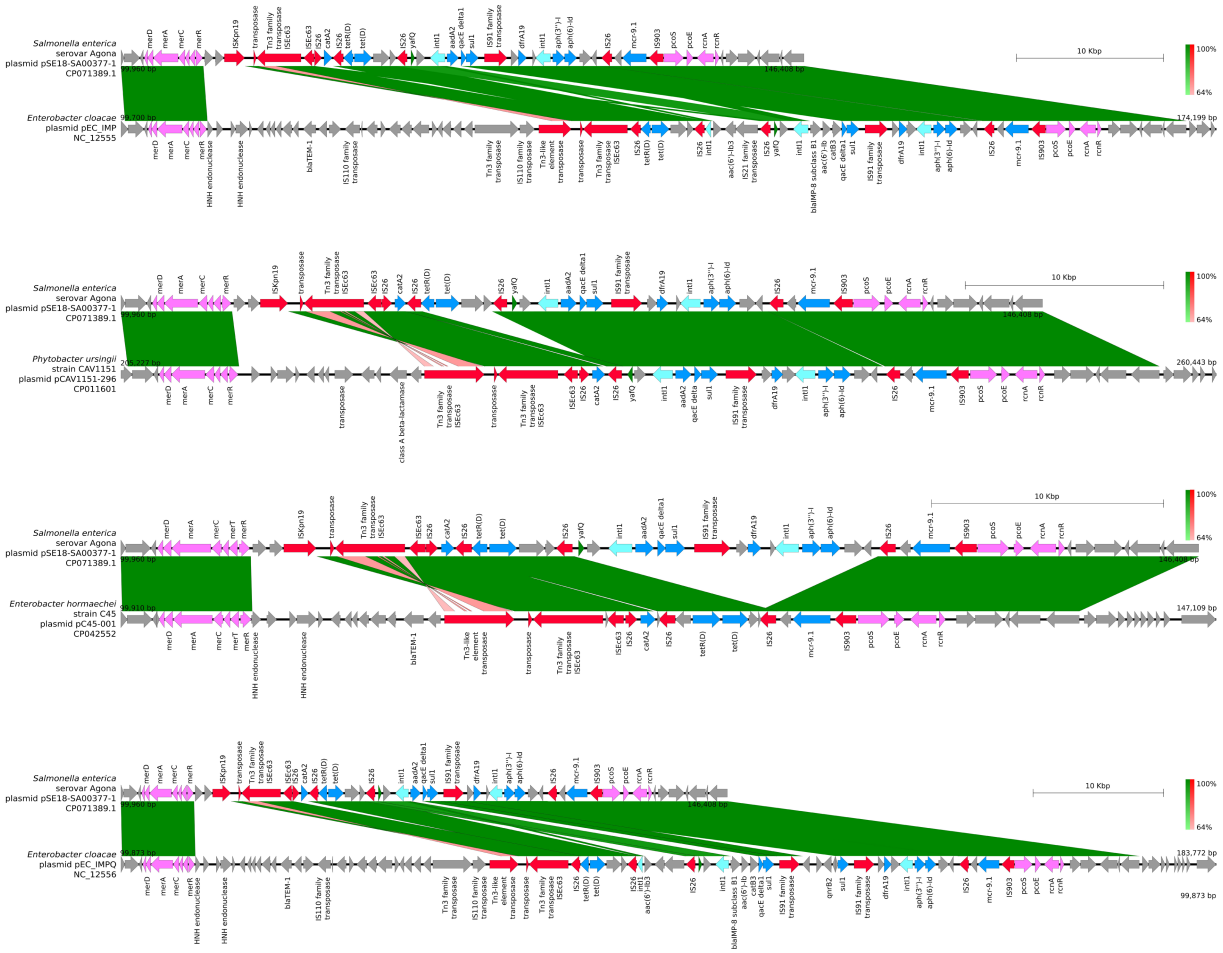

FIGURE 4
Visualization of comparison of the structural features between the 100–146 kb region of pSE18-SA00377-1 and four plasmids with the highest nucleotide coverages (NC_012555, CP011601, CP042552, and NC_012556). Annotations of coding sequences are only colored if BLAST results indicated homology between pSE18-SA00377-1 and the other plasmid. Annotation coloring is based on the same eight categories as in Figures 2, 3, with insertion sequences and transposases in red, toxin-antitoxin system in dark green, integrases in turquoise, heavy metal resistances in pink, and antibiotic resistances in light blue. Coding sequences present in the plasmids belonging to one of the eight colored annotation categories, but showing no homology by blast to pSE18-SA00377-1, were labeled but left in gray.

All four of these plasmids harbor minimum eight of the nine ARGs constituent of this region, *catA2*, *tetD*, *aadA2*, *qacEΔ1*, *sul1*, *dfrA19*, *aph(3″)-Ib*, *aph(6)-Id*, and *mcr-9.1*. In two plasmids, CP016526 and CP012170, the region with its nine ARGs is present but in an inverted state and only partially. Four ARGs are located toward the end of the region, *dfrA19*, *aph(3″)-Ib*, *aph(6)-Id*, and *mcr-9.1*. They are present in both plasmids, while the remaining ARGs, *catA2*, *tetD*, *aadA2*, *qacEΔ1*, *sul1* have merged with the second cluster of ARGs (220–260 kb). In the five remaining plasmids, the organization of the ARGs is considerably changed, with the order of the ARGs split or shuffled (CP022696, CP022533, CP042552, CP032842, and KX810825) or merged partially (CP042552) with the second cluster of ARGs.

Similar to the first region, the second cluster of ARGs at 220–260 kb of pSE18-SA00377-1 was compared to four of the 11 aforementioned high-similarity plasmids and visualized using easyfig (Figure 5). For this second cluster, two plasmids (NC_012555 and NC_012556) harbor the same structural features in the same organizational matter as pSE18-SA00377-1, while the remainder of the 11 plasmids only carry a partial ARG load.

The mapping of these 11 plasmids to pSE18-SA00377-1 indicates that the pSE18-SA00377-1 contains a conserved plasmid backbone which frequently occurs in other plasmids of other genera. However, the widespread geographical origin of these 11 plasmids and their occurrence in a wide range of bacterial genera does not allow for ascertaining a potential common origin. Nevertheless, the high number of ARGs as well as their associations with composite transposons indicates that multiple insertion events had occurred and led to this accumulation of resistance genes in two distinct clusters. This accumulation of resistance genes is also present in the other 11 plasmids, although the first cluster of ARGs seems to be more stable, being present in its entirety and same organizational structure in four other plasmids.

This analysis showed that plasmids with high nucleotide similarity to pSE18-SA00377-1 and similar antibiotic resistance

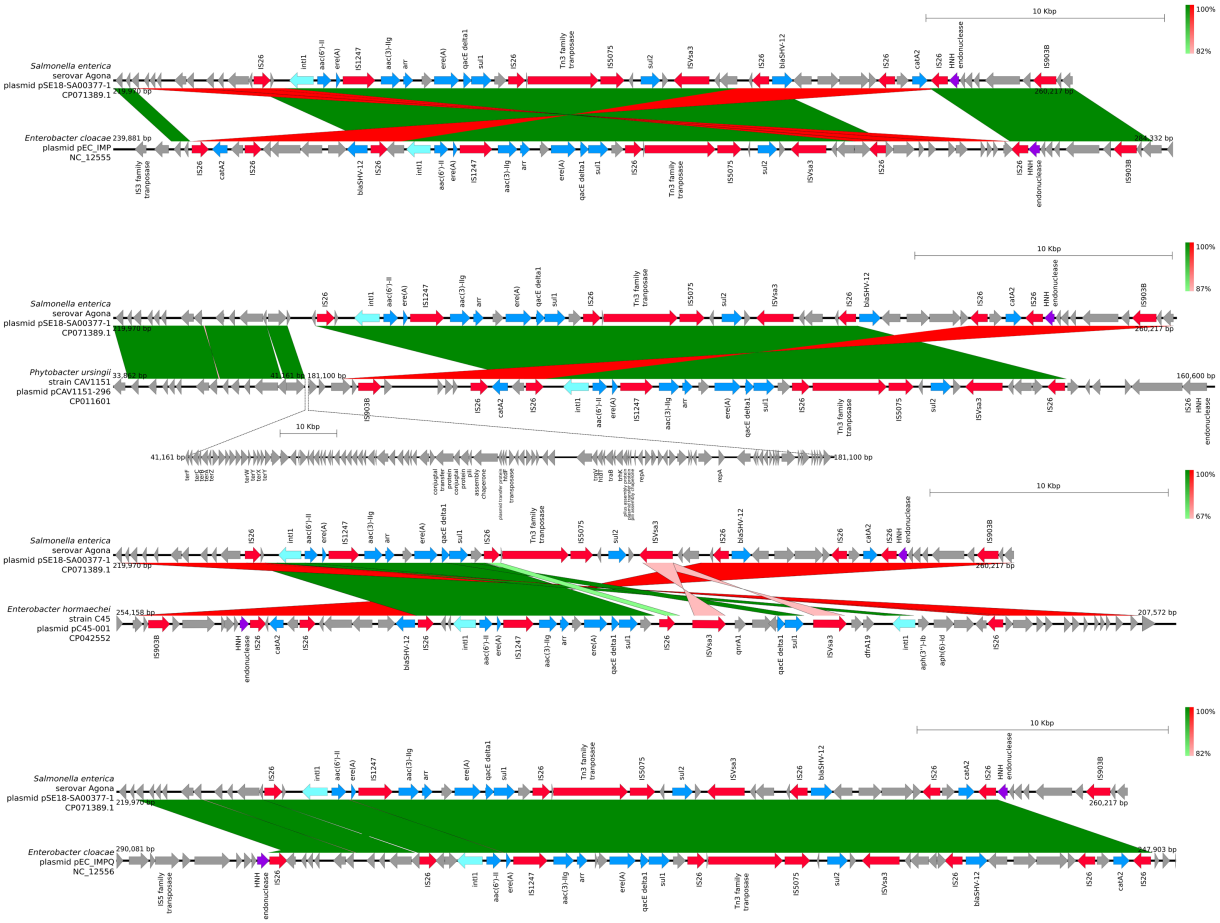

FIGURE 5
Visualization of comparison of the structural features between the 220–260 kb region of pSE18-SA00377-1 and four plasmids with the highest nucleotide coverages (NC_012555, CP011601, CP042552, and NC_012556). Annotations of coding sequences are only colored if blast results indicated homology between pSE18-SA00377-1 and the other plasmid's. Annotation coloring is based on the same eight categories as in Figures 2, 3, with insertion sequences and transposases in red, toxin-antitoxin system in dark green, integrases in turquoise, heavy metal resistances in pink, and antibiotic resistances in light blue. Coding sequences present in the plasmids belonging to one of the eight colored annotation categories, but showing no homology by blast to pSE18-SA00377-1, were labeled but left in gray.

profiles can be found. However, no plasmid with a mapping coverage of >80% had been isolated from *S.* Agona isolates, with the highest mapping coverages all belonging to plasmids isolated either from *Enterobacter* or *Phytobacter* isolates. As high-similarity plasmids to pSE18-SA00377-1 seem to be found in a wide variety of genera and different geographical origins, it can be speculated that pSE18-SA00377-1 had been taken up from other sources, potentially in a clinical environment as a majority of the closely-related plasmids were isolated from clinical sources. Alternatively, the pSE18-SA00377-1 plasmid could have been taken up from environmental sources, for example when wastewater is re-used for irrigation in agriculture, as the SE18-SA00377 isolate was isolated from a dietary supplement consisting of plant-based ingredients. These hypotheses are also supported by the potential origins of the other constituent plasmids of the 18-SA00377 isolate. Firstly, the second largest plasmid, pSE18-SA00377-2 (94,574 bp), is a P1-like phage plasmid, carrying the p0111 plasmid replication gene, which was first identified from an enterohemorrhagic *E. coli* strain (Ogura et al., 2009) and is still frequently found in *E. coli* isolates, including clinical and food isolates (Balbuena-Alonso et al., 2022). Moreover, it carries two prophage-like elements pp1 and pp2, which were first identified in the core genome of *E. faecalis* isolates (Matos et al., 2013).

As an IncX3 plasmid and carrier of the bla_{TEM-1} gene, the second smallest plasmid, pSE18-SA00377-3 (50,931 bp), plays a role in the dissemination of carbapenemase resistance genes. The IncX plasmid family has been reported in a wide variety of *Enterobacteriaceae* from different sources (Guo et al., 2022) and thus the plasmid is a cause of concern due to its additional ARG load of *aac(3)-IIe*, *qnrS1*, and *floR*. Furthermore, pSE18-SA00377-3 also carries *tmrB*, a gene encoding the tunicamycin resistance protein which confers resistance to tunicamycin in *Bacillus subtilis* (Noda et al., 1992).

3.6. Plasmids with high similarity antibiotic resistance profiles to 18-SA00377

In order to limit the search to closely related isolates of *Salmonella* but also other members of the *Enterobacteriaceae* family, several NCBI

Pathogen Detection databases were queried for isolates with highly similar antibiotic resistance profiles (80% overlap in ARGs with 18-SA00377). This resulted in short-read sequences from the following four databases: *S. enterica*, *E. coli* and *Shigella*, *Klebsiella*, and *Citrobacter* (Table 4).

The nine *Salmonella* isolates of different serovars, geographical origins, and isolation sources (Table 1) were compared by mapping of contigs to p18-SA00377-1 and mapping results visualized (Figure 6).

Four isolates were isolated from animal sources, three isolates from clinical sources, and one isolate from a porcine food source. The animal samples were all isolated from the United States, but their isolation types range from domestic pigs (*Sus scrofa domesticus*) to wild boar (*Sus scrofa*) and cattle (*Bos taurus*). However, these environmental samples harboring these plasmids were either of serovar Agona or 4,12:i:-, the monophasic variant of *S.* Typhimurium.

However, based on available isolates' metadata, it was impossible to infer if the mapping against the pSE18-SA00377-1 occurred within their chromosomes or in constituent plasmids, due to draft character of the used short-read derived assemblies. Nevertheless, based on the numerous breaks in coverage of mapped sections and the large number of very small mapped sections (e.g., IS*26*), it might be that the ARGs are located in the isolates' chromosomes and not on a plasmid and the two described ARG clusters. Furthermore, multiple occurrences of these areas could become merged into repetitive regions during assembly of the short-read data.

In conclusion, this SE18-SA00377 isolate belongs to a sublineage of *S. enterica* serovar Agona that is multidrug-resistant and might be plant-associated. Along with its four plasmids, pSE18-SA00377-1, pSE18-SA00377-2, pSE18-SA00377-3, and pSE18-SA00377-4, the isolate carries a total of 23 different ARGs, conferring resistance to 12 different classes of antibiotics, with its largest plasmid of 295,499 kb in size, pSE18-SA00377-1, conferring the majority of them. Moreover, the pSE18-SA00377-1 plasmid is not only the main carrier of antibiotic resistance genes but also of heavy metal resistances. The structure of this plasmid is striking as its ARGs have accumulated in two distinct regions. This accumulation of ARGs as well as the presence of these clusters and a large part of its backbone in plasmids isolated from a wide range of genera, matrices, years of isolation and geographical origins suggest that this plasmid has a complex history with numerous transmission events.

Further analysis of plasmids from human, veterinary, and environmental sources may provide further insights into the evolution of this plasmid. In particular, due to the highly drug-resistant nature of this plasmid, identifying potential reservoirs of multidrug-resistant isolates is crucial, as they have the capacity to disseminate antibiotic and metal resistance genes.

Here, we present an in-depth characterization of a multidrug-resistant *S.* Agona, isolated from dietary supplements in 2018. Its phylogeny to other *S.* Agona isolates from Germany was established and supplemented with available sequences of *S.* Agona that have been reported globally and are available in the NCBI database. Detailed annotation of its largest constituent plasmid included antimicrobial resistance genes on mobile genetic elements. Closely related plasmids were queried through a two-pronged approach: MOB-typing and taxonomic classification of plasmids. Lastly, structural comparisons with high-similarity plasmids revealed a composite plasmid structure, found in isolates from numerous other genera, geographic origins and isolation matrices. These analyses showed that this plasmid is a potential reservoir for antimicrobial and heavy metal resistance determinants and has the potential to adapt to various hosts and environments. Thus, highlighting the need for continued surveillance to prevent future outbreaks.

TABLE 4 Table showing the overview of isolates resulting from querying the respective NCBI Pathogen Detection databases with custom R script.

Isolates' species	Countries of origin [a)]	Years of collection	Isolation source [a)]	Median mapped coverage (%)	Assembly levels	Median % of matching AMR genes	NCBI Pathogen detection database
Salmonella enterica	USA (7), Germany (1), NA (1)	2014–2019	environmental/other (5), clinical (3), NA (1)	92	All contigs	83	*Salmonella enterica*
Klebsiella ssp.	Montenegro (4), Canada (4), USA (3), Romania/Taiwan/France/Mexico/Australia (all 1)	2001–2021	Clinical (13), environmental/other (2)	91	All contigs	61	*Klebsiella pneumoniae*
Escherichia coli	USA (11), Russia (3), Germany/France (2), Australia/Canada/Chile/China/Czech Republic/Rwanda (all 1)	2007–2021	Clinical (11), environmental/other (7)	90	Contigs (19), scaffolds (5)	61	*E.coli* and *Shigella*
Citrobacter ssp.	USA (4), Australia (3), Canada (2), United Kingdom/China (both 1)	2010–2021	Clinical (11), environmental/other (1)	89	All contigs	61	*Citrobacter freundii*

[a)] Numbers in brackets indicate the number of occurrences, NA, data not available.

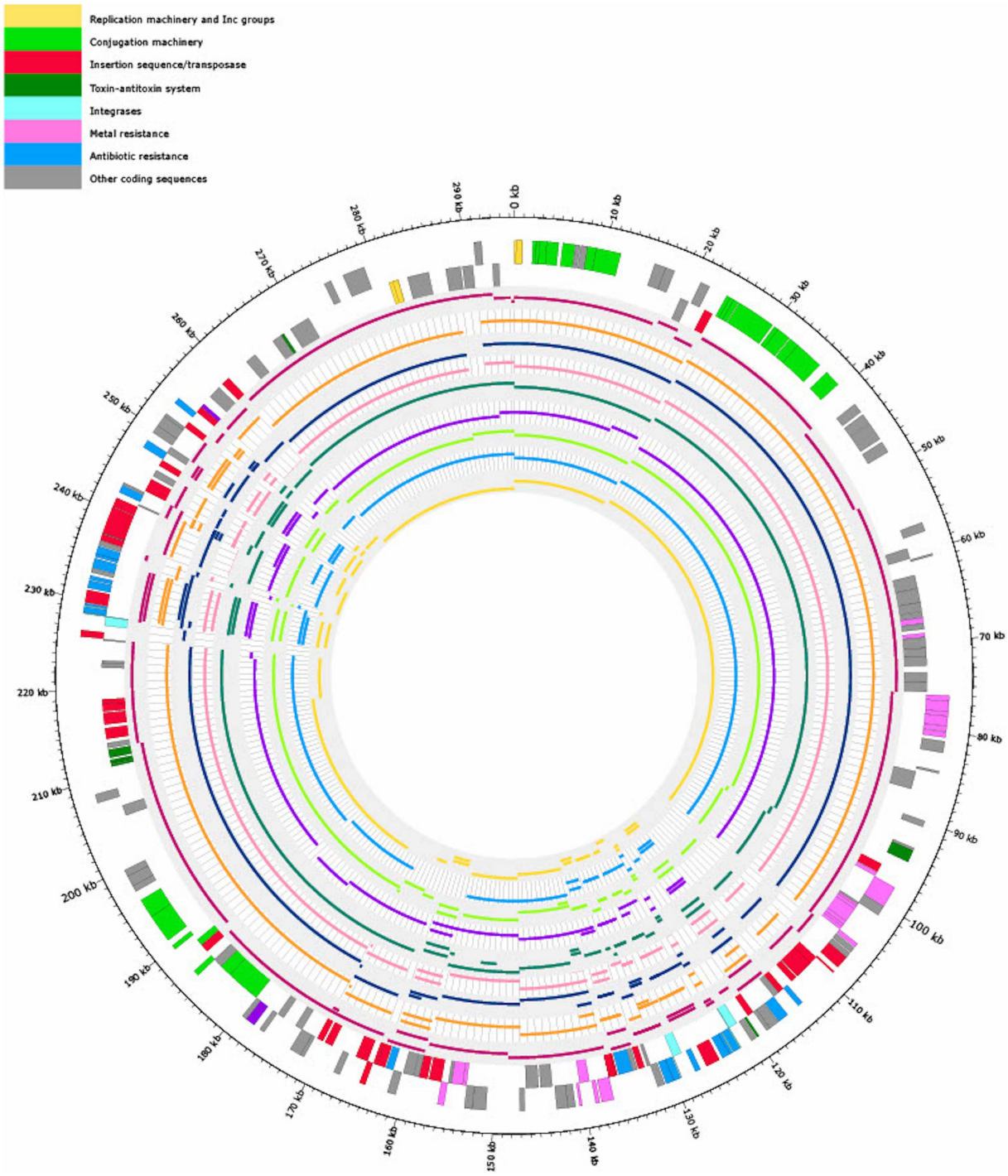

FIGURE 6
Visualization of mapping the contigs of nine *Salmonella* isolates (Table 1) with similar antibiotic resistance profiles (80% overlap of ARGs with 18-SA00377) to pSE18-SA00377-1. Regions of the nine *Salmonella* isolates that did not map to pSE18-SA00377-1 are not shown. The outermost black circle designates the base positions around the plasmid. Farthest two tracks on the outside represents the pSE18-SA00377-1 plasmid annotations, with replication machinery and incompatibility groups colored in yellow, conjugation machinery in light green, insertion sequences and transposases in red, toxin-antitoxin system in dark green, integrases in turquoise, heavy metal resistances in pink, and antibiotic resistances in light blue. Coding sequences for other gene products are colored in gray. For labeling of annotations refer to Figure 2.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

LB: Formal analysis, Investigation, Methodology, Visualization, Writing – review & editing. MB: Formal analysis, Investigation, Methodology, Visualization, Writing – review & editing. CD: Data curation, Formal analysis, Investigation, Software, Writing – review & editing. JG: Conceptualization, Supervision, Writing – review & editing, Software. J-AH: Methodology, Writing – review & editing. BM: Project administration, Resources, Supervision, Writing – review & editing, Funding acquisition. IS: Project administration, Resources, Writing – review & editing. TA: Project administration, Supervision, Writing – review & editing. KN: Resources, Writing – review & editing. JF: Conceptualization, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing, Data curation, Formal analysis, Funding acquisition, Visualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the German Federal Institute for Risk Assessment, Grant no. 60_0103_01.P543. LB received funding by the FARMED project, which is part of the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement no. 773830: One Health European Joint Program.

Acknowledgments

We would like to express our thanks to the whole team at the NRL for *Salmonella* for serotyping and at the NRL for Antimicrobial Resistance for MIC testing, as well as to Beatrice Baumann, Katharina Thomas, and Angelina Groger for their tireless support in the laboratory, in particular for sequencing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1284929/full#supplementary-material

References

Abraham, S., O'Dea, M., Trott, D. J., Abraham, R. J., Hughes, D., Pang, S., et al. (2016). Isolation and plasmid characterization of carbapenemase (IMP-4) producing *Salmonella enterica* typhimurium from cats. *Sci. Rep.* 6:35527. doi: 10.1038/srep35527

Akhter, S., Aziz, R. K., and Edwards, R. A. (2012). PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res.* 40:e126. doi: 10.1093/nar/gks406

Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., et al. (2019). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 48, D517–D525. doi: 10.1093/nar/gkz935

Ao, T. T., Feasey, N. A., Gordon, M. A., Keddy, K. H., Angulo, F. J., and Crump, J. A. (2015). Global burden of invasive nontyphoidal Salmonella disease, 2010(1). *Emerg. Infect. Dis.* 21, 941–949. doi: 10.3201/eid2106.140999

Balbuena-Alonso, M. G., Cortés-Cortés, G., Kim, J. W., Lozano-Zarain, P., Camps, M., and del Carmen Rocha-Gracia, R. (2022). Genomic analysis of plasmid content in food isolates of *E. coli* strongly supports its role as a reservoir for the horizontal transfer of virulence and antibiotic resistance genes. *Plasmid* 123-124:102650. doi: 10.1016/j.plasmid.2022.102650

Barton, B. M., Harding, G. P., and Zuccarelli, A. J. (1995). A general method for detecting and sizing large plasmids. *Anal. Biochem.* 226, 235–240. doi: 10.1006/abio.1995.1220

Bauer, A. P., Dieckmann, S. M., Ludwig, W., and Schleifer, K.-H. (2007). Rapid identification of *Escherichia coli* safety and laboratory strain lineages based on multiplex-PCR. *FEMS Microbiol. Lett.* 269, 36–40. doi: 10.1111/j.1574-6968.2006.00594.x

Blanc-Potard, A. B., Solomon, F., Kayser, J., and Groisman, E. A. (1999). The SPI-3 pathogenicity island of *Salmonella enterica*. *J. Bacteriol.* 181, 998–1004. doi: 10.1128/jb.181.3.998-1004.1999

Borowiak, M., Fischer, J., Hammerl, J. A., Hendriksen, R. S., Szabo, I., and Malorny, B. (2017). Identification of a novel transposon-associated phosphoethanolamine transferase gene, *mcr-5*, conferring colistin resistance in d-tartrate fermenting *Salmonella enterica* subsp. *enterica* serovar Paratyphi B. *J. Antimicrob. Chemother.* 72, 3317–3324. doi: 10.1093/jac/dkx327

Brouard, C., Espie, E., Weill, F. X., Kerouanton, A., Brisabois, A., Forgue, A. M., et al. (2007). Two consecutive large outbreaks of *Salmonella enterica* serotype Agona infections in infants linked to the consumption of powdered infant formula. *Pediatr. Infect. Dis. J.* 26, 148–152. doi: 10.1097/01.inf.0000253219.06258.23

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinform.* 10:421. doi: 10.1186/1471-2105-10-421

Carattoli, A. (2003). Plasmid-mediated antimicrobial resistance in *Salmonella enterica*. *Curr. Issues Mol. Biol.* 5, 113–122. doi: 10.21775/cimb.005.113

Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., et al. (2014). *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* 58, 3895–3903. doi: 10.1128/AAC.02412-14

Chen, L., Zheng, D., Liu, B., Yang, J., and Jin, Q. (2015). VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res.* 44, D694–D697. doi: 10.1093/nar/gkv1239

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Clark, G. M., Kaufmann, A. F., Gangarosa, E. J., and Thompson, M. A. (1973). Epidemiology of an international outbreak of *Salmonella Agona*. *Lancet* 2, 490–493. doi: 10.1016/s0140-6736(73)92082-5

Cummins, M. L., Sanderson-Smith, M., Newton, P., Carlile, N., Phalen, D. N., Maute, K., et al. (2020). Whole-genome sequence analysis of an extensively drug-resistant *Salmonella enterica* serovar Agona isolate from an Australian silver Gull (*Chroicocephalus novaehollandiae*) reveals the acquisition of multidrug resistance plasmids. *mSphere* 5:e00743-20. doi: 10.1128/mSphere.00743-20

Dangel, A., Berger, A., Messelhäußer, U., Konrad, R., Hörmansdorfer, S., Ackermann, N., et al. (2019). Genetic diversity and delineation of *Salmonella* Agona outbreak strains by next generation sequencing, Bavaria, Germany, 1993 to 2018. *Euro Surveill.* 24:1800303. doi: 10.2807/1560-7917.Es.2019.24.18.1800303

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34, 2666–2669. doi: 10.1093/bioinformatics/bty149

Deneke, C., Brendebach, H., Uelze, L., Borowiak, M., Malorny, B., and Tausch, S. H. (2021a). Species-specific quality control, assembly and contamination detection in microbial isolate sequences with AQUAMIS. *Genes* 12:644. doi: 10.3390/genes12050644

Deneke, C., Uelze, L., Brendebach, H., Tausch, S. H., and Malorny, B. (2021b). Decentralized investigation of bacterial outbreaks based on hashed cgMLST. *Front. Microbiol.* 12:649517. doi: 10.3389/fmicb.2021.649517

Doublet, B., Praud, K., Weill, F. X., and Cloeckaert, A. (2009). Association of IS*26*-composite transposons and complex In4-type integrons generates novel multidrug resistance loci in *Salmonella* genomic island 1. *J. Antimicrob. Chemother.* 63, 282–289. doi: 10.1093/jac/dkn500

European Food Safety Authority and European Centre for Disease Prevention and Control (2022a). The European union one health 2021 Zoonoses report. *EFSA J.* 20:e07666. doi: 10.2903/j.efsa.2022.7666

European Food Safety Authority and European Centre for Disease Prevention and Control (2022b). The European union summary report on antimicrobial resistance in zoonotic and indicator bacteria from humans, animals and food in 2019–2020. *EFSA J.* 20:e07209. doi: 10.2903/j.efsa.2022.7209

European Food Safety Authority and European Centre for Disease Prevention and Control (2023). The European union summary report on antimicrobial resistance in zoonotic and indicator bacteria from humans, animals and food in 2020/2021. *EFSA J.* 21:e07867. doi: 10.2903/j.efsa.2023.7867

Feldgarden, M., Brover, V., Gonzalez-Escalona, N., Frye, J. G., Haendiges, J., Haft, D. H., et al. (2021). AMRFinderPlus and the reference gene catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci. Rep.* 11:12728. doi: 10.1038/s41598-021-91456-0

Fernández, J., Guerra, B., and Rodicio, M. R. (2018). Resistance to Carbapenems in non-Typhoidal *Salmonella enterica* Serovars from humans, animals and food. *Vet. Sci.* 5:40. doi: 10.3390/vetsci5020040

Gerlach, R. G., Jäckel, D., Geymeier, N., and Hensel, M. (2007). *Salmonella* pathogenicity island 4-mediated adhesion is coregulated with invasion genes in *Salmonella enterica*. *Infect. Immun.* 75, 4697–4709. doi: 10.1128/iai.00228-07

Grimont, P. A. D., and Weill, F.. (2007). *Antigenic formulae of the Salmonella serovars*. WHO Collaborating Centre for Reference and Research on Salmonella. Paris

Guo, X., Chen, R., Wang, Q., Li, C., Ge, H., Qiao, J., et al. (2022). Global prevalence, characteristics, and future prospects of IncX3 plasmids: a review. *Front. Microbiol.* 13:979558. doi: 10.3389/fmicb.2022.979558

Hadziabdic, S., Fischer, J., Malorny, B., Borowiak, M., Guerra, B., Kaesbohrer, A., et al. (2018). In vivo transfer and microevolution of avian native IncA/C(2)Bla(NDM-1)-carrying plasmid pRH-1238 during a broiler chicken infection study. *Antimicrob. Agents Chemother.* 62:e02128-17. doi: 10.1128/AAC.02128-17

Ilyas, B., Tsai, C. N., and Coombes, B. K. (2017). Evolution of *Salmonella*-host cell interactions through a dynamic bacterial genome. *Front. Cell. Infect. Microbiol.* 7:428. doi: 10.3389/fcimb.2017.00428

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114. doi: 10.1038/s41467-018-07641-9

Jennings, E., Thurston, T. L. M., and Holden, D. W. (2017). *Salmonella* SPI-2 type III secretion system effectors: molecular mechanisms and physiological consequences. *Cell Host Microbe* 22, 217–231. doi: 10.1016/j.chom.2017.07.009

Johansson, M. H. K., Bortolaia, V., Tansirichaiya, S., Aarestrup, F. M., Roberts, A. P., and Petersen, T. N. (2020). Detection of mobile genetic elements associated with antibiotic resistance in *Salmonella enterica* using a newly developed web tool: MobileElementFinder. *J. Antimicrob. Chemother.* 76, 101–109. doi: 10.1093/jac/dkaa390

Jourdan-da Silva, N., Fabre, L., Robinson, E., Fournet, N., Nisavanh, A., Bruyand, M., et al. (2018). Ongoing nationwide outbreak of *Salmonella* Agona associated with internationally distributed infant milk products, France, December 2017. *Euro Surveill.* 23:17-00852. doi: 10.2807/1560-7917.ES.2018.23.2.17-00852

Juraschek, K., Borowiak, M., Tausch, S. H., Malorny, B., Kasbohrer, A., Otani, S., et al. (2021). Outcome of different sequencing and assembly approaches on the detection of plasmids and localization of antimicrobial resistance genes in commensal *Escherichia coli*. *Microorganisms* 9:598. doi: 10.3390/microorganisms9030598

Killalea, D., Ward, L. R., Roberts, D., de Louvois, J., Sufi, F., Stuart, J. M., et al. (1996). International epidemiological and microbiological study of outbreak of *Salmonella agona* infection from a ready to eat savoury snack--I: England and Wales and the United States. *BMJ* 313, 1105–1107. doi: 10.1136/bmj.313.7065.1105

Koch, J., Schrauder, A., Alpers, K., Werber, D., Frank, C., Prager, R., et al. (2005). *Salmonella* Agona outbreak from contaminated aniseed, Germany. *Emerg. Infect. Dis.* 11, 1124–1127. doi: 10.3201/eid1107.041022

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Li, I. C., Yu, G. Y., Huang, J. F., Chen, Z. W., and Chou, C. H. (2022). Comparison of reference-based assembly and *De novo* assembly for bacterial plasmid reconstruction and AMR gene localization in *Salmonella enterica* Serovar Schwarzengrund isolates. *Microorganisms* 10:227. doi: 10.3390/microorganisms10020227

Lordan, R. (2021). Dietary supplements and nutraceuticals market growth during the coronavirus pandemic - implications for consumers and regulatory oversight. *PharmaNutrition* 18:100282. doi: 10.1016/j.phanu.2021.100282

Lou, L., Zhang, P., Piao, R., and Wang, Y. (2019). *Salmonella* Pathogenicity Island 1 (SPI-1) and its complex regulatory network. *Front. Cell. Infect. Microbiol.* 9:270. doi: 10.3389/fcimb.2019.00270

Marcus, S. L., Brumell, J. H., Pfeifer, C. G., and Finlay, B. B. (2000). *Salmonella* pathogenicity islands: big virulence in small packages. *Microbes Infect.* 2, 145–156. doi: 10.1016/S1286-4579(00)00273-2

Matos, R. C., Lapaque, N., Rigottier-Gois, L., Debarbieux, L., Meylheuc, T., Gonzalez-Zorn, B., et al. (2013). *Enterococcus faecalis* prophage dynamics and contributions to pathogenic traits. *PLoS Genet.* 9:e1003539. doi: 10.1371/journal.pgen.1003539

Mba-Jonas, A., Culpepper, W., Hill, T., Cantu, V., Loera, J., Borders, J., et al. (2018). A multistate outbreak of human *Salmonella* Agona infections associated with consumption of fresh, whole papayas imported from Mexico-United States, 2011. *Clin. Infect. Dis.* 66, 1756–1761. doi: 10.1093/cid/cix1094

Meinen, A., Simon, S., Banerji, S., Szabo, I., Malorny, B., Borowiak, M., et al. (2019). Salmonellosis outbreak with novel *Salmonella enterica* subspecies *enterica* serotype (11,z41,e,n,z15) attributable to sesame products in five European countries, 2016 to 2017. *Euro Surveill.* 24:1800543. doi: 10.2807/1560-7917.Es.2019.24.36.1800543

Ng, L. K., Martin, I., Alfa, M., and Mulvey, M. (2001). Multiplex PCR for the detection of tetracycline resistant genes. *Mol. Cell. Probes* 15, 209–215. doi: 10.1006/mcpr.2001.0363

Nicolay, N., Thornton, L., Cotter, S., Garvey, P., Bannon, O., McKeown, P., et al. (2011). *Salmonella enterica* serovar Agona European outbreak associated with a food company. *Epidemiol. Infect.* 139, 1272–1280. doi: 10.1017/S0950268810002360

Noda, Y., Yoda, K., Takatsuki, A., and Yamasaki, M. (1992). TmrB protein, responsible for tunicamycin resistance of *Bacillus subtilis*, is a novel ATP-binding membrane protein. *J. Bacteriol.* 174, 4302–4307. doi: 10.1128/jb.174.13.4302-4307.1992

Ogura, Y., Ooka, T., Iguchi, A., Toh, H., Asadulghani, M., Oshima, K., et al. (2009). Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 106, 17939–17944. doi: 10.1073/pnas.0903585106

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2015). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189

Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., et al. (2016). Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 17:132. doi: 10.1186/s13059-016-0997-x

Parkhill, J., Dougan, G., James, K. D., Thomson, N. R., Pickard, D., Wain, J., et al. (2001). Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* 413, 848–852. doi: 10.1038/35101607

Robertson, J., and Nash, J. H. E. (2018). MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb. Genom.* 4:e000206. doi: 10.1099/mgen.0.000206

Roer, L., Hendriksen, R. S., Leekitcharoenphon, P., Lukjancenko, O., Kaas, R. S., Hasman, H., et al. (2016). Is the evolution of *Salmonella enterica* subsp. *enterica* linked to restriction-modification systems? *mSystems* 1:e00009-16. doi: 10.1128/mSystems.00009-16

Rodríguez, I. W., Barownick, R., Helmuth, M. C., Mendoza, M. R., Rodicio, A., Schroeter, B., et al. (2009). Extended-spectrum β-lactamases and AmpC β-lactamases in ceftiofur-resistant *Salmonella enterica* isolates from food and livestock obtained in Germany during 2003-07. *J. Antimicrob. Chemother.* 64, 301–309. doi: 10.1093/jac/dkp195

Rozwandowicz, M., Brouwer, M. S. M., Fischer, J., Wagenaar, J. A., Gonzalez-Zorn, B., Guerra, B., et al. (2018). Plasmids carrying antimicrobial resistance genes in *Enterobacteriaceae*. *J. Antimicrob. Chemother.* 73, 1121–1137. doi: 10.1093/jac/dkx488

Russo, E. T., Biggerstaff, G., Hoekstra, R. M., Meyer, S., Patel, N., Miller, B., et al. (2013). A recurrent, multistate outbreak of *Salmonella* serotype agona infections associated with dry, unsweetened cereal consumption, United States, 2008. *J. Food Prot.* 76, 227–230. doi: 10.4315/0362-028X.JFP-12-209

Schwengers, O. (2020). *Platon database*. Geneva, Switzerland: Zenodo.

Schwengers, O. (2021). *Bakta database*. Geneva, Switzerland: Zenodo.

Schwengers, O., Barth, P., Falgenhauer, L., Hain, T., Chakraborty, T., and Goesmann, A. (2020). Platon: identification and characterization of bacterial plasmid contigs in short-read draft assemblies exploiting protein sequence-based replicon distribution scores. *Microb. Genom.* 6:mgen000398. doi: 10.1099/mgen.0.000398

Schwengers, O., Jelonek, L., Dieckmann, M. A., Beyvers, S., Blom, J., and Goesmann, A. (2021). Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb. Genom.* 7:000685. doi: 10.1099/mgen.0.000685

Shohat, T., Green, M. S., Merom, D., Gill, O. N., Reisfeld, A., Matas, A., et al. (1996). International epidemiological and microbiological study of outbreak of *Salmonella* Agona infection from a ready to eat savoury snack--II: Israel. *BMJ* 313, 1107–1109. doi: 10.1136/bmj.313.7065.1107

Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34, D32–D36. doi: 10.1093/nar/gkj014

Simon, S., Trost, E., Bender, J., Fuchs, S., Malorny, B., Rabsch, W., et al. (2018). Evaluation of WGS based approaches for investigating a food-borne outbreak caused by *Salmonella enterica* serovar Derby in Germany. *Food Microbiol.* 71, 46–54. doi: 10.1016/j.fm.2017.08.017

Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinform* 27, 1009–1010. doi: 10.1093/bioinformatics/btr039

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44, 6614–6624. doi: 10.1093/nar/gkw569

Taylor, J. P., Barnett, B. J., del Rosario, L., Williams, K., and Barth, S. S. (1998). Prospective investigation of cryptic outbreaks of *Salmonella agona* salmonellosis. *J. Clin. Microbiol.* 36, 2861–2864. doi: 10.1128/JCM.36.10.2861-2864.1998

Technical University of Denmark. (2022). *Annual report on Zoonoses in Denmark 2021*. Denmark, Technical University of Denmark.

The European Committee on Antimicrobial Susceptibility Testing. (2021). *Breakpoint tables for interpretation of MICs and zone diameters*. EUCAST. Denmark

Uelze, L., Becker, N., Borowiak, M., Busch, U., Dangel, A., Deneke, C., et al. (2021). Toward an integrated genome-based surveillance of *Salmonella enterica* in Germany. *Front. Microbiol.* 12:626941. doi: 10.3389/fmicb.2021.626941

van Asten, A. J. A. M., and van Dijk, J. E. (2005). Distribution of "classic" virulence factors among *Salmonella* spp. *FEMS Immunol. Med. Microbiol.* 44, 251–259. doi: 10.1016/j.femsim.2005.02.002

Velásquez, J. C., Hidalgo, A. A., Villagra, N., Santiviago, C. A., Mora, G., and Fuentes, J. A. (2016). SPI-9 of *Salmonella enterica* serovar Typhi is constituted by an operon positively regulated by RpoS and contributes to adherence to epithelial cells in culture. *Microbiology* 162, 1367–1378. doi: 10.1099/mic.0.000319

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 13:e1005595. doi: 10.1371/journal.pcbi.1005595

Yoshida, C. E., Kruczkiewicz, C. R., Laing, E. J., Lingohr, V. P. J., Gannon, V. P., Nash, J. H., et al. (2016). The *Salmonella in silico* typing resource (SISTR): an open web-accessible tool for rapidly typing and subtyping draft *Salmonella* genome assemblies. *PLoS One* 11:e0147101. doi: 10.1371/journal.pone.0147101

Zhang, L. J., Gu, X. X., Zhang, J., Yang, L., Lu, Y. W., Fang, L. X., et al. (2020). Characterization of a *fosA3* carrying IncC-IncN plasmid from a multidrug-resistant ST17 *Salmonella* Indiana isolate. *Front. Microbiol.* 11:1582. doi: 10.3389/fmicb.2020.01582

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 15 November 2023
DOI 10.3389/fmicb.2023.1261261

OPEN ACCESS

EDITED BY
Benjamin Andrew Evans,
University of East Anglia, United Kingdom

REVIEWED BY
Fupin Hu,
Fudan University, China
Shangshang Qin,
Zhengzhou University, China

*CORRESPONDENCE
Jiachang Cai
✉ caijiachang@zju.edu.cn

†These authors have contributed equally to this work

RECEIVED 19 July 2023
ACCEPTED 01 November 2023
PUBLISHED 15 November 2023

CITATION
Wang L, Shen W and Cai J (2023) Mobilization of the *bla*$_{KPC-14}$ gene among heterogenous plasmids in extensively drug-resistant hypervirulent *Klebsiella pneumoniae*. *Front. Microbiol.* 14:1261261. doi: 10.3389/fmicb.2023.1261261

Mobilization of the *bla*$_{KPC-14}$ gene among heterogenous plasmids in extensively drug-resistant hypervirulent *Klebsiella pneumoniae*

Lin Wang†, Weiyi Shen† and Jiachang Cai*

Clinical Microbiology Laboratory, The Second Affiliated Hospital of Zhejiang University School of Medicine, Zhejiang University, Hangzhou, China

Introduction: Ceftazidime/avibactam (CZA) is an effective alternative for the treatment of infections caused by KPC-producing carbapenem-resistant *Klebsiella pneumoniae* (CRKP). However, KPC variants with CZA resistance have been observed in clinical isolates, further limiting the treatment options of clinical use.

Methods: In this study, we isolated three KPC-14-producing CRKP from two patients in intensive care units without CZA therapy. The antimicrobial susceptibility was determined using the broth microdilution method. Three CRKP were subjected to whole-genome sequencing to analyze the phylogenetic relatedness and the carriage of antimicrobial resistance genes and virulence factors. Long-read sequencing was also performed to obtain the complete sequences of the plasmids. The horizontal transfer of the *bla*$_{KPC-14}$ gene was evaluated by conjugation experiments.

Results: Three CRKP displayed resistance or reduced susceptibility to ceftazidime/avibactam, colistin, and tigecycline. Single-nucleotide polymorphism (SNP) analysis demonstrated the close phylogenetic distance between these strains. A highly similar IncFII/IncR plasmid encoding *bla*$_{KPC-14}$ was shared by three CRKP, with *bla*$_{KPC-14}$ located in an NTE$_{KPC}$-Ib element with the core region of IS*Kpn27*-*bla*$_{KPC-14}$-IS*Kpn6*. This structure containing *bla*$_{KPC-14}$ was also observed in another *tet*(A)-carrying plasmid that belonged to an unknown Inc-type in two out of three isolates. The horizontal transferability of these integrated plasmids to Escherichia coli EC600 was confirmed by the cotransmission of *tet*(A) and *bla*$_{KPC-14}$ genes, but the single transfer of *bla*$_{KPC-14}$ on the IncFII/IncR plasmid failed. Three CRKP expressed yersiniabactin and carried a hypervirulence plasmid encoding *rmpA2* and aerobactin-related genes, and were thus classified as carbapenem-resistant hypervirulent *K. pneumoniae* (hvKP).

Discussion: In this study, we reported the evolution of a mosaic plasmid encoding the *bla*$_{KPC-14}$ gene via mobile elements in extensively drug-resistant hvKP. The *bla*$_{KPC-14}$ gene is prone to integrate into other conjugative plasmids via the NTE$_{KPC}$-Ib element, further facilitating the spread of ceftazidime/avibactam resistance.

KEYWORDS

***bla*$_{KPC-14}$ gene, ceftazidime/avibactam, CR-hvKp, *Klebsiella pneumoniae*, gene transfer**

1. Introduction

The widespread of carbapenem-resistant *Klebsiella pneumoniae* (CRKP) is considered an urgent threat to public health, as it complicates patient care and increases morbidity and mortality in cases of infection (Wang M. et al., 2022; Pérez-Galera et al., 2023). Available data from the China Antimicrobial Surveillance Network (CHINET[1]) showed that the prevalence of CRKP has rapidly increased in China, from 2.9% in 2005 to 24.2% in 2022. Colistin and tigecycline constitute some of the last resorts for the treatment of CRKP infections, however, resistance to these antibiotics in CRKP strains has also been reported recently, further reducing the repertoire of useful antibiotics (Chen et al., 2021; Tian et al., 2021).

Ceftazidime/avibactam (CZA), a novel β-lactam/β-lactamase inhibitor combination, is an effective alternative for the treatment of CRKP infections (Van Duin and Bonomo, 2016). This combination shields ceftazidime from breakdown by Ambler class A, class C, and some class D β-lactamases and thus exhibits potent inhibition of strains producing KPC and OXA-48-like carbapenemases (Criscuolo and Trecarichi, 2020). KPC-producing CRKP is widespread globally and is the predominant type of CRKP in China, which is frequently related to nosocomial outbreaks (Findlay et al., 2021; Wang L. et al., 2022; Wang M. et al., 2022). Although recent studies have shown evidence for CZA as a promising option for such infections, resistance to this antibiotic has rapidly evolved, mainly due to the production of variants of KPC-2 or KPC-3 enzymes (Humphries and Hemarajata, 2017; El-Kady et al., 2022). The single amino acid substitution that confers CZA resistance was commonly encountered in the omega loop (positions 164–179), particularly for the Asp179Tyr (D179Y) mutation in KPC-3 (KPC-31) and KPC-2 (KPC-33) (Livermore et al., 2015; Barnes et al., 2017). Additionally, KPC variants with CZA-resistance mediated by amino acid changes outside the omega loop region (e.g., KPC-41, KPC-23, KPC-14, KPC-8, KPC-123, and KPC-93) were also observed in the clinical isolates from patients following CZA therapy and those who were not treated with CZA (Bianco et al., 2021; Liu et al., 2022b; Wang L. et al., 2022). KPC-14, the variant with a deletion of two amino acids (Δ242-GT-243) of KPC-2 that exhibits CZA resistance, has been sporadically detected in clinical isolates of CRKP (Bianco et al., 2020; Niu et al., 2020; Linh et al., 2021; Jiang et al., 2022).

Here, we investigated the genetic relationship of CRKP harboring two structurally distinct bla_{KPC-14}-encoding plasmids and analyzed the evolution of one plasmid that was able to undergo horizontal transfer between *Enterobacterales*.

2. Materials and methods

2.1. Patients and bacterial strains

Three CRKP (strains SP1023 and F1025 from Patient A, and strain SP1030 from Patient B) were isolated from two patients admitted to the neurology intensive care unit (NICU) of a tertiary hospital in Hangzhou City in 2022. Species identification was determined by MALDI-TOF MS (Bruker Daltonik GmbH, Bremen, Germany). This study was approved by the Ethics Committee of The Second Affiliated Hospital of Zhejiang University School of Medicine.

Patient A, a 38-year-old male, had poorly controlled hypertension for several years. He underwent three consecutive intracranial hematoma evacuations at a local hospital due to cerebellar hemorrhage. Blood cultures revealed carbapenem-resistant *Acinetobacter baumannii* (CRAB) and a combination of cefoperazone/sulbactam (1,1, 2 g IV every 6 h) and polymyxin B (750,000 IU IV every 12 h) was administered (Figure 1A). The patient was in a coma and was transferred to the NICU of our hospital for further treatment. Upon admission, the patient received supportive treatments, including mechanical ventilation, blood transfusion, fluid replacement, and appropriate medications. The original antimicrobial therapy regimen was continued. The subsequent sputum culture revealed the growth of *Klebsiella aerogenes* and cefoperazone/sulbactam- and carbapenem-resistant *A. baumannii*. Therefore, cefoperazone/sulbactam was replaced with tigecycline (100 mg IV every 12 h) on Day 7. Nine days later (Day 16), both organisms were cleared. However, CRKP (strain SP1023) exhibiting ceftazidime/avibactam and polymyxin B resistance was isolated from the sputum sample (Figure 1A). On Day 18, fecal screening for CRE yielded pandrug-resistant *K. pneumoniae* (strain F1025), which was resistant to ceftazidime/avibactam, polymyxin B, and tigecycline (Figure 1A). In addition to CRKP, *Pseudomonas aeruginosa* and carbapenem-resistant *A. baumannii* (CRAB) were detected in the sputum sample on Day 32. Meropenem was used (1 g IV every 12 h) instead of the previous antimicrobials, and only CRKP was cleared after 9 days of treatment. Tigecycline was used again, and amikacin (400 mg nasogastric feeding every 12 h) was added to the treatment regimen 5 days later. However, *P. aeruginosa* that developed carbapenem resistance and CRAB persisted in the patient's respiratory tract. Due to his extremely poor condition and acute exacerbation of chronic renal failure, the patient died of multiple organ failure on Day 68.

Patient B, a 40-year-old male with spontaneous intracerebral hemorrhage, underwent surgical evacuation of the intracranial hematoma at a local hospital. Six days later, he remained in a coma and was transferred to the NICU of our hospital. The CT scans showed postoperative hematoma in the surgical area and scattered infiltrates in both lungs. The patient received supportive treatments to reduce intracranial pressure, sedation, pain management, enteral nutrition, endotracheal intubation, and mechanical ventilation. Additionally, piperacillin/tazobactam (8:1, 4.5 g IV every 6 h) was administered for antimicrobial therapy throughout the hospitalization period (Figure 1B). On the second day of admission, CRAB was isolated from the sputum sample and persisted until discharge. On Day 17 and Day 19, CRKP (strain SP1030) was detected in the sputum culture that was resistant to ceftazidime/avibactam and colistin (Figure 1B). After 8 days of mechanical ventilation, the patient was weaned off the ventilator (Day 15). Since the patient's vital signs stabilized and his mental function recovered, he was discharged for further treatment at a rehabilitation hospital on Day 20.

2.2. Antimicrobial susceptibility tests

The minimal inhibitory concentrations (MICs) of 18 antimicrobial agents, including imipenem, meropenem, ertapenem, ceftazidime/avibactam, ceftazidime, cefotaxime, cefepime, piperacillin/tazobactam, cefoperazone/sulbactam, cefmetazole, aztreonam, ciprofloxacin, amikacin, chloramphenicol, fosfomycin, tetracycline, tigecycline, and

1 http://www.chinets.com/

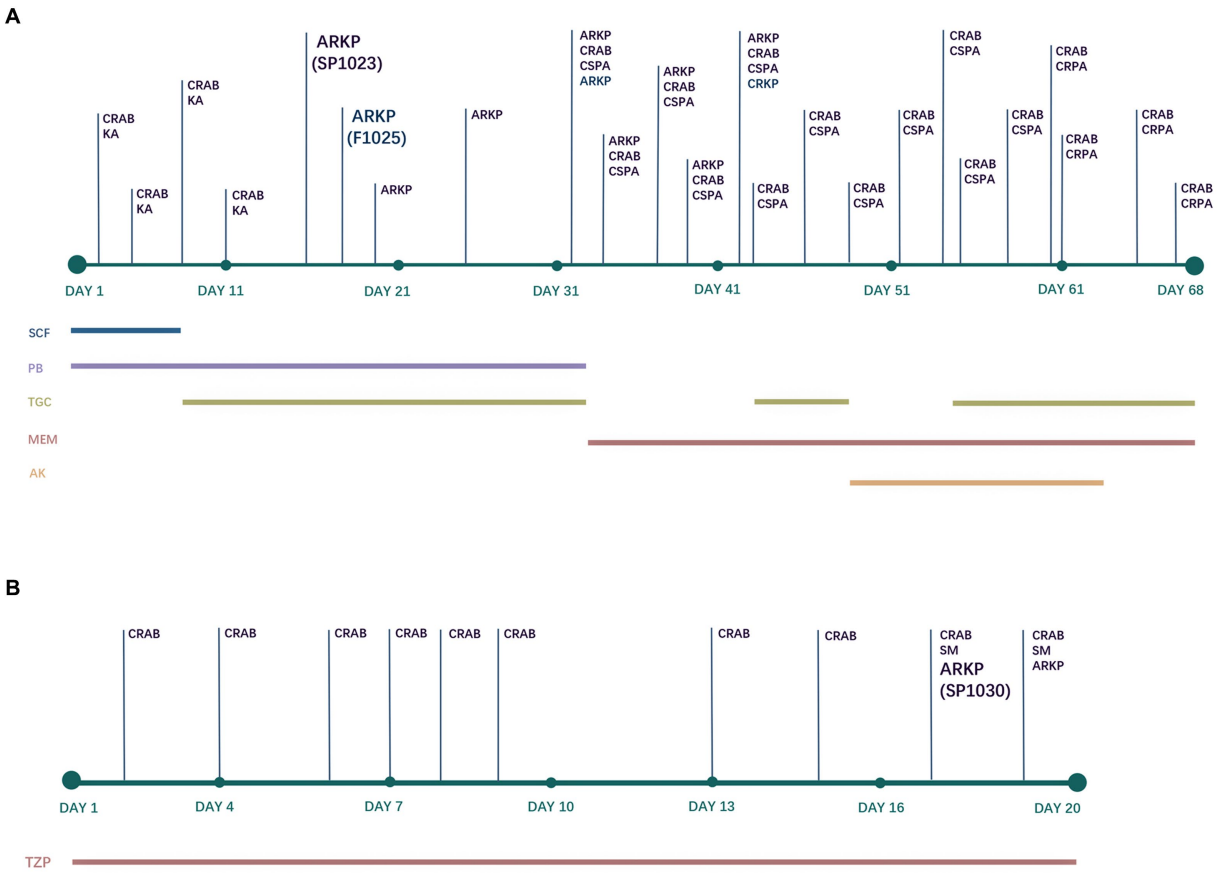

FIGURE 1
Antimicrobial treatments and bacterial isolation of Patient A **(A)** and Patient B **(B)**. CRAB, carbapenem-resistant *Acinetobacter baumannii*; KA, *Klebsiella aerogenes*; ARKP, ceftazidime/avibactam-resistant *Klebsiella pneumoniae*; CSPA, carbapenem-susceptible *Pseudomonas aeruginosa*; CRPA, carbapenem-resistant *Pseudomonas aeruginosa*; CRKP, carbapenem-resistant *Klebsiella pneumoniae* (ceftazidime/avibactam-susceptible); SM, *Stenotrophomonas maltophilia*; SCF, cefoperazone/sulbactam; PB, polymyxin B; TGC, tigecycline; MEM, meropenem; AK, amikacin; TZP, piperacillin/tazobactam.

colistin, were determined using the broth microdilution method (Clinical and Laboratory Standards Institute, 2018) and interpreted according to the Clinical and Laboratory Standards Institute (CLSI) guidelines (Clinical and Laboratory Standards Institute, 2021). Tigecycline susceptibility was interpreted using breakpoints recommended by the US Food and Drug Administration.[2] *Escherichia coli* ATCC 25922, *K. pneumoniae* 700603, and *Pseudomonas aeruginosa* ATCC 27583 were used as the quality control strains in parallel.

2.3. Whole genome sequencing and genome analysis

To investigate the evolution and genetic relatedness of these CRKP, the genomic DNA of three CRKP (strains SP1023, F1025, and SP1030) was subjected to WGS by both the short-read Illumina NovaSeq 6000 platform and the hybrid long-read Oxford Nanopore PromethION 48 platform. The complete genome was assembled by Flye assembler v2.9.2 (Kolmogorov et al., 2020) and polished by Pilon v1.24 (Walker et al., 2014). The antimicrobial resistance genes and the plasmid types for the assembly scaffolds were identified by ResFinder 4.1 and PlasmidFinder 2.0, respectively, at the Center for Genomic Epidemiology.[3] The sequence types and virulence factors were identified using Kleborate v0.3.0 (Lam et al., 2021). A pairwise comparison of genomes and variant callings for single-nucleotide polymorphisms (SNPs) was conducted using Snippy v4.4.5 with default settings. The plasmids encoding bla_{KPC-14} and virulence-associated genes were annotated by the RAST server (Overbeek et al., 2014) and BLASTN program. The comparison of plasmids was visualized and annotated by BRIG v0.95 (Alikhan et al., 2011).

2.4. Conjugation experiment

The transferability of bla_{KPC-14} genes was estimated by conjugation experiments with filter mating methods (Cai et al., 2008). Rifampin-resistant *E. coli* EC600 was used as the recipient strain. The putative transconjugants grown on selective media supplemented with 8 mg/L CZA or 30 mg/L tetracycline were identified by MALDI-TOF MS and

2 https://www.fda.gov/drugs/developmentresources/tigecycline-injection-products

3 https://www.genomicepidemiology.org/

screened for the presence of bla_{KPC-14} genes. The conjugation frequency equaled the number of transconjugants divided by the number of recipients.

2.5. Virulence testing in the *Galleria mellonella* infection model

The *G. mellonella* (wax moth larvae) infection model was used to confirm the hypervirulent phenotype of the CRKP strains as previously described (McLaughlin et al., 2014). Overnight cultures of *K. pneumoniae* were diluted in sterile phosphate-buffered saline to obtain a concentration of 10^8 CFU/mL. Wax moth larvae weighing 250–300 mg (Tianjin Huiyude Biotech Company, Tianjin, China) were injected with 10 μL bacterial suspension and incubated for 48 h at 35°C. The survival rate of *G. mellonella* was recorded at 12 h, 24 h, 36 h, and 48 h. ST11 *K. pneumoniae* FJ8 without virulence factors and the hypervirulent *K. pneumoniae* 4 were used as the negative and positive controls, respectively (Gu et al., 2018). All experiments were performed in triplicate. Kaplan–Meier survival curves were plotted using Prism 9.

2.6. Nucleotide sequence accession numbers

The complete genome of the chromosome and plasmids for *K. pneumoniae* SP1023, F1025, and SP1030 were downloaded with BioSample accession numbers SAMN36464959, SAMN36465167, and SAMN36465169, respectively.

3. Results

3.1. Antimicrobial susceptibility results

As Table 1 illustrated, three CRKP shared a high-level resistance to CZA with MIC values of >64/4 mg/L and showed resistance or decreased susceptibility to meropenem and ertapenem. These strains also shared an extensive drug resistance (XDR) profile (Magiorakos et al., 2012) to ceftazidime, cefotaxime, cefepime, aztreonam, ciprofloxacin, amikacin, chloramphenicol, fosfomycin, and tetracycline but retained susceptibility to imipenem. More worrisome, resistance to colistin was observed in these CRKP, and *K. pneumoniae* F1024 exhibited additional resistance to another clinically important antibiotic, tigecycline, which was interpreted as resistance to almost all the antimicrobial agents frequently used in clinical settings.

3.2. Whole genome analysis of KPC-14-producing isolates

All three CRKP were identified as the ST11 type, which was the predominant ST type of CRKP in China (Liu et al., 2022a). Moreover, these three strains all belonged to the K64 serotype. Genome-based phylogenetic analysis suggested that these CRKP were closely related within 27 SNPs (with strain F1025 as the reference genome), indicating that these XDR strains originated from the same clone.

Whole-genome analysis demonstrated that three CRKP exhibited a similar carriage profile of β-lactamase genes, including bla_{KPC-14}, bla_{SHV-11}, bla_{TEM-1}, and bla_{LAP-2}, while *K. pneumoniae* F1025 and SP1030 additionally expressed SHV-12 (Table 2). However, the amplification of the bla_{KPC} gene failed both in *A.baumannii* and *P. aeruginosa* isolated from Patient A. Multiple antimicrobial resistance genes were also identified in three CRKP, conferring resistance to quinolones (*qnrS1*), phenicols (*catA2*), sulfonamides (*sul1* and *sul2*), tetracyclines [*tet*(A)], fosfomycin (*fosA* and *fosA3*), and aminoglycosides (*aadA2b* and *rmtB*). Detailed analysis showed three CRKP shared that mutations for type 1 Tet(A) variants (I5R, V55M, I75V, T84A, S201A, F202S, and V203F) and the insertion of IS*Kpn26* elements at position 75 in the *acrR* gene, both of which contributed to the elevated MICs of tigecycline (Chiu et al., 2017). Insertional inactivation by IS*Kpn18* was also detected in *ramR*, another tigecycline resistance determinant gene, further driving the generation of resistance to this antibiotic in *K. pneumoniae* SP1025. The *mgrB* genes was interrupted by IS*Kpn26* at the same position (nucleotide 75) in three CRKP, thus accounting for the resistance to colistin.

3.3. Transferability of bla_{KPC-14}-carrying plasmids

The bla_{KPC-14} gene could be conjugated into *E. coli* EC600 from *K. pneumoniae* F1025 and SP1030 with similar conjugation efficiencies of 2.1×10^{-5} and 3.5×10^{-5}, respectively, but conjugation failed in strain SP1023, suggesting a different location of the bla_{KPC-14} gene among these strains. Elevated CZA and several β-lactam MIC values were observed in *E. coli* transconjugants for *K. pneumoniae* F1025 and SP1030, confirming the functionality of bla_{KPC-14} (Table 1). Notably, these *E. coli* transconjugants also showed resistance to tetracycline and slightly higher tigecycline MIC values, indicating the possible cotransfer of *tet*(A) and bla_{KPC-14} genes. Therefore, selective media containing tetracycline were used to screen for putative transconjugants. Each of the three CRKP was able to transfer the *tet*(A) gene to the recipient *E. coli* EC600 with a similar efficiency at approximately 10^{-5}; however, *E. coli* transconjugants with different donors displayed heterogeneity in antimicrobial susceptibility profiles. Unlike the *E. coli* transconjugants F1025-TE and SP1030-TE (with *K. pneumoniae* F1025 and SP1030 as donors, respectively), which displayed similar profiles to those of their counterparts for bla_{KPC-14}, the *E. coli* transconjugant SP1023-TE (with *K. pneumoniae* SP1023 as the donor) showed decreased susceptibility to tetracycline and tigecycline but retained the same MIC values for CZA and other β-lactams as the recipient strain. The amplification of bla_{KPC-14} was also carried out in *E. coli* transconjugants F1025-TE and SP1030-TE, but failed in SP1023-TE, further supporting the dissimilar bla_{KPC-14} gene locations and plasmid carriage of the three CRKP.

3.4. Molecular analysis of bla_{KPC-14}-carrying plasmids

To clarify the location and genetic platforms of bla_{KPC-14} genes, hybrid long-read sequencing of all three strains was performed. This yielded the complete genome for three CRKP with similar sizes of approximately 6 Mbp, consisting of one chromosome and varied numbers of plasmids (Table 3). A similar IncFII/IncR hybrid plasmid encoding bla_{KPC-14} was harbored by all three CRKP, designated as plasmids pSP1023-KPC, pF1025-KPC, and pSP1030-KPC for *K. pneumoniae* SP1023, F1025, and SP1030, respectively. These plasmids carried a variety of additional antimicrobial resistance determinants, including the bla_{TEM-1}, *fosA3*, and

TABLE 1 Antimicrobial susceptibility results of *K. pneumoniae* isolates and their *E. coli* transconjugants.

Strain	MICs (mg/L)													
	IPM[b]	MEM	ETP	CZA	CAZ	CTX	FEP	TZP	SCF	CMZ	ATM	CIP	AK	CHL	FOS	TE	TGC	COL
K. pneumoniae SP1023	0.25	2	4	>64/4	>128	64	>64	32/4	32/16	32	>128	>32	>128	>128	>256	>64	2	4
K. pneumoniae F1025	0.5	4	16	>64/4	>128	>128	>64	>256/4	64/32	128	>128	>32	>128	>128	>256	>64	8	4
K. pneumoniae SP1030	0.5	2	8	>64/4	>128	>128	>64	>256/4	128/64	32	>128	>32	>128	>128	>256	>64	2	4
Transconjugant SP1023-TE[a]	0.25	0.06	≤0.03	≤0.5/4	≤0.5	≤0.5	≤0.5	≤8/4	≤8/4	≤2	≤1	1	≤4	>64	≤8	>64	0.125	≤0.5
Transconjugant F1025-CZA	0.25	0.06	0.25	16/4	>128	32	16	≤8/4	≤8/4	≤2	>128	1	≤4	>64	≤8	>64	0.125	≤0.5
Transconjugant F1025-TE	0.25	0.06	0.25	16/4	>128	32	16	≤8/4	≤8/4	≤2	>128	1	≤4	>64	≤8	>64	0.125	≤0.5
Transconjugant SP1030-CZA	0.25	0.06	0.5	16/4	>128	32	32	≤8/4	≤8/4	≤2	>128	1	≤4	>64	≤8	>64	0.125	≤0.5
Transconjugant SP1030-TE	0.25	0.06	0.5	16/4	>128	32	32	≤8/4	≤8/4	≤2	>128	1	≤4	>64	≤8	>64	0.125	≤0.5
E. coli EC600	0.25	0.06	≤0.03	≤0.5/4	≤0.5	≤0.5	≤0.5	≤8/4	≤8/4	≤2	≤1	≤0.25	≤4	≤4	≤8	≤1	0.06	1

[a]*E. coli* transconjugants were selected by two different agar plates containing ceftazidime/avibactam or tetracycline, respectively.

[b]IPM, imipenem; MEM, meropenem; ETP, ertapenem; CZA, ceftazidime/avibactam; CAZ, ceftazidime; CTX, cefotaxime; FEP, cefepime; TZP, piperacillin/tazobactam; SCF, cefoperazone/sulbactam; CMZ, cefmetazole; ATM, aztreonam; CIP, ciprofloxacin; AK, amikacin; CHL, chloramphenicol; FOS, fosfomycin; TE, tetracycline; TGC, tigecycline; COL, colistin. For piperacillin/tazobactam and ceftazidime/avibactam, the tazobactam and avibactam were tested at a fixed concentration of 4 mg/L. For cefoperazone/sulbactam, the combination was tested with concentrations of 2:1 ratio (antibiotic: inhibitor).

rmtB genes, with sizes ranging from 112,363–112,364 bp. Blast analysis demonstrated that the backbone of this plasmid showed the highest similarity to plasmid pCRKP66R-3 with 100% identity (100% coverage, GenBank accession CP063835), which harbored the bla_{KPC-2} gene. These plasmids also displayed 100% identity to two other bla_{KPC-2}-encoding plasmids named plasmid pC76-KPC (86% coverage, GenBank accession CP080299) and pCT77-KPC (87% coverage, GenBank accession CP080305), both of which were previously discovered in CRKP in our hospital (Figure 2) (Chen et al., 2021). The bla_{KPC-14} gene was in the non-Tn*4401* element as an NTE_{KPC}-Ib-like transposon with a genetic array of "IS*26*-ΔTn*3*-IS*Kpn8*-bla_{KPC-14}-IS*Kpn6*-*korC*-*klcA*-*rep*-orf-IS*26*," which was commonly and uniquely identified in ST11 KPC-producing CRKP (Yang et al., 2021). Consistent with many IncFIIK/IncR bla_{KPC}-harboring plasmids widely disseminated in China, this plasmid was nonconjugative, which was probably due to the absence of relaxase, thus explaining the failure of the single transfer of bla_{KPC-14} in the conjugation experiment.

The type 1 Tet(A) variant colocalized with the *catA2*, bla_{LAP-2}, *qnrS1*, and *sul2* genes on unknown Inc-type plasmids with sizes of 84,754 bp, 101,767 bp, and 101,768 bp for *K. pneumoniae* SP1023, F1025, and SP1030, respectively (Figure 2A). This backbone was also observed in *tet*(A)-harboring plasmids in *K. pneumoniae* C76 and *K. pneumoniae* CT77 with 100% identity (91% coverage) but integrated with a 4,956 bp DNA fragment (based on *K. pneumoniae* SP1030) containing the bla_{LAP-2} and *qnrS1* genes. Unlike plasmid pSP1023-tetA in *K. pneumoniae* SP1023, the bla_{KPC-14} locus of 8,248 bp was inserted into the *tet*(A)-carrying plasmids in *K. pneumoniae* F1025 and SP1030, which were further designated as plasmid pF1025-KPC-tetA and pSP1030-KPC-tetA, respectively. Another resistance locus harboring the bla_{SHV-12} gene was also observed in these two plasmids but was absent in plasmid pSP1023-tetA. The structural discrepancy in these *tet*(A)-carrying plasmids echoed the differences in the genetic and phenotypic profiles of the transconjugants we noticed. These observations indicated that the plasmids carrying both *tet*(A) and bla_{KPC-14} genes possibly evolved from those with a single occurrence of the *tet*(A) gene, such as pSP1023-tetA, which could also be further traced back to previously identified plasmids. The bla_{KPC-14} genes were flanked by an NTE_{KPC}-Ib-like transposon similar to that on the IncFII/IncR plasmids in our study, indicating that these nonconjugative plasmids might be the source of the bla_{KPC-14}-carrying fragments for plasmids pF1025-KPC-tetA and pSP1030-KPC-tetA. As previously reported, the diversity of mobile elements and transposons in NTE_{KPC}-I elements actively promoted the transposition of the bla_{KPC} genes to various genetic locations (Yang et al., 2021); thus, we speculated that the mobilization event mediated the integration of the bla_{KPC-14} locus into the backbones of *tet*(A)-carrying plasmids, thus creating the binary carriage profile of bla_{KPC-14}-harboring plasmids in *K. pneumoniae* F1025 and SP1030. The integrated plasmids retained the fully functional conjugative genes homologous to the previous *tet*(A)-carrying plasmids (Figure 2), further facilitating the cotransfer and dissemination of KPC-14 and the type 1 Tet(A) variant among *Enterobacterales*.

3.5. Virulence analysis

The hypervirulent phenotype of three CRKP was observed in the *G. mellonella* infection model (Figure 3). At 48 h post-infection, the

TABLE 2 Clinical and genetic characteristics of three KPC-14-producing *K. pneumoniae* isolates.

Strain	Patient	Gender	Age	Diagnosis	Specimen	Antibiotic resistance genes	Virulence factors	*mgrB* mutation	*ramR* mutation	*acrR* mutation
K. pneumoniae SP1023	Patient A	Male	38	Cerebral hemorrhage	Sputum	bla_{KPC-14}, bla_{SHV-11}, bla_{TEM-1}, bla_{LAP-2}, *qnrS1*, *catA2*, *tet*(A), *fosA*, *fosA3*, *sul1*, *sul2*, *aadA2b*, *rmtB*	RmpA2, aerobactin, yersiniabactin	IS*Kpn26* insertion at nt 75	Wild type	IS*Kpn26* insertion at nt 281
K. pneumoniae F1025	Patient A	Male	38	Cerebral hemorrhage	Feces	bla_{KPC-14}, bla_{SHV-11}, bla_{SHV-12}, bla_{TEM-1}, bla_{LAP-2}, *qnrS1*, *catA2*, *tet*(A), *fosA*, *fosA3*, *sul1*, *sul2*, *aadA2b*, *rmtB*	RmpA2, aerobactin, yersiniabactin	IS*Kpn26* insertion at nt 75	IS*Kpn18* insertion at nt 396	IS*Kpn26* insertion at nt 281
K. pneumoniae SP1030	Patient B	Male	40	Cerebral hemorrhage	Sputum	bla_{KPC-14}, bla_{SHV-11}, bla_{SHV-12}, bla_{TEM-1}, bla_{LAP-2}, *qnrS1*, *catA2*, *tet*(A), *fosA*, *fosA3*, *sul1*, *sul2*, *aadA2b*, *rmtB*	RmpA2, aerobactin, yersiniabactin	IS*Kpn26* insertion at nt 75	Wild type	IS*Kpn26* insertion at nt 281

TABLE 3 Location of antimicrobial resistance genes and virulence genes.

Strain	Patient	Chromosome	bla_{KPC}-carrying plasmid	*tet*(A)-carrying plasmid	Virulence plasmid
K. pneumoniae SP1023	Patient A	4,747,758 bp, bla_{SHV-11}, *fosA*, *sul1*, *aadA2b*, *irp1*, *irp2* No. of SNPs: 3[a]	112,364 bp, nonconjugative, bla_{KPC-14}, bla_{TEM-1}, *fosA3*, *rmtB* No. of SNPs: 0	84,754 bp, conjugative, *tet*(A), *catA2*, bla_{LAP-2}, *qnrS1*, *sul2* No. of SNPs: 1	204,778 bp, *rmpA2*, *iutA*, *iucABCD* No. of SNPs: 0
K. pneumoniae F1025	Patient A	5,424,169 bp, bla_{SHV-11}, *fosA*, *sul1*, *aadA2b*, *irp1*, *irp2* No. of SNPs: NA	112,364 bp, nonconjugative, bla_{KPC-14}, bla_{TEM-1}, *fosA3*, *rmtB* No. of SNPs: NA	101,767 bp, conjugative, *tet*(A), *catA2*, bla_{LAP-2}, *qnrS1*, *sul2*, bla_{KPC-14}, bla_{SHV-12} No. of SNPs: NA	204,770 bp, *rmpA2*, *iutA*, *iucABCD* No. of SNPs: NA
K. pneumoniae SP1030	Patient B	5,477,323 bp, bla_{SHV-11}, *fosA*, *sul1*, *aadA2b*, *irp1*, *irp2* No. of SNPs: 4	112,363 bp, nonconjugative, bla_{KPC-14}, bla_{TEM-1}, *fosA3*, *rmtB* No. of SNPs: 1	101,768 bp, conjugative, *tet*(A), *catA2*, bla_{LAP-2}, *qnrS1*, *sul2*, bla_{KPC-14}, bla_{SHV-12} No. of SNPs: 1	204,770 bp, *rmpA2*, *iutA*, *iucABCD* No. of SNPs: 0

[a]The SNP numbers was estimated by Snippy with the reference sequences of *K. pneumoniae* F1025.
No., numbers; NA, not applicable.

survival rates of larvae infected by strains SP1023, F1025, and SP1030 were 12.5, 29.2, and 16.7%, respectively, which were lower than that of larvae infected by the negative control strain *K. pneumoniae* FJ8 at 79.2%. WGS analysis showed that these strains expressed the same virulence factor profile specific to hypervirulent *K. pneumoniae* (hvKP), including RmpA2, aerobactin, and yersiniabactin (Table 2); thus, the CRKP in our study were classified as CR-hvKP. Long-read sequencing revealed that the *rmpA2* and the *iutAiucABCD* gene cluster (aerobactin siderophore) were located on the IncHI1B/IncFIB-type pLVPK-like plasmid with sizes ranging from 204,770 to 204,778 bp (Table 3, Figure 2C). These plasmids showed high identity (>99%) to a variety of other hypervirulence plasmids of *K. pneumoniae* in the NCBI database, indicating the wide dissemination of these plasmids conferring hypervirulent phenotypes.

4. Discussion

The rapid and wide dissemination of KPC-producing CRKP represents a serious threat to public health and a serious challenge for healthcare workers. Most of the KPC-producing CRKP also harbor determinants that confer resistance to a variety of antimicrobial agents thus further limiting the clinical options for treatment. CZA exhibited great activity against these multidrug-resistant pathogens; however, many concerns have been raised over the emergence of resistant KPC variants with various genetic landscapes, demonstrating the substantial evolutionary potential of this enzyme (Findlay et al., 2021; Liu et al., 2022b; Wu et al., 2022).

In this study, we described the emergence of the CZA-resistant CRKP harbored the bla_{KPC-14} gene on two structurally different plasmids. Three CRKP collectively harbored a bla_{KPC-14}-carrying IncFII/IncR plasmid, which was widespread and commonly identified in KPC-producing CRKP in China (Chen et al., 2014; Dong et al., 2018). There were a few studies that described the emergence and *in vivo* selection of KPC-14 rendering resistance following CZA treatment (Bianco et al., 2020; Niu et al., 2020). However, due to the absence of CZA therapeutic regimens during the hospitalization of the two patients, the source of the bla_{KPC-14} gene in our hospital was still unclear. As recently shown, the KPC-14 enzyme demonstrated the loss of carbapenemase activity (Compain and Arthur, 2017), which was true for the low-level resistance or susceptibility to the carbapenems of KPC-14 producers in our study. This finding reminded us that these resistance determinants may silently spread and be easily ignored during routine surveillance in clinical settings. Although these bla_{KPC-14}-encoding plasmids were nonconjugative, they could provide translocatable fragments as reservoirs for the mobilization of the bla_{KPC-14} gene into other plasmid backbones. In this study, we also identified a conjugative plasmid integrated with the bla_{KPC-14}-containing NTE_{KPC}-I element, which was relatively prevalent among clinical strains in China (Cerdeira et al., 2019). Structurally similar plasmids were previously reported in our hospital with the colocation of the type 1 Tet(A) variant (Chen et al., 2021), which was associated with resistance to another clinically important antibiotic, tigecycline. Moreover, the evolution of these plasmids was observed, as they could additionally capture various other resistance loci, further contributing to the coselection and persistence of these determinants of resistance to last-line antibiotics. The considerable genetic plasticity of these plasmids enabled the acquisition of further resistance-encoding and hypervirulence-encoding genetic elements; thus, these isolates can better adapt to various environments to stimulate the spread of bla_{KPC-14} among *Enterobacterales*.

The multidimensional transmission and potential silent spread of bla_{KPC-14} in CRKP is concerning, especially in those that also carry hypervirulent phenotypes. The KPC-14-producing CRKP in our study originated from the same ST11 clone, which was the dominant clone of CRKP in China and served as a salient example of the evolutionary acquisition of resistance genes and virulence factors for a newly emerged superbug (Liao et al., 2020). The pLVPK-like virulence plasmids commonly converted normal ST11 strains to ST11 hvKP and were first reported in our hospital in 2018 (Gu et al., 2018). These plasmids enhanced the environmental survival and the rapid dissemination of the hypervirulent phenotype with a limited fitness cost in ST11 CRKP (Zhou et al., 2020), consistent with the persistence of these plasmids during spread and evolution observed in our study. More worrisome, these superbugs still have the

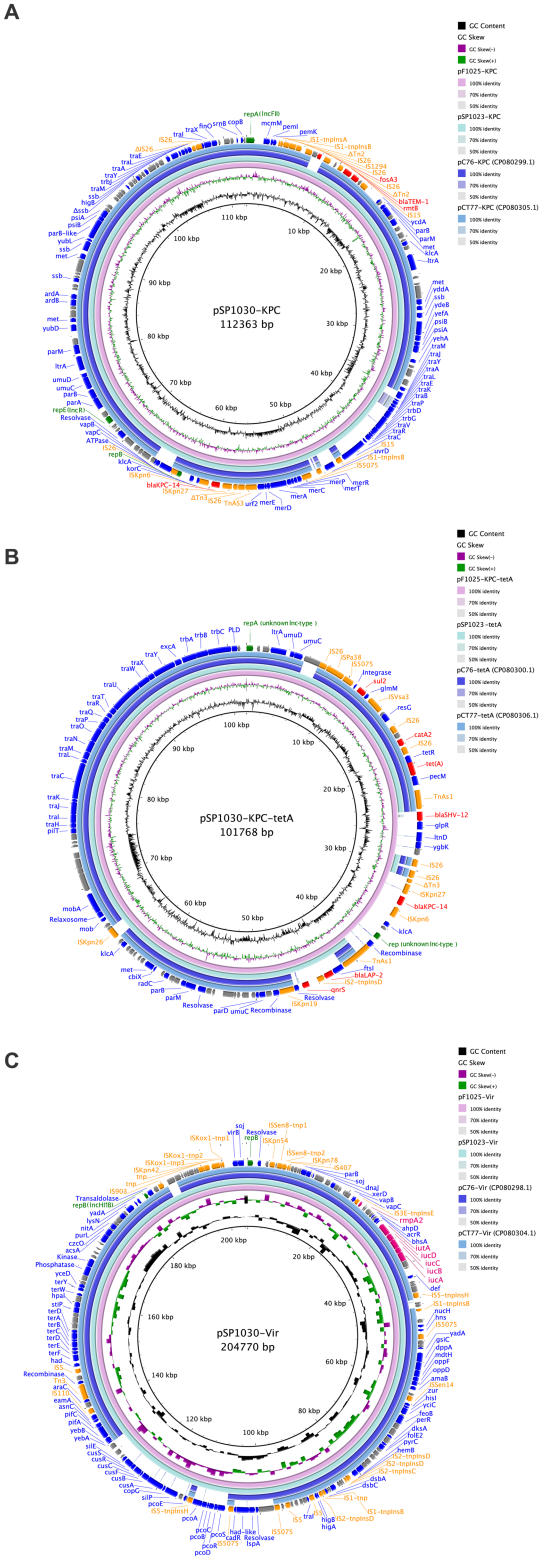

FIGURE 2
Comparisons of bla_{KPC}-carrying plasmid **(A)**, *tet*(A)-carrying plasmid **(B)**, and the virulence plasmid **(C)** with the reference plasmids. The circles from inside to outside display the scale in kilobase pairs, the GC skew, the GC content, the similarity to the reference plasmids, and the annotation of the plasmid in our study, respectively. All insertion sequences are labeled orange, antimicrobial resistance genes are labeled red, genes encoding proteins with known functions are labeled blue, *rep* genes are labeled green, and the hypothetical proteins are labeled gray.

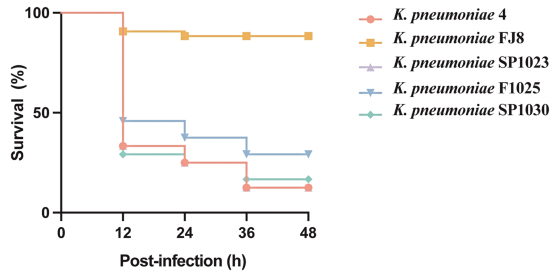

FIGURE 3
Virulence potential of *K. pneumoniae* strains in a *G. mellonella* infection model.

exceptional ability to attain extra resistance to clinically important drugs such as colistin and tigecycline via *in vivo* selection following antibiotic treatment. This could have affected *K. pneumoniae* F1025 in our study, which displayed resistance to almost all antibiotics, including CZA, colistin, and tigecycline. The emergence of such XDR strains may cause severe infections that are difficult to treat with current antibiotics, especially for ICU patients with complicated diseases.

5. Conclusion

In this study, we described the evolution of a conjugative mosaic plasmid encoding the bla_{KPC-14} gene via mobile elements in CR-hvKP. The nonconjugative IncFII/IncR plasmid could serve as the reservoir of the mobilizable fragment of the bla_{KPC-14} gene, similar to NTE_{KPC}-I elements, allowing it to integrate into other conjugative plasmid backbones, further facilitating the spread of bla_{KPC-14}. Therefore, constant surveillance to control the development and further spread of CZA resistance is of great significance.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Ethics statement

The studies involving humans were approved by the Ethics Committee of The Second Affiliated Hospital of Zhejiang University School of Medicine. The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from primarily isolated as part of your previous study for which ethical approval was obtained. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

LW: Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Writing – original draft. WS: Data curation, Formal analysis, Investigation, Validation, Writing – original draft, Software, Visualization. JC: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Funding

The work was supported by the Zhejiang Provincial Natural Science Foundation of China (LY22H200001).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alikhan, N.-F., Petty, N. K., Ben Zakour, N. L., and Beatson, S. A. (2011). BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. doi: 10.1186/1471-2164-12-402

Barnes, M. D., Winkler, M. L., Taracila, M. A., Page, M. G., Desarbre, E., Kreiswirth, B. N., et al. (2017). *Klebsiella pneumoniae* Carbapenemase-2 (KPC-2), substitutions at ambler position Asp179, and resistance to ceftazidime-avibactam: unique antibiotic-resistant phenotypes emerge from β-lactamase protein engineering. *mBio* 8, e00528–e00517. doi: 10.1128/mBio.00528-17

Bianco, G., Boattini, M., Iannaccone, M., Bondi, A., Ghibaudo, D., Zanotto, E., et al. (2021). Carbapenemase detection testing in the era of ceftazidime/avibactam-resistant KPC-producing *Enterobacterales*: a 2-year experience. *J. Glob. Antimicrob. Resist.* 24, 411–414. doi: 10.1016/j.jgar.2021.02.008

Bianco, G., Boattini, M., Iannaccone, M., Cavallo, R., and Costa, C. (2020). Bloodstream infection by two subpopulations of *Klebsiella pneumoniae* ST1685 carrying KPC-33 or KPC-14 following ceftazidime/avibactam treatment: considerations regarding acquired heteroresistance and choice of carbapenemase detection assay. *J. Antimicrob. Chemother.* 75, 3075–3076. doi: 10.1093/jac/dkaa283

Cai, J. C., Zhou, H. W., Zhang, R., and Chen, G.-X. (2008). Emergence of *Serratia marcescens*, *Klebsiella pneumoniae*, and *Escherichia coli* isolates possessing the plasmid-mediated carbapenem-hydrolyzing β-lactamase KPC-2 in intensive care units of a Chinese hospital. *Antimicrob. Agents Chemother.* 52, 2014–2018. doi: 10.1128/AAC.01539-07

Cerdeira, L. T., Lam, M. M. C., Wyres, K. L., Wick, R. R., Judd, L. M., Lopes, R., et al. (2019). Small IncQ1 and col-like plasmids harboring bla_{KPC-2} and non-Tn *4401* elements (NTE_{KPC}-IId) in high-risk lineages of *Klebsiella pneumoniae* CG258. *Antimicrob. Agents Chemother.* 63, e02140–e02118. doi: 10.1128/AAC.02140-18

Chen, L., Mathema, B., Chavda, K. D., DeLeo, F. R., Bonomo, R. A., and Kreiswirth, B. N. (2014). Carbapenemase-producing *Klebsiella pneumoniae*: molecular and genetic decoding. *Trends Microbiol.* 22, 686–696. doi: 10.1016/j.tim.2014.09.003

Chen, J., Zeng, Y., Zhang, R., and Cai, J. (2021). *In vivo* emergence of colistin and tigecycline resistance in carbapenem-resistant hypervirulent *Klebsiella pneumoniae* during antibiotics treatment. *Front. Microbiol.* 12:702956. doi: 10.3389/fmicb.2021.702956

Chiu, S.-K., Huang, L.-Y., Chen, H., Tsai, Y.-K., Liou, C.-H., Lin, J.-C., et al. (2017). Roles of *ramR* and *tet*(a) mutations in conferring tigecycline resistance in carbapenem-resistant *Klebsiella pneumoniae* clinical isolates. *Antimicrob. Agents Chemother.* 61, e00391–e00317. doi: 10.1128/AAC.00391-17

Clinical and Laboratory Standards Institute (2018). *Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically*, *11*. Wayne, PA: CLSI standard M07.

Clinical and Laboratory Standards Institute (2021). *Performance standards for antimicrobial susceptibility Ttesting*, *31*. Wayne, PA: CLSI supplement M100.

Compain, F., and Arthur, M. (2017). Impaired inhibition by avibactam and resistance to the ceftazidime-avibactam combination due to the D179Y substitution in the KPC-2 β-lactamase. *Antimicrob. Agents Chemother.* 61, e00451–e00417. doi: 10.1128/AAC.00451-17

Criscuolo, M., and Trecarichi, E. M. (2020). Ceftazidime/avibactam and ceftolozane/tazobactam for multidrug-resistant gram hegatives in patients with hematological malignancies: current experiences. *Antibiotics* 9:58. doi: 10.3390/antibiotics9020058

Dong, N., Yang, X., Zhang, R., Chan, E. W.-C., and Chen, S. (2018). Tracking microevolution events among ST11 carbapenemase-producing hypervirulent *Klebsiella pneumoniae* outbreak strains. *Emerg. Microbes Infect.* 7, 1–8. doi: 10.1038/s41426-018-0146-6

El-Kady, R. A. E.-H., Elbaiomy, M. A., and Elnagar, R. M. (2022). Molecular mechanisms mediating ceftazidime/avibactam resistance amongst carbapenem-resistant *Klebsiella pneumoniae* isolates from cancer patients. *Infect. Drug Resist.* 15, 5929–5940. doi: 10.2147/IDR.S384972

Findlay, J., Poirel, L., Juhas, M., and Nordmann, P. (2021). KPC-mediated resistance to ceftazidime-avibactam and collateral effects in *Klebsiella pneumoniae*. *Antimicrob. Agents Chemother.* 65:e0089021. doi: 10.1128/AAC.00890-21

Gu, D., Dong, N., Zheng, Z., Lin, D., Huang, M., Wang, L., et al. (2018). A fatal outbreak of ST11 carbapenem-resistant hypervirulent *Klebsiella pneumoniae* in a Chinese hospital: a molecular epidemiological study. *Lancet Infect. Dis.* 18, 37–46. doi: 10.1016/S1473-3099(17)30489-9

Humphries, R. M., and Hemarajata, P. (2017). Resistance to ceftazidime-avibactam in *Klebsiella pneumoniae* due to porin mutations and the increased expression of KPC-3. *Antimicrob. Agents Chemother.* 61, e00537–e00517. doi: 10.1128/AAC.00537-17

Jiang, M., Sun, B., Huang, Y., Liu, C., Wang, Y., Ren, Y., et al. (2022). Diversity of ceftazidime-avibactam resistance mechanism in KPC-2-producing *Klebsiella pneumoniae* under antibiotic selection pressure. *Infect. Drug Resist.* 15, 4627–4636. doi: 10.2147/IDR.S371285

Kolmogorov, M., Bickhart, D. M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S. B., et al. (2020). metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat. Methods* 17, 1103–1110. doi: 10.1038/s41592-020-00971-x

Lam, M. M. C., Wick, R. R., Watts, S. C., Cerdeira, L. T., Wyres, K. L., and Holt, K. E. (2021). A genomic surveillance framework and genotyping tool for *Klebsiella pneumoniae* and its related species complex. *Nat. Commun.* 12:4188. doi: 10.1038/s41467-021-24448-3

Liao, W., Liu, Y., and Zhang, W. (2020). Virulence evolution, molecular mechanisms of resistance and prevalence of ST11 carbapenem-resistant *Klebsiella pneumoniae* in China: a review over the last 10 years. *J. Glob. Antimicrob. Resist.* 23, 174–180. doi: 10.1016/j.jgar.2020.09.004

Linh, T. D., Thu, N. H., Shibayama, K., Suzuki, M., Yoshida, L., Thai, P. D., et al. (2021). Expansion of KPC-producing *Enterobacterales* in four large hospitals in Hanoi, Vietnam. *J. Glob. Antimicrob. Resist.* 27, 200–211. doi: 10.1016/j.jgar.2021.09.007

Liu, C., Dong, N., Chan, E. W. C., Chen, S., and Zhang, R. (2022a). Molecular epidemiology of carbapenem-resistant *Klebsiella pneumoniae* in China, 2016–20. *Lancet Infect. Dis.* 22, 167–168. doi: 10.1016/S1473-3099(22)00009-3

Liu, C., Wu, Y., Huang, L., Zhang, Y., Sun, Q., Lu, J., et al. (2022b). The rapid emergence of ceftazidime-avibactam resistance mediated by KPC variants in carbapenem-resistant *Klebsiella pneumoniae* in Zhejiang province, China. *Antibiotics* 11:731. doi: 10.3390/antibiotics11060731

Livermore, D. M., Warner, M., Jamrozy, D., Mushtaq, S., Nichols, W. W., Mustafa, N., et al. (2015). *In vitro* selection of ceftazidime-avibactam resistance in Enterobacteriaceae with KPC-3 carbapenemase. *Antimicrob. Agents Chemother.* 59, 5324–5330. doi: 10.1128/AAC.00678-15

Magiorakos, A.-P., Srinivasan, A., Carey, R. B., Carmeli, Y., Falagas, M. E., Giske, C. G., et al. (2012). Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin. Microbiol. Infect.* 18, 268–281. doi: 10.1111/j.1469-0691.2011.03570.x

McLaughlin, M. M., Advincula, M. R., Malczynski, M., Barajas, G., Qi, C., and Scheetz, M. H. (2014). Quantifying the clinical virulence of *Klebsiella pneumoniae* producing carbapenemase *Klebsiella pneumoniae* with a *galleria mellonellamodel* and a pilot study to translate to patient outcomes. *BMC Infect. Dis.* 14:31. doi: 10.1186/1471-2334-14-31

Niu, S., Chavda, K. D., Wei, J., Zou, C., Marshall, S. H., Dhawan, P., et al. (2020). A ceftazidime-avibactam-resistant and carbapenem-susceptible *Klebsiella pneumoniae* strain harboring bla_{KPC-14} isolated in New York City. *mSphere* 5, e00775–e00720. doi: 10.1128/mSphere.00775-20

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* 42, D206–D214. doi: 10.1093/nar/gkt1226

Pérez-Galera, S., Bravo-Ferrer, J. M., Paniagua, M., Kostyanev, T., De Kraker, M. E. A., Feifel, J., et al. (2023). Risk factors for infections caused by carbapenem-resistant *Enterobacterales*: an international matched case-control-control study (EURECA). *eClinicalMedicine* 57:101871. doi: 10.1016/j.eclinm.2023.101871

Tian, Y., Zhang, Q., Wen, L., and Chen, J. (2021). Combined effect of polymyxin B and tigecycline to overcome heteroresistance in carbapenem-resistant *Klebsiella pneumoniae*. *Microbiol. Spectr.* 9, e00152–e00121. doi: 10.1128/Spectrum.00152-21

Van Duin, D., and Bonomo, R. A. (2016). Ceftazidime/avibactam and ceftolozane/tazobactam: second-generation β-lactam/β-lactamase inhibitor combinations. *Clin. Infect. Dis.* 63, 234–241. doi: 10.1093/cid/ciw243

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Wang, M., Earley, M., Chen, L., Hanson, B. M., Yu, Y., Liu, Z., et al. (2022). Clinical outcomes and bacterial characteristics of carbapenem-resistant *Klebsiella pneumoniae* complex among patients from different global regions (CRACKLE-2): a prospective, multicentre, cohort study. *Lancet Infect. Dis.* 22, 401–412. doi: 10.1016/S1473-3099(21)00399-6

Wang, L., Shen, W., Zhang, R., and Cai, J. (2022). Identification of a novel ceftazidime-avibactam-resistant KPC-2 variant, KPC-123, in *Citrobacter koseri* following ceftazidime-avibactam treatment. *Front. Microbiol.* 13:930777. doi: 10.3389/fmicb.2022.930777

Wu, Y., Yang, X., Liu, C., Zhang, Y., Cheung, Y. C., Chan, W. C., et al. (2022). Identification of a KPC variant conferring resistance to ceftazidime-avibactam from ST11 carbapenem-resistant *Klebsiella pneumoniae* strains. *Microbiol. Spectr.* 10, e02655–e02621. doi: 10.1128/spectrum.02655-21

Yang, X., Dong, N., Chan, E. W.-C., Zhang, R., and Chen, S. (2021). Carbapenem resistance-encoding and virulence-encoding conjugative plasmids in *Klebsiella pneumoniae*. *Trends Microbiol.* 29, 65–83. doi: 10.1016/j.tim.2020.04.012

Zhou, K., Xiao, T., David, S., Wang, Q., Zhou, Y., Guo, L., et al. (2020). Novel subclone of carbapenem-resistant *Klebsiella pneumoniae* sequence type 11 with enhanced virulence and transmissibility. *China. Emerg. Infect. Dis.* 26, 289–297. doi: 10.3201/eid2602.190594

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 17 November 2023
DOI 10.3389/fmicb.2023.1281473

OPEN ACCESS

EDITED BY
Octavio Luiz Franco,
Catholic University of Brasilia (UCB), Brazil

REVIEWED BY
Ueli Von Ah,
Agroscope (Switzerland), Switzerland
Li Weicheng,
Inner Mongolia Agricultural University, China

*CORRESPONDENCE
Eszter Kaszab
✉ kaszab.eszter@vmri.hun-ren.hu

†These authors have contributed equally to this work and share first authorship

RECEIVED 22 August 2023
ACCEPTED 03 November 2023
PUBLISHED 17 November 2023

CITATION
Kaszab E, Laczkó L, Kardos G and Bányai K (2023) Antimicrobial resistance genes and associated mobile genetic elements in Lactobacillales from various sources. *Front. Microbiol.* 14:1281473. doi: 10.3389/fmicb.2023.1281473

Antimicrobial resistance genes and associated mobile genetic elements in Lactobacillales from various sources

Eszter Kaszab[1,2,3*†], Levente Laczkó[2,4†], Gábor Kardos[2,3,5,6] and Krisztián Bányai[1,3,7]

[1]HUN-REN Veterinary Medical Research Institute, Budapest, Hungary, [2]One Health Institute, Faculty of Health Sciences, University of Debrecen, Debrecen, Hungary, [3]National Laboratory of Infectious Animal Diseases, Antimicrobial Resistance, Veterinary Public Health and Food Chain Safety, Veterinary Medical Research Institute, Budapest, Hungary, [4]HUN-REN-DE Conservation Biology Research Group, Debrecen, Hungary, [5]National Public Health Center, Budapest, Hungary, [6]Department of Gerontology, Faculty of Health Sciences, University of Debrecen, Debrecen, Hungary, [7]Department of Pharmacology and Toxicology, University of Veterinary Medicine, Budapest, Hungary

Lactobacillales are commonly used in food products and as probiotics in animal and human medicine. Despite being generally recognized as safe, lactic acid bacteria may harbor a variety of antimicrobial resistance genes (ARGs), which may be transferable to human or veterinary pathogens, thus, may pose veterinary and public health concerns. This study investigates the resistome of Lactobacillales. A total of 4,286 whole-genome sequences were retrieved from NCBI RefSeq database. We screened ARGs in whole genome sequences and assessed if they are transmissible by plasmid transfer or by linkage to integrative mobile genetic elements. In the database, 335 strains were found to carry at least one ARG, and 194 strains carried at least one potentially transferable ARG. The most prevalent transferable ARG were *tet*M and *tet*W conferring antibiotic resistance to tetracycline. This study highlights the importance of the One Health concept by demonstrating the potential for Lactobacillales, commonly used in food products, to serve as reservoirs and vectors for ARGs.

KEYWORDS

Lactobacillales, antimicrobial resistance genes, one health, iMGE, plasmid

1. Introduction

Lactic acid bacteria (LAB) are frequent and abundant members of the human and animal microbiota and at the same time have a long history of utilization by humans for various purposes. Fermented food and beverages are ancient, and their versatility is considerable. Furthermore, many of these are considered to carry health benefits due to favorable nutritional properties and pathogen exclusion. Utilization of fermented or putrified food has been documented as early as more than 8,000 years ago (Boethius, 2016), and it is widely hypothesized based on archeological and evolutionary evidence that their preference predated the appearance of modern humans (Amato et al., 2021).

In addition, the traditional wisdom that such food fermented by lactic acid bacteria are beneficial for health incited their usage as probiotics, rapidly gaining popularity in nutrition as well as in medicine. Studies have shown that LAB supplementation can have a positive impact on various health outcomes, including gastrointestinal health, immune function, and even

mental health (Marco et al., 2017). Therefore Hill et al. (2014) reworked the definition of probiotics to the following: "live microorganisms which when administered in adequate amounts confer a health benefit on the host." The International Scientific Association for Probiotics and Prebiotics (ISAPP) recommends that the term probiotic be used only on products that deliver live microorganisms with a suitable viable count of well-defined strains with a reasonable expectation of delivering benefits for the wellbeing of the host (Hill et al., 2014). However, it is important to note that the effects of LAB can vary depending on the specific strain and the individual's health status. Probiotics are recommended widely by doctors and pharmacists and have gained fame and popularity in the general public, which brought about their increasing consumption in human as well as in veterinary medicine (Sharma et al., 2014). Overall, LAB play a crucial role in the development of food products with living flora, and are thought to be a component of a healthy diet due to their potential health benefits. The size of the probiotic market is estimated at $65 billion and is projected to increase (Abid and Koh, 2019), despite that measurable health benefit has been proven only in case of few of the many indications in which they are widely used (Abid and Koh, 2019). This increase anticipates the addition of novel strains to the probiotic arsenal, most of which is expected to come from the group of LAB.

Lactic acid bacteria have been shown to harbor acquired antibiotic resistance genes (Sharma et al., 2014). Considering their frequent association with the microbiota of humans and animals, playing a role in the food chain as live components of fermented food together with the frequency of probiotic administration both to humans and to animals, these bacteria have ample opportunity to serve as important sources, vehicles, and targets for exchanging mobile genetic elements, including those harboring antibiotic resistance genes. Thus, they may infest the microbiome we intend to improve using them with antibiotic resistance genes (Montassier et al., 2021). Ironically, probiotics are also frequently recommended to diminish side effects of taking antibiotics (Ouwehand et al., 2016), and frequently prescribed together with them (Rodgers et al., 2013) or to spare them.

Though antibiotic resistance among commercial probiotic products is started to be surveyed (Wong et al., 2015; Cui et al., 2020; Rozman et al., 2020; Toth et al., 2021), approaches for comprehensive understanding of the resistance gene array occurring in LAB, thus an overview of the potential threats they pose, are scarce. This study aims at reviewing the genomes of LAB deposited in the NCBI reference sequence database in order to determine the abundance and distribution of resistance genes in the available lactic acid bacterium genomes.

2. Materials and methods

2.1. Data

We screened publicly available reference genomes of Lactobacillales for the presence and identity of antimicrobial resistance genes (ARG). Genome sequences were obtained from the NCBI RefSeq, a well-curated database of high-quality reference genome sequences. All accessions of Lactobacillales ($n = 4,286$) were downloaded on 04/10/2021, including the genome sequences in fasta format and the corresponding metadata (Supplementary Table S1). Accessions were divided into nine categories by their isolation source to make the interpretation of results easier (animals and humans, animal source food, dairy products, foods and beverages of plant origin, plant and environment, other).

2.2. Bioinformatic analysis

Mass screening of individual genomes was achieved using ABRicate (Seemann, 2020; https://github.com/tseemann/abricate) which software relies on BLAST (Altschul et al., 1990) to match the nucleotide sequences of ARGs against genome sequences. ABRicate used the CARD database (database version 3.1.4, 2021-10-05; McArthur et al., 2013) to detect acquired ARGs in the genome sequences. Contigs bearing at least one ARG were screened for integrative mobile genetic elements (iMGE) using MobileElementFinder 1.0.5 (Johansson et al., 2021). Then, we checked the distance between the coordinates of ARGs and iMGEs located on the same contig and defined them as linked if their distance was less than 10,000 base pairs (bps). The average gene length of prokaryotes is approximately 1,000 bp. We hypothesized that there is a high probability of linkage if a ARG can be found within a maximum distance of about 10 ORF distance (*ca.* 10,000 bp). We modified the concept of Johansson et al. (2021), according to which ARGs are considered linked to AMRs if they can be found within a distance defined by the length of the longest iMGE. Our threshold is more conservative to avoid identifying false positives. Toth et al. (2021), also following the concept of Johansson et al. (2021) found this threshold for *Lactococcus* to be 11,256 bps, which is close to our universal threshold defined for the sake of simplicity. We acknowledge that a threshold-based linkage assessment may not accurately describe the actual molecular events, but argue that in this case the majority of ARGs linked to iMGEs would be identified. The threshold of 10,000 bps is about twice longer the average length of all identified iMGEs (4260.4 bps) and is 25% of the length of the longest iMGE (52,209 bps). Accessions were subject to phylogenetic reconstruction to evaluate if the presence/absence of ARGs is condensed in the phylogeny of Lactobacillales. First, we created a core alignment of samples. We obtained the amino acid sequences of 402 genes characteristic of Lactobacillales from the BUSCO (version 5.2.2; Simao et al., 2015) single-copy orthologous gene database (odb10). Then, we used tblastn 2.10+ (Altschul et al., 1990) to match each core gene amino acid sequence against the database of available Lactobacillales genomes created with makeblastdb 2.10+ (Altschul et al., 1990). The best hits with a query coverage larger than 90% and an identity higher than 85% were retained and aligned using MAFFT 7.490 (Katoh et al., 2002) with the --auto option turned on. We used AMAS.py 1.0 (Borowiec, 2016) to concatenate the gene alignments and to retrieve alignment statistics. Then, the accession's phylogenetic relationships were reconstructed with FastTree 2.1.11 (Price et al., 2010), explicitly designed for large alignments, using default values. Phylogenetic trees were rooted using the minimal ancestor deviation method as implemented in MAD 2.2 (Wade et al., 2020). In the next step, we reconstructed the phylogeny using the abovementioned methods of genome accessions harboring at least on ARG. The phylogenetic tree and the distribution of ARGs with linked iMGEs were visualized using the ggplot2 3.3.5 (Wickham, 2016) and ggtreeExtra 1.0.4 (Xu et al., 2021) R 4.0.4 packages (R Core Team,

2022). The classification of genera followed the guidelines of Zheng et al. (2020). To evaluate the chance of spreading ARGs, the plasmid origin of contigs with ARGs linked to iMGEs was predicted by using PlasFlow 1.1 (Krawczyk et al., 2018).

3. Results

3.1. ARG diversity in Lactobacillales

Out of 4,286 available genomes, 334 (7.8%) harbored at least one ARG. In total, we discovered 42 different ARGs in accessions of Lactobacillales. Accessed Lactobacillales genomes contained up to eight ARGs (Figure 1). The *tet*W gene conferring antibiotic resistance to tetracycline showed the highest frequency with 151 occurrences, followed by *tet*M harbored by 109 accessions (Supplementary Tables S1, S2; Figure 2A). The genes *lnu*A, *lnu*C, and *lsa*C conferring resistance to lincosamides; *Erm*A, *Erm*B, and *Erm*T conferring antibiotic resistance to lincosamides and macrolides; *ANT*(6)-Ia conferring resistance to streptomycin; *vat*E conferring resistance to dalfopristin and *tet*(L) conferring resistance to tetracyline, could be found in a moderate number of genomes ($n = 22–65$). The remaining ARGs were found in less than 10 accessions (Supplementary Table S2). We discovered 193 accessions harboring ARGs linked to iMGEs. The *tet*M and *tet*W gene were associated with an iMGE in 40 and 39 accessions, respectively. The genes *Inu*C, *tet*(L), and *vat*E were associated with iMGEs in more than 30 Lactobacillales genomes. The resistance genes *Inu*A, *Erm*B, *Erm*A, *ANT*(6)-Ia, and *Erm*T were found on iMGE-bearing contigs in a moderate number of accessions ($n = 13–26$). The rest of the ARGs (Figure 1; Supplementary Figure S1B) were linked to iMGEs in up to five genomes (Supplementary Table S3; Figure 2B). The most common ARG associations were *tet*W, *tet*M, *Erm*B, and ANT(6)-Ia. Some genomes contained multiple iMGE associated ARGs, for instance *Latilactobacillus sakei*, *Ligilactobacillus salivarius*, and *Lactobacillus amylovorus* genomes among others.

3.2. ARG abundance by source of isolation

Antimicrobial resistance genes could be discovered in all isolation sources (Table 1). The highest number and ratio of ARG positive

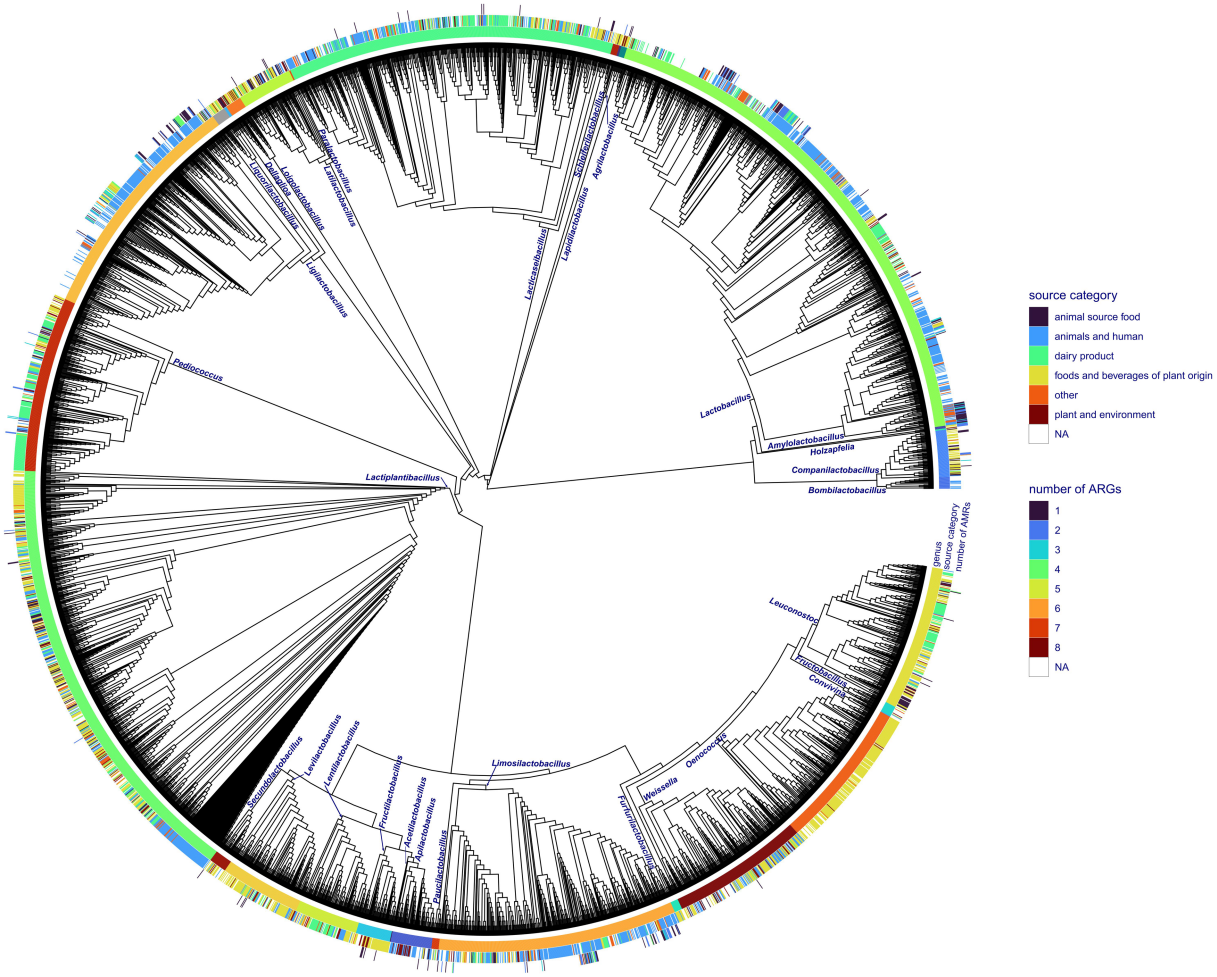

FIGURE 1
Phylogenetic reconstruction of all the Lactobacillales genome accessions obtained from NCBI RefSeq with the corresponding isolation source category and number of ARGs discovered in particular genomes. The phylogenetic tree is shown as a cladogram.

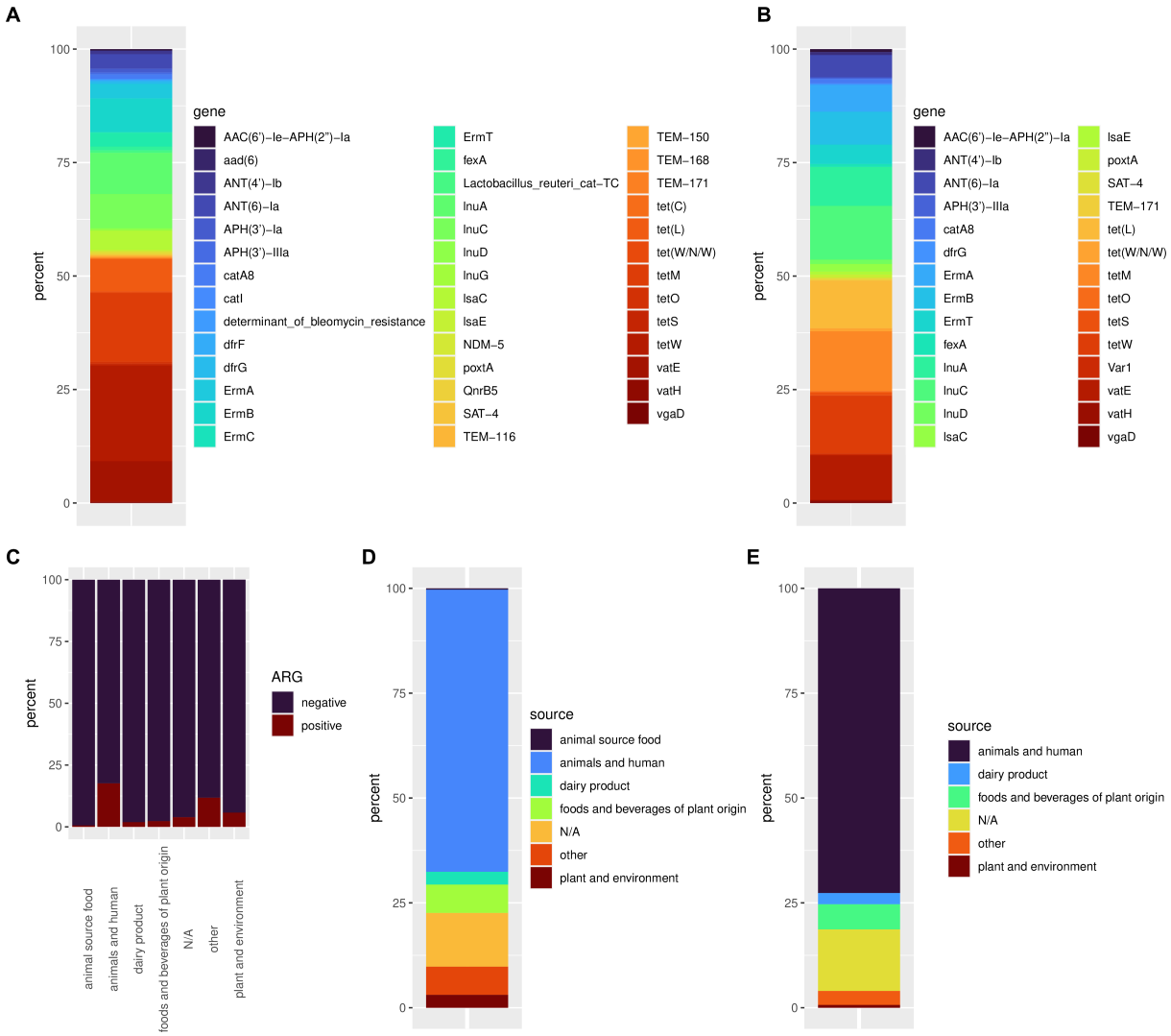

FIGURE 2
The relative frequency of all discovered ARGs **(A)** and the relative frequency of ARGs associated with iMGEs **(B)**. Ratio of ARG positive samples by isolations sources **(C)**, isolation source frequency of accessions showing at least one ARG **(D)**, and accessions showing ARGs linked to iMGEs **(E)**.

TABLE 1 **Frequency of ARG positive samples by different isolation sources.**

Isolation source	Total accessions	ARG positive accessions (%)	Accessions with ARGs linked to iMGEs (%)
N/A	1,069	49 (4.5)	22 (2.1)
Animals and human	1,247	220 (17.6)	109 (8.7)
Animal source food	167	1 (0.6)	0 (0)
Dairy product	515	10 (1.9)	4 (0.8)
Foods and beverages of plant origin	929	22 (2.4)	9 (0.9)
Other	186	22 (11.8)	5 (2.7)
Plant and environment	173	10 (5.8)	1 (0.6)

genomes (17.6%) could be linked to isolation source categories animals and humans (Table 1; Figures 2C,D). Samples isolated from dairy products, food, and beverages of plant origin, plants and the environment showed a lower ratio of ARG positive accessions ranging from 0.6 to 5.8%. We discovered only one genome (*Ligilactobacillus salivarius*, isolation source: ground beef) harboring an ARG isolated from food of animal origin. Almost 4% of ARG positive genome sequences had an unknown isolation source, and 11.8% of such accessions could not be associated with the isolation source categories defined in this study. Similarly, bacterial genomes from animals and

humans showed the highest frequency of iMGE-associated ARGs (Table 1; Figure 2E), followed by foods and beverages of plant origin and dairy products. The only ARG identified from food of animal origin (0.6% of accessions in the category) was not linked to iMGEs. Over 14% of iMGE positive accessions had an unknown origin, and 3.3% of such genomes were associated with the "other" isolation source.

3.3. ARGs in different taxa of Lactobacillales

The core gene amino acid alignment of all genomes consisted of 207,592 positions (Supplementary Table S4) with 18.6% of missing characters. The alignment contained 166,530 variable sites, of which 156,870 appeared to be informative. The phylogenetic reconstruction of the whole dataset (Figure 2) resulted in a well-resolved phylogenetic tree. All internal branches leading to distinct genera received statistical support higher than 95%, except the branch separating *Lactiplantibacillus* from *Pediococcus* (SH-like support value, 78%).

The core gene alignment of ARG harboring accessions showed 139,132 variable and 132,339 informative sites out of 197,279 positions. This alignment lacked 13.889% of positions (Supplementary Table S4). The phylogeny of the ARG-harboring genome accessions reconstructed using core gene alignments showed a support value higher than 99% for all the internal branches. Lower support values could only be observed for short branches separating strains within the same genus.

Seventeen out of 31 genera showed at least one accession harboring ARGs (Figure 1). The *tet*M gene was present in the highest number of genera with the highest frequency in *Ligilactobacillus* (Supplementary Figure S1A). *Erm*B could be identified in a lower number of genera and appeared to be the most characteristic of *Lactobacillus* and *Pediococcus*. The ARGs *vat*E and *tet*W were found in six and three genera, respectively, and their frequency appeared to be relatively high in *Lactobacillus*. The rest of the ARGs could be identified in up to five genera of Lactobacillales, and some of them appeared to be rare and characteristic of certain genera (Supplementary Figure S1A; Figure 3). The *tet*M and *Erm*B genes found in several genera were frequently associated with iMGEs (Supplementary Figure S1B). Almost all *Lactobacillus* accessions harboring *Erm*T were associated with the presence of iMGEs. More than 40% of the *tet*W observations in *Lactobacillus* were linked to iMGEs and this ARG was also relatively frequent in *Ligilactobacillus*. The *Inu*C, *tet*(L), and *vat*E genes were linked to iMGEs in *Lactobacillus* and *Ligilactobacillus* with *vat*E occasionally occurring in *Leuconostoc*, *Ligilactobacillus*, *Limosilactobacillus*, and *Pediococcus*. *Ligilactobacillus* had a relatively high number of accessions with the ARGs *Erm*A and *ANT*(6)-Ia linked to iMGEs, of which *ANT*(6)-Ia were also observed in *Lactobacillus* and *Lactiplantibacillus* and *Companilactobacillus*. The rest of the ARGs were much more rarely associated with iMGEs (Supplementary Figure S1B; Figure 3).

PlasFlow identified 24 contigs of plasmid origin that harbored ARGs linked to iMGEs. The majority of the remaining contigs ($n = 142$) were identifiable as bacterial chromosomes alongside 27 unidentified contigs (i.e., neither identifiable as plasmids nor as chromosomes). For the sake of simplicity, below, we refer to this category as bacterial chromosomes. When located on a plasmid, ARGs were associated with up to four iMGEs and up to 10 iMGEs when located on a bacterial chromosome. One frequently discovered ARG, *tet*W was associated with iMGEs on bacterial chromosomes. Some genes were exclusive to plasmids or to chromosomes (Supplementary Figure S2). The *tet*M gene was linked to iMGEs when located on plasmids and in approximately half of the observations located on chromosomes, similarly to *tet*(L). The latter had less observations on chromosomes (Supplementary Figure S2). When located on a chromosome, *Inu*C and *Inu*A could be characterized with a high frequency of iMGEs, whereas mobile elements linked to these ARGs appeared less frequently on plasmids. The gene *Erm*T was linked to iMGEs only on bacterial chromosomes, and *ANT*(6)-Ia was associated with iMGEs with a high frequency both on chromosomes and plasmids.

4. Discussion

Historically, LAB are thought to be beneficial to health and have a positive effect on human and animal diet. Their fermentative metabolism has been exploited since modern humans arose and started processing animal and plant-based foods. Based on this experience many years old, fermented food is promoted and advertised with the claimed health-preserving properties of such food and the "living flora" therein (Leeuwendaal et al., 2022; Shah et al., 2023). Such claims are increasingly used to foster the multibillion dollar food and beverage industry (Abid and Koh, 2019). In addition, recommending probiotics based frequently on LAB is a common practice in humans to restore integrity of the microbiota of the gastrointestinal tract as well as in veterinary applications to achieve production advantage, sometimes without firm evidence on their utility. Members of Lactobacillales are used extensively for purposes of human and veterinary medicine and nutrition (Hill et al., 2014; Garcia-Hernandez et al., 2016; Dowarah et al., 2017; Al-Yami et al., 2022).

Besides the benefits, these bacteria may harbor ARGs and have the capacity to transfer these ARGs to other bacteria colonizing the individual or the animal receiving probiotic formulas or consuming fermented food through iMGEs. When pathogenic bacteria acquire such iMGE-linked ARGs, the risk that resistant pathogenic bacterial strains evolve significantly increases. This is especially important because these bacteria are generally recognized as safe, therefore, used excessively in human nutrition, feeds for companion animals and in animal husbandry (Montassier et al., 2021; Radovanovic et al., 2023).

Plasmid transmissibility has been studied extensively in LAB, particularly in species commonly used in food fermentation or as probiotics, such as *Lactobacillus, Lactococcus, Enterococcus*, and *Streptococcus* (Schleifer et al., 1985; Igimi et al., 1996; Ammann et al., 2008). These bacteria are known to exchange plasmids both within and between species (Bolotin et al., 2004), leading to the spread of beneficial as well as potentially harmful traits. The mechanisms of plasmid transfer can also vary depending on the specific LAB species and the type of plasmid involved (Igimi et al., 1996; Ammann et al., 2008; Ortiz Charneco et al., 2021). Such plasmid transfer events may trigger acquisition of resistance genes from the host microbiota from pathogens or commensals, which may then be transmitted, for example, in an animal flock to microbiota of other individuals, or these LAB may be the source for resistance genes.

As outlined by the results above, understanding plasmid transmissibility in LAB is important for the development of safe and effective probiotic products, as well as for the control of antibiotic

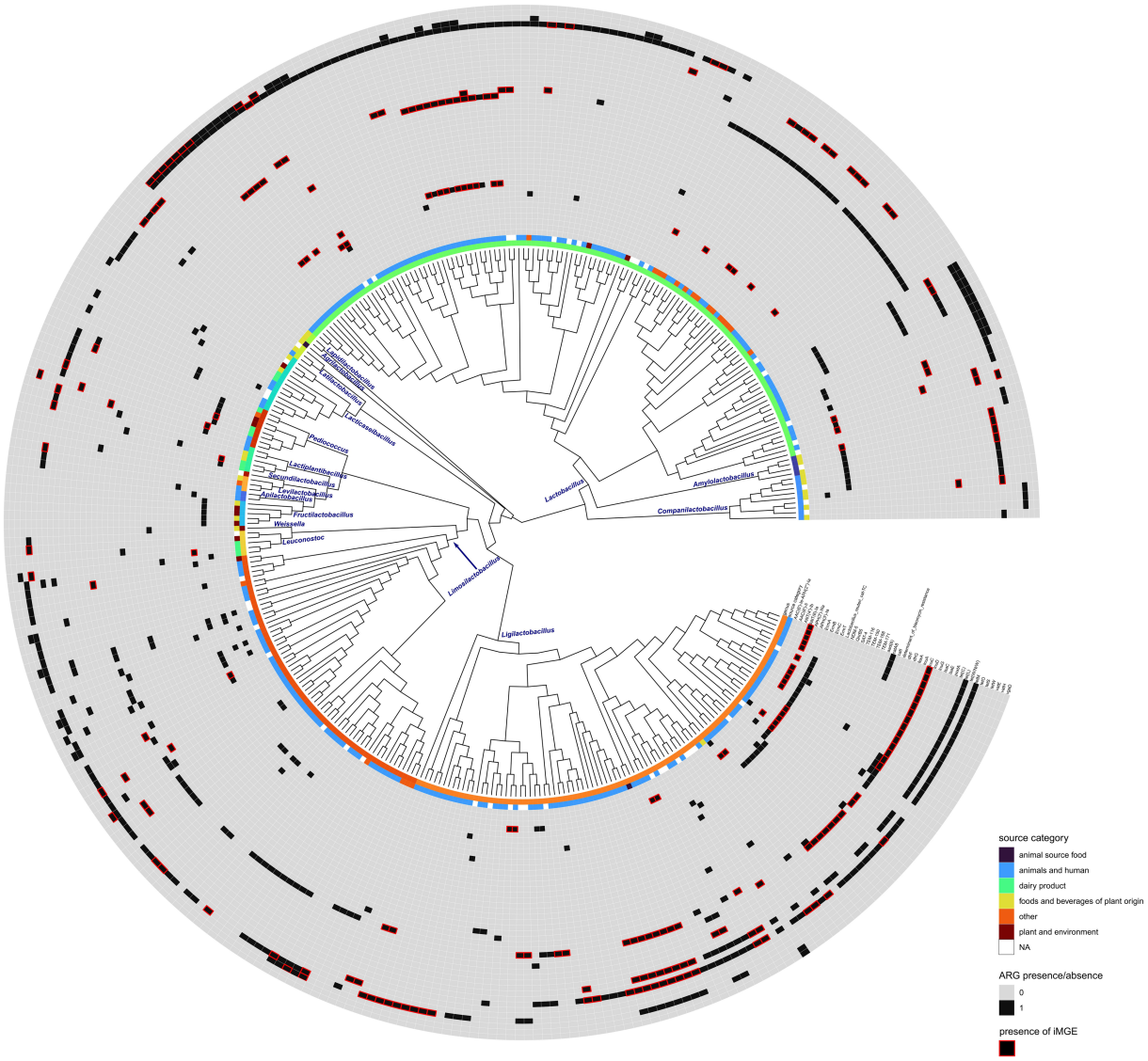

FIGURE 3
Phylogenetic reconstruction of ARG harboring genome accessions. Isolation source category and the presence of each ARG are displayed as different tracks. Red rectangles show ARGs that were found within 10,000 bps of iMGEs. The phylogenetic tree is shown as a cladogram.

resistance in foodborne pathogens (Ammor et al., 2007; van Reenen and Dicks, 2011). Such a transfer may occur from the LAB to members of the human or animal microbiota, then eventually to potential pathogens. LABs in probiotics or in food may acquire the genes from pathogens or commensals and then serve as a vector to environmental bacteria when passed to the environment. This risk is highlighted by the higher occurrence of resistance genes in LABs from human and animal sources. For the safe development of probiotics the mechanisms of plasmid transfer in LAB and the factors that influence the spread of plasmids between LAB and other bacteria in complex microbial environments and community (such as the gut microbiota) should be extensively studied (Mayo et al., 2014).

In this study, we evaluated the ARG repertoire of lactobacilli using the RefSeq database as source of annotated high-quality genomes. Although this approach may have prevented the identification of a more complex landscape of ARG repertoire by omitting a fairly large amount of data (e.g., fragmented genomes, data from metagenomic surveys, etc.), we felt that the ARG diversity in LAB of the order Lactobacillales is more than satisfactorily illustrated with the analysis of over 4,000 bacterial reference genomes, and sacrificed exhaustive analysis to avoiding use of unconfirmed and potentially misleading data in the study. In similar studies, metadata (such as host and isolation source) are of paramount importance, since the prevalence and spreading potential of ARGs can be only assessed in their presence. The accessions downloaded from RefSeq had a high proportion of missing metadata (4% for ARG-positive samples and 14% for iMGE-positive samples). This observation aligns with the findings of Yarmosh et al. (2022), who also reported a lack of metadata in RefSeq and called for the deep curation of metadata to aid the reusability of the data and increase the reproducibility of research. The imbalanced number of samples classified by their isolation source (Table 1) highlights the importance of data generation in microbial ecology to study microbial diversity (Mony et al., 2020) and the need for high-quality data sharing (see Wilkinson et al., 2016).

In addition to plasmid transmissibility, the linkage of ARGs to iMGEs can be another potential threat. Not surprisingly, Lactobacillales harboring ARG and even ARG linked to iMGEs, were 10–100 times more commonly found in isolates originating from animal and human sources than those isolated from other sources. As production animals, pets and humans are frequently treated with antibiotics, their bacteria are markedly more exposed, in general, than strains from other sources. This higher selective pressure may result in higher risk of horizontal transfer of resistance genes from human or veterinary pathogens. Probiotics are consumed frequently by animals and humans, the increased opportunity for such transfers may also be due to consumption of large amounts of these bacteria as probiotics (Imperial and Ibana, 2016; Liu et al., 2020).

These results draw attention yet again to the One Health aspects of ARGs. LAB present in fermented plant-derived food or in dairy products consumed may exchange ARGs with LAB as well as other bacteria associated with the human or animal microbiomes, possibly acting as their sources or vectors along the food chain, e.g., probiotic strains used in animal feed or their genes in other host bacteria may persist in the animal product and reach consumers. ARGs harbored in Lactobacillales warn for caution when creating and using probiotics, especially in veterinary practice where use of probiotics is expected to rise.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

EK: Formal Analysis, Methodology, Writing – original draft. LL: Formal Analysis, Methodology, Visualization, Writing – original draft. GK: Conceptualization, Funding acquisition, Supervision, Writing – original draft. KB: Conceptualization, Funding acquisition, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The work is supported by project GINOP-2.3.4-15-2020-00008. The project is cofinanced by the European Union and the European Regional Development Fund. Additional support was provided by the National Laboratory for Infectious Animal Diseases, Antimicrobial Resistance, Veterinary Public Health, and Food Chain Safety, RRF-2.3.1-21-2022-00001 and the SA-27/2021 grant of the Eötvös Loránd Research Network.

Acknowledgments

The financial support of the funders are gratefully acknowledged.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1281473/full#supplementary-material

References

Abid, M. B., and Koh, C. J. (2019). Probiotics in health and disease: fooling mother nature? *Infection* 47, 911–917. doi: 10.1007/s15010-019-01351-0

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1006/jmbi.1990.9999

Al-Yami, A. M., Al-Mousa, A. T., Al-Otaibi, S. A., and Khalifa, A. Y. (2022). Lactobacillus species as probiotics: isolation sources and health benefits. *J. Pure Appl. Microbiol.* 16, 2270–2291. doi: 10.22207/Jpam.16.4.19

Amato, K. R., Mallott, E. K., Maia, P. D., and Sardaro, M. L. S. (2021). Predigestion as an evolutionary impetus for human use of fermented food. *Curr. Anthropol.* 62, S207–S219. doi: 10.1086/715238

Ammann, A., Neve, H., Geis, A., and Heller, K. J. (2008). Plasmid transfer via transduction from *Streptococcus thermophilus* to *Lactococcus lactis*. *J. Bacteriol.* 190, 3083–3087. doi: 10.1128/JB.01448-07

Ammor, M. S., Florez, A. B., and Mayo, B. (2007). Antibiotic resistance in non-enterococcal lactic acid bacteria and bifidobacteria. *Food Microbiol.* 24, 559–570. doi: 10.1016/j.fm.2006.11.001

Boethius, A. (2016). Something rotten in Scandinavia: the world's earliest evidence of fermentation. *J. Archaeol. Sci.* 66, 169–180. doi: 10.1016/j.jas.2016.01.008

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, D. S. (2004). Recent genetic transfer between Lactococcus lactis and enterobacteria. *J. Bacteriol.* 186, 6671–6677. doi: 10.1128/Jb.186.19.6671-6677.2004

Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4:e1660. doi: 10.7717/peerj.1660

Cui, Y., Wang, S., Ding, S., Shen, J., and Zhu, K. (2020). Toxins and mobile antimicrobial resistance genes in Bacillus probiotics constitute a potential risk for one health. *J. Hazard. Mater.* 382:121266. doi: 10.1016/j.jhazmat.2019.121266

Dowarah, R., Verma, A. K., and Agarwal, N. (2017). The use of Lactobacillus as an alternative of antibiotic growth promoters in pigs: a review. *Anim. Nutr.* 3, 1–6. doi: 10.1016/j.aninu.2016.11.002

Garcia-Hernandez, Y., Perez-Sanchez, T., Boucourt, R., Balcazar, J. L., Nicoli, J. R., Moreira-Silva, J., et al. (2016). Isolation, characterization and evaluation of probiotic lactic acid bacteria for potential use in animal production. *Res. Vet. Sci.* 108, 125–132. doi: 10.1016/j.rvsc.2016.08.009

Hill, C., Guarner, F., Reid, G., Gibson, G. R., Merenstein, D. J., Pot, B., et al. (2014). Expert consensus document. The international scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. *Nat. Rev. Gastroenterol. Hepatol.* 11, 506–514. doi: 10.1038/nrgastro.2014.66

Igimi, S., Ryu, C. H., Park, S. H., Sasaki, Y., Sasaki, T., and Kumagai, S. (1996). Transfer of conjugative plasmid pAMβ1 from *Lactococcus lactis* to mouse intestinal bacteria. *Lett. Appl. Microbiol.* 23, 31–35. doi: 10.1111/j.1472-765X.1996.tb00023.x

Imperial, I. C., and Ibana, J. A. (2016). Addressing the antibiotic resistance problem with probiotics: reducing the risk of its double-edged sword effect. *Front. Microbiol.* 7:1983. doi: 10.3389/fmicb.2016.01983

Johansson, M. H. K., Bortolaia, V., Tansirichaiya, S., Aarestrup, F. M., Roberts, A. P., and Petersen, T. N. (2021). Detection of mobile genetic elements associated with antibiotic resistance in *Salmonella enterica* using a newly developed web tool: MobileElementFinder. *J. Antimicrob. Chemother.* 76, 101–109. doi: 10.1093/jac/dkaa390

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Krawczyk, P. S., Lipinski, L., and Dziembowski, A. (2018). PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Res.* 46:e35. doi: 10.1093/nar/gkx1321

Leeuwendaal, N. K., Stanton, C., O'Toole, P. W., and Beresford, T. P. (2022). Fermented foods, health and the gut microbiome. *Nutrients* 14:1527. doi: 10.3390/nu14071527

Liu, Y., Tong, Z., Shi, J., Jia, Y., Yang, K., and Wang, Z. (2020). Correlation between exogenous compounds and the horizontal transfer of plasmid-borne antibiotic resistance genes. *Microorganisms* 8:1211. doi: 10.3390/microorganisms8081211

Marco, M. L., Heeney, D., Binda, S., Cifelli, C. J., Cotter, P. D., Foligne, B., et al. (2017). Health benefits of fermented foods: microbiota and beyond. *Curr. Opin. Biotechnol.* 44, 94–102. doi: 10.1016/j.copbio.2016.11.010

Mayo, B., Rachid, C. T., Alegria, A., Leite, A. M., Peixoto, R. S., and Delgado, S. (2014). Impact of next generation sequencing techniques in food microbiology. *Curr. Genomics* 15, 293–309. doi: 10.2174/1389202915666140616233211

McArthur, A. G., Waglechner, N., Nizam, F., Yan, A., Azad, M. A., Baylay, A. J., et al. (2013). The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* 57, 3348–3357. doi: 10.1128/AAC.00419-13

Montassier, E., Valdes-Mas, R., Batard, E., Zmora, N., Dori-Bachash, M., Suez, J., et al. (2021). Probiotics impact the antibiotic resistance gene reservoir along the human GI tract in a person-specific and antibiotic-dependent manner. *Nat. Microbiol.* 6, 1043–1054. doi: 10.1038/s41564-021-00920-0

Mony, C., Vandenkoornhuyse, P., Bohannan, B. J. M., Peay, K., and Leibold, M. A. (2020). A landscape of opportunities for microbial ecology research. *Front. Microbiol.* 11:561427. doi: 10.3389/fmicb.2020.561427

Ortiz Charneco, G., Kelleher, P., Buivydas, A., Streekstra, H., van Themaat, E. V. L., de Waal, P. P., et al. (2021). Genetic dissection of a prevalent plasmid-encoded conjugation system in *Lactococcus lactis*. *Front. Microbiol.* 12:680920. doi: 10.3389/fmicb.2021.680920

Ouwehand, A. C., Forssten, S., Hibberd, A. A., Lyra, A., and Stahl, B. (2016). Probiotic approach to prevent antibiotic resistance. *Ann. Med.* 48, 246–255. doi: 10.3109/07853890.2016.1161232

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490

R Core Team (2022). "R: A language and environment for statistical computing.". R Foundation for Statistical Computing, Vienna, Austria.

Radovanovic, M., Kekic, D., Gajic, I., Kabic, J., Jovicevic, M., Kekic, N., et al. (2023). Potential influence of antimicrobial resistance gene content in probiotic bacteria on the gut resistome ecosystems. *Front. Nutr.* 10:1054555. doi: 10.3389/fnut.2023.1054555

Rodgers, B., Kirley, K., and Mounsey, A. (2013). PURLs: prescribing an antibiotic? Pair it with probiotics. *J. Fam. Pract.* 62, 148–150.

Rozman, V., Lorbeg, P. M., Accetto, T., and Matijasic, B. B. (2020). Characterization of antimicrobial resistance in lactobacilli and bifidobacteria used as probiotics or starter cultures based on integration of phenotypic and in silico data. *Int. J. Food Microbiol.* 314:108388. doi: 10.1016/j.ijfoodmicro.2019.108388

Schleifer, K. H., Kraus, J., Dvorak, C., Kilpperbalz, R., Collins, M. D., and Fischer, W. (1985). Transfer of Streptococcus-Lactis and related streptococci to the genus Lactococcus gen-Nov. *Syst. Appl. Microbiol.* 6, 183–195. doi: 10.1016/S0723-2020(85)80052-7

Seemann, T. (2020). Abricate. Github. Available at: https://github.com/tseemann/abricate

Shah, A. M., Tarfeen, N., Mohamed, H., and Song, Y. D. (2023). Fermented foods: their health-promoting components and potential effects on gut microbiota. *Fermentation* 9:1527. doi: 10.3390/fermentation9020118

Sharma, P., Tomar, S. K., Goswami, P., Sangwan, V., and Singh, R. (2014). Antibiotic resistance among commercially available probiotics. *Food Res. Int.* 57, 176–195. doi: 10.1016/j.foodres.2014.01.025

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Toth, A. G., Csabai, I., Judge, M. F., Maroti, G., Becsei, A., Spisak, S., et al. (2021). Mobile antimicrobial resistance genes in probiotics. *Antibiotics* 10:1287. doi: 10.3390/antibiotics10111287

van Reenen, C. A., and Dicks, L. M. (2011). Horizontal gene transfer amongst probiotic lactic acid bacteria and other intestinal microbiota: what are the possibilities? *Arch. Microbiol.* 193, 157–168. doi: 10.1007/s00203-010-0668-3

Wade, T., Rangel, L. T., Kundu, S., Fournier, G. P., and Bansal, M. S. (2020). Assessing the accuracy of phylogenetic rooting methods on prokaryotic gene families. *PLoS One* 15:e0232950. doi: 10.1371/journal.pone.0232950

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3:160018. doi: 10.1038/sdata.2016.18

Wong, A., Saint Ngu, D. Y., Dan, L. A., Ooi, A., and Lim, R. L. H. (2015). Detection of antibiotic resistance in probiotics of dietary supplements. *Nutr. J.* 14:95. doi: 10.1186/s12937-015-0084-2

Xu, S., Dai, Z., Guo, P., Fu, X., Liu, S., Zhou, L., et al. (2021). ggtreeExtra: compact visualization of richly annotated phylogenetic data. *Mol. Biol. Evol.* 38, 4039–4042. doi: 10.1093/molbev/msab166

Yarmosh, D. A., Lopera, J. G., Puthuveetil, N. P., Combs, P. F., Reese, A. L., Tabron, C., et al. (2022). Comparative analysis and data provenance for 1,113 bacterial genome assemblies. *mSphere* 7:e0007722. doi: 10.1128/msphere.00077-22

Zheng, J., Wittouck, S., Salvetti, E., Franz, C., Harris, H. M. B., Mattarelli, P., et al. (2020). A taxonomic note on the genus Lactobacillus: description of 23 novel genera, emended description of the genus Lactobacillus Beijerinck 1901, and union of Lactobacillaceae and Leuconostocaceae. *Int. J. Syst. Evol. Microbiol.* 70, 2782–2858. doi: 10.1099/ijsem.0.004107

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 27 November 2023
DOI 10.3389/fmicb.2023.1293443

OPEN ACCESS

EDITED BY
Yuji Morita,
Meiji Pharmaceutical University, Japan

REVIEWED BY
Sara Domingues,
University of Coimbra, Portugal
Alexandro Rodríguez-Rojas,
University of Veterinary Medicine
Vienna, Austria

*CORRESPONDENCE
Mengnan Jiang
jiangmn@chinacdc.cn
Duochun Wang
wangduochun@icdc.cn
Qiang Wei
weiqiang@chinacdc.cn

RECEIVED 13 September 2023
ACCEPTED 26 October 2023
PUBLISHED 27 November 2023

CITATION
Mei L, Song Y, Liu D, Li Y, Liu L, Yu K, Jiang M, Wang D and Wei Q (2023) Characterization of a mobilizable megaplasmid carrying multiple resistance genes from a clinical isolate of *Pseudomonas aeruginosa*. *Front. Microbiol.* 14:1293443. doi: 10.3389/fmicb.2023.1293443

Characterization of a mobilizable megaplasmid carrying multiple resistance genes from a clinical isolate of *Pseudomonas aeruginosa*

Li Mei[1], Yang Song[2], Dongxin Liu[1], Yixiao Li[1], Li Liu[1], Keyi Yu[3], Mengnan Jiang[1]*, Duochun Wang[3]* and Qiang Wei[1]*

[1]National Pathogen Resource Center, Chinese Center for Disease Control and Prevention, Beijing, China, [2]Division of Infectious Disease, Key Laboratory of Surveillance and Early-Warning on Infectious Disease, Chinese Center for Disease Control and Prevention, Beijing, China, [3]National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China

Introduction: The horizontal transfer of antibiotic resistance genes mediated by plasmids seriously hinders the effectiveness of modern medical treatment, and thus has attracted widespread attention. Additionally, the co-selection mechanism of antibiotic resistance genes (ARGs) and heavy metal resistance genes (MRGs) on mobile elements may further exacerbate the horizontal transfer of resistance genes.

Methods: In this study, a multidrug-resistant *Pseudomonas aeruginosa* strain, termed BJ86 (CHPC/NPRC1.4142), was isolated from a patient's sputum specimen. *In vitro* tests for antimicrobial susceptibility, conjugation, whole-genome sequencing, and bioinformatics analysis were used to explore the potential mechanisms of resistance and its spread.

Results and discussion: Sequencing analysis indicates that *P. aeruginosa* BJ86 carries an amazing 522.5 kb-length megaplasmid, pBJ86, which contained a 93.5 kb-length multiple resistance region (MRR); 18 kinds of genes were identified as ARGs in this region, including *tmexCD-oprJ*, bla_{DIM-1}, *qnrVC6* that mediate resistance to multiple antibiotics and the operons *mer* that mediates heavy metal mercury resistance. In addition, there is also an 80 kb variable region (VR) on the plasmid pBJ86, and the genes encoding relaxase and type IV coupling protein (T4CP) were determined in this region, both of which are related to the conjugation and transfer ability of the plasmid. Bioinformatics analysis shows that many functional genes have insertion sequences and transposases on their flanks, which may have accumulated in the plasmid pBJ86 after multiple acquisition events. Conjugated transfer and *in vitro* tests for antimicrobial susceptibility verified the mobility and plasmid pBJ86-mediated resistance. To our knowledge, we are the first to report a mobilizable megaplasmid that simultaneously carried *tmexCD-oprJ*, bla_{DIM-1}, *qnrVC6*, and the operons *mer* and can be transferred with frequencies of 6.24×10^{-7} transconjugants per donor cell.

KEYWORDS

***Pseudomonas aeruginosa*, mobilizable plasmid, conjugation transfer, *tmexCD-oprJ*, bla_{DIM-1}, *qnrVC6*, *mer* genes**

1 Introduction

Pseudomonas aeruginosa is a Gram-negative opportunistic pathogen that is one of the most common causes of acute infection in patients with immune dysfunction or other susceptible diseases, often leading to poor prognosis in critically ill patients (Sievert et al., 2013; Nathwani et al., 2014; Bassetti et al., 2018a,b). In the United States, *P. aeruginosa* has

been listed by the Infectious Diseases Society of America as one of the six most dangerous clinical pathogens worldwide. In China, *P. aeruginosa* has also been one of the most common Gram-negative bacteria in clinical practice in the past 20 years, with a prevalence rate third only to *Klebsiella pneumoniae* and *Escherichia coli* (Hu et al., 2016).

Due to the inherent structure and characteristics of *P. aeruginosa*, it has a high intrinsic resistance to many antibiotics (Pang et al., 2019). Meanwhile, bacteria can also obtain antibiotic resistance genes (ARGs) through spontaneous genetic mutations or horizontal gene transfer (HGT). The former can mainly be transmitted to offspring through vertical gene transfer (VGT), while the latter has the potential to spread ARGs faster and more extensively, significantly promoting the emergence and development of multidrug-resistant strains (Li and Zhang, 2022). For *P. aeruginosa*, *tmexCD-oprJ, bla*$_{DIM-1}$, *and qnrVC6,* can encode resistance to different types of antibiotics. The *tmexCD-oprJ* Operon encodes an efflux pump, and its expression is usually affected by the regulatory factor *nfxB*. When it is overexpressed, it is usually related to the resistance of levofloxacin, ciprofloxacin, and other fluoroquinolones, but it can also pump out other antibiotics, such as macrolide and tetracyclines. In addition, mutations in *mexD* may also alter the substrate specificity of efflux pumps, which is related to changes in bacterial resistance to antibiotics such as carbicillin and ceflozan-tazobactam (Gomis-Font et al., 2021; Lorusso et al., 2022). As a class B metallo-β-lactamase (MBL) gene, *bla*$_{DIM}$ mainly exists on plasmids and mediates resistance to almost all β-lactams, including broad-spectrum cephalosporin and carbapenem, but does not affect monobactams (Tada et al., 2017). *qnrVC6* is a newly emerging quinolone resistance gene in *Pseudomonas* sp., which may have an additive effect of quinolone resistance together with other genes in the *qnr* family, thus helping to obtain full quinolone resistance (Liu et al., 2018). Therefore, the coexistence of *mexCD-oprJ* with *bla*$_{DIM}$, *qnrVC6*, and other ARGs in mobile elements can accelerate the spread of ARGs and pose a serious threat to the effectiveness of current clinical treatment.

Recent research has shown that many substances, including antibiotics, non-antibiotic drugs, and environmental pollutants, can promote HGT (Alav and Buckner, 2023). In addition, the selective pressure exerted by heavy metals on pathogenic microorganisms is increasingly being recognized as a critical driving factor in promoting the selection and spread of antibiotic resistance in human and animal food chains (Capita and Alonso-Calleja, 2013). Heavy metals with sublethal concentrations can induce bacterial resistance (Di Cesare et al., 2016) and promote the horizontal transfer of resistance genes (Wang et al., 2015), a direct or indirect factor affecting antibiotic resistance. Heavy metal resistance genes (MRGs) have been found to often coexist with ARGs on plasmids, and it allows bacteria to obtain MRGs while also obtaining ARGs through co-selection mechanisms and vice versa (Baker-Austin et al., 2006). This potential mechanism may further exacerbate the horizontal transfer of ARGs.

This study describes the gene structure and characteristics of a mobilizable multiple resistance megaplasmid pBJ86 carried from a clinical isolate of *P. aeruginosa*, which can transfer via conjugation. The plasmid simultaneously carries ARGs *tmexCD-oprJ, bla*$_{DIM-1}$, *qnrVC6,* and MRGs in the operon *mer*.

2 Materials and methods

2.1 Bacterial strain and identification

In 2019, *P. aeruginosa* strain BJ86 (CHPC/NPRC1.4142) was isolated from a patient's sputum at the Friendship Hospital in Beijing, China. The strain is stored at the National Pathogen Resource Center (NPRC). Bacterial identification was performed using matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS).

2.2 Susceptibility testing

Minimum inhibitory concentrations (MICs) of 14 antimicrobial agents (Aztreonam, Cefepime, Ceftazidime, Piperacillin-zazobactam, Imipenem, Meropenem, Ciprofloxacin, Levofloxacin, Norfloxacin, Amikacin, Gentamicin, Tobramycin, Fosfomycin w/G6P, and Colistin) were detected by BD Phoenix™ M50 with NMIC-413. The results of Fosfomycin w/G6P refer to EUCAST (The European Committee on Antimicrobial Susceptibility Testing, 2020) and the results of other antibiotics were interpreted following the CLSI guidelines (Clinical Laboratory Standards Institute, 2023). *P. aeruginosa* ATCC27853 and *Escherichia coli* ATCC 25922 were used as control strains.

2.3 Genome sequencing and assembly

The genomes of the strains involved in the study were extracted by the Wizard Genomic DNA Extraction Kit (Promega, Madison, WI, USA) following the manufacturer's instructions and sequenced using the Illumina NovaSeq6000 and Oxford Nanopore Technologies MinION platforms. In order to improve the reliability of subsequent information analysis results, Sickle (https://githubcom/najoshi/sickle) was first used to process the raw data to obtain clean data. After sample quality control, clean data was assembled using Unicycler (https://github.com/rrwick/Unicycler) with the hybrid assembly strategy (Wick et al., 2017).

2.4 Bioinformatics analysis

Based on whole genome sequencing (WGS) results, further verification of the bacterial strain using EzBioCloud databases (https://www.ezbiocloud.net/tools/ani) was carried out to calculate Average Nucleotide Identity (ANI). The MLST2.0 databases was used to perform Multilocus sequence typing (MLST), the Resfinder4.1 databases (https://cge.food.dtu.dk/services/ResFinder/) to search for resistance genes, and the PlasmidFinder database (https://cge.food.dtu.dk/services/PlasmidFinder/) to identify plasmind type. All three databases are located in the Center for Genomic Epidemiology (CGE) server. The TAfinder database (https://bioinfo-mml.sjtu.edu.cn/TAfinder/index.php) was used to predict the Type II toxin-antitoxin (TA) systems, the tRNAscan SE 1.3.1 databases (http://lowelab.ucsc.edu/tRNAscan-SE/) to predict tRNA, codon 1.4.4-4 to analyze codon bias, the

CRISPRCasFinder database (https://crisprcas.i2bc.paris-saclay.fr/) to predict CRISPR, and the PHASTER database (http://phaster.ca) to carry out annotation phase sequences. Prokka1.14.6 was used for annotation, using default parameters. BLAST was utilized in the NCBI database to search for sequences with higher plasmid coverage and consistency in GenBank compared to this study. Sequence comparisons and map generation were performed using BRIG and Easyfig (Sullivan et al., 2011).

2.5 Plasmid conjugation method

The conjugation assay was performed by using BJ86 as the donor and rifampicin-resistant *P. aeruginosa* PAO1 as the recipient to identify the self-transfer ability of the plasmid in this study. Rifampicin (32 μg/mL) and meropenem (8 μg/mL) were the antibiotics and concentrations used for selection. The donor and recipient bacteria were mixed in a 1:1 volume ratio and shaken at 37°C and 180 r/min for 10 h. Next, 200 μL of the mixed liquid after cultivation was taken and dropped onto a dual-resistance plate [the Brain-Heart Infusion (BHI) agar plates containing 32 μg/mL rifampicin and 8 μg/mL meropenem]. After this, the mixed liquid was inverted and cultured at 37°C for 10–24 h. Successful conjugation was confirmed by detecting ARGs, susceptibility testing, and MLST of suspected transconjugants. The calculation formula for conjugation frequency is as follows: conjugation frequency = number of transconjugant bacterial colonies × 10^x/number of donor bacterial colonies × 10^y (x, y is the dilution factor). The growth kinetics of stains were detected using Bioscreen fully automated microbial growth curve analyzer under conditions of 37°C and OD_{600}.

2.6 Nucleotide sequence accession number

The complete sequences of the plasmid pBJ86 and the chromosomes of BJ86 were submitted to GenBank under accession numbers CP133755 and CP133756, respectively. In addition, the complete genome sequence of strain BJ86 can be obtained through the National Microbiology Data Center (www.nmdc.cn) using the number NMDC60134482.

3 Results

3.1 General features of the stain BJ86 and plasmid pBJ86 sequence

Genome *P. aeruginosa* BJ86, including a complete circular chromosome sequence (6,865,712 bp, GC accounting for 65.89%, MLST sequence type, ST2446 type) and a circular plasmid (pBJ86) sequence (522,477 bp, GC accounting for 56.62%), were identified. The plasmid pBJ86 has 599 predicted coding sequences (CDS), and 80.1% (480/599) of the CDS encode proteins of undetermined function (Figure 1; Supplementary Table S1). Comparing the complete sequence of pBJ86 with the plasmid sequence in Genbank, it was found that this sequence has the highest similarity to the plasmid pNY11173-DIM (Sequence ID: CP096957.1) from *P. aeruginosa* strain NY11173 (99.84% identity at 77% coverage), plasmid pBJP69-DIM (Sequence ID: MN208064.1) from *Pseudomonas sp.* strain BJP69 (99.87% identity at 77% coverage), and plasmid unnamed2 (Sequence ID: CP027170.1) from *P. aeruginosa* strain AR_0356 (98.07% identity at 68% coverage) (Figure 1).

3.2 Susceptibility testing of BJ86 and resistance genes

P. aeruginosa BJ86 is resistant to five categories and 10 agents of antibiotics, including two aminoglycosides (gentamicin and tobramycin), two antipseudomonal carbapenems (imipenem and meropenem), two antipseudomonal cephalosporins (cefepime and ceftazidime), three antipseudomonal fluoroquinolones (ciprofloxacin, levofloxacin, and norfloxacins), and one antipseudomonal penicillin + β-lactamase inhibitors (piperacillin-tazobactam) (Table 1). Due to the strain BJ86 being non-susceptible to at least one agent in ≥3 antimicrobial categories listed in Table 1, it belongs to a multidrug-resistant (MDR) strain (Magiorakos et al., 2012).

There are six kinds of ARGs [*aph (3′)-IIb, fosA, crpP,* bla_{PAO}, bla_{OXA-50}, and *catB7*] carried on the chromosome. The plasmid carries 16 kinds of ARGs, including six aminoglycoside resistance genes [*aph (3″)-Ib, aph (6″)-Id, aac (6′)-Ib cr, aadA1, aac (6′)-Ib Hangzhou, aac (6′)* and- *IIa*], two β-lactamase-encoding genes (bla_{DIM-1} and bla_{OXA-4}), one quinolone resistance gene (*qnrVC6*), one sulfonamide resistance gene (*sul1*), two trimethoprim resistance genes (*dfrA1 two dfrB7*), one quaternary ammonium compound resistance gene (*qacE*), one flufenicol resistance gene (*floR*), one chloramphenicol resistance genes (*catB*), and one operon *tmexCD-OprJ* encoding the antibiotic efflux pump. There are also two kinds of MRGs, including two structurally similar operons to *mer* that mediate resistance to mercury and one operon, *ter*, that mediates resistance to the oxygen ion form of tellurium.

3.3 Genetic contexts of multiple resistance region

There is a multiple resistance region (MRR) with a length of 93.5 kb on plasmid pPBJ86, which contains 107 CDS and a GC content of 58%. In the MRR, a total of 18 resistance genes were detected (Figure 2A). Among them, the operon *tmexCD-oprJ* encoding an efflux pump and the ARGs bla_{DIM-1}, *qnrVC6* can mediate the resistance of *P. aeruginosa* to multiple antibiotics, and the *mer* genes mediate the resistance to mercury.

Compared to pNY11173-DIM (Sequence ID: CP096957.1) and p12969-DIM (Sequence ID: KU130294.1), pBJ86 also has a complete genome structure encoding efflux pump and its regulatory factors *nfxB*, but it is evident that pBJ86 has a truncated *nfxB* (Figure 2B). The bla_{DIM} gene in pBJ86 was located upstream of the transposable element Tn*3* and in a Class 1 integron, with the gene arrangement *intI1-*bla_{DIM-1}*-aac(6′)-lb-*bla_{OXA-4}*-aadA1-*Δ3′-CS. The truncation of 74 bp 3′-CS is likely due

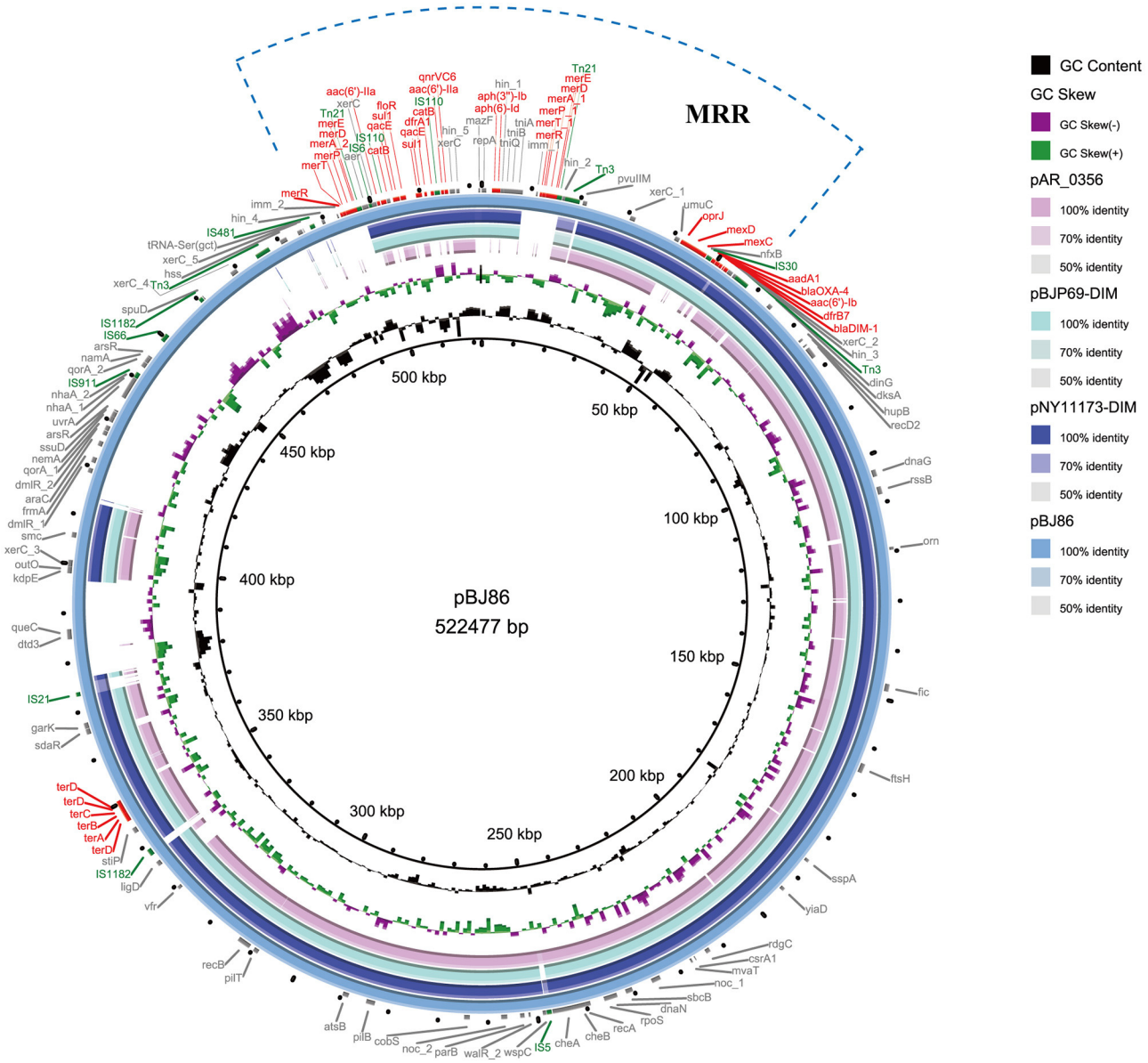

FIGURE 1
Comparative structural analysis of pBJ86 with other similar plasmids available in the NCBI nr database. The outermost circle represents the reference plasmid pBJ86. Genes involved in antimicrobial and heavy metal resistance are red, mobile elements are green, and genes involved in other functions are gray.

to the IS*30* family transposase IS*Pa59* insertion. The genetic environment downstream of bla_{DIM} is relatively conservative, while upstream there are many other ARGs inserted, such as the chloramphenicol resistance gene *catB* in pNY11173-DIM and p12969-DIM (Figure 2B). *QnrVC6* mediates quinolone resistance through targeted protection. Highly similar to p12969-DIM, the *qnrVC6* gene was located upstream of *intI1* and in a complex Class 1integron, with the gene arrangement *intI1-qnrVC6-aacA3-IS*110-*catB-qacEΔ1-sul1-IS*CR1-*floR-sul1-qacEΔ1*. The resistance gene upstream of *qnrVC6* has a duplicate copy of the gene arrangement *sul1-qacEΔ1-catB-IS*110-*aac(6′)-Ia*, with insertion sequences *IS*110 and *IS*CR1 on both sides. This phenomenon may be caused by inserting sequence events. Through BLASTn, we also found that the sequence composed of *qnrVC6* and its nearby upstream and downstream genes is highly similar to a segment on the AM23 chromosome of *Pseudomonas* sp. (Sequence ID: CP113432.1), indicating that this sequence may have undergone exchanges between chromosomes and plasmids (Figure 2C).

The MRR not only carries multiple ARGs but contains two similar operons responsible for mercury resistance and their regulator, *merRTPADE*. Through BLASTn, it can be found that there are highly similar regions on both plasmids and chromosomes of *Pseudomonas* sp. to the region containing mercury resistant genes on pBJ86, and their similarity can reach 100%. For example, a fragment on the chromosome of *Pseudomonas* sp. 2022CK-00068 (sequence ID: CP124658.1) and *Pseudomonas* sp.

10.3389/fmicb.2023.1293443

TABLE 1 Antimicrobial resistance pattern and resistant genes identified in *Pseudomonas aeruginosa* strain BJ86.

Antimicrobial class/agent	MIC(μg/mL) (Susceptibility)	Resistance genes	
		Plasmid	Chromosome
Aminoglycosides			
Amikacin	8[S]		
Gentamicin	>8[R]	*aac(6′)-IIa*	
Tobramycin	>8[R]	*aac(6′)-Ib-Hangzhou, aac(6′)-IIa*	
Antipseudomonal carbapenems			
Imipenem	>8[R]	bla_{DIM-1}	
Meropenem	>8[R]	bla_{DIM-1}	
Antipseudomonal cephalosporins			
Cefepime	>16[R]	bla_{OXA-4}	bla_{PAO}
Ceftazidime	>32[R]	bla_{DIM-1}	bla_{PAO}
Antipseudomonal fluoroquinolones			
Ciprofloxacin	>4[R]	*qnrVC6*	*crpP*
Levofloxacin	>8[R]	*tmexCD-OprJ*	
Norfloxacins	>8[R]	*tmexCD-OprJ*	
Antipseudomonal penicillins + β-lactamase inhibitors			
Piperacillin-Tazobactam	>64/4[R]	bla_{DIM-1}	
Monobactams			
Aztreonam	8[S]		
Phosphonic acids			
Fosfomycin w/G6P	64[<ECOFF]		*fosA*
Polymyxins			
Colistin	≤1[S]		

R, resistance; I, intermediate; S, susceptible; ECOFF, epidemiological break point based on *in vitro* drug sensitivity data.

BAMCP07-48 (sequence ID: CPU15377.1), and a segment on the plasmid of *Pseudomonas* sp. AR_0353 (sequence ID: CP027173.1). Additionally, *Tn21* can always be found on the flank of the mercury resistance gene (Figure 2D).

3.4 Conjugative system on plasmids

At the position of plasmid pBJ86 420 to 500 kb, a variable region (VR) with a length of ~80 kb exhibits low alignment with its similar plasmids. This region does not contain any resistance genes, but genes encoding relaxase and type IV coupling protein (T4CP) can be found in this region, which is closely related to the mobility of the plasmid. The sequence containing both the genes encoding Relaxase and T4CP (*traG*) were highly similar to the sequence on the chromosome of *P.aeruginosa* L00-a (Sequence ID: CP097383.1), with 100% identity at 95% coverage (Figure 3A). In contrast with the chromosome of *P.aeruginosa* L00-a, there are some mobile elements on this sequence of pBJ86, like *IS*66, *IS*1182, and *Tn*3, which may mediate the insertion of genes encoding the conjugative system.

3.5 Genetic contexts of ICE*BJ86* on chromosome

On the chromosome of BJ86, there is an integrative and conjugative element (ICE), ICE*BJ86*, which carries a fluoroquinolone resistance gene *crpP* and a gene region encoding type IV secretion system (T4SS). The GC content of ICEs is usually lower than their host strains. The typical insertion sites are tRNA or other highly-conserved genes (Sun et al., 2023). In this study, the GC content of ICE*BJ*86 is 60.45%, 5.44% less than that of chromosome and inserted into the $tRNA^{Lys}$ locus. Additionally, there are 54 bp direct repeats on both sides of it. Comparing the sequence of ICE*BJ*86 to those of NCBI, it was found that multiple strains of *Pseudomonas sp.* all carried this mobile element, such as *Pseudomonas sp.* SGVI (GenBank: KT887560.1) and *Pseudomonas sp.* 2021CK-01197 (GenBank: CP124643.1) (Figure 3B). However, *Pseudomonas* sp. SGVI was isolated from France in 1992 as ST111 type, and *Pseudomonas* sp. 2021CK-01197 from the United States in 2021 as ST308 type, with significant differences in time, location, and ST type, indicating that this type of ICE tends to be inherited through horizontal rather than vertical transmission.

10.3389/fmicb.2023.1293443

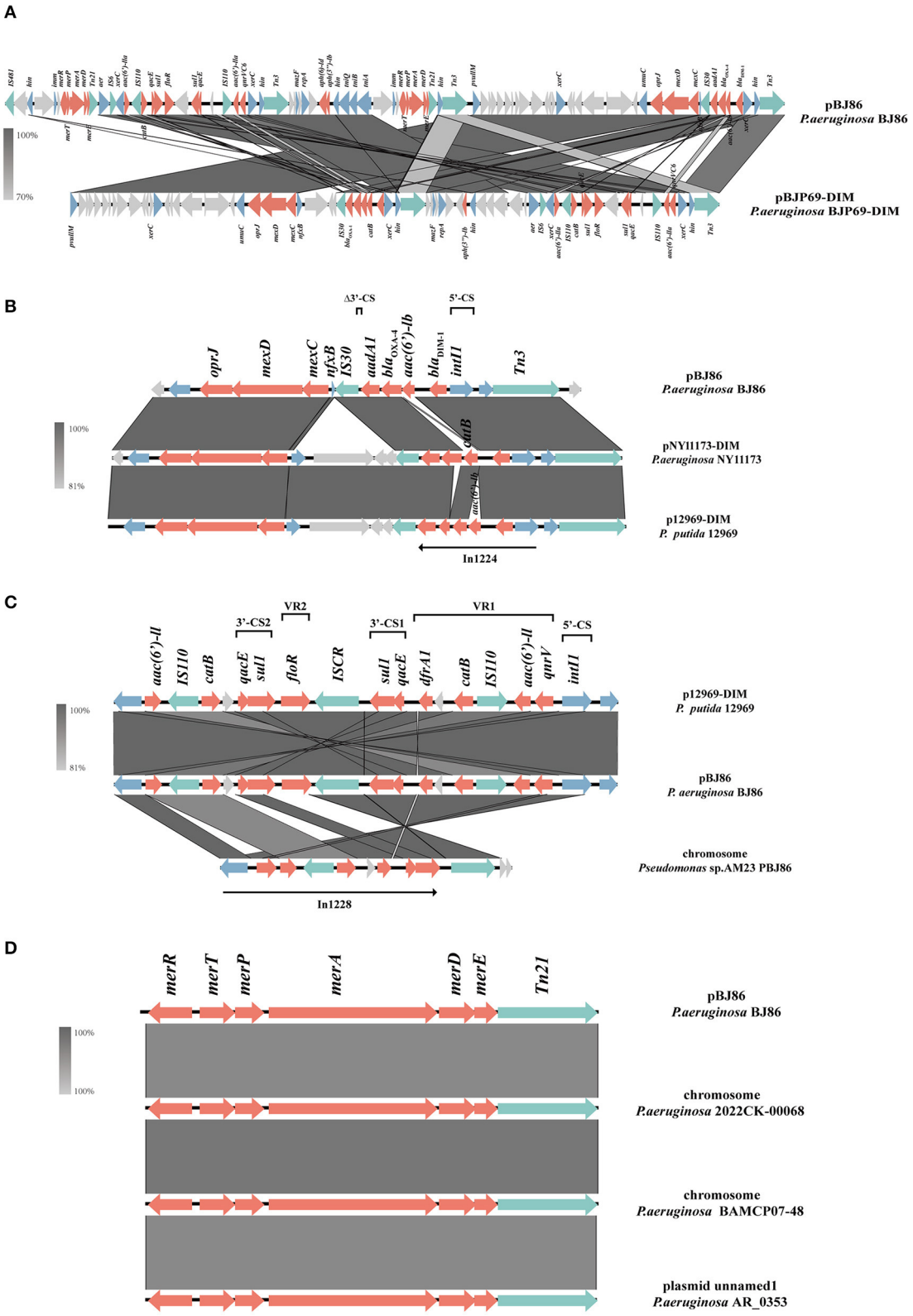

FIGURE 2
Schematic presentation of multiple resistance regions. Genes are presented as arrows, with the arrowhead indicating the direction of transcription. Genes involved in antimicrobial resistance are in red, mobile elements are in green, genes involved in other functions are in blue, and genes involved in undetermined coding functions are in light gray. Homologous segments generated by a BLASTn comparison (>70% identity) are gray boxes. Alignment of genetic contexts of the MRR with a length of 93.5 kb **(A)**, *tmexCD-oprJ* **(B)**, bla_{DIM-1} **(B)**, *qnrVC6* **(C)**, and *merRTPADE* **(D)** in the plasmid pBJ86-MRR are with similar sequences.

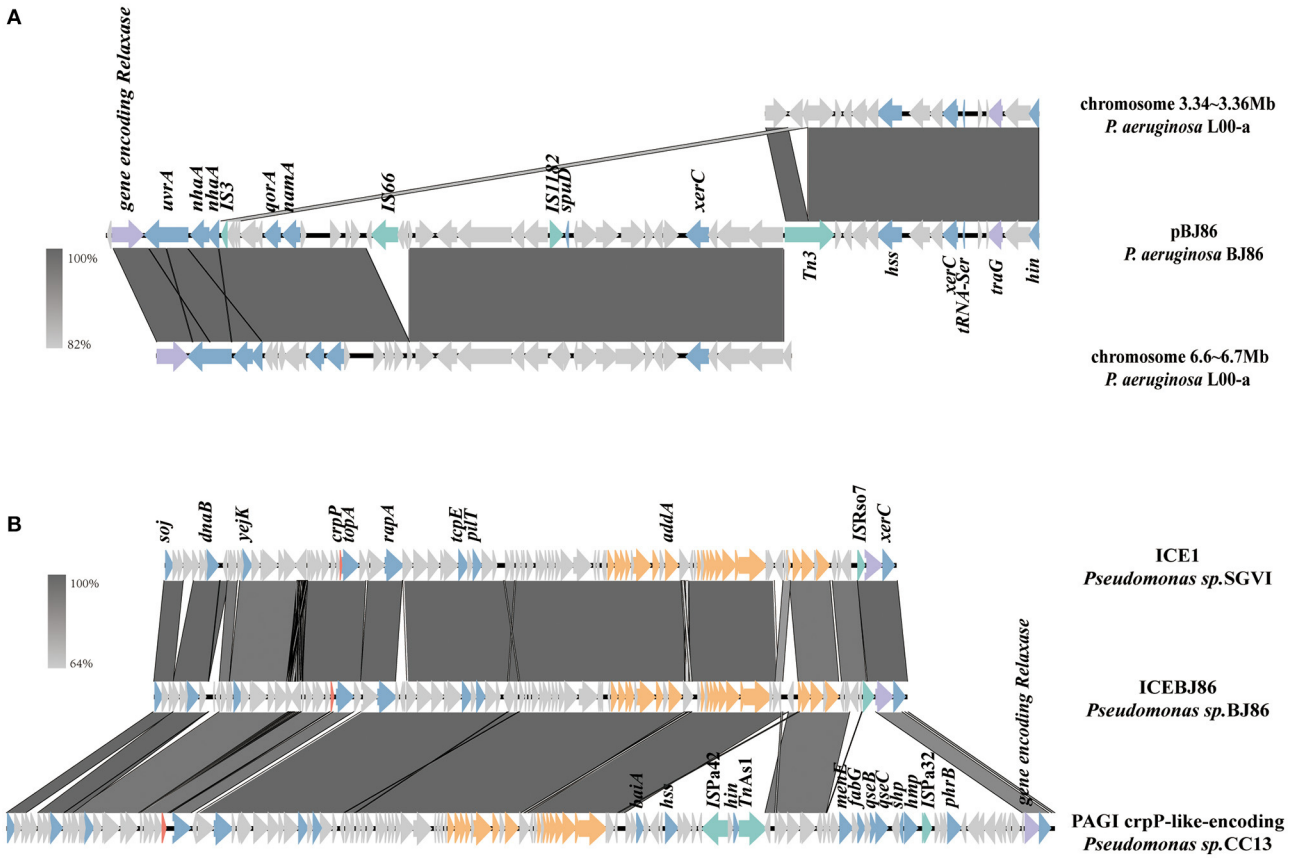

FIGURE 3
Schematic presentation of conjugative systems. Genes are presented as arrows, with the arrowhead indicating the direction of transcription. Genes involved in antimicrobial and heavy metal resistance are in red, mobile elements are in green, genes involved in the Type 4 secretion system are in orange, genes encoding relaxase and T4CP are in purple, genes involved in other functions are in blue, and genes involved in undetermined coding function are in light gray. Homologous segments generated by a BLASTn comparison (>64% identity) are shown in gray boxes. The alignment of genetic contexts of the conjugative system in plasmid pBJ86 **(A)** and ICEBJ86 **(B)** on the chromosome of BJ86 are with similar sequences.

3.6 Conjugation transfer experiment

Screening for resistance genes and MLST typing through second-generation sequencing data susceptibility testing confirmed that the plasmid pBJ86 could be transferred to rifampicin-resistant *Pseudomonas aeruginosa* PAO1, with frequencies of 6.24×10^{-7} transconjugants per donor cell. Firstly, the MLST typing of the donor strain BJ86 is ST2446, while the MLST typing of the recipient strain PAO1 is consistent with that of the transconjugant T86, both ST549. Secondly, based on the screening results of resistance genes, all 13 resistance genes carried by pBJ86 can also be screened on the transconjugant T86. However, we did not identify the fluoroquinolone resistance gene *crpP* carried by ICE*BJ86* on the chromosome of BJ86 on T86. Lastly, the sensitivity test further confirmed the success of the conjugation transfer experiment. As shown in Table 2, in comparison with PAO1, the transconjugant T86 showed increased MICs of cefepime, ceftazidime, ciprofloxacin, gentamicin, imipenem, levofloxacin, meropenem, norfloxacin, piperacillin-tazobactam, and tobramycin. Among them, the antibiotic sensitivity level of cefepime, ceftazidime, gentamicin, meropenem, and piperacillin-tazobactam has changed from sensitivity to resistance. The antibiotic sensitivity level of imipenem and norfloxacin has changed from intermediate to resistant. The antibiotic sensitivity level of tobramycin has changed from sensitivity to intermediate.

In addition, the growth curve of the receptor strain PAO1 before and after conjugation showed that there is no significant difference in the lag phase and logarithmic phase, and the OD value of PAO1 in the logarithmic phase is slightly higher than T86 (Figure 4). However, compared to the donor strain BJ86, which has a longer stationary phase, they both rapidly enter the decline phase after a brief stationary phase (about 1 h).

4 Discussion

As of 2022, the largest plasmid found to be carried by *P. aeruginosa* is a multidrug resistance plasmid pPAG5, which was found in a clinical strain of *P. aeruginosa* PAG5 isolated from a patient's urine, with a size of 513.3 kb (Li et al., 2022). Compared to pPAG5, pBJ86 has a larger size of 522.5 kb, and they both belong to megaplasmids. The emergence of megaplasmids is related to coping with different selection pressures, as they typically carry more resistance genes to enhance the adaptability of host microorganisms or (and) more effectively maintain their stability

10.3389/fmicb.2023.1293443

TABLE 2 Comparison of antimicrobial susceptibility test results before and after conjugation transfer experiment.

Strains	MIC (mg/L)													
	AN	ATM	FEP	CAZ	CIP	CL	GM	IMP	LVX	MEM	NOR	TZP	TM	FOS	TGC
BJ86	≤8(S)	8(S)	>16(R)	>32(R)	>4(R)	≤1(S)	>8(R)	>8(R)	>8(R)	>8(R)	>8(R)	>64/4(R)	>8(R)	64[a]	8
PAO1	≤8(S)	≤2(S)	≤1(S)	2(S)	2(R)	≤1(S)	≤2(S)	4(I)	4(R)	1(S)	8(I)	≤4/4(S)	≤2(S)	128	4
T86	≤8(S)	≤2(S)	>16(R)	>32(R)	>4(R)	≤1(S)	>8(R)	>8(R)	>8(R)	>8(R)	>8(R)	>64/4(R)	8(I)	64	8

AN, Amikacin; ATM, Aztreonam; FEP, Cefepime; CAZ, Ceftazidime; CIP, Ciprofloxacin; CL, Colistin; GM, Gentamicin; IMP, Imipenem; LVX, Levofloxacin; MEM, Meropenem; NOR, Norfloxacin; TZP, Piperacillin-Tazobactam; TM, Tobramycin; FOS, Fosfomycin w/G6P; TGC, Tigecycline.
S, susceptible; R, resistance; I, intermediate.
[a]The sensitivity of the donor bacteria BJ86, recipient bacteria PAO1, and transconjugant T86 to fosfomycin is lower than the ECOFF value (ECOFF ≤ 128 mg/L).

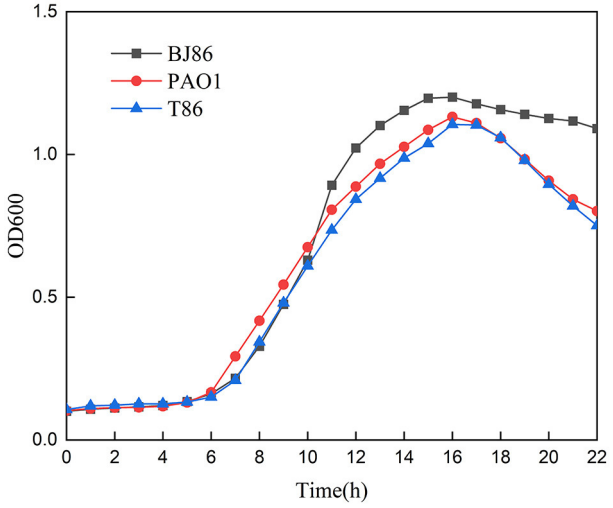

FIGURE 4
Growth curve of the donor stain BJ86, the recipient stain PAO1, and the transconjugant T86.

and transmission. In the *Pseudomonas*, these megaplasmids have the characteristics of high stability, low adaptation cost, and efficient transmission, making them effective vehicles for gene exchange. In *Salmonella Enterica*, *Agrobacterium tumefaciens*, and *Acinetobacter*, the megaplasmids carried by them also play an important role in the spread of AMR and tumor formation (Gordon and Christie, 2014; Cazares et al., 2020; Hall et al., 2022).

Plasmid pBJ86 carries many resistance-related genes, such as the genes encoding an efflux pump *tmexCD-oprJ*, which can mediate multiple antibiotic resistance, and the repressor *nfxB*, which mainly regulates its expression. Several mutations in *nfxB* have been related to the increased expression of *tmexCD-oprJ*, such as nucleotide deletions and missense and non-sense mutations. When overexpressed, it is associated with resistance to several antibiotics. In addition, recent studies have shown that *tmexCD-oprJ* is directly associated with tigecycline resistance and reduced sensitivity to other antibiotics under laboratory conditions (Lv et al., 2020; Lorusso et al., 2022). There is a significant truncation of *nfxB* on the plasmid pBJ86, so we tested the MIC of tigecycline against donor bacteria, recipient bacteria, and transconjugant by an antimicrobial susceptibility test *in vitro*. The results showed that the resistance to tigecycline increased from 4 to 8 μg/mL after conjugation. Compared with tetracycline, tigecycline has a broader antimicrobial spectrum and more potent antibacterial activity and can overcome the generation of a tetracycline resistance mechanism in most bacteria. Tigecycline has good antibacterial activity against most Gram-positive and Gram-negative aerobic and anaerobic bacteria *in vitro*, including *S. aureus, Enterococcus* spp., *S. pneumoniae, Haemophilus influenzae, Enterobacteriaceae*, and *Bacteroides* spp. (Pankey, 2005). For multidrug-resistant Gram-negative bacteria, especially carbapenem-resistant *Enterobacteriaceae*, tigecycline is a last line of defense drug together with polymyxin. *Pseudomonas aeruginosa* is naturally resistant to tigecycline (Dean et al., 2003), and tigecycline is usually not included in the clinical treatment of

Pseudomonas aeruginosa infection (Pulmonary Infection Assembly of Chinese Thoracic Society, 2022). A *tmexCD-oprJ*-like gene cluster has been identified on the plasmids of several *Pseudomonas sp.* strains in the Genbank sequence (Bassetti et al., 2018a). If *tmexCD-oprJ* can become widespread among different species strains through plasmids and other mobile elements and mediate tigecycline resistance of bacteria coexisting with it, then we may need to re-examine the public health significance of the *tmexCD-oprJ* carried by a plasmid of *Pseudomonas* sp. In addition, pBJ86 also carries ARGs that mediate multiple antibiotic resistance, such as $bla_{\mathrm{DIM-1}}$, $bla_{\mathrm{OXA-4}}$, *qnrVC6,* and *aac(6′)-la,* which mediate β-lactams, quinolone, and aminoglycoside antibiotic resistance. Although the bla_{DIM} gene has been detected worldwide (Leski et al., 2013; Sun et al., 2016; Tran et al., 2021; Delgado-Blas et al., 2022), compared to other commonly detected MBL genes such as bla_{IMP}, bla_{VIM}, and bla_{NDM}„ pathogenic microorganisms carrying the bla_{DIM} gene are still less reported. Many studies have not included the bla_{DIM} gene in screening strains carrying the MBL gene using Polymerase Chain Reaction (PCR) technology (Murugan et al., 2018; Chen et al., 2022; Delgado-Blas et al., 2022). However, $bla_{\mathrm{DIM-1}}$, like other MBL genes, can also hydrolyze almost all β-lactams, which seriously hinders the effectiveness of clinical treatment (Tada et al., 2017). We should strengthen the monitoring of this MBL gene in future research.

In addition to possessing many ARGs, pBJ86 carries two structurally similar HRGs, *merRTPADE*. These two MRGs are located in the MRR region of the plasmid pBJ86, adjacent to the aminoglycoside resistance genes. Elements carrying HRGs are often detected in contaminated environments, and sublethal concentrations of heavy metals can promote the horizontal transfer of plasmids carrying ARGs (Zhang et al., 2018). Unlike antibiotics, metals are not easily degraded, so that they can bring sustained and stable selection pressure to microorganisms. Some pathogenic strains of *Pseudomonas* have already established environmental hosts, and resistance genes can be horizontally transferred from environmental organisms to human symbionts, which can bring direct public health impacts (Baker-Austin et al., 2006). In recent years, more and more studies have detected heavy metal resistance genes in clinical isolates, which typically coexist with various ARGs on mobile elements (Perez-Palacios et al., 2021; Li et al., 2022), and may be related to the co-selection mechanism between resistance genes. The co-selection mechanism includes co-resistance and cross-resistance. They all lead to the same consequence: the development of resistance to one antibacterial agent accompanied by resistance to another agent (Baker-Austin et al., 2006). Wireman et al. (1997) demonstrated that strains with the *mer* genes are more likely to develop multiple drug resistance than strains without the *mer* genes. Other studies have shown that *mer* operons that can co-transfer genes related to β-lactams, aminoglycosides, quinolones, sulfonamides, trimethoprim antibiotics, and disinfectants are of great significance for the transmission of several determinants of antimicrobial resistance (Perez-Palacios et al., 2021). Their studies have again demonstrated the correlation between metal and antibiotic resistance. In the future, we can continue to explore the mechanisms by which metal pollutants enhance the occurrence and transmission of antibiotic resistance in pathogenic microorganisms.

Compared to other plasmids carrying the genes *tmexCD-oprJ*, bla_{DIM}, and *qnrVC6,* such as the plasmids containing DIM, including the first one discovered in *Pseudomonas stutzeri* in 2007 and one found to carry the DIM-1 variant DIM-2 in *Pseudomonas putida* in 2013, these plasmids have not been successfully transferred to the recipient strains. Similarly, the IncpRBL16 plasmid carrying *tmexCD3-toprJ3* found in *P. aeruginosa* could not conjugate transfer successfully (Dong et al., 2022). The plasmid pBM413 carrying *qnrVC6* was preliminarily determined immobile due to its lack of genes encoding relaxase and T4SS. To our knowledge, pBJ86 is the first plasmid to carry *tmexCD-oprJ*, bla_{DIM}, and *qnrVC6* simultaneously that can be successfully conjugated.

Compared to the other three methods of HGT [natural transformation, transduction, and vesicle-mediated transfer (Jiang et al., 2022)], conjugation is the most common mechanism of HGT and that with the broadest host range (Amábile-Cuevas and Chicurel, 1992). Conjugative systems involve an origin of transfer (oriT), a relaxase, a type IV coupling protein (T4CP), and a type IV secretion system (T4SS) (Getino and De La Cruz, 2018). According to the ability and characteristics of horizontal transfer, plasmids can be divided into three categories: conjugative plasmids, mobilizable plasmids, and non-conjugative plasmid. Among them, mobilizable plasmids only contain DNA transfer and replication systems but do not contain mating pair formation (Mpf). That is, they only contain Relaxase and T4CP. This type of plasmid often requires channels encoded by Mpf of another genetic element present in the cell for conjugation transfer, and Mpf is a form of T4SS (Smillie et al., 2010). Through gene level analysis, we can screen genes encoding T4CP and Relaxase in the low alignment region of plasmid pBJ86, but no genes encoding T4SS were found. Therefore, pBJ86 belongs to the mobilizable plasmid. Due to the successful conjugation transfer of pBJ86 and the absence of other conjugative plasmids in *Pseudomonas aeruginosa* BJ86, we subsequently analyzed the chromosomes of BJ86. The analysis results show that there are two T4SSs on the chromosome of BJ86, one located in ICE and the other in the chromosome genome outside of ICE. Further experimental research is needed to determine which T4SS-mediated plasmid conjugation is involved. Comparison of growth curves of receptor strain PAO1 before and after conjugation to the plasmid pBJ86 indicates that pBJ86 did not increase fitness costs or gains to the host bacteria. The rapid decline of T86 and PAO1 after the logarithmic phase may be a characteristic of the receptor strain PAO1 itself. Besides that, the toxin-antitoxin (TA) systems on plasmids can perform post segregational killing (PSK), thereby ensuring plasmid stability by killing plasmid-free daughter cells. However, we did not annotate the known TA systems on the plasmid pBJ86, indicating that strains carrying plasmid pBJ86 are likely to lose their plasmids after multiple subcultures (Hernández-Ramírez et al., 2017).

Like plasmids, ICE conjugation transfer also requires a complete system. There is an ICE*BJ86* with a complete conjugation system on the chromosome of BJ86. However, WGS analysis of the transconjugant T86 shows that its genome does not contain the ARG *crpP* carried by ICE*BJ86*, which means that ICE*BJ86* did not enter the receptor bacteria through conjugation, which may be because ICEs are quiescent elements located on the host chromosome in most cases, as activating ICEs can bring additional

stress to the host strain. Only under certain conditions can ICEs be induced, such as SOS response cell-cell signaling (quorum sensing), a selective advantage that the ICE provides to the host (Johnson and Grossman, 2015).

The selective pressure of the environment is forcing the evolution rate of microorganisms to accelerate day by day, and resistance mechanisms that can render multiple drugs ineffective have emerged and spread at an astonishing rate. This is a source of concern and corresponding strategies must be devised. In recent years, more and more research and reviews have focused on the factors that affect the process of conjugative transmission, and we find that many endogenous and exogenous factors have an impact on the complex process. From the perspective of endogenous factors, the characteristics of bacteria themselves, such as host bacterial pilus morphology, restriction modification (RM), and CRISPR Cas systems, as well as the exclusion and susceptibility inhibition of plasmids, are common factors affecting the conjugation process (Getino and De La Cruz, 2018). From the perspective of exogenous factors, antibiotics [ampicillin, gentamicin, and tetracycline (Wang et al., 2023)], non-antibiotic pharmaceuticals [antipyretic analgesics acetaminophen, antiepileptic drug carbamazepine, and anticancer drug paclitaxel (Yang et al., 2022)], chemicals [copper ions and other ionic liquids (Wang et al., 2015; Song et al., 2021)], pollutants generated by human activities in the external environment [microplastics and disinfection by-products (DBPs) (He et al., 2022; Weise et al., 2022)], and even substances once thought to be able to sterilize and inhibit bacteria in the external environment, will also accelerate the spread of ARGs. Indeed, current research suggests that many substances can inhibit this process. Conjugation inhibitors (COINs) have been used to target specific components of conjugation systems to inhibit the process of conjugation transfer, such as blocking the activity of relaxase in receptor bacteria by using relaxase inhibitors, inhibiting the formation of conjugation transfer channels by affecting Mpf, and preventing contact between donor bacteria and receptor bacteria by interfering with the function of conjugative pilus (Getino and De La Cruz, 2018). Synthetic fatty acids and MoS2-decorated nanocomposite Fe2O3@MoS2 are among the current strategies (Getino et al., 2015; Wang et al., 2019).

5 Conclusion

In conclusion, this study identified and characterized a 522.5 kb mobilizable megaplasmid, pBJ86, from a clinical isolate *P. aeruginosa* strain, BJ86. To our knowledge, this is the largest plasmid found so far in *P. aeruginosa*. Plasmid pBJ86 carries multiple resistance genes, like the antibiotic resistance-related genes *tmexCD-oprJ*, bla_{DIM-1}, and *qnrVC6,* and the mer operon system with resistance to mercury. We analyzed the genetic context of the abovementioned genes and the related genes involved in pBJ86 conjugation transfer. So far as we know, pBJ86 is the first plasmid to carry *tmexCD-oprJ*, bla_{DIM-1}, *qnrVC6,* and the *mer* operon system simultaneously and that can be successfully conjugated. The emergence and transfer of the abovementioned ARGs pose a severe threat to the effectiveness of clinical treatment. We should adopt a more systematic approach to monitor mobile elements carrying ARGs and take adequate measures to curb the horizontal spread of ARGs. This research may provide discoveries and insights into the genomic diversity and molecular evolution of *P. aeruginosa*., which helps to understand the emergence of multidrug-resistant bacteria and the transmission mechanism of resistance genes.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Ethics statement

Ethical approval was not required for the studies on humans in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used.

Author contributions

LM: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing—original draft. YS: Conceptualization, Investigation, Methodology, Supervision, Writing—review & editing. DL: Methodology, Supervision, Writing—review & editing. YL: Methodology, Supervision, Writing—review & editing. LL: Methodology, Visualization, Writing—review & editing. KY: Writing—review & editing, Methodology, Validation. MJ: Project administration, Writing—review & editing, Funding acquisition. DW: Project administration, Resources, Writing—review & editing, Conceptualization, Methodology. QW: Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Writing—review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research project was funded by the National Key R&D Program of China (2022YFC2602200), the National Science and Technology Infrastructure of China (Project No. National Pathogen Resource Center-NPRC-32), and the Natural Science Foundation of Shenzhen (JCYJ20210324130009024).

Acknowledgments

The NPRC of China is a national scientific and technological resource-sharing service platform focusing on preserving and sharing pathogenic microbial resources. The strain resources used in this study are all preserved in the NPRC, and this study is also an exploration and utilization of the strain resources in the resource center.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1293443/full#supplementary-material

References

Alav, I., and Buckner, M. M. C. (2023). Non-antibiotic compounds associated with humans and the environment can promote horizontal transfer of antimicrobial resistance genes. *Crit. Rev. Microbiol.* 18, 1–18. doi: 10.1080/1040841X.2023.2233603

Amábile-Cuevas, C. F., and Chicurel, M. E. (1992). Bacterial plasmids and gene flux. *Cell* 70, 189–199. doi: 10.1016/0092-8674(92)90095-T

Baker-Austin, C., Wright, M. S., Stepanauskas, R., and McArthur, J. V. (2006). Co-selection of antibiotic and metal resistance. *Trends Microbiol.* 14, 176–182. doi: 10.1016/j.tim.2006.02.006

Bassetti, M., Vena, A., Croxatto, A., Righi, E., and Guery, B. (2018a). How to manage *Pseudomonas aeruginosa* infections. *DIC* 7, 1–18. doi: 10.7573/dic.212527

Bassetti, M., Vena, A., Russo, A., Croxatto, A., Calandra, T., Guery, B., et al. (2018b). rational approach in the management of *Pseudomonas aeruginosa* infections: *Curr. Opin. Infect. Dis.* 31, 578–586. doi: 10.1097/QCO.0000000000000505

Capita, R., and Alonso-Calleja, C. (2013). Antibiotic-resistant bacteria: a challenge for the food industry. *Crit. Rev. Food Sci. Nutr.* 53, 11–48. doi: 10.1080/10408398.2010.519837

Cazares, A., Moore, M. P., Hall, J. P. J., Wright, L. L., Grimes, M., Emond-Rhéault, J., et al. (2020). Megaplasmid family driving dissemination of multidrug resistance in pseudomonas. *Nat. Commun.* 11, 1370. doi: 10.1038/s41467-020-15081-7

Chen, D., Xiao, L., Hong, D., Zhao, Y., Hu, X., Shi, S., et al. (2022). Epidemiology of resistance of carbapenemase-producing *Klebsiella Pneumoniae* to ceftazidime-avibactam in a Chinese Hospital. *J. Appl. Microbiol.* 132, 237–243. doi: 10.1111/jam.15166

Clinical and Laboratory Standards Institute (2023). *Performance Standards for Antimicrobial Susceptibility Testing, 33rd Edn.* [EB/OL]. Available online at: https://clsi.org/standards/products/microbiology/documents/m100/ (accessed August 5, 2023).

Dean, C. R., Visalli, M. A., Projan, S. J., Sum, P.-E., and Bradford, P. A. (2003). Efflux-mediated resistance to tigecycline (GAR-936) in *Pseudomonas Aeruginosa* PAO1. *Antimicrob. Agents Chemother.* 47, 972–978. doi: 10.1128/AAC.47.3.972-978.2003

Delgado-Blas, J. F., Valenzuela Agüi, C., Marin Rodriguez, E., Serna, C., Montero, N., Saba, C. K. S., et al. (2022). Dissemination routes of carbapenem and pan-aminoglycoside resistance mechanisms in hospital and urban wastewater canalizations of Ghana. *mSystems* 7, e01019–e01021. doi: 10.1128/msystems.01019-21

Di Cesare, A., Eckert, E. M., D'Urso, S., Bertoni, R., Gillan, D. C., Wattiez, R., et al. (2016). Co-Occurrence of of integrase 1, antibiotic and heavy metal resistance genes in municipal wastewater treatment plants. *Water Res.* 94, 208–214. doi: 10.1016/j.watres.2016.02.049

Dong, N., Liu, C., Hu, Y., Lu, J., Zeng, Y., Chen, G., et al. (2022). Emergence of an extensive drug resistant *Pseudomonas aeruginosa* strain of chicken origin carrying *Bla* $_{IMP-45}$, *Tet* (X6), and *tmexCD3 - toprJ3* on an Inc $_{pRBL16}$ plasmid. *Microbiol. Spectr.* 10, e02283–e02222. doi: 10.1128/spectrum.02283-22

Getino, M., and De La Cruz, F. (2018). Natural and artificial strategies to control the conjugative transmission of plasmids. *Microbiol. Spectr.* 6, 25. doi: 10.1128/microbiolspec.MTBP-0015-2016

Getino, M., Sanabria-Ríos, D. J., Fernández-López, R., Campos-Gómez, J., Sánchez-López, J. M., Fernández, A., et al. (2015). Prevent plasmid-mediated horizontal gene transfer. *MBio* 6, e01032–e01015. doi: 10.1128/mBio.01032-15

Gomis-Font, M. A., Pitart, C., Del Barrio-Tofiño, E., Zboromyrska, Y., Cortes-Lara, S., Mulet, X., et al. (2021). Emergence of resistance to novel cephalosporin–β-Lactamase inhibitor combinations through the modification of the *Pseudomonas aeruginosa* MexCD-OprJ Efflux Pump. *Antimicrob. Agents Chemother.* 65, e00089–e00021. doi: 10.1128/AAC.00089-21

Gordon, J. E., and Christie, P. J. (2014). The *Agrobacterium* Ti plasmids. *Microbiol. Spectr.* 2, 18. doi: 10.1128/microbiolspec.PLAS-0010-2013

Hall, J. P. J., Botelho, J., Cazares, A., and Baltrus, D. A. (2022). What makes a megaplasmid? *Phil. Trans. R. Soc. B* 377, 20200472. doi: 10.1098/rstb.2020.0472

He, K., Xue, B., Yang, X., Wang, S., Li, C., Zhang, X., et al. (2022). Low-concentration of trichloromethane and dichloroacetonitrile promote the plasmid-mediated horizontal transfer of antibiotic resistance genes. *J. Hazard. Mater.* 425, 128030. doi: 10.1016/j.jhazmat.2021.128030

Hernández-Ramírez, K. C., Chávez-Jacobo, V. M., Valle-Maldonado, M. I., Patiño-Medina, J. A., Díaz-Pérez, S. P., Jácome-Galarza, I. E., et al. (2017). Plasmid pum505 encodes a toxin–antitoxin system conferring plasmid stability and increased *Pseudomonas aeruginosa* virulence. *Microb. Pathog.* 112, 259–268. doi: 10.1016/j.micpath.2017.09.060

Hu, F.-P., Guo, Y., Zhu, D.-M., Wang, F., Jiang, X.-F., Xu, Y.-C., et al. (2016).W. Resistance trends among clinical isolates in China Reported from CHINET surveillance of bacterial resistance, 2005–2014. *Clin. Microbiol. Infect.* 22, S9–S14. doi: 10.1016/j.cmi.2016.01.001

Jiang, Q., Feng, M., Ye, C., and Yu, X. (2022). Effects and relevant mechanisms of non-antibiotic factors on the horizontal transfer of antibiotic resistance genes in water environments: a review. *Sci. Total Environ.* 806, 150568. doi: 10.1016/j.scitotenv.2021.150568

Johnson, C. M., and Grossman, A. D. (2015). Integrative and conjugative elements (ICEs): what they do and how they work. *Annu. Rev. Genet.* 49, 577–601. doi: 10.1146/annurev-genet-112414-055018

Leski, T. A., Bangura, U., Jimmy, D. H., Ansumana, R., Lizewski, S. E., Li, R. W., et al. (2013). Identification of *Bla* $_{OXA-51-like}$, *Bla* $_{OXA-58}$, *Bla* $_{DIM-1}$, and *Bla* $_{VIM}$ carbapenemase genes in hospital enterobacteriaceae isolates from Sierra Leone. *J. Clin. Microbiol.* 51, 2435–2438. doi: 10.1128/JCM.00832-13

Li, M., Guan, C., Song, G., Gao, X., Yang, W., Wang, T., et al. (2022). Characterization of a conjugative multidrug resistance IncP-2 megaplasmid, pPAG5, from a clinical *Pseudomonas aeruginosa* isolate. *Microbiol. Spectr.* 10, e01992–e01921. doi: 10.1128/spectrum.01992-21

Li, W., and Zhang, G. (2022). Detection and various environmental factors of antibiotic resistance gene horizontal transfer. *Environ. Res.* 212, 113267. doi: 10.1016/j.envres.2022.113267

Liu, J., Yang, L., Li, L., Li, B., Chen, D., Xu, Z., et al. (2018). Comparative genomic analyses of two novel qnrVC6 carrying multidrug-resistant pseudomonas spp. strains. *Microb. Pathog.* 123, 269–274. doi: 10.1016/j.micpath.2018.07.026

Lorusso, A. B., Carrara, J. A., Barroso, C. D. N., Tuon, F. F., and Faoro, H. (2022). Role of efflux pumps on antimicrobial resistance in *Pseudomonas aeruginosa*. *IJMS* 23, 15779. doi: 10.3390/ijms232415779

Lv, L., Wan, M., Wang, C., Gao, X., Yang, Q., Partridge, S. R., et al. (2020). Emergence of a plasmid-encoded resistance-nodulation-division efflux pump conferring resistance to multiple drugs, including tigecycline, in *Klebsiella pneumoniae*. *MBio* 11, e02930–e02919. doi: 10.1128/mBio.02930-19

Magiorakos, A.-P., Srinivasan, A., Carey, R. B., Carmeli, Y., Falagas, M. E., Giske, C. G., et al. (2012). Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin. Microbiol. Infect.* 18, 268–281. doi: 10.1111/j.1469-0691.2011.03570.x

Murugan, N., Malathi, J., Therese, K. L., and Madhavan, H. N. (2018). Application of six multiplex PCR's among 200 Clinical isolates of *Pseudomonas Aeruginosa* for the detection of 20 drug resistance encoding genes. *Kaohsiung J. Med. Sci.* 34, 79–88. doi: 10.1016/j.kjms.2017.09.010

Nathwani, D., Raman, G., Sulham, K., Gavaghan, M., and Menon, V. (2014). Clinical and economic consequences of hospital-acquired resistant and multidrug-resistant *Pseudomonas aeruginosa* infections: a systematic review and meta-analysis. *Antimicrob. Resist. Infect. Control* 3, 32. doi: 10.1186/2047-2994-3-32

Pang, Z., Raudonis, R., Glick, B. R., Lin, T.-J., and Cheng, Z. (2019). Antibiotic Resistance in *Pseudomonas aeruginosa*: mechanisms and alternative therapeutic strategies. *Biotechnol. Adv.* 37, 177–192. doi: 10.1016/j.biotechadv.2018.11.013

Pankey, G. A. (2005). Tigecycline. *J. Antimicrob. Chemother.* 56, 470–480. doi: 10.1093/jac/dki248

Perez-Palacios, P., Delgado-Valverde, M., Gual-de-Torrella, A., Oteo-Iglesias, J., Pascual, Á., and Fernández-Cuenca, F. (2021). Co-transfer of plasmid-encoded bla carbapenemases genes and mercury resistance operon in high-risk clones of *Klebsiella pneumoniae*. *Appl. Microbiol. Biotechnol.* 105, 9231–9242. doi: 10.1007/s00253-021-11684-2

Pulmonary Infection Assembly of Chinese Thoracic Society (2022). Chinese expert consensus on the management of lower respiratory tract infections of *Pseudomonas aeruginosa* in adults. *Chin. J. Tuberc. Respir. Dis.* 45, 739–752. doi: 10.3760/cma.j.cn112147-20220407-00290

Sievert, D. M., Ricks, P., Edwards, J. R., Schneider, A., Patel, J., Srinivasan, A., et al. (2013). National Healthcare Safety Network (NHSN) Team and Participating NHSN Facilities. Antimicrobial-resistant pathogens associated with healthcare-associated infections summary of data reported to the national healthcare safety network at the centers for disease control and prevention, 2009–2010. *Infect. Control Hosp. Epidemiol.* 34, 1–14. doi: 10.1086/668770

Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C., and De La Cruz, F. (2010). Mpbility of plasmids. *Microbiol. Mol. Biol. Rev.* 74, 434–452. doi: 10.1128/MMBR.00020-10

Song, Z., Zuo, L., Li, C., Tian, Y., and Wang, H. (2021). Copper ions facilitate the conjugative transfer of SXT/R391 integrative and conjugative element across bacterial genera. *Front. Microbiol.* 11, 616792. doi: 10.3389/fmicb.2020.616792

Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics* 27, 1009–1010. doi: 10.1093/bioinformatics/btr039

Sun, F., Zhou, D., Wang, Q., Feng, J., Feng, W., Luo, W., et al. (2016). Genetic characterization of a novel *Bla* $_{DIM-2}$ -carrying megaplasmid P12969-DIM from Clinical *Pseudomonas Putida*. *J. Antimicrob. Chemother.* 71, 909–912. doi: 10.1093/jac/dkv426

Sun, H., Yang, Y., Yi, K., Zhang, M., Luo, X., He, D., et al. (2023). ICEGpa1804, a novel integrative and conjugative element carrying eight resistance genes, identified in *Glaesserella parasuis*. *Int. J. Antimicrob. Agents* 61, 106740. doi: 10.1016/j.ijantimicag.2023.106740

Tada, T., Shimada, K., Satou, K., Hirano, T., Pokhrel, B. M., Sherchand, J. B., et al. (2017). *Pseudomonas aeruginosa* clinical isolates in Nepal coproducing metallo-β-lactamases and 16S rRNA methyltransferases. *Antimicrob. Agents Chemother.* 61, e00694–e00617. doi: 10.1128/AAC.00694-17

The European Committee on Antimicrobial Susceptibility Testing (2020). *Breakpoint Tables for Interpretation of MICs and Zone Diameters, Version 10.0.* Available online at: http://www.eucast.org/clinical_breakpoints/ (accessed August 5, 2023).

Tran, H. A., Vu, T. N. B., Trinh, S. T., Tran, D. L., Pham, H. M., Ngo, T. H. H., et al. (2021). Resistance mechanisms and genetic relatedness among carbapenem-resistant *Pseudomonas aeruginosa* isolates from three major hospitals in Hanoi, Vietnam (2011–15). *Antimicrob. Resist.* 3, dlab103. doi: 10.1093/jacamr/dlab103

Wang, H., Qi, H., Zhu, M., Gong, S., Huang, Z., Zhang, Y., et al. (2019). MoS2 Decorated nanocomposite: Fe2O3@MoS2 inhibits the conjugative transfer of antibiotic resistance genes. *Ecotoxicol. Environ. Saf.* 186, 109781. doi: 10.1016/j.ecoenv.2019.109781

Wang, Q., Mao, D., and Luo, Y. (2015). Ionic liquid facilitates the conjugative transfer of antibiotic resistance genes mediated by plasmid RP4. *Environ. Sci. Technol.* 49, 8731–8740. doi: 10.1021/acs.est.5b01129

Wang, S., Li, S., Du, D., Abass, O. K., Nasir, M. S., Yan, W., et al. (2023). Stimulants and donors promote megaplasmid pND6-2 horizontal gene transfer in activated sludge. *J. Environ. Sci.* 126, 742–753. doi: 10.1016/j.jes.2022.03.011

Weise, K., Winter, L., Fischer, E., Kneis, D., La Cruz Barron, D., e., et al. (2022). Promote bacterial conjugative plasmid transfer. *Microbiol. Spectr.* 10, e00410–e00422. doi: 10.1128/spectrum.00410-22

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 13, e1005595. doi: 10.1371/journal.pcbi.1005595

Wireman, J., Liebert, C. A., Smith, T., and Summers, A. O. (1997). Association of mercury resistance with antibiotic resistance in the gram-negative fecal bacteria of primates. *Appl. Environ. Microbiol.* 63, 4494–4503. doi: 10.1128/aem.63.11.4494-4503.1997

Yang, B., Wang, Z., Jia, Y., Fang, D., Li, R., Liu, Y., et al. (2022). Paclitaxel and its derivative facilitate the transmission of plasmid-mediated antibiotic resistance genes through conjugative transfer. *Sci. Total Environ.* 810, 152245. doi: 10.1016/j.scitotenv.2021.152245

Zhang, Y., Gu, A. Z., Cen, T., Li, X., He, M., Li, D., et al. (2018). Sub-inhibitory concentrations of heavy metals facilitate the horizontal transfer of plasmid-mediated antibiotic resistance genes in water environment. *Environ. Pollut.* 237, 74–82. doi: 10.1016/j.envpol.2018.01.032

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 21 March 2024
DOI 10.3389/fmicb.2024.1374708

OPEN ACCESS

EDITED BY
Dwayne R. Roach,
San Diego State University, United States

REVIEWED BY
Swapnil Ganesh Sanmukh,
Université Clermont Auvergne, France
Michael Benedik,
Texas A&M University, United States

*CORRESPONDENCE
Anja Poehlein
✉ apoehle3@gwdg.de

RECEIVED 22 January 2024
ACCEPTED 27 February 2024
PUBLISHED 21 March 2024

CITATION
Schüler MA, Daniel R and Poehlein A (2024) Novel insights into phage biology of the pathogen *Clostridioides difficile* based on the active virome.
Front. Microbiol. 15:1374708.
doi: 10.3389/fmicb.2024.1374708

Novel insights into phage biology of the pathogen *Clostridioides difficile* based on the active virome

Miriam A. Schüler, Rolf Daniel and Anja Poehlein*

Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August-University Göttingen, Göttingen, Germany

The global pathogen *Clostridioides difficile* is a well-studied organism, and researchers work on unraveling its fundamental virulence mechanisms and biology. Prophages have been demonstrated to influence *C. difficile* toxin expression and contribute to the distribution of advantageous genes. All these underline the importance of prophages in *C. difficile* virulence. Although several *C. difficile* prophages were sequenced and characterized, investigations on the entire active virome of a strain are still missing. Phages were mainly isolated after mitomycin C-induction, which does not resemble a natural stressor for *C. difficile*. We examined active prophages from different *C. difficile* strains after cultivation in the absence of mitomycin C by sequencing and characterization of particle-protected DNA. Phage particles were collected after standard cultivation, or after cultivation in the presence of the secondary bile salt deoxycholate (DCA). DCA is a natural stressor for *C. difficile* and a potential prophage-inducing agent. We also investigated differences in prophage activity between clinical and non-clinical *C. difficile* strains. Our experiments demonstrated that spontaneous prophage release is common in *C. difficile* and that DCA presence induces prophages. Fourteen different, active phages were identified by this experimental procedure. We could not identify a definitive connection between clinical background and phage activity. However, one phage exhibited distinctively higher activity upon DCA induction in the clinical strain than in the corresponding non-clinical strain, although the phage is identical in both strains. We recorded that enveloped DNA mapped to genome regions with characteristics of mobile genetic elements other than prophages. This pointed to mechanisms of DNA mobility that are not well-studied in *C. difficile* so far. We also detected phage-mediated lateral transduction of bacterial DNA, which is the first described case in *C. difficile*. This study significantly contributes to our knowledge of prophage activity in *C. difficile* and reveals novel aspects of *C. difficile* (phage) biology.

KEYWORDS

***Clostridioides difficile*, temperate phage, phage induction, secondary bile salt, deoxycholate, mobile genetic element**

1 Introduction

The pathogen *Clostridioides difficile* significantly contributes to nosocomial infections worldwide (Balsells et al., 2019). A *C. difficile* infection mainly establishes after antibiotic treatment due to diverse resistances in *C. difficile* strains combined with the disturbed intestinal microflora (Spigaglia, 2016). An intact microbiome usually provides resistance to *C. difficile* colonization and disease manifestation by producing different secondary bile salts such as

lithocholate and deoxycholate (Sorg and Sonenshein, 2008). These compounds primarily aid the human intestines in digesting lipid nutrients due to their detergent character (Ridlon et al., 2006). However, they also impede *C. difficile* spore germination and cell growth (Thanissery et al., 2017). Symptoms of an established *C. difficile* infection are caused by its toxins, predominantly toxin A and B, which are encoded in the pathogenicity locus (Chandrasekaran and Borden Lacy, 2017). Symptom severity ranges from mild to severe manifestation, which can also include the death of the infected individual (Balsells et al., 2019). The personal health condition significantly affects resistance against a *C. difficile* infection, but the infecting strain is of relevance as well (Czepiel et al., 2019). Different *C. difficile* strains are linked to divergent virulence, and various aspects such as toxin production levels or secondary bile salt resistance were shown to correlate with disease severity (Lewis et al., 2017; Alonso et al., 2022). However, these studies also partially contradict the concluded relevance of specific features on virulence. In addition to general virulence factors, mobile genetic elements (MGEs) also contribute to *C. difficile* virulence (Govind et al., 2009; Sekulovic et al., 2011; Goh et al., 2013; Smits et al., 2022). Those elements play a key role in fast adaptation to environmental conditions via horizontal gene transfer (HGT) (la Cruz Fernando and Davies, 2000). *C. difficile* genomes harbor various MGEs, including prophages (Sebaihia et al., 2006). Prophages are widespread among the species *C. difficile*, and multiple prophages can exist within the same host (Fortier, 2018). Primarily, prophages were assumed to affect *C. difficile* virulence solely by encoding advantageous traits such as antibiotic resistance, which was demonstrated by phage-mediated transduction of an erythromycin resistance (Goh et al., 2013). Prophages were shown to influence toxin production in *C. difficile* strains by down- or upregulation of toxin gene expression via encoded transcriptional regulators (Govind et al., 2009; Sekulovic et al., 2011). Moreover, phages can influence the production of a protective biofilm or support the release of virulence-associated proteins and DNA (Fortier and Sekulovic, 2013; Nanda et al., 2015). These findings drew further attention to the influence of phages on *C. difficile* virulence. Phage research commonly works with prophage induction by introducing stressors such as UV radiation or mitomycin C. Meanwhile, studies confirmed spontaneous prophage release from *C. difficile* isolates, and clinically relevant antibiotics were also investigated for their phage-inducing effect (Meessen-Pinard et al., 2012; Nale et al., 2012). All experiments on *C. difficile* phages, however, worked with cultivation conditions that do not represent the actual habitat. Some components of the intestinal environment are stressful for *C. difficile*, such as the secondary bile salts. One of the prominent secondary bile salts is deoxycholate (DCA), which can promote biofilm formation in *C. difficile* (Dubois et al., 2019). DCA thereby supports bacterial persistence in the host and facilitates disease manifestation and relapse. A biofilm not only protects the bacterial cells from antimicrobial substances but also supports the adaptation of the bacterial population via HGT among the participating cells that are in close contact with the biofilm structure (Ghigo, 2001; Lécuyer et al., 2018). It was further demonstrated that DCA induces the bacterial SOS response (Kandell and Bernstein, 1991). Activation of the SOS response in turn induces prophages, leads to phage particle production, and releases via host cell lysis (Hu et al., 2021). It is therefore likely that DCA induces prophages as well, which would be a critical aspect in *C. difficile* biology and shed new light on genetic transfer *in vivo*.

In this study, we examined prophage activity in different *C. difficile* strains under spontaneous conditions and in the presence of DCA. The analyzed *C. difficile* strains were of non-clinical and clinical origin and corresponded pairwise to each other based on their sequence type (ST). Active prophage regions in these strains were identified by sequencing particle-protected DNA and were analyzed for DCA-induced activity. Possible differences between clinical and non-clinical strains, which could contribute to virulence, were also examined. The sequencing approach is more sensitive than electron microscopy or plaque assays, the commonly used detection methods for active *C. difficile* phages. We could therefore detect active prophages that might otherwise be missed due to insufficient activity but could contribute to HGT.

2 Methods

2.1 Strains and cultivation conditions

The *C. difficile* strains used in this study were of clinical or non-clinical background (personal communication), with four pairs of clinical and non-clinical strains corresponding to ST and one additional non-clinical strain (Table 1). Strains were routinely cultivated under anaerobic conditions in supplemented Brain Heart Infusion Broth (BHIS; supplemented with 0.5% yeast extract, 0.05% L-cysteine, 0.0001% Na-resazurin, purged with nitrogen) at 37°C. Putative prophage regions of the strains were predicted with PHASTEST (Wishart et al., 2023).

TABLE 1 *Clostridioides difficile* strains used in this study.

Strain	ST (Clade)	Toxin profile	Background	GenBank accession
TS3_3	1 (2)	$A^+B^+CDT^+$	Non-clinical	CP134872
DSM 28196	1 (2)	$A^+B^+CDT^+$	Clinical	CP012320
B1_2	3 (1)	$A^+B^+CDT^-$	Non-clinical	CP132141-43
SC084-01-01	3 (1)	$A^+B^+CDT^-$	Clinical	CP132146-48
J2_1	8 (1)	$A^+B^+CDT^-$	Non-clinical	CP134690-1
SC083-01-01	8 (1)	$A^+B^+CDT^-$	Clinical	CP132144-45
MA_1	11 (5)	$A^+B^+CDT^+$	Non-clinical	CP132139-40
DSM 29747	11 (5)	$A^+B^+CDT^+$	Clinical	CP019864
MA_2	340 (C-III)	$A^-B^-CDT^-$	Non-clinical	CP129431-32

Strain information on ST (Clade), profile of toxins A, B, and CDT, clinical or non-clinical background, and GenBank accession of the genome is listed.

2.2 DCA tolerance assessment

The stress effect of DCA on the various strains was assessed via a minimum inhibitory concentration (MIC) assay and relative growth determinations at different concentrations. Cultivation was performed in cell culture plates (24 well for suspension cells, Sartorius AG, Göttingen, Germany) in an anaerobic tent (Coy Laboratory, Grass Lake, United States). Overnight cultures of *C. difficile* strains were cultivated as described above and used to inoculate 2 mL medium to a final OD_{600} of 0.05, with either BHIS medium alone or supplemented with DCA (sodium deoxycholate, Sigma-Aldrich Chemie Gmbh, Taufkirchen, Germany) in concentrations ranging from 0.255 mM to 1.2 mM, which cover the physiological concentration range in humans (Hamilton et al., 2007). Culture plates were incubated at 37°C for 22 h and kept anaerobic during the OD_{600} measurement in a Synergy 2 Plate Reader (Biotek Agilent Technologies Deutschland GmbH, Böblingen, Germany). Relative growth in the presence of DCA was calculated in relation to the untreated cultures. Each experiment was performed in triplicate.

Relative growth was visualized with ggplot2 (v3.4.2; Wickham, 2016) in RStudio (v2022.06.0; RStudio Team, 2020), and significance was determined with the Tukey's 'honest significant difference' method implemented in the stats package (v4.2.0; Team, R Core, 2013).

2.3 Prophage induction and phage particle isolation

Phage particles were isolated from untreated and DCA-induced cultures. Two pre-warmed flasks of 55 mL BHIS for each isolate were inoculated 1:100 from an overnight culture and incubated at 37°C. Growth was monitored via OD_{600} measurements until an OD of ~0.6 (0.5–0.7). One culture for each isolate was induced with 0.255 mM (0.01%) DCA final concentration. The physiological concentration of DCA varies between individuals (Hamilton et al., 2007). We, therefore, used this concentration that is within the physiological range and was also used in various *C. difficile* studies regarding growth behavior or spore germination (Theriot et al., 2016; Lewis et al., 2017). The DCA solution was freshly prepared under anaerobic conditions, with DCA suspended in distilled water so that 100 µL inducing solution was required per 10 mL culture. The solution was sterilized by filtration (Filtropur S 0.2 µm, Sarstedt AG & Co. KG, Nümbrecht, Germany) and anaerobically added to the induction cultures. The second culture was not induced for analysis of spontaneous phage activity. Induced and non-induced cultures were further incubated until 22-h total incubation. The final OD_{600} of each culture was determined before isolating phage particles.

For phage particle isolation, the cells were pelleted via centrifugation at 4°C and 3,000 x *g* for 15 min. The remaining cells were removed by filtration of the supernatant with a 0.45-µm Filtropur S filter (Sarstedt). Phage particles were pelleted via centrifugation at 8°C and 20,000 x *g* for 1 h. The pellet was suspended in 1 mL SM buffer (50 mM Tris–HCl, 100 mM NaCl, 8 mM $MgSO_4$, 7 H_2O, pH 7.4) supplemented with 0.5 mM $CaCl_2$ (supporting phage stability and upcoming DNase treatment) and let soak overnight at 4°C. Particle suspension was further supported the next day by shaking at 150 rpm (LT-V Lab-Shaker, Adolf Kühner AG, Birsfelden, Germany) at room temperature for 2 h. The suspended samples were finally transferred to 2 mL DNA Low Binding microtubes (Sarstedt) for the following treatments using cut filter tips to reduce possible shearing. The samples were stored at 4°C.

2.4 Isolation of particle-protected DNA from phage samples

The phage DNA was isolated using the MasterPure Gram Positive DNA Purification Kit (Epicenter, Madison, WI, United States) with modifications. Prior to the phage DNA isolation, phage samples were supplemented with 2 µL of 100 mg/mL lysozyme solution (lysozyme from chicken egg white 177,000 U/mg; Serva, Heidelberg, Germany) suspended in SM buffer to remove remaining host cell debris and with 50 µg/mL final concentration RNase A (Biozym Scientific GmbH, Hess. Oldendorf, Germany) and 10 U Baseline-ZERO DNase (Biozym Scientific GmbH) to digest host nucleic acids. The samples were incubated at 37°C for 6 h with gentle inversion every 30 min. Fragments of host nucleic acids resulting from the digestion were removed by phage particle pelleting via centrifugation at 4°C and 20,000 x *g* for 1 h. The recovered pellet was suspended in 150 µL SM buffer. The suspension was supported by shaking at 150 rpm (LT-V Lab-Shaker) at room temperature for 10 min and slight flicking. EDTA (0.5 M, pH 8.0) was added to 10 mM final concentration for complete DNase inhibition.

Phage particles were lysed by adding 1% SDS (10% solution) and 2 µL Proteinase K (50 µg/µl; Biozym Scientific GmbH), incubation at 56°C for 1.5 h, and gentle inversion every 30 min. Subsequently, the samples were completely cooled down on ice before the addition of 130 µL MPC protein precipitation reagent (pre-cooled to −20°C). After mixing by gentle inversion, the proteins were pelleted via centrifugation at 4°C and 10,000 x *g* for 10 min. The DNA-containing supernatant was transferred to 1.5 mL DNA Low Binding microtubes (Sarstedt). The DNA was precipitated by addition of 0.3 M Na-acetate (3 M, pH 5.2), 10 mM $MgCl_2$ (2 M), and 0.8 volume isopropanol at room temperature. The samples were inverted 40 times and incubated at room temperature for 10 min before centrifugation at 4°C and 15,000 x *g* for 30 min. The supernatant was removed carefully, and the DNA pellet was washed twice with 150 µL 70% ethanol (pre-cooled to −20°C) and centrifugation at 4°C and 15,000 x *g* for 5 to 10 min. The supernatant was removed, and the sample was briefly centrifuged again to collect all residual ethanol for final removal. DNA pellets were air-dried under a sterile bench and immediately suspended in 20 µL TE buffer. Complete DNA elution was supported by brief storage at 4°C and slight flicking, before final storage at −20°C. The DNA concentration was assessed with the Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, United States) using the HS dsDNA assay kit.

2.5 Phage DNA sequencing and sequencing read processing

The phage DNA was subjected to Illumina sequencing for dsDNA by paired-end library preparation with the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, United States) as recommended by the manufacturer. Libraries were sequenced using an Illumina MiSeq system and MiSeq Reagent Kit version 3 (2 × 300 bp, 600 cycles) as recommended by the manufacturer.

All following software was used in default mode. Sequencing raw reads were quality processed with fastp (v0.23.4; Chen et al., 2018) before removing the sequencing adapters with Trimmomatic (v0.39; Bolger et al., 2014). Processed reads were mapped onto the corresponding host genome using bowtie2 (v2.5.0), and the resulting SAM file was converted to a TDS file for bioinformatics analysis with the TraV software (Dietrich et al., 2014).

2.6 Data analysis of phage sequencing reads

TDS files of processed reads for the same isolate were together inspected in TraV (Dietrich et al., 2014) for read coverage, and reads were normalized by calculation of nucleotide activity per kilobase of exon model per million mapped reads (NPKM). This results in a value for each CDS corresponding to its read coverage in reference to the overall read amount. NPKM values were further normalized to account for the differing growth behavior under the induction conditions by transforming values to an OD_{600} of 2.0. In this way, NPKM values reflected phage abundance under the different conditions at the same cell density, which allows a better qualitative estimation of phage activity. OD normalization and visualization of NPKMs values were done in Rstudio (v2022.06.0; RStudio Team, 2020) using the packages tidyverse (v2.0.0; Wickham et al., 2019), ggforce (v0.4.1; Pedersen, 2022), and ggplot2 (v3.4.2; Wickham, 2016). NPKM values were plotted against the host genome with regard to the sequence start of the corresponding CDS, and original and normalized NPKM values were plotted together to visualize the effect of OD normalization. Phage regions predicted with PHASTEST (Wishart et al., 2023) were also implemented.

2.7 Phage genome annotation and gene content analysis

Active regions identified via sequence read mapping were extracted for a new genome annotation. Sequence ranges were thereby adopted from PHASTEST (Wishart et al., 2023) predictions or selected based on read mapping in TraV (Dietrich et al., 2014), if the predicted prophage region did not cover the entire mapped region. Annotation was customized for phage genomes using Pharokka (v1.3.2; Bouras et al., 2023) in default mode with sequence re-orientation to the large terminase subunit. If the large terminase subunit was not annotated automatically, it was determined via BLAST analysis (Johnson et al., 2008) and manually annotated. For specific genes and their encoded protein, additional analyses with protein BLAST (Johnson et al., 2008) and InterProScan (v5.63–95.0; Jones et al., 2014) were performed.

2.8 Phage genome-based classifications

Genome-based classifications were done with different bioinformatic analysis tools. An average nucleotide identity analysis (ANI) with pyani (v0.2.12; Pritchard, 2014) and MUMmer3 alignment (Kurtz et al., 2004) (ANIm) was used in default mode to compare the phages among each other. The assessment of the DNA-packaging strategy was performed based on the study of Rashid et al. (2016). The large terminase subunit was aligned at the protein sequence level with ClustalW, and a maximum-likelihood phylogenetic tree was constructed with the Whelan and Goldman (WAG) substitution model and otherwise default parameter with the software MEGA (v11.0.13; Tamura et al., 2021). The branches were collapsed in MEGA (Tamura et al., 2021) if none of our phages clustered within, and final modifications for visualization were done in Inkscape (v0.48).[1] A nucleotide BLAST analysis (Johnson et al., 2008) with default parameters was performed to check for similarity to already known phage genomes and the prevalence in genomes of other *C. difficile* strains. BLAST results were ordered based on query coverage and hits with a query coverage below 10% were neglected unless relevant matches with higher coverages were not obtained. For assigning the phages to a morphological family of the order *Caudovirales*, phage genomes were inspected for the presence of baseplate proteins and sequence length of the tail length tape measure protein (Zinke et al., 2022). If the tail length tape measure protein was not annotated, it was identified via protein BLAST analysis (Johnson et al., 2008) and manually curated in the genome.

3 Results and discussion

3.1 DCA tolerance is linked to the genetic but not clinical strain background

Before investigating the effect of the secondary bile salt DCA on prophage activity in *C. difficile*, individual DCA tolerance of the various strains was assessed in the form of a MIC assay with relative growth determinations (Figure 1). DCA concentrations ranged from 0.255 mM to 1.2 mM, thereby comprising the physiological human concentration of DCA (Hamilton et al., 2007). All strains already exhibited reduced growth at the lowest concentration, which further decreased with increasing concentration. At all concentrations, strains of the same ST showed no significant difference in DCA tolerance, which implied comparable stress levels. Similar stress levels in turn might imply similar cellular strategies to cope with DCA-associated cellular damage but could also hint at similar DCA-induced prophage activity. In contrast, ST-specific tolerance differences were apparent across the DCA concentration range, with ST1 exhibiting the highest tolerance, followed by ST8, ST11, ST3, and, lastly, ST340, which was the most susceptible ST. Consequently, DCA tolerance correlated with the ST but not with the clinical background. The tolerance difference between the STs was most distinct at the lowest concentration, which was also used in the prophage induction experiments. Determined MICs ranged from 1 mM (ST11 and ST340) to 1.2 mM (ST1, ST3, ST8).

3.2 Sequencing-based assessments of prophage activity

Prophage prediction of all analyzed genomes exhibited various putative prophage regions, often with multiple incomplete and intact

1 https://inkscape.org/de/

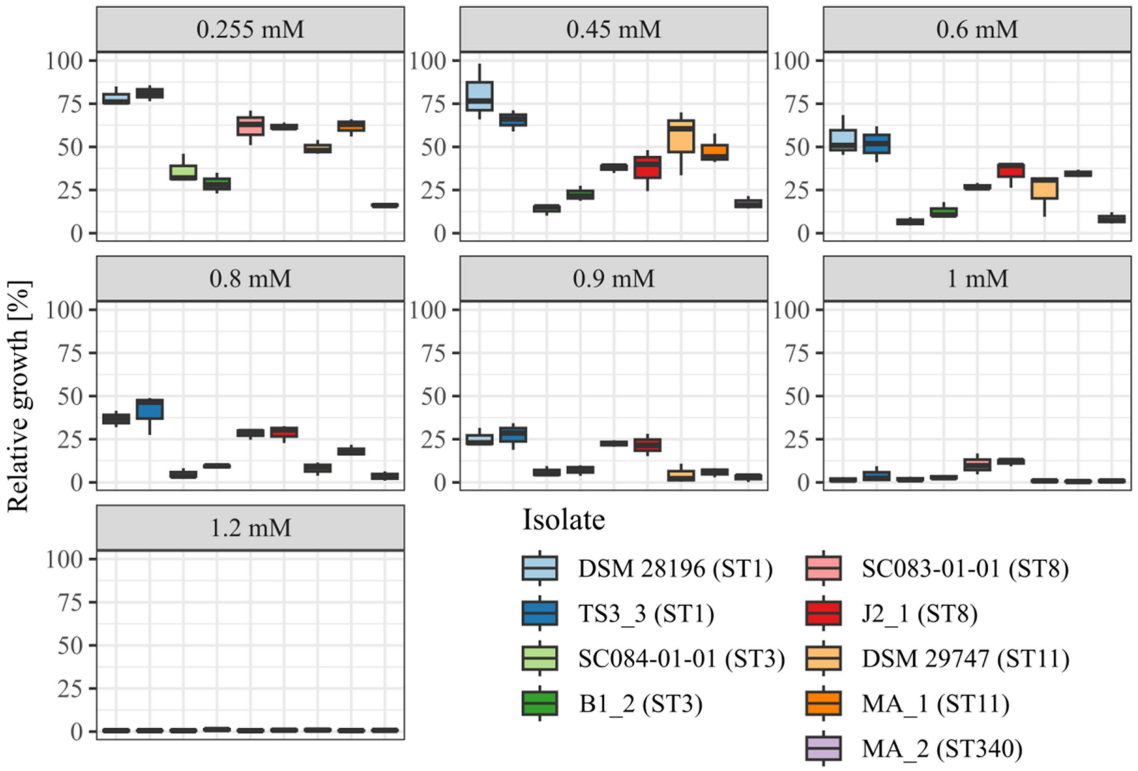

FIGURE 1
DCA tolerance at various concentrations. Tolerance of the strains to DCA was determined as relative growth compared to the reference culture (regarded as 100%) and MIC assay at concentrations from 0.255 mM to 1.2 mM. Strains of the same ST are depicted next to each other. No significant tolerance differences were observed within an ST at all concentrations.

predicted regions in one genome (Supplementary Table S1). Active prophages were determined by sequencing of particle-protected DNA. Sequencing reads were mapped to the corresponding host genome. Normalized read coverage (NPKM values) represented phage abundance as a relative measure indicative of phage activity (Hertel et al., 2015). In the following, the term phage activity describes the production and resulting abundance of DNA-containing particles, and the mentioned NPKM values refer to the OD-normalized data. Sequencing of phage DNA libraries was successful for all samples except for strain DSM 29747, which was the only strain without a predicted intact prophage genomic region (Supplementary Table S1). This strain was therefore missing in further analyses.

The mapping of phage DNA sequencing reads onto respective host genomes is depicted in Figures 2–6. The overall examination of the mapping results revealed distinct activity of at least one region in all strains and under both induction conditions. This demonstrated spontaneous phage activity in all strains and simultaneous activity of multiple phages within the same host. Almost all regions matched well with the predicted and intact prophage regions. All these regions are summarized in Table 2 for name assignment to facilitate the following descriptions. As apparent in Table 2, regions with positive prophage prediction accorded in size with typical genome sizes of *C. difficile* phages (Heuler et al., 2021), while those without were significantly smaller.

Comparison of corresponding strains showed no apparent differences in carriage or location of active phages. Correspondingly, a correlation to the clinical background of a strain was not detected. As active prophages of corresponding strains resided at corresponding genome positions, we performed an ANIm analysis on extracted sequences of all active regions to assess their similarity among each other (Supplementary Figure S1). This confirmed identical sequences of the analogous phages TS3_3_phi/DSM28196_phi1 of ST1-strains and high similarity of both phages J2_1_phi1/SC083-01-01_phi1 and J2_1_phi2/SC083-01-01_phi2 among the ST8-strains, while the other phages exhibited only little or no similarity to the others.

All spontaneously active regions (Table 2) were inspected for their activities under DCA induction. The overall NPKM transformation to an OD of 2 revealed distinctly higher signals under DCA induction in most active regions and all strains. This verified the phage-inducing effect of DCA, which varied apparently between the different regions even within the same host and thereby implied phage dependency.

The genomes of both ST1-strains carried an identical phage (Supplementary Figure S1; TS3_3_phi/ DSM28196_phi1) as determined by ANIm analysis (Supplementary Figure S1) and similar genome position. The analogous phages showed distinct spontaneous activity with approximate magnitude, but their DCA-induced activity differed substantially. Phage DSM28196_phi1 (Figure 2B) exhibited ~3.5x higher signal increase under DCA than TS3_3_phi in the non-clinical strain (Figure 2A). Since the phages were identical, the differing reactions seemed to be host-related. The genome of strain DSM 28196 possessed another active region DSM28196_phi2, which showed minor activity under both conditions and was phage-atypical by comprising only four genes and missing a prophage prediction (Figure 2B).

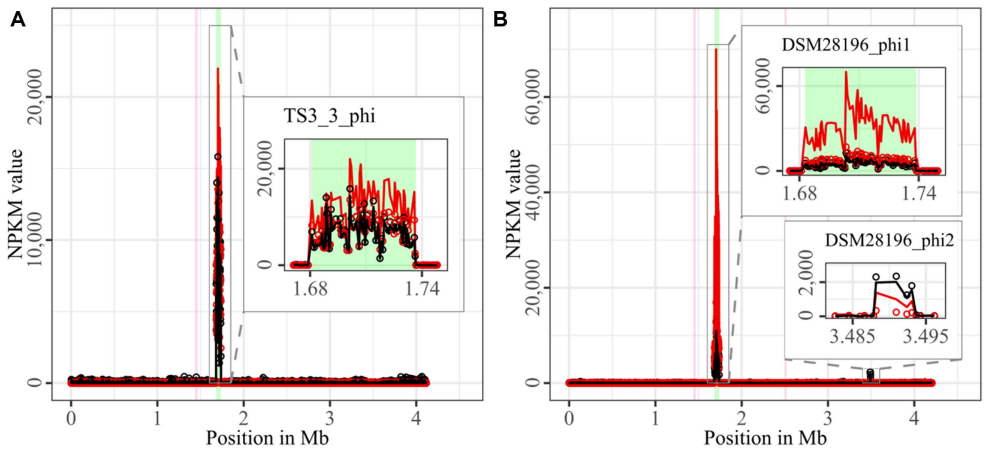

FIGURE 2
Coverage of phage sequencing reads of ST1-strains TS3_3 and DSM 28196. Sequencing reads of phage particles from spontaneous (black) and DCA-induced (red) release of strains **(A)** TS3_3 and **(B)** DSM 28196 were normalized to NPKM values (circles) with TraV (Dietrich et al., 2014), and NPKM values were additionally OD-normalized to $OD_{600} = 2$ (line graphs), before plotted against the chromosome (position in Mb) of the respective strain. Active regions were magnified for better visualization. Prophage regions predicted by PHASTEST (Wishart et al., 2023) are highlighted in the background (intact = green, incomplete = pink).

The genomes of ST3-strains possessed several active regions (Figure 3), which were not similar to each other (Supplementary Figure S1). In the genome of the non-clinical strain B1_2 (Figure 3A), two of the four active regions comprised the two ECEs. B1_2_phi4 on ECE2 exhibited the highest NPKM values under DCA induction. Remarkably, B1_2_phi4 is another phage-atypical but active region without corresponding prophage prediction, as previously detected for DSM28196_phi2 (Figure 2B). B1_2_phi3 on ECE1 also showed increased activity under DCA, but both spontaneous and induced activity were not particularly high. B1_2_phi2 on the chromosome showed spontaneous activity and a substantial increase in DCA-induced activity. B1_2_phi1 on the chromosome exhibited almost no signal under spontaneous conditions, which slightly increased upon DCA induction. This might indicate true DCA induction of this phage without prior spontaneous activity. In contrast to strain B1_2, the corresponding clinical strain SC084-01-01 possessed no prominently active region on the chromosome (Figure 3B). The activity could be observed within the region SC084-01-01_phi1, but NPKM values were very low under both conditions and signals did not cover the whole phage region. The abundancy of this phage was probably insufficient to capture distinct activity by the sequencing approach. Similar activity was observed for SC084-01-01_phi2 on ECE1 of SC084-01-01. SC084-01-01_phi3 on ECE2 was the only region in SC084-01-01 with prominent activity under spontaneous conditions, and the activity further increased apparently under DCA induction.

The genomes of both ST8-strains exhibited activity for their analogous chromosomal and extrachromosomal regions (Figure 4), which were similar phages according to ANIm analysis (Supplementary Figure S1) and similar genomic location. J2_1_phi2 on the ECE of strain J2_1 showed distinct spontaneous activity and a DCA-induced activity increase (Figure 4A), whereas SC083-01-01_phi1 on the ECE of SC083-01-01 was only slightly active under spontaneous conditions, and the signal increase upon DCA induction was only little. The activity of the chromosomal phages differed apparently as well. J2_1_phi1 on the J2_1 chromosome was spontaneously active and showed increased activity under DCA induction (Figure 4A). The left part of the prophage region started with minor activity, which drastically increased at the terminase genes. Strikingly, the left part of the prophage region started with minor activity, which drastically increased at the terminase genes. Further remarkable, sequencing reads mapped beyond the predicted prophage region and spread upstream (~30 kb) and downstream (~135 kb). The downstream region adjoined the phage activity with similar NPKM values that gradually decreased over the entire section. The chromosomal phage SC083-01-01_phi1 of the corresponding clinical ST8-strain SC083-01-01 did not exhibit these peculiarities (Figure 4B). It was spontaneously active, and the activity increased substantially under DCA treatment. This increase was strikingly twice as high as observed for the analogous phage J2_1_phi1 (Figure 4A), despite their similarity (Supplementary Figure S1). Such dissimilar activity increase among analogous phages was already observed in the ST1-strains (Figure 2). In both ST8- and ST1-pairs, a distinctly stronger increase upon DCA induction was connected to the clinical background of the strains. Since DCA-stress levels did not significantly differ between corresponding clinical and non-clinical strains (Figure 1) and thereby suggested similar induction levels, the question of the underlying mechanism of these diverging activities in analogous phages arose. This prompted an undefined regulation of phage induction involved in the clinical strains.

Strain MA_1 could not be compared to its corresponding clinical strain DSM 29747, but distinct spontaneous activity was visible for MA_1_phi on the ECE, which increased under DCA treatment (Figure 5).

Strain MA_2 is, to our knowledge, the first cryptic *C. difficile* strain with detailed phage examination. This strain possessed four active regions (Figure 6). MA_2_phi2 on the chromosome and MA_2_phi4 on the ECE both showed prominent activity under spontaneous conditions and a distinct activity increase upon DCA induction.

Interestingly, MA_2_phi4 was another active region without phage prediction, as observed previously for DSM28196_phi2 (Figure 2B) and B1_2_phi4 (Figure 3A). This was also true for MA_2_phi3 on the chromosome, where the activity was, however, very low under spontaneous conditions. The activity of this region increased upon DCA induction. Chromosomal region MA_2_phi1 exhibited the least activity in this genome even under DCA treatment. Read mapping for MA_2_phi1 and MA_2_phi2 did not cover the entire predicted prophage regions, but sections without read coverage contained only bacterial genes and were therefore evidently mispredicted.

Almost all identified active prophage regions could be induced by the secondary bile salt DCA. Mentionable, the phage activity measured after DCA treatment might be influenced by a direct effect of DCA on the phage. A study on different bacteriophages in *Escherichia coli* investigated the effect of bile salts on the host–phage interaction and observed varying survival rates of the phages (Scanlan et al., 2017). Consequently, DCA might not only induce but also damage phages, thereby reducing the abundance of induced phage particles and the measured phage activity, respectively. The measured activity in our experiments might therefore be lower than the actual activity after induction.

DCA is not the only secondary bile salt present in the human intestines that can stress *C. difficile* cells (Sorg and Sonenshein, 2008; Thanissery et al., 2017). It might therefore be assumed that the other secondary bile salts induce prophages as well. Overall, the analysis of prophage induction and activity with all secondary bile

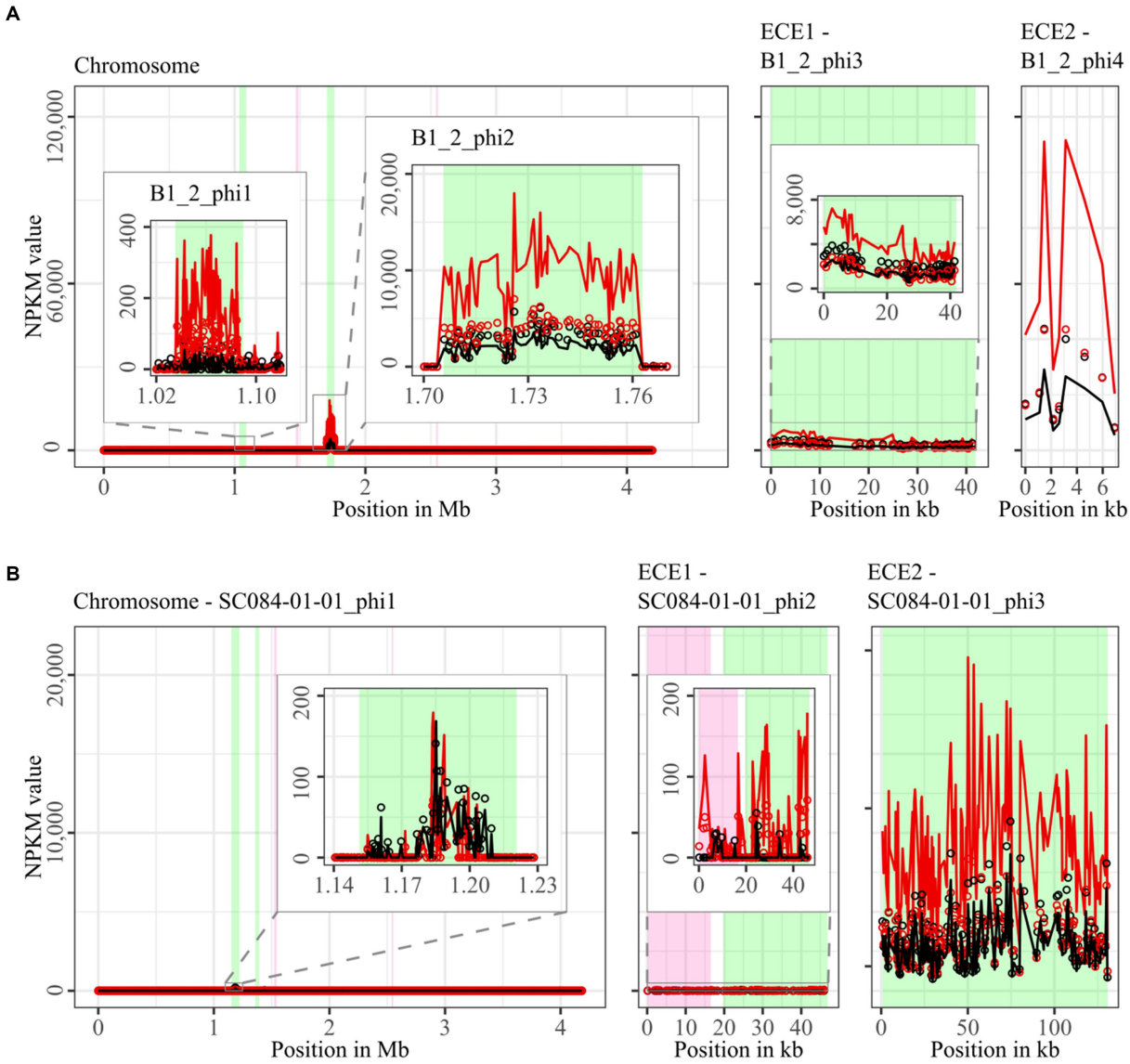

FIGURE 3
Coverage of phage sequencing reads of ST3-strains B1_2 and SC084-01-01. Sequencing reads of phage particles from spontaneous (black) and DCA-induced (red) release of strains **(A)** B1_2 and **(B)** SC084-01-01 were normalized to NPKM values (circles) with TraV (Dietrich et al., 2014), and NPKM values were additionally OD-normalized to $OD_{600} = 2$ (line graphs), before plotted against the chromosome (position in Mb) and ECE (position in kb) of the respective strain. Active regions were magnified for better visualization. Prophage regions predicted by PHASTEST (Wishart et al., 2023) are highlighted (intact = green, incomplete = pink).

salts—individually and combined—is advisable to further understand *C. difficile* prophage activity *in vivo*. In this regard, the corresponding toxin production of the analyzed strains would be another interesting aspect for investigations. The secondary bile salts were demonstrated to negatively affect toxin production (Thanissery et al., 2017) and function (Tam et al., 2020). In contrast, prophages can affect toxin production both negatively (Govind et al., 2009) and positively (Sekulovic et al., 2011). As such, phage induction in the presence of secondary bile salts can not only promote phage-mediated HGT and drive adaptation but also influence toxin levels upon lysogenization of other toxigenic strains, thereby further affecting strain virulence and disease severity.

Moreover, almost all ECEs were detected as active phages. Although extrachromosomal prophages have already been described in *C. difficile* strains (Garneau et al., 2018; Ramírez-Vargas et al., 2018), only a few of these have been isolated and characterized so far (Heuler et al., 2021).

3.3 Phage genome annotation and gene content analysis

3.3.1 Phage genomes harbor virulence-relevant genes

All active regions identified via sequence read mapping (Figures 2–6; Table 2) were inspected after a new annotation with Pharokka (Bouras et al., 2023) for genes that might increase the virulence of the host (genomic information in Supplementary Data File S1). All of them exhibited characteristic phage genes in a modular organization according to the different encoded functions, as typically seen in *C. difficile* phages (Govind et al., 2006; Meessen-Pinard et al., 2012; Garneau et al., 2018). In some genomes, the protein characteristics for plasmids were found, such as genes encoding partition proteins and replication initiation factors (Filutowicz, 2009). Genes encoding plasmid-related proteins in addition to phage-typical ones are common in a certain type of MGE, so-called phage-plasmids (Pfeifer et al., 2021). These phage-plasmid features were especially recorded for extrachromosomal prophages but were also present on the chromosomally integrated prophage MA_2_phi1. Further frequently observed genes in the phage genomes encoded proteins with potential involvement in cellular metabolism and growth, such as metallo-proteases, kinases, a phosphatase, and most of all putative rhodanese-related sulfurtransferases. These genes might be advantageous for bacterial fitness, thereby indirectly contributing to host virulence. Two phage genomes (B1_2_phi1 and SC084-01-01_phi1) carried genes encoding hemolysin XhlA, an established virulence factor in other bacterial species (Thomas et al., 2021), capable of lysing mammalian erythrocytes (Cowles and Goodrich-Blair, 2004). In the opportunistic pathogen *Mannheimia haemolytica*, temperate phages were induced that encoded hemolysin XhlA and discussed to contribute to bacterial pathogenicity and transfer of this virulence factor (Niu et al., 2015). Hemolysis in *C. difficile* is not commonly known, but few studies demonstrated its hemolytic capability (Alkudmani, 2018). However, XhlA is also present in other prophage genomes as part of the lysis module (Krogh et al., 1998). Indeed, the gene encoding XhlA was found next to the lysis-relevant genes encoding holin and endolysin in phages B1_2_phi1 and SC084-01-01_phi1. Thus, the actual role of hemolysin XhlA in these phages and its potential impact on host virulence remains unclear. Phage SC084-01-01_phi1 possessed a gene encoding an ABC transporter, which might contribute to antibiotic resistance of the host (Orelle et al., 2019). Phage SC084-01-01_phi3 harbored a putative spore protease, which could influence bacterial sporulation or germination ability and, as a consequence, alter bacterial fitness (Garneau et al., 2018). Other genes conferring antibiotic resistance or encoding known virulence factors were not identified in the active phage genomes. Six phages (TS3_3_phi, DSM28196_phi1, B1_2_phi2, J2_1_phi1, SC083-01-01_phi, MA_2_phi2) carried arrays of clustered regularly interspaced short palindromic repeats (CRISPR), which is similar to other *C. difficile* phages with described CRISPRs (Hargreaves et al., 2014a; Rashid et al., 2016; Garneau et al., 2018). Temperate phages carrying CRISPR arrays increase host immunity against other invading phages (Barrangou et al., 2007). The corresponding host genomes were verified to encode Cas proteins required for CRISPR-Cas-mediated phage immunity (Koonin and Makarova, 2009). CRISPRs in prophages represent horizontally transferrable immunity against phages, which is specifically relevant in the context of phage therapy, an alternative treatment approach for bacterial infections with growing importance in view of increasing multidrug-resistances (Gutiérrez and Domingo-Calap, 2020).

3.3.2 Active non-phage elements likely belong to so far undescribed HGT mechanisms in *Clostridioides difficile*

The regions without corresponding prophage prediction (Table 2) did not possess a phage-typical genome accordingly. Therefore, further gene analysis of these non-phage elements was performed based on the original genome annotation with Prokka (Seemann, 2014; genomic information in Supplementary Data File S1). No proteins involved in capsid or tail production, DNA packaging, or host lysis were present. The lack of structural proteins was striking since these DNA elements were enveloped according to the DNA isolation procedure. This indicated the involvement of unrelated particles. Even additional analysis of hypothetical proteins with InterProScan (Jones et al., 2014) and BLASTp (Johnson et al., 2008) could not identify further functions. All other proteins were assigned to functions with DNA-binding activity, such as helicases, integrases, relaxases, and transcriptional regulators. These genes are typical for phage genomes but also for MGEs such as transposons as Integrative and Conjugative or Mobilizable Elements (ICE/IME; Bellanger et al., 2014). Screening for these MGEs with ICEscreen (Lao et al., 2022) validated DSM28196_phi2 as complete IME, while MA_2_phi3 was detected as invalid ICE. Interestingly, the peculiar upstream region of J2_1_phi1 (Figure 4A) was also identified as an ICE, although it was incomplete. These integrative MGEs do not encode proteins for the production of particles that carry the respective mobile sequence (Bellanger et al., 2014). Transposons were found to hitchhike co-residing phages in several bacteria such as *Staphylococcus aureus* (Lindsay et al., 1998), *Vibrio cholerae* (Seed et al., 2013), and *Enterococcus faecalis* (Duerkop et al., 2012), enabling the phage-mediated transduction of virulence-relevant genes. This type of transduction was demonstrated once in *C. difficile* with the transfer of a conjugative transposon carrying an antibiotic resistance gene (Goh et al., 2013). Transduction can be either generalized, specialized, or lateral (Borodovich et al., 2022). They all imply the "headful" DNA packaging, in which the terminase starts DNA packaging at a bacterial homolog to the phage packaging site until the capsid is full, which consequently implies the transduced DNA to be of similar phage genome size (Borodovich et al., 2022). The

transduction mechanisms differ in transduced DNA and frequency (Kleiner et al., 2020). Specialized and lateral transduction involves host DNA adjacent to the prophage, while random host DNA is packaged in generalized transduction (Kleiner et al., 2020). Generalized and specialized transduction are processes of erroneous DNA packaging, which results in low transduction frequency detectable by sequencing read coverage (Kleiner et al., 2020; Borodovich et al., 2022). In contrast, lateral transduction results in high sequencing read coverage comparable to actual phage activity, as this mechanism is considered a natural phage trait instead of a mistaken process (Kleiner et al., 2020; Borodovich et al., 2022). This trait comprises phage genome replication and simultaneous DNA packaging before excision from the chromosome, whereby a substantial amount of adjacent host DNA is packaged as well (Chiang et al., 2019; Fillol-Salom et al., 2021). All these characteristics of lateral transduction accorded with the observed sequencing read mapping downstream of phage J2_1_phi1 (Figure 4A), indicating lateral transduction of this DNA segment by phage J2_1_phi1. This observation is thereby the first description of lateral transduction in *C. difficile*. The downstream region of J2_1_phi1 (Figure 4A) did not comprise characteristic genes for MGEs. Instead, several genes encoded proteins with potential relevance for strain virulence, such as genes for ABC transporters, a multidrug efflux system ATP-binding protein, stress-related proteins, proteins involved in resistance to vancomycin and daunorubicin, and the putative virulence factor BrkB. Therefore, mobilization and transfer of this region are critical regarding the spread of antibiotic resistance or virulence-related genes. The transfer of such advantageous genes via lateral transduction involves a high transfer frequency of the DNA (Kleiner et al., 2020), which can boost the efficacy of gene dissemination in *C. difficile* and

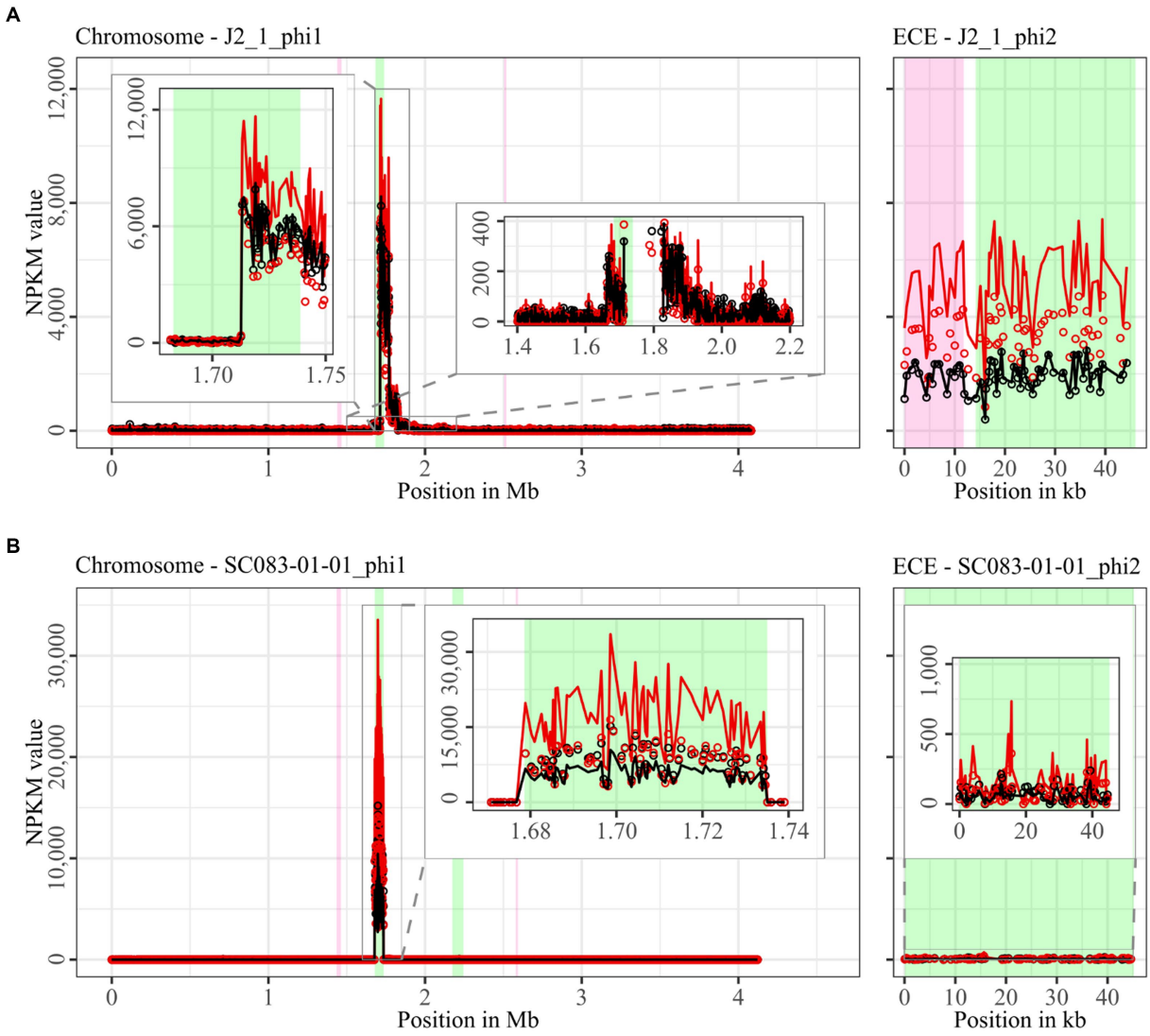

FIGURE 4
Coverage of phage sequencing reads of ST8-strains J2_1 and SC083-01-01. Sequencing reads of phage particles from spontaneous (black) and DCA-induced (red) release of strains **(A)** J2_1 and **(B)** SC083-01-01 were normalized to NPKM values (circles) with TraV (Dietrich et al., 2014), and NPKM values were additionally OD-normalized to $OD_{600} = 2$ (line graphs), before plotted against the chromosome (position in Mb) and ECE (position in kb) of the respective strain. Active regions were magnified for better visualization. Prophage regions predicted by PHASTEST (Wishart et al., 2023) are highlighted (intact = green, incomplete = pink).

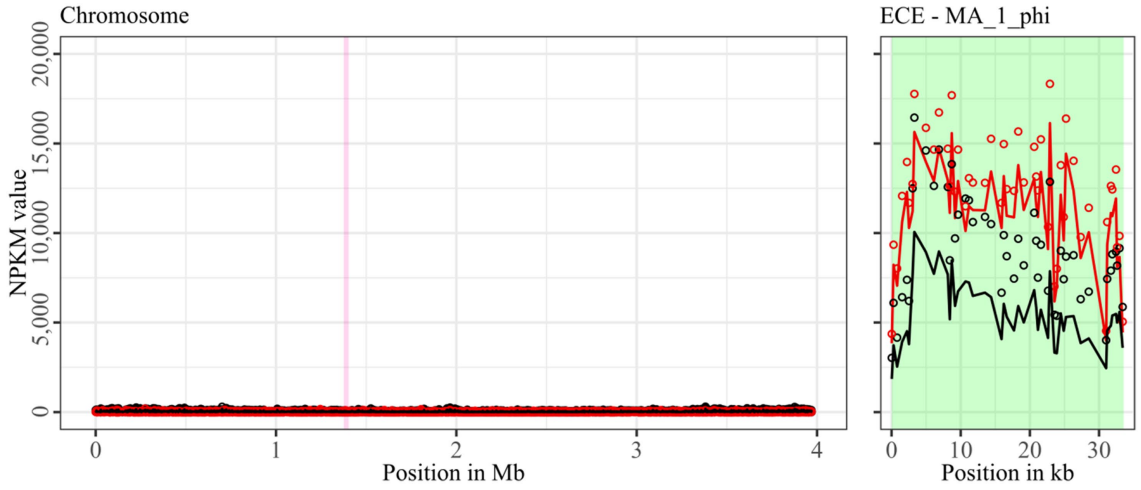

FIGURE 5
Coverage of phage sequencing reads of ST11-strain MA_1. Sequencing reads of phage particles from spontaneous (black) and DCA-induced (red) release of strain MA_1 were normalized to NPKM values (circles) with TraV (Dietrich et al., 2014), and NPKM values were additionally OD-normalized to OD_{600} = 2 (line graphs), before plotted against the chromosome (position in Mb) of MA_1. Active regions were magnified for better visualization. Prophage regions predicted by PHASTEST (Wishart et al., 2023) are highlighted (intact = green, incomplete = pink).

consequently promote evolutionary adaptation of the bacterial population.

The action of lateral transduction might also explain the drastic NPKM difference observed for the sequencing read mapping within the genome phage J2_1_phi1 (Figure 4A) as inaccurate excision of the phage genome after *in situ* replication leads to phage genome truncation.

Since no evidence for lateral transduction was observed for phage J2_1_phi1's analog SC083-01-01_phi1 (Figure 4B) despite their high similarity (Supplementary Figure S1), the question about underlying differences arose. Direct phage genome comparison revealed diverging excisionases and integrases as well as four additional amino acids in the large terminase protein of J2_1_phi1. All these proteins perform activities destined for lateral transduction (Borodovich et al., 2022).

The extrachromosomal non-phage elements were significantly smaller than the co-existing phages (Table 2), which are in contrast with the headful packaging mechanism required in transduction. This indicated a form of DNA-protecting particle other than phages, such as gene transfer agents (GTA). These phage-like particles carry DNA between 4 and 14 kb (Lang et al., 2012), which is similar to the sizes of the detected non-phage elements (Table 2). However, GTAs package bacterial DNA randomly (Lang et al., 2012), which does not fit the detected distinct activity of specific regions, making GTAs unlikely as a mode of action. Extracellular vesicles are known in various bacteria and described to carry and transfer genetic content, e.g., plasmids, between cells (Fulsundar et al., 2014; Bitto et al., 2017; Tran and Boedicker, 2017). This type of alternative HGT is not well characterized so far but was demonstrated to allow interspecies gene transfer (Fulsundar et al., 2014), which underlines the importance of vesicle-mediated DNA exchange. Noteworthy, vesicle-driven HGT in *C. difficile* has not been described so far.

The envelopment of diverse MGEs could imply a more effective transfer of these DNA elements since the enveloped DNA is protected from degradation outside of the bacterial cell. This would promote the evolutionary adaptation of *C. difficile* via the spread of genes advantageous for bacterial fitness or virulence, e.g., by conferring resistance against antimicrobial substances such as antibiotics (Goh et al., 2013) or by encoding virulence-related proteins (as, for example, found in the laterally transduced chromosomal region downstream of J2_1_phi1). Many genes present in the enveloped MGEs remain of unknown function, making their potential influence on the strain's fitness or virulence unclear. However, the data revealed the activity of these MGEs in about half of the analyzed strains and showed the considerable prevalence of these enveloped, mobile DNA elements in *C. difficile*. This indicated a significant contribution of those mechanisms in DNA transfer between *C. difficile* organisms that should be further investigated in the context of *C. difficile* adaptation and evolution. In this regard, HGT of virulence-related genes via these DNA transfer modes can be supported by close cell contact in biofilm structures that can be triggered by secondary bile salts (Dubois et al., 2019) or phage activity (Nanda et al., 2015). Moreover, phage-mediated cell lysis likely promotes the release of these various DNA transfer agents (Nanda et al., 2015), which partially showed higher activity under DCA treatment (Figures 3A, 6). A complex and dynamic interplay between the effect of DCA (or secondary bile salts in general) and the corresponding activity of phages and the other MGEs on the evolutionary adaptation of *C. difficile* can be assumed.

3.4 Classification of the active phages

3.4.1 Terminase-based determination of the phage DNA-packaging strategy

The assessment of the phage DNA-packaging mechanisms was performed to validate the above-hypothesized transduction events. The large terminase subunit was analyzed via protein sequence alignment and phylogenetic tree construction referring to Rashid et al. (2016). This assigned the phages to different phage DNA-packaging mechanisms (Figure 7). All our phages were assigned to clusters comprising other *C. difficile* phages, which predominantly represented the P22-like headful packaging mechanism, followed by the

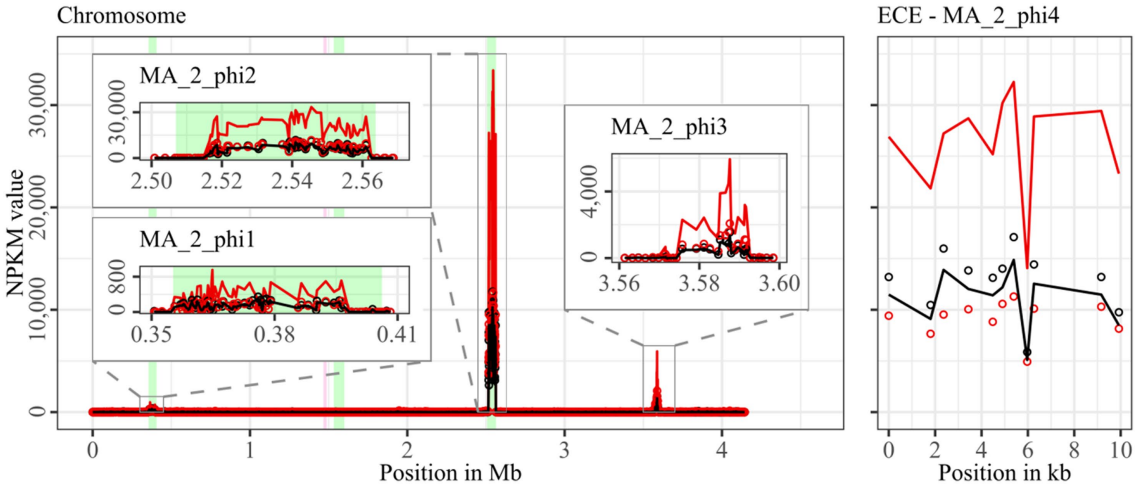

FIGURE 6
Coverage of phage sequencing reads of cryptic ST340-strain MA_2. Sequencing reads of phage particles from spontaneous (black) and DCA-induced (red) release of strain MA_2 were normalized to NPKM values (circles) with TraV (Dietrich et al., 2014), and NPKM values were additionally OD-normalized to $OD_{600} = 2$ (line graphs), before plotted against the chromosome (position in Mb) and ECE (position in kb) of MA_2. Active regions were magnified for better visualization. Prophage regions predicted by PHASTEST (Wishart et al., 2023) are highlighted (intact = green, incomplete = pink).

TABLE 2 Overview of active regions in all strains.

Strain	Location	Region name	Activity certain	Phage predicted	Size (bp)
TS3_3	Chromosome	TS3_3_phi	✓	✓	55,976
DSM 28196	Chr. region 1	DSM28196_phi1	✓	✓	55,976
	Chr. region 2	DSM28196_phi2	✓	—	5,484
B1_2	Chr. region 1	B1_2_phi1	~	✓	54,334
	Chr. region 2	B1_2_phi2	✓	✓	57,163
	ECE1	B1_2_phi3	✓	✓	42,358
	ECE2	B1_2_phi4		—	7,624
SC084-01-01	Chromosome	SC084-01-01_phi1	~	✓	69,503
	ECE1	SC084-01-01_phi2	~	✓	47,363
	ECE2	SC084-01-01_phi3	✓	✓	130,799
J2_1	Chromosome	J2_1_phi1	✓	✓	55,958
	ECE	J2_1_phi2	✓	✓	46,261
SC083-01-01	Chr. region 1	SC083-01-01_phi1	✓	✓	56,419
	ECE	SC083-01-01_phi2	✓	✓	45,313
MA_1	ECE	MA_1_phi	✓	✓	33,670
MA_2	Chr. region 1	MA_2_phi1	✓	✓	42,327
	Chr. region 2	MA_2_phi2	✓	✓	46,234
	Chr. region 3	MA_2_phi3	✓	—	16,820
	ECE	MA_2_phi4	✓	—	10,144

All active regions identified based on sequencing read coverage (Figures 2–6) were renamed according to a numbered scheme (strain_phiX). Regions were validated for certain (✓) or uncertain (~) activity based on coverage signal strength. Additionally, prophage prediction with PHASTEST (Wishart et al., 2023) is included by stating positive (✓) or negative (—) prediction, and region size in bp is stated as well.

3′-extended COS ends and an unknown strategy. Consequently, most of the phages were predicted to utilize the headful packaging mechanism and would, thus, be indeed capable of transducing host DNA. The mechanism "P22-like headful" originates from the packaging strategy employed by phage P22 of *Salmonella enterica*. Phage P22 was originally described to perform generalized transduction (Ebel-Tsipis et al., 1972), but recent evidence demonstrated also specialized as well as lateral transduction activity (Fillol-Salom et al., 2021). These terminase analysis results supported the assumption of lateral transduction of the phage J2_1_phi1 downstream region.

3.4.2 Nucleotide BLAST analyses assess phage prevalence and novelty

A nucleotide BLAST analysis (Johnson et al., 2008) was performed on all active regions in stated Table 2 to check for similar phages and

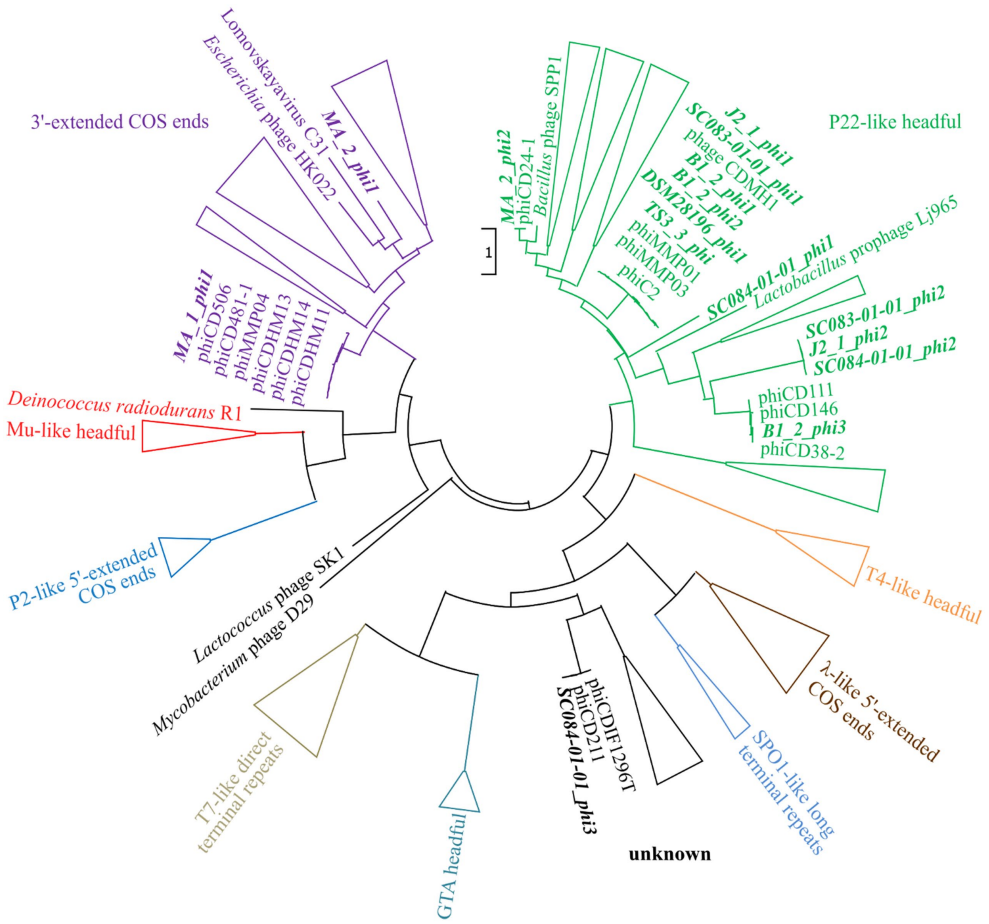

FIGURE 7
Maximum-likelihood phylogenetic tree of large terminase protein sequences. The large terminase of the active phages (highlighted in bold and italic) were aligned on the protein level to the reference sequences from Rashid et al. (2016). The branches of the different DNA-packaging strategies were colored according to Rashid et al. (2016), and the branches were collapsed for better visualization if none of our phages was included.

elements and to assess prevalence among *C. difficile* strains. The results are available in Supplementary Data File S2 and summarized in Table 3. Most of the phages matched against *C. difficile* phages with query coverages between 4 and 90% and percent identities between 86.26 and 99.86%. This confirmed that our phages are indeed similar to known phages but still represent novel types, which underlines the contribution of this study to the general knowledge of *C. difficile* phages. Furthermore, chromosomal phages also matched against a multitude of *C. difficile* chromosomes, while the extrachromosomal phages often corresponded to *C. difficile* plasmids and few chromosomes. This demonstrated the prevalence of the identified phages in other *C. difficile* strains. The non-phage elements B1_2_phi4 and MA_2_phi4 yielded no significant BLAST hit against a phage but matched against *C. difficile* plasmids. Both matching plasmids belong to classes of *C. difficile* plasmids with similar organization that are present in diverse *C. difficile* strains (Smits et al., 2018; Roseboom et al., 2023). All these plasmids might therefore likewise be inducible and particle-protected, which implies a different mechanism of HGT than currently assumed.

3.4.3 Genome-based phage assignment to *Myoviridae* and *Siphoviridae*

Finally, we classified our phages morphologically. All known *C. difficile* phages so far belong to the *Caudovirales* family of *Myoviridae* or *Siphoviridae*, which are distinguished by tail appearance (Heuler et al., 2021). Genome inspection for the presence of baseplate protein characteristics for *Myoviridae* and the length of the tail length tape measure protein as indicator for *Siphoviridae* could assign eleven phages to *Myoviridae* and four phages to *Siphoviridae* (Supplementary Table S2).

4 Conclusion

We aimed to investigate prophage activity in different clinical and non-clinical *C. difficile* strains and unravel the potential relationships between phage activity and the clinical background of the strain. Our investigations did not find specific connections to the clinical background, although we observed stronger DCA-related activity with clinical background for phages that were similar between clinical and non-clinical strains. We further revealed several interesting findings with relevance for future *C. difficile* phage research. We identified and characterized several active prophages in various *C. difficile* strains with a sequencing-based approach. This sensitive approach allowed the detection of multiple co-existing prophages with diverse activity. Most of these phages were distinctly active without specific induction, but they showed increased activity when induced with the secondary bile salt DCA. This proved that spontaneous activity is common in

TABLE 3 Nucleotide BLAST results of the identified active regions.

	First best *C. difficile*-phage BLAST hit				
Phage	Entry	Query cover %	Percent identity %	*C. difficile* matches preceding/ succeeding	Family
TS3_3_phi	CD2301	36	91.64	98 chromosomes / 32 chromosomes/assemblies 5 phages	M (Whittle et al., 2022)
DSM28196_phi1	phiC2	36	91.64	98 chromosomes / 32 chromosomes/assemblies 5 phages	M (Goh et al., 2007)
DSM28196_phi2	—	< 79	< 92.44	Only chromosomes	—
B1_2_phi1	CDMH1	60	88.33	35 chromosomes / 102 chromosomes 13 phages	M (Hargreaves et al., 2014b)
B1_2_phi2	phiC2	43	97.80	80 chromosomes / 55 chromosomes 3 phages	M (Goh et al., 2007)
B1_2_phi3	phiCD111	84	95.62	0 / 3 chromosomes 9 plasmids 16 phages	S (Sekulovic et al., 2014)
B1_2_phi4	plasmid pJMR5-4[a]	9	86.26	0 / 2 chromosomes 1 plasmid	—
SC084-01-01_phi1	phiCD418	54	92.42	29 chromosomes / 137 chromosomes 8 phages	M (Whittle et al., 2022)
SC084-01-01_phi2	HGP05	45	93.09	20 plasmids, 2 chromosomes / 145 chromosomes 14 phages	—
SC084-01-01_phi3	phiCD211	89	99.86	0 / 10 plasmids 6 phages	S (Garneau et al., 2018)
J2_1_phi1	phiC2	37	92.19	112 chromosomes / 27 chromosomes 3 phages	M (Goh et al., 2007)
J2_1_phi2	HGP05	46	91.63	20 plasmids, 2 chromosomes / 93 chromosomes 8 phages	—
SC083-01-01_phi1	phiC2	37	92.19	98 chromosomes / 43 chromosomes 4 phages	M (Goh et al., 2007)
SC083-01-01_phi2	HGP05	47	91.64	20 plasmids, 2 chromosomes / 140 chromosomes 8 phages	—
MA_1_phi	phiCD506	90	99.60	2 plasmids / 172 chromosomes 27 plasmids 20 phages	M (Sekulovic et al., 2014)
MA_2_phi1	phiCD24-1	4	88.74	17 chromosomes or genome assemblies / 0	S (Fortier and Moineau, 2007)
MA_2_phi2	phiCDKH01	74	94.21	1 chromosome / 89 chromosomes 1 phage	S (Hinc et al., 2021)
MA_2_phi3	—	< 44	< 97.72	Only chromosomes	—
MA_2_phi4	plasmid pCD-WTSI1[a]	76	91.95	0 / 25 plasmids	—

[a]Since ECEs B1_2_phi4 and MA_2_phi4 were identified to be no phage, BLAST results were checked for *C. difficile* plasmids instead.
Results were summarized regarding the first phage-related *C. difficile* BLAST hit by stating its description, query coverage, percent identity, and a number of preceding and succeeding chromosome/plasmid entries. The few hits of metagenome-assembled phages were not listed. If available, information on the phage family of the respective hit is indicated (M, *Myoviridae*; S, *Siphoviridae*).

C. difficile prophages and that the natural stressor DCA triggers prophage induction. These findings are crucial for investigating *C. difficile* biology since secondary bile salts and phage activity evidently affect *C. difficile* fitness and virulence. Both are known to influence toxin production or promote the exchange of clinically relevant genes by triggering biofilm structures or enabling HGT. We also found genes with potential connection to virulence in some phage genomes. In this context, research on actual *in vivo* phage mobility should increasingly resemble *C. difficile*'s natural habitat. The sequencing approach further revealed active regions without phage identity. Based on genomic examinations, these regions were identified as another form of MGE, in most cases possibly integrative elements. These elements apparently participated in a strategy of mobilization that involves some kind of DNA envelopment, which pointed to

phage-mediated DNA transduction, GTAs, or bacterial vesicles. This phenomenon was observed in several of the analyzed strains, which indicated that mobilization of enveloped DNA other than phage lysogeny might be a frequent mechanism in *C. difficile*. Since only one example of transduction in *C. difficile* is known so far, these mechanisms of DNA transfer via envelopment should be further investigated. One of these observations was likely the result of phage-mediated lateral transduction, thereby enabling the inter-cellular transfer of large chromosomal DNA segments. This specific type of transduction in *C. difficile* has not been described so far and opens up a new perspective on *C. difficile* phage research and HGT. Moreover, it can be assumed that the analyzed and observed aspects of DCA treatment and the associated activity of phages and enveloped MGEs interact in a complex dynamic that affects HGT and evolutionary adaptation in *C. difficile in vivo*, thereby encouraging further research.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Table S3.

Author contributions

MS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft. RD: Conceptualization, Funding acquisition, Supervision, Writing – review & editing. AP: Conceptualization, Project administration, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was funded by the Federal State of Lower Saxony, Niedersächsisches Vorab CDiff, and CDInfect projects (VWZN2889/3215/3266). The authors acknowledge support by the Open Access Publication Funds of the Göttingen University. This study was partly supported by the Göttingen Graduate Center for Neurosciences, Biophysics, and Molecular Biosciences at the Georg-August-Universität Göttingen. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

The authors thank Melanie Heinemann for technical assistance. A version of the manuscript is available as a preprint (Schüler et al., 2023).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2024.1374708/full#supplementary-material

References

Alkudmani, Zeina Subhi B. (2018). "The identification and characterization of novel Haemolysin genes from Clostridium Difficile." University College London. Available at: https://discovery.ucl.ac.uk/id/eprint/10078163/7/Alkudmani_00_Thesis_edited.pdf.

Alonso, C. D., Kelly, C. P., Garey, K. W., Gonzales-Luna, A. J., Williams, D., Daugherty, K., et al. (2022). Ultrasensitive and quantitative toxin measurement correlates with baseline severity, severe outcomes, and recurrence among hospitalized patients with *Clostridioides difficile* infection. *Clin. Infect. Dis.* 74, 2142–2149. doi: 10.1093/cid/ciab826

Balsells, E., Shi, T., Leese, C., Lyell, I., Burrows, J., Wiuff, C., et al. (2019). Global burden of *Clostridium Difficile* infections: a systematic review and Meta-analysis. *J. Glob. Health* 9:010407. doi: 10.7189/jogh.09.010407

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712. doi: 10.1126/science.1138140

Bellanger, X., Payot, S., Leblond-Bourget, N., and Guédon, G. (2014). Conjugative and Mobilizable Genomic Islands in Bacteria: evolution and diversity. *FEMS Microbiol. Rev.* 38, 720–760. doi: 10.1111/1574-6976.12058

Bitto, N. J., Chapman, R., Pidot, S., Costin, A., Lo, C., Choi, J., et al. (2017). Bacterial membrane vesicles transport their DNA cargo into host cells. *Sci. Rep.* 7:7072. doi: 10.1038/s41598-017-07288-4

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Borodovich, T., Shkoporov, A. N., Paul Ross, R., and Hill, C. (2022). Phage-mediated horizontal gene transfer and its implications for the human gut microbiome. *Gastroenterology Report* 10, 1–12. doi: 10.1093/gastro/goac012

Bouras, G., Nepal, R., Houtak, G., Psaltis, A. J., Wormald, P.-J., and Vreugde, S. (2023). Pharokka: a fast scalable bacteriophage annotation tool. *Bioinformatics* 39:btac776. doi: 10.1093/bioinformatics/btac776

Chandrasekaran, R., and Borden Lacy, D. (2017). The role of toxins in *Clostridium Difficile* infection. *FEMS Microbiol. Rev.* 41, 723–750. doi: 10.1093/femsre/fux048

Chen, S., Zhou, Y., Chen, Y., and Jia, G. (2018). Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Chiang, Y. N., Penadés, J. R., and Chen, J. (2019). Genetic transduction by phages and Chromosomal Islands: the new and noncanonical. *PLoS Pathog.* 15:e1007878. doi: 10.1371/journal.ppat.1007878

Cowles, K. N., and Goodrich-Blair, H. (2004). Expression and activity of a *Xenorhabdus Nematophila* Haemolysin required for full virulence towards *Manduca Sexta* insects. *Cell. Microbiol.* 7, 209–219. doi: 10.1111/j.1462-5822.2004.00448.x

Czepiel, J., Dróżdż, M., Pituch, H., Kuijper, E. J., Perucki, W., Mielimonka, A., et al. (2019). *Clostridium Difficile* infection: review. *Eur. J. Clin. Microbiol. Infect. Dis.* 38, 1211–1221. doi: 10.1007/s10096-019-03539-6

Dietrich, S., Wiegand, S., and Liesegang, H. (2014). TraV: a genome context sensitive transcriptome browser. *PLoS One* 9:e93677. doi: 10.1371/journal.pone.0093677

Dubois, T., Tremblay, Y. D. N., Hamiot, A., Martin-Verstraete, I., Deschamps, J., Monot, M., et al. (2019). A microbiota-generated bile salt induces biofilm formation in *Clostridium Difficile*. *Npj Biofilms and Microbiomes* 5:14. doi: 10.1038/s41522-019-0087-4

Duerkop, B. A., Clements, C. V., Rollins, D., Rodrigues, J. L. M., and Hooper, L. V. (2012). A composite bacteriophage alters colonization by an intestinal commensal bacterium. *Proc. Natl. Acad. Sci.* 109, 17621–17626. doi: 10.1073/pnas.1206136109

Ebel-Tsipis, J., Fox, M. S., and Botstein, D. (1972). Generalized transduction by bacteriophage P22 in *Salmonella Typhimurium*. *J. Mol. Biol.* 71, 449–469. doi: 10.1016/0022-2836(72)90362-2

Fillol-Salom, A., Bacigalupe, R., Humphrey, S., Chiang, Y. N., Chen, J., and Penadés, J. R. (2021). Lateral transduction is inherent to the life cycle of the archetypical *Salmonella* phage P22. *Nat. Commun.* 12:6510. doi: 10.1038/s41467-021-26520-4

Filutowicz, M. (2009). "Plasmids, bacterial" in *Encyclopedia of microbiology*. Ed. M. Schaechter (Netherland: Elsevier), 644–665.

Fortier, L.-C. (2018). Bacteriophages contribute to shaping *Clostridioides (Clostridium) difficile* species. *Front. Microbiol.* 9:2033. doi: 10.3389/fmicb.2018.02033

Fortier, L.-C., and Moineau, S. (2007). Morphological and genetic diversity of temperate phages in *Clostridium Difficile*. *Appl. Environ. Microbiol.* 73, 7358–7366. doi: 10.1128/AEM.00582-07

Fortier, L.-C., and Sekulovic, O. (2013). Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* 4, 354–365. doi: 10.4161/viru.24498

Fulsundar, S., Harms, K., Flaten, G. E., Johnsen, P. J., Chopade, B. A., and Nielsen, K. M. (2014). Gene transfer potential of outer membrane vesicles of *Acinetobacter Baylyi* and effects of stress on Vesiculation. *Appl. Environ. Microbiol.* 80, 3469–3483. doi: 10.1128/AEM.04248-13

Garneau, J. R., Sekulovic, O., Dupuy, B., Soutourina, O., Monot, M., and Fortier, L.-C. (2018). High prevalence and genetic diversity of large PhiCD211 (PhiCDIF1296T)-like prophages in *Clostridioides difficile*. *Appl. Environ. Microbiol.* 84, e02164–e02117. doi: 10.1128/AEM.02164-17

Ghigo, J.-M. (2001). Natural conjugative plasmids induce bacterial biofilm development. *Nature* 412, 442–445. doi: 10.1038/35086581

Goh, S., Hussain, H., Chang, B. J., Emmett, W., Riley, T. V., and Mullany, P. (2013). Phage ϕC2 mediates transduction of Tn6215, encoding erythromycin resistance, between *Clostridium difficile* strains. *MBio* 4, e00840–e00813. doi: 10.1128/mBio.00840-13

Goh, S., Ong, P. F., Song, K. P., Riley, T. V., and Chang, B. J. (2007). The complete genome sequence of *Clostridium Difficile* phage ΦC2 and comparisons to ΦCD119 and inducible prophages of CD630. *Microbiology* 153, 676–685. doi: 10.1099/mic.0.2006/002436-0

Govind, R., Fralick, J. A., and Rolfe, R. D. (2006). Genomic organization and molecular characterization of *Clostridium Difficile* bacteriophage ΦCD119. *J. Bacteriol.* 188, 2568–2577. doi: 10.1128/JB.188.7.2568-2577.2006

Govind, R., Vediyappan, G., Rolfe, R. D., Dupuy, B., and Fralick, J. A. (2009). Bacteriophage-mediated toxin gene regulation in *Clostridium Difficile*. *J. Virol.* 83, 12037–12045. doi: 10.1128/JVI.01256-09

Gutiérrez, B., and Domingo-Calap, P. (2020). Phage therapy in gastrointestinal diseases. *Microorganisms* 8:1420. doi: 10.3390/microorganisms8091420

Hamilton, J. P., Xie, G., Raufman, J.-P., Hogan, S., Griffin, T. L., Packard, C. A., et al. (2007). Human Cecal bile acids: concentration and Spectrum. *American J. Physiology-Gastrointestinal and Liver Physiol.* 293, G256–G263. doi: 10.1152/ajpgi.00027.2007

Hargreaves, K. R., Flores, C. O., Lawley, T. D., and Clokie, M. R. J. (2014a). Abundant and diverse clustered regularly interspaced short palindromic repeat spacers in *Clostridium Difficile* strains and prophages target multiple phage types within this pathogen. *MBio* 5, e01045–e01013. doi: 10.1128/mBio.01045-13

Hargreaves, K. R., Kropinski, A. M., and Clokie, M. R. J. (2014b). What does the talking?: quorum sensing Signalling genes discovered in a bacteriophage genome. *PLoS One* 9:e85131. doi: 10.1371/journal.pone.0085131

Hertel, R., Rodríguez, D. P., Hollensteiner, J., Dietrich, S., Leimbach, A., Hoppert, M., et al. (2015). Genome-based identification of active prophage regions by next generation sequencing in *Bacillus Licheniformis* DSM13. *PLoS One* 10:e0120759. doi: 10.1371/journal.pone.0120759

Heuler, J., Fortier, L.-C., and Sun, X. (2021). *Clostridioides difficile* phage biology and application. *FEMS Microbiol. Rev.* 45:fuab012. doi: 10.1093/femsre/fuab012

Hinc, K., Kabała, M., Iwanicki, A., Martirosian, G., Negri, A., and Obuchowski, M. (2021). Complete genome sequence of the newly discovered temperate *Clostridioides difficile* bacteriophage PhiCDKH01 of the family Siphoviridae. *Arch. Virol.* 166, 2305–2310. doi: 10.1007/s00705-021-05092-0

Hu, J., Ye, H., Wang, S., Wang, J., and Han, D. (2021). Prophage activation in the intestine: insights into functions and possible applications. *Front. Microbiol.* 12:785634. doi: 10.3389/fmicb.2021.785634

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T. L. (2008). NCBI BLAST: A Better Web Interface. *Nucleic Acids Res.* 36, W5–W9. doi: 10.1093/nar/gkn201

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Kandell, R. L., and Bernstein, C. (1991). Bile salt/acid induction of DNA damage in bacterial and mammalian cells: implications for Colon Cancer. *Nutr. Cancer* 16, 227–238. doi: 10.1080/01635589109514161

Kleiner, M., Bushnell, B., Sanderson, K. E., Hooper, L. V., and Duerkop, B. A. (2020). Transductomics: sequencing-based detection and analysis of transduced DNA in pure cultures and microbial communities. *Microbiome* 8:158. doi: 10.1186/s40168-020-00935-5

Koonin, E. V., and Makarova, K. S. (2009). CRISPR-Cas: an adaptive immunity system in prokaryotes. *F1000 Biology Reports* 1, 1–6. doi: 10.3410/B1-95

Krogh, S., Jørgensen, S. T., and Devine, K. M. (1998). Lysis genes of the *Bacillus Subtilis* defective prophage PBSX. *J. Bacteriol.* 180, 2110–2117. doi: 10.1128/JB.180.8.2110-2117.1998

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. doi: 10.1186/gb-2004-5-2-r12

la Cruz Fernando, D., and Davies, J. (2000). Horizontal gene transfer and the origin of species: lessons from Bacteria. *Trends Microbiol.* 8, 128–133. doi: 10.1016/S0966-842X(00)01703-0

Lang, A. S., Zhaxybayeva, O., and Thomas Beatty, J. (2012). Gene transfer agents: phage-like elements of genetic exchange. *Nat. Rev. Microbiol.* 10, 472–482. doi: 10.1038/nrmicro2802

Lao, J., Lacroix, T., Guédon, G., Coluzzi, C., Payot, S., Leblond-Bourget, N., et al. (2022). ICEscreen: a tool to detect Firmicute ICEs and IMEs, isolated or enclosed in composite structures. *NAR Genomics and Bioinformatics* 4:lqac079. doi: 10.1093/nargab/lqac079

Lécuyer, F., Bourassa, J.-S., Gélinas, M., Charron-Lamoureux, V., Burrus, V., and Beauregard, P. B. (2018). Biofilm formation drives transfer of the conjugative element ICE *Bs1* in *Bacillus Subtilis*. *MSphere* 3:e00473. doi: 10.1128/mSphere.00473-18

Lewis, B. B., Carter, R. A., Ling, L., Leiner, I., Taur, Y., Kamboj, M., et al. (2017). Pathogenicity locus, Core genome, and accessory gene contributions to *Clostridium Difficile* virulence. *MBio* 8, e00885–e00817. doi: 10.1128/mBio.00885-17

Lindsay, J. A., Ruzin, A., Ross, H. F., Kurepina, N., and Novick, R. P. (1998). The gene for toxic shock toxin is carried by a family of Mobile Pathogenicity Islands in *Staphylococcus Aureus*. *Mol. Microbiol.* 29, 527–543. doi: 10.1046/j.1365-2958.1998.00947.x

Meessen-Pinard, M., Sekulovic, O., and Fortier, L.-C. (2012). Evidence of *in vivo* prophage induction during *Clostridium Difficile* infection. *Appl. Environ. Microbiol.* 78, 7662–7670. doi: 10.1128/AEM.02275-12

Nale, J. Y., Shan, J., Hickenbotham, P. T., Fawley, W. N., Wilcox, M. H., and Clokie, M. R. J. (2012). Diverse temperate bacteriophage carriage in *Clostridium Difficile* 027 strains. *PLoS One* 7:e37263. doi: 10.1371/journal.pone.0037263

Nanda, A. M., Thormann, K., and Frunzke, J. (2015). Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. *J. Bacteriol.* 197, 410–419. doi: 10.1128/JB.02230-14

Niu, Y. D., Cook, S. R., Wang, J., Klima, C. L., Hsu, Y.-h., Kropinski, A. M., et al. (2015). Comparative analysis of multiple inducible phages from *Mannheimia Haemolytica*. *BMC Microbiol.* 15:175. doi: 10.1186/s12866-015-0494-5

Orelle, C., Mathieu, K., and Jault, J.-M. (2019). Multidrug ABC transporters in Bacteria. *Res. Microbiol.* 170, 381–391. doi: 10.1016/j.resmic.2019.06.001

Pedersen, Thomas Lin. (2022). "Ggforce: Accelerating 'ggplot2' Title." Available at: https://github.com/thomasp85/ggforce.

Pfeifer, E., Jorge, A., de Sousa, M., Touchon, M., and Rocha, E. P. C. (2021). Bacteria have numerous distinctive groups of phage-plasmids with conserved phage and variable plasmid gene repertoires. *Nucleic Acids Res.* 49, 2655–2673. doi: 10.1093/nar/gkab064

Pritchard, Leighton. (2014). "PYANI Script." 2014. Available at: https://github.com/widdowquinn/pyani.

Ramírez-Vargas, G., Goh, S., and Rodríguez, C. (2018). The novel phages PhiCD5763 and PhiCD2955 represent two groups of big Plasmidial Siphoviridae phages of *Clostridium Difficile*. *Front. Microbiol.* 9:26. doi: 10.3389/fmicb.2018.00026

Rashid, S., Barylski, J., Hargreaves, K., Millard, A., Vinner, G., and Clokie, M. (2016). Two novel Myoviruses from the north of Iraq reveal insights into *Clostridium Difficile* phage diversity and biology. *Viruses* 8:310. doi: 10.3390/v8110310

Ridlon, J. M., Kang, D.-J., and Hylemon, P. B. (2006). Bile salt biotransformations by human intestinal Bacteria. *J. Lipid Res.* 47, 241–259. doi: 10.1194/jlr.R500013-JLR200

Roseboom, A. M., Ducarmon, Q. R., Hornung, B. V. H., Harmanus, C., Crobach, M. J. T., Kuijper, E. J., et al. (2023). Carriage of three plasmids in a single human clinical isolate of Clostridioides difficile. *Plasmid* 125:102669. doi: 10.1016/j.plasmid.2022.102669

RStudio Team. (2020). "RStudio: Integrated Development Environment for R." Available at: http://www.rstudio.com/.

Scanlan, P. D., Bischofberger, A. M., and Hall, A. R. (2017). Modification of *Escherichia Coli* –bacteriophage interactions by surfactants and antibiotics *in vitro*. *FEMS Microbiol. Ecol.* 93:fiw211. doi: 10.1093/femsec/fiw211

Schüler, M. A., Daniel, A., and Poehlein, A. (2023). Novel insights into phage biology of the pathogen clostridioides difficile based on the active virome. *BioRxiv*. doi: 10.1101/2023.09.27.559748

Sebaihia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R., et al. (2006). The multidrug-resistant human pathogen *Clostridium Difficile* has a highly Mobile, mosaic genome. *Nat. Genet.* 38, 779–786. doi: 10.1038/ng1830

Seed, K. D., Lazinski, D. W., Calderwood, S. B., and Camilli, A. (2013). A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* 494, 489–491. doi: 10.1038/nature11927

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Sekulovic, O., Garneau, J. R., Néron, A., and Fortier, L.-C. (2014). *Clostridium Difficile* isolates of human and animal origins. *Appl. Environ. Microbiol.* 80, 2555–2563. doi: 10.1128/AEM.00237-14

Sekulovic, O., Meessen-Pinard, M., and Fortier, L.-C. (2011). Prophage-stimulated toxin production in *Clostridium Difficile* NAP1/027 Lysogens. *J. Bacteriol.* 193, 2726–2734. doi: 10.1128/JB.00787-10

Smits, W. K., Roseboom, A. M., and Corver, J. (2022). Plasmids of *Clostridioides difficile*. *Curr. Opin. Microbiol.* 65, 87–94. doi: 10.1016/j.mib.2021.10.016

Smits, W. K., Scott Weese, J., Roberts, A. P., Harmanus, C., and Hornung, B. (2018). A helicase-containing module defines a family of PCD630-like plasmids in *Clostridium Difficile*. *Anaerobe* 49, 78–84. doi: 10.1016/j.anaerobe.2017.12.005

Sorg, J. A., and Sonenshein, A. L. (2008). Bile salts and Glycine as Cogerminants for *Clostridium Difficile* spores. *J. Bacteriol.* 190, 2505–2512. doi: 10.1128/JB.01765-07

Spigaglia, P. (2016). Recent advances in the understanding of antibiotic resistance in *Clostridium Difficile* infection. *Therapeutic Advan. Infectious Dis.* 3, 23–42. doi: 10.1177/2049936115622891

Tam, J., Icho, S., Utama, E., Orrell, K. E., Gómez-Biagi, R. F., Theriot, C. M., et al. (2020). Intestinal bile acids directly modulate the structure and function of *C. difficile* TcdB toxin. *Proc. Natl. Acad. Sci.* 117, 6792–6800. doi: 10.1073/pnas.1916965117

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Team, R Core. (2013). "*R: A language and environment for statistical computing*." Vienna, Austria: R Foundation for Statistical Computing.

Thanissery, R., Winston, J. A., and Theriot, C. M. (2017). Inhibition of spore germination, growth, and toxin activity of clinically Relevant *C. difficile* strains by gut microbiota derived secondary bile acids. *Anaerobe* 45, 86–100. doi: 10.1016/j.anaerobe.2017.03.004

Theriot, C. M., Bowman, A. A., and Young, V. B. (2016). Antibiotic-induced alterations of the gut microbiota Alter secondary bile acid production and allow for *Clostridium Difficile* spore germination and outgrowth in the large intestine. *MSphere* 1, e00045–e00015. doi: 10.1128/mSphere.00045-15

Thomas, P., Abdel-Glil, M. Y., Eichhorn, I., Semmler, T., Werckenthin, C., Baumbach, C., et al. (2021). Genome sequence analysis of *Clostridium Chauvoei* strains of European origin and evaluation of typing options for outbreak investigations. *Front. Microbiol.* 12:732106. doi: 10.3389/fmicb.2021.732106

Tran, F., and Boedicker, J. Q. (2017). Genetic cargo and bacterial species set the rate of vesicle-mediated horizontal gene transfer. *Sci. Rep.* 7:8813. doi: 10.1038/s41598-017-07447-7

Whittle, M. J., Bilverstone, T. W., van Esveld, R. J., Lücke, A. C., Lister, M. M., Kuehne, S. A., et al. (2022). A novel bacteriophage with broad host range against *Clostridioides difficile* Ribotype 078 supports SlpA as the likely phage receptor. *Microbiology Spectrum* 10:e0229521. doi: 10.1128/spectrum.02295-21

Wickham, Hadley. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al. (2019). Welcome to the Tidyverse. *J. Open Source Software* 4:1686. doi: 10.21105/joss.01686

Wishart, D. S., Han, S., Saha, S., Oler, E., Peters, H., Grant, J. R., et al. (2023). PHASTEST: faster than PHASTER, better than PHAST. *Nucleic Acids Res.* 51, W443–W450. doi: 10.1093/nar/gkad382

Zinke, M., Schröder, G. F., and Lange, A. (2022). Major tail proteins of bacteriophages of the order Caudovirales. *J. Biol. Chem.* 298:101472. doi: 10.1016/j.jbc.2021.101472

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 09 April 2024
DOI 10.3389/fmicb.2024.1286822

OPEN ACCESS

EDITED BY
Alberto Danielli,
University of Bologna, Italy

REVIEWED BY
Robert Martin Blumenthal,
University of Toledo, United States
Ankan Banerjee,
University of Marburg, Germany

*CORRESPONDENCE
Matthias Bochtler
✉ mbochtler@iimcb.gov.pl
Shuang-yong Xu
✉ xus2007@gmail.com

†These authors have contributed equally to this work

RECEIVED 31 August 2023
ACCEPTED 08 February 2024
PUBLISHED 09 April 2024

CITATION
Helbrecht I, Heiter D, Yang W, Vincze T, Hanneman A, Lutz T, Ettwiller L, Bochtler M and Xu S-y (2024) Characterization of winged helix domain fusion endonucleases as N6-methyladenine-dependent type IV restriction systems. *Front. Microbiol.* 15:1286822. doi: 10.3389/fmicb.2024.1286822

Characterization of winged helix domain fusion endonucleases as N6-methyladenine-dependent type IV restriction systems

Igor Helbrecht[1,2,3†], Daniel Heiter[1†], Weiwei Yang[1], Tamas Vincze[1], Andrew Hanneman[1], Thomas Lutz[1], Laurence Ettwiller[1], Matthias Bochtler[2,3]* and Shuang-yong Xu[1]*

[1]New England Biolabs, Inc., Ipswich, MA, United States, [2]Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland, [3]International Institute of Molecular and Cell Biology, Warsaw, Poland

Winged helix (wH) domains, also termed winged helix-turn-helix (wHTH) domains, are widespread in all kingdoms of life and have diverse roles. In the context of DNA binding and DNA modification sensing, some eukaryotic wH domains are known as sensors of non-methylated CpG. In contrast, the prokaryotic wH domains in DpnI and HhiV4I act as sensors of adenine methylation in the 6mApT (N6-methyladenine, 6mA, or N6mA) context. DNA-binding modes and interactions with the probed dinucleotide are vastly different in the two cases. Here, we show that the role of the wH domain as a sensor of adenine methylation is widespread in prokaryotes. We present previously uncharacterized examples of PD-(D/E)XK—wH (FcyTI, Psp4BI), PUA—wH—HNH (HtuIII), wH—GIY-YIG (Ahi29725I, Apa233I), and PLD—wH (Aba4572I, CbaI) fusion endonucleases that sense adenine methylation in the Dam$^+$ Gm6ATC sequence contexts. Representatives of the wH domain endonuclease fusion families with the exception of the PLD—wH family could be purified, and an *in vitro* preference for adenine methylation in the Dam context could be demonstrated. Like most other modification-dependent restriction endonucleases (MDREs, also called type IV restriction systems), the new fusion endonucleases except those in the PD-(D/E)XK—wH family cleave close to but outside the recognition sequence. Taken together, our data illustrate the widespread combinatorial use of prokaryotic wH domains as adenine methylation readers. Other potential 6mA sensors in modified DNA are also discussed.

KEYWORDS

winged helix fusion endonucleases, HNH endonuclease, GIY-YIG endonuclease, PLD family endonuclease, N6mA-dependent restriction system, genome conflict

Introduction

In most restriction–modification (R-M) scenarios, nucleobase modification serves as a mark of self and provides protection against endonuclease digestion. In some cases, however, phages have learned to exploit this principle by modifying their own DNA, either by incorporation of non-standard nucleoside triphosphates or by post-replicative modifications catalyzed either by host or phage enzymes. Modification-dependent restriction endonucleases (MDREs) specifically target such modified DNA (modified base or backbone). The MDREs

come in two main groups, distinguished by the presence or absence of nucleoside triphosphate (NTP)-consuming motor proteins. The NTP-independent proteins are typically modular, with separate modification sensing and DNA cleavage domains. Because of this architecture, DNA cleavage typically takes place at a distance from the site of modification. For some enzymes, a single site is sufficient, but typically, cleavage is most efficient when it is directed by appropriately spaced modifications, which cooperate to position an endonuclease dimer for a double strand (ds) cut in the DNA.

The catalytic domains present in restriction can be grouped into the almost universally used hydrolases (Orlowski and Bujnicki, 2008) and the very rarely used lyases (Miyazono et al., 2014). The hydrolases, in turn, can be grouped into a surprisingly small set of phylogenetically unrelated enzyme groups. PD-(D/E)XK enzymes are named for characteristic amino acids (aa) built around a central β-sheet, which harbors one or two catalytic Mg^{2+} ions (Pingoud et al., 2005). The metal ions are held in place in part by the D and D or E (abbreviated as D/E) interacting residues, which, together with a K residue, activate a water molecule for direct inline attack on the scissile phosphate (Bujnicki and Rychlewski, 2001; Kosinski et al., 2005). HNH enzymes, also called ββα-Me enzymes or His-Me finger enzymes (Jablonska et al., 2017; Wu et al., 2020), harbor a single metal cation in their active site. Metal identity requirements are less strict than for PD-(D/E)XK enzymes. Many divalent transition metal ions are acceptable (Pommer et al., 2001). Like PD-(D/E)XK enzymes, the HNH enzymes are believed to catalyze attack on the scissile phosphate by a water molecule. However, water activation is not by a lysine residue but by the first histidine of the HNH motif (Flick et al., 1998; Sokolowska et al., 2009). GIY-YIG enzymes (Dunin-Horkawicz et al., 2006; Kaminska et al., 2008) also bind a single metal cation in the active site. These enzymes activate the water molecule with a tyrosine residue, most likely from the GIY motif (Sokolowska et al., 2011). Finally, there are also completely metal-independent endonuclease domains. They resemble phospholipase D; therefore, the enzymes containing them are known as PLD endonucleases (Grazulis et al., 2005; Chan et al., 2007). The PLD enzymes are believed to catalyze phosphodiester cleavage via a covalent intermediate (Sasnauskas et al., 2010).

The modification sensor domains, like the endonuclease domains, are now understood to be classifiable into only a few groups of phylogenetically unrelated sensors. The largest group of sensors is the PUA (PseudoUridine synthase and Archaeosine transglycosylase) superfamily (Lutz et al., 2019). PUA superfamily sensors comprise SRA (SET and Ring finger Associated) domains with specificity for 5mC, as in MspJI (Cohen-Karni et al., 2011), and related domains (Kazrani et al., 2014; Shao et al., 2014), originally also termed SRA domains, with specificity for 5-hydroxymethylcytosine (5hmC) and glucosyl-5-hydroxymethyl-cytosine (g5hmC), as in the PvuRts1I family of restriction endonucleases (Janosi et al., 1994; Borgaro and Zhu, 2013). The PUA superfamily also comprises EVE (according to the PDB identifier 2eve for a prototypical protein; Bertonati et al., 2009) domains specific for 5mC and 5hmC, as found in VcaM4I (Pastor et al., 2021), and YTH (YT521-B Homologs) domains (YTH-McrB/NTPase fusion) specific for 6-methyladenine (6mA) (Iyer et al., 2006; Hosford et al., 2020; Xu et al., 2020), as well as ASCH (ASC-1 Homology) domains. Bioinformatic analysis has suggested that ASCH domains might be 6mA readers (Iyer et al., 2016), but this prediction has not been confirmed by experimental data so far. Instead, it has been shown that the *E. coli* YqfB, an ASCH domain protein, is able to hydrolyze various N4-acylated cytosines (4acC) and cytidines (Stanislauskiene et al., 2020). All PUA superfamily domains are engaged in nucleotide flipping (Cerrudo et al., 2014; Pastor et al., 2021). Irrespective of their detailed specificity, they scrutinize the modified base in a dedicated pocket of the enzyme (Roberts and Cheng, 1998; Horton et al., 2014). It has been shown that *E. coli* McrBC endonuclease also recognizes modified cytosines by base flipping (Sukackaite et al., 2012).

Apart from the PUA superfamily, other modification sensor domains may also be involved in restriction, such as the NEco domain in EcoKMcrA with affinity for 5mC and 5hmC (Czapinska et al., 2018). Unlike the PUA superfamily domains, the NEco domain senses 5mC or 5hmC without nucleotide flipping in the context of dsDNA (Slyvka et al., 2019). Finally, a winged helix (wH) domain has been described as a 6mA sensor in DpnI. Like the NEco domain, the wH (winged helix) domain senses nucleobase modifications in the context of dsDNA without flipping (Mierzejewska et al., 2014). However, in contrast to the NEco domain, it has specificity for 6mA rather than 5mC. Also, in contrast to NEco, which recognizes methyl groups of fully methylated CpG in two separate pockets, the wH domain recognizes methyl groups of fully methylated ApT in a single pocket, exploiting their proximity in space. The wH domain in DpnI is unusual in being fused to a nuclease domain, which has a separate sequence (GATC) and modification (6mA) specificity (Siwek et al., 2012). Therefore, it acts more like an effector domain in type IIE enzymes (Senesac and Allen, 1995; Roberts et al., 2003), except that both the nuclease and sensor/effector domain are specific for methylated rather than non-methylated DNA.

Winged helix (wH) domains are a group of DNA-binding domains that belong to the superfamily of helix-turn-helix (HTH) proteins (Brennan, 1993; Lai et al., 1993; Gajiwala and Burley, 2000). Structurally, canonical winged helix domains consist of an N-terminal α-helices and β-strand, the HTH motif, and a β-hairpin. The "wings" of the motif are the loops connecting the strands of the β-hairpin and immediately downstream of it (Iyer et al., 2016; Figure 1A). Winged helix motifs were first found in transcription factors, but it is now clear that they also have roles in transcription initiation complexes (Teichmann et al., 2012), in the binding of left-handed Z-DNA (Schwartz et al., 2001) or RNA (Tang et al., 2021), or in protein–protein interactions (Wah et al., 1997). In transcription factors, wH domains tend to interact with DNA, just as would be expected for the HTH motif that is embedded within them. In other words, they insert the second helix of the HTH motif, which is the third helix of the wH domain, into the major groove of DNA (Gajiwala and Burley, 2000). However, other DNA-binding modes are also possible in special cases (Gajiwala et al., 2000; Wolberger and Campbell, 2000). A recent example of such alternative binding modes is the complexes of eukaryotic winged helix domains with dsDNA containing non-methylated CpG (Stielow et al., 2021; Becht et al., 2023; Weber et al., 2023). A winged helix motif in a restriction endonuclease (REase) was first noticed in the DNA-binding domain of FokI (Wah et al., 1997), but this particular wH domain does not appear to be involved in interactions with DNA.

The role of the winged helix domain in adenine methylation sensing was first noticed in DpnI. DpnI is a G6mATC-specific endonuclease that cleaves within the recognition sequence and has a strong preference for DNA that is adenine-methylated in both strands (Siwek et al., 2012). In DpnI, the winged helix domain plays the role

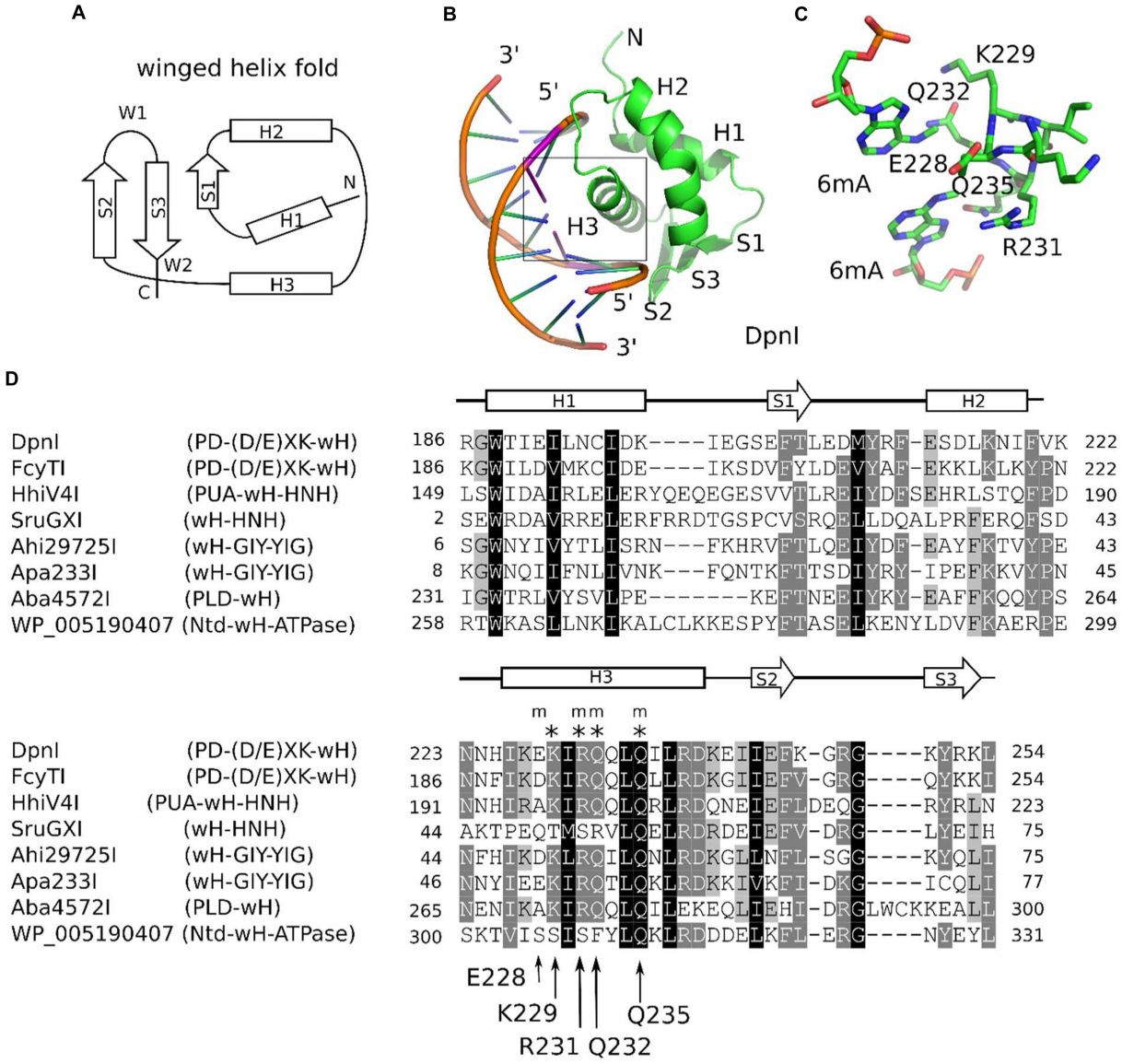

FIGURE 1
Winged helix (wH) domain fold and role as a methylation sensor. **(A)** Canonical wH fold. **(B)** Cartoon representation of the DpnI wH domain bound to fully methylated DNA, based on the crystal structure (Siwek et al., 2012). **(C)** Methyl binding region of the DpnI wH domain. **(D)** Alignment of representative winged helix domains in endonuclease or NTPase fusions. The secondary structure annotation is based on the DpnI experimental structure, analyzed for secondary structure elements using DSSP (Kabsch and Sander, 1983). DpnI residues that are involved in the formation of the pocket for the methyl groups (of DNA methylated in both strands) are marked by an "m," and those that are involved in hydrogen bonding with the nucleobases of the GATC target sequence of DpnI are marked by an asterisk ("*"). Their identities and residue numbers in case of DpnI are indicated below the alignment (with reference to the DpnI structure with PDB accession 4kyw; Mierzejewska et al., 2014). Note the strong overlap between methyl pocket-forming residues and residues that are involved in target sequence selection.

of an effector domain that senses 6mA separately from and with slightly relaxed sequence specificity compared to the PD-(D/E)XK nuclease domain (Mierzejewska et al., 2014; Figures 1B,C). More recently, a winged helix domain was also implicated in the sensing of adenine methylation, also in the G6mATC context. HHPV4I (also called HhiV4I; Lu et al., 2023) is a three-domain enzyme, with a PUA (SRA)-like domain at the N-terminus, a winged helix domain in the middle, and an HNH endonuclease domain at the C-terminus. The PUA superfamily domain, described as an SRA domain by Lu et al., appears not to be involved in DNA modification sensing (Lu et al., 2023). By contrast, the winged helix domain directed preferential cleavage of Dam$^+$ over Dam$^-$ DNA, and it has much higher affinity to Dam$^+$ than to Dam$^-$ DNA in gel shift experiments. In contrast to DpnI, HHPV4I (HhiV4I) cleaves at a distance from the site of adenine methylation, suggesting that the endonuclease domain is directed by the winged helix domain and does not sense adenine methylation on its own (Lu et al., 2023).

In this study, we show that the winged helix domain is widely used as an adenine methylation sensor in MDREs (Figure 1D). We present additional examples of proteins that share the PD-(D/E)XK—wH architecture with DpnI or the PUA—wH—HNH architecture with HHPV4I (HhiV4I). Additionally, we show that the wH domain is can

also be naturally paired with an HNH domain in the absence of a PUA superfamily domain, with a GIY-YIG endonuclease domain, with a PLD (phospholipase D) nuclease domain, or with an NTPase (GTPase/ATPase) domain. For the PD-(D/E)XK—wH, PUA—wH—HNH, wH—HNH, wH—GIY-YIG, and PLD—wH enzymes, we detect Dam$^+$-dependent toxicity in *E. coli* cells, either by tight binding or digestion near the modified sites. For the fusion endonucleases PD-(D/E)XK—wH, PUA—wH—HNH, wH—HNH, and wH—GIY-YIG, but not the PLD—wH enzymes, we find representatives that are active also *in vitro*, and we show that their preferred substrate is fully methylated DNA with one enzyme exception. Unlike DpnI, many of the new wH fusion endonucleases cleave DNA outside the G6mATC recognition sequence and two modified sites in tandem with a short spacer, enhance their cleavage activity.

Materials and methods

Materials

E. coli T7 expression strains C2566 (Dam$^+$) and its isogenic Dam-deficient strain ER2948 [constructed and provided by Dr. Lise Raleigh, New England Biolabs (NEB)], expression vector pTXB1, pBR322, phage λ DNA (Dam$^+$ or Dam$^-$), 2-log (1 kb plus) DNA ladder, chitin beads, restriction enzymes, EcoGII methylase (frequent adenine methylase), Q5 DNA polymerase PCR master mix and cloning kit (Hi-Fi DNA assembly enzyme mix), NEBExpress Ni-NTA magnetic beads, and dZTP (2-aminoadenine triphosphate is abbreviated as base Z in the literature, here we use dZ to denote the 2'-deoxynucleoside) were provided by Michael Kuska (NEB Organic Synthesis Division). Fast-flow Ni-agarose beads were from Qiagen or NEB. The T7 expression vector pET21b with C-terminal 6× His tag was originally purchased from Novagen (NdeI-XhoI). 5hmdCTP/dGTP/dATP/dTTP mix was purchased from Zymo Research. FPLC DEAE and Heparin columns (5 mL) were purchased from GE HealthCare or Cytiva.

Endonuclease assays

For restriction of modified DNA, Dam$^+$ pBR322 and λ DNA were used. In some cases, we also used M.EcoGII-modified pBR322 (Dam$^-$), with all modified adenine bases except in polyA tracks. Restriction buffers used were: NEB buffer 2.1 (medium salt, 50 mM NaCl) or CutSmart buffer (with 50 mM potassium acetate). Endonucleases: we usually perform an enzyme titration in endonuclease activity assays; enzyme concentrations are indicated in each digestion. Buffer compositions in 1× NEB restriction buffers: buffer 1.1 (10 mM Bis-Tris-Propane-HCl, 10 mM $MgCl_2$, 100 µg/mL BSA or recombinant albumin, pH 7.0 at 25°C); buffer 2.1 (50 mM NaCl, 10 mM Tris–HCl, 10 mM $MgCl_2$, 100 µg/mL BSA or recombinant albumin, pH 7.9 at 25°C); buffer 3.1 (100 mM NaCl, 10 mM Tris–HCl, 10 mM $MgCl_2$, 100 µg/mL BSA or recombinant albumin, pH 7.9 at 25°C); CutSmart buffer (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 100 µg/mL BSA or recombinant albumin, pH 7.9 at 25°C). For restriction digestions in different divalent cations, a medium salt buffer (50 mM NaCl, 20 mM Tris–HCl, pH 7.5) was supplemented with divalent cations in 0.1, 1, and 10 mM final concentrations.

Synthetic oligos with modified and unmodified GATC sites

Single-stranded DNA oligos were synthesized by the NEB organic synthesis division:

Top strand Gm6ATC (top M+).

5′/56FAM/ACTCATGCAGGCATGCAGG/m6A/TCGCAGTCAGATTTATGTGTCATATAGT ACGTGATTCAAG-3′.

Bottom strand Gm6ATC (bottom M+).

5′-CTTGAATCACGTACTATATGACACATAAATCTGACTGCG/m6A/TC CTGCATGCCTGCATGAGT-3′.

Top strand GATC (unmodified, top M−).

5′/56FAM/ACTCATGCAGGCATGCAGGATCGCAGTCAGATTTATGTGTCATATAGTACGTGATTCAAG-3′.

Bottom strand GATC (unmodified, bottom M−).

5′-CTTGAATCACGTACTATATGACACATAAATCTGACTGCGATCCTGCATGCCTGCATGAGT-3′.

Duplex oligos abbreviation:

Fully modified = M+ top/M+ bottom = M+/M+.

Unmodified = M− top/M− bottom = M−/M−.

Hemi-modified = M+ top/M− bottom = M+/M−.

Hemi-modified = M− top/M+ bottom = M−/M+.

Duplex oligos were digested by DpnI (2 U), MboI (5 U), FcyTI (0.1 µg), Ahi29725I and Apa233I (1 µg) in NEB buffer 2.1 at 37°C for 1 h. For HhiV4I digests, reactions were carried out in NEB buffer 2.1 supplemented with 1 mM $MnCl_2$ (this enzyme is a Mn^{2+}-dependent REase; see below).

Protein expression and purification

C2566 (Dam$^+$) competent cells (cloning grade) were provided by NEB. ER2948 (Dam$^-$) competent cells were prepared by a modified rubidium chloride method. *E. coli* cells were cultured at 37°C to mid-log phase, and IPTG was added to the culture at 0.5 mM final concentration for protein production (at 18°C overnight). Cells were lysed by sonication in chitin column buffer or Ni-agarose column buffer. Clarified cell lysates with over-expressed proteins (target protein-intein-CBD) or C-terminal 6× His-tagged proteins were loaded onto chitin or Ni-agarose columns, respectively, for affinity purification. The protein purification protocols were used as recommended by the manufacturers. In some cases, the partially purified proteins were further purified by chromatography through DEAE (flow-through to remove nucleic acids at 0.3 M NaCl concentration) and Hi-Trap Heparin (5 mL). BigDye® Terminator v3.1 Cycle Sequencing Kit was purchased from Thermo-Fisher (Applied Biosystems). Restriction gene inserts in plasmids were sequenced to verify the correct sequences. Dam$^+$ pBR322 DNA fragments after restriction digestion were sequenced to determine the cut sites. DNA sequence edits were carried out using DNAStar or Geneious software packages. BlastP searches in the GenBank and UniProtKB databases were performed using the respective web servers. NCBI Pfam and conserved domains were used to visualize protein domains of REase homologs.

Plasmid preparation and transformation

Plasmid mini-prep kits and competent cells were provided by NEB. Plasmid mini-preparation and bacterial transformation were done according to the manufacturer's recommendation.

Clustering locus-specific annotations analysis

For clustering locus-specific annotations (CLANS) analysis, the homologs of the five groups of wH-containing enzymes (sequences of the reference proteins are listed in Supplementary material) were obtained by blast (BlastP, default settings in the UniProtKB website) using the UniProtKB reference proteomes and Swiss-Prot database. A total of 1,258 wH fusion endonuclease homolog proteins were analyzed. The resulting homolog protein fasta files were combined and subjected to CLANS analysis using the MPI Bioinformatics Toolkit. The result of the CLANS analysis is visualized with the CLANS Java application (Frickey and Lupas, 2004). CLANS analyzes all-against-all pairwise sequence similarities to establish relationships within a protein family. In the CLANS network figure, each node (a small colored dot) represents a full-length protein (or wH domain in the lower box), and the line connects two proteins/domains that share sequence similarities. The lengths of the lines represent the degree of sequence similarities, with short lines representing close similarities and vice versa. Each node is the same size, but the size of the color "blob" is related to the number of nodes clustered together.

Phylogenetic analysis

To perform phylogenetic analysis, we first built a hmm (hidden Markov model) profile with the wH domains of the six representative wH-containing enzymes (note: wH domains only, not full-length proteins). The resulting profile was used to search homologs using hmmsearch (HMMER v3.1b2) on the combined fasta file (see CLANS analysis). We extracted the wH domain in each homolog protein and performed multiple alignments with the five representative wH domains using MAFFT (v7.508). The maximum likelihood tree was constructed using RAxML (v8.2.12) with option -f a to enable rapid bootstrapping for 100 times. Other options used were -p 1237, −x 1,237, −m PROTGAMMAAUTO. The best-scoring maximum likelihood tree with bootstrap values was visualized with iTOL.[1]

Proteome analysis

The four samples (10 µg each) were digested with trypsin using S-Trap micro-columns as recommended by the manufacturer (PROTIFI). LC–MS/MS of the digests was conducted on the Thermo Orbitrap Exploris, with two injections made per sample. Data were searched using Proteome Discoverer v2.5 against a combined FASTA database containing the Uniprotkb_*E. coli*-B strain proteins and the four winged helix (wH) endonuclease protein sequences (FcyTI, HhiV4I, Ahi29725I, and Apa233I). Results are presented in 8 Excel spreadsheets, one file for each injection (see raw data on the proteome of the purified proteins). The most confident protein ID is determined by Score Sequest HT (the last column in each table—the higher the score, the more confident the identification). Proteins identified are reported with >1 unique peptide per protein and a 1% false discovery rate.

Results

Bioinformatic screen of wH domain endonucleases

Some known wH domains are adenine methylation (6mA)-specific. This notion was originally suggested by the demonstration of 6mA specificity of the DpnI wH domain and further strengthened by the observation of adenine specificity of the wH domain of HHPV4I (Lu et al., 2023), which was reported when this study was being finalized. In the hope of finding new adenine methylation-specific endonucleases, we searched for fusions of wH domains with endonuclease domains known to play roles in R-M (Pingoud and Jeltsch, 2001). Apart from additional PD-(D/E)XK—wH and PUA—wH—HNH endonucleases, we also identified fusion proteins with a wH—HNH, wH—GIY-YIG, and PLD—wH architecture. In addition, we found PD-(D/E)XK—wH—NTPase cases that can be considered as fusions of a DpnI-like protein with an McrB-like NTPase (GTPase/ATPase) domain and NTD—wH—NTPase cases, apparently without a nuclease domain and with an N-terminal domain of unknown function. One additional example is the wH—Mrr catalytic domain (PD-QXK)—NTPase fusion endonuclease. For this study, we concentrated on the NTP-independent enzymes with wH and endonuclease domains. To better understand the sequence relationships between the new wH fusion proteins, we carried out a CLANS (Frickey and Lupas, 2004) analysis. CLANS determines all-against-all pairwise sequence similarities to establish relationships within a protein family. It is not intended to find sequence motifs within protein sequences, which are better detected using other software. In CLANS analysis of the full-length proteins [with BamHI (GGATCC) and related enzymes as a control], the wH domain-containing endonucleases segregated clearly into separate groups, driven by the sequence similarity between endonuclease domains of the same group (Figure 2). However, when we limited the CLANS analysis to the wH domains alone, the segregation into groups according to the endonuclease domain was not so clear. In particular, wH domains from PD-(D/E)XK and GIY-YIG endonucleases were fully intermingled, possibly suggesting multiple separate fusion events (Figure 2, box). The finding suggests that fusions of the same type may have arisen independently several times in evolution. Nevertheless, much of the diversity of the new fusion proteins is clearly due to divergent evolution. This is also supported by the phylogenetic tree of the wH domains (Supplementary Figure S1). Note, however, that bootstrap values for most branches of the tree are very low, making the tree very tentative overall.

Rare occurrence of methyltransferase genes in the immediate genetic neighborhood of the new fusion proteins

wH proteins frequently bind DNA, but only some of them are methylation-dependent (e.g., DpnI), whereas others are not (e.g.,

1 https://itol.embl.de/

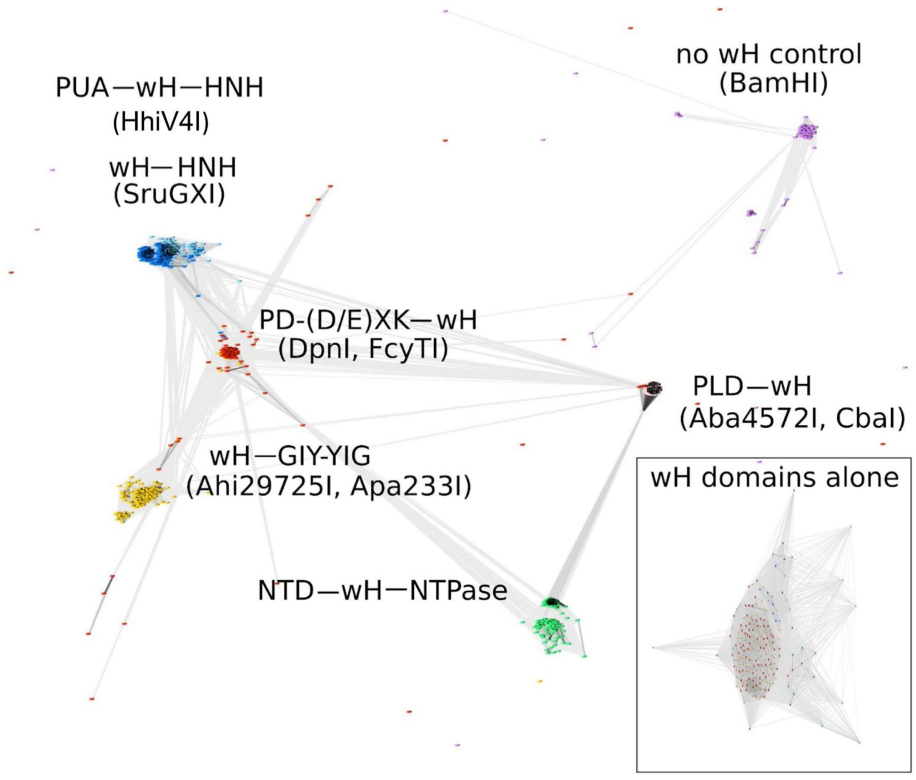

FIGURE 2
CLANS analysis of the full-length wH fusion proteins with PD-(D/E)XK (marked as red), PUA and HNH (dark blue), HNH (light blue), GIY-YIG endonuclease domains (yellow), and NTD (N-terminal domain)-NTPase (green) found in this study, with BamHI and related isoschizomers as a control group (purple). BOX: CLANS analysis of the wH domains alone, color-coded as in full-length proteins.

FokI). If the new wH fusion proteins were modification-specific, they should occur as stand-alone enzymes. Otherwise, they should be associated with a host genome protecting DNA methyltransferase of any type (N6mA, N4mC, C5). Typically, such a methyltransferase would be located in the immediate genomic neighborhood of the endonuclease, so that the entire system could work as a defense island (Makarova et al., 2011). To test for possible association with DNA methyltransferases, we inspected the genomic neighborhoods of over 1,000 wH fusion proteins. In 87% of cases, none of the three genes adjacent upstream or downstream to the wH fusion gene encoded a DNA methyltransferase, suggesting that most of the new wH fusion proteins acted as stand-alone endonucleases, possibly as type IV restriction systems.

Likely specificity of at least some of the wH fusion proteins for 6mA in the GATC context

Adenine methylation in bacteria occurs frequently in the GATC context i.e. the target sequence of the Dam methyltransferase (MTase) (Marinus and Casadesus, 2009), which is widely distributed in bacteria because of its diverse house-keeping roles, including DNA replication (Boye and Lobner-Olesen, 1990) and mismatch repair (Au et al., 1992; Josephs et al., 2015). Hence, it was likely that the putative MDREs with the wH domain might detect adenine methylation in this sequence context. This idea was further supported by the precedent of the wH domain in DpnI, which is known to be specific for adenine methylation in the Dam context (with some leeway for the outer bases S6mATS, where S is G or C). Inspection of the crystal structure of the DpnI domain in complex with target DNA revealed that the same residues contribute to both the methyl binding pocket and the sequence specificity, suggesting that methyl sensing and detection of the G6mATC target sequence are intricately linked (Mierzejewska et al., 2014). Large-scale analysis of the wH domain fusion proteins showed that the motif for 6mA and GATC recognition (see arrows in Figure 1D), or closely related motifs, were present in approximately 10% of the new fusion proteins. With the exception of SruGXI as a representative of the wH-HNH endonucleases, we selected for further characterization the fusion proteins that had the motif for 6mA and GATC specificity. Such fusion proteins are very likely to recognize m6A in the GATC context. The reminder is that stand-alone wH fusion proteins are likely to recognize modified DNA (otherwise they would be toxic to the host). However, it is currently unclear whether the modification is m6A, and, if so, whether the sequence context is GATC.

Avoidance of genomic conflict

Inspection of a 10 kb interval around genes encoding the new fusion proteins revealed association with methyltransferases in some cases. PD-(D/E)XK—wH domains co-occurred with predicted C5 methyltransferases in 40 cases. In some cases, they also co-occurred

with a predicted 4mC (N4mC) or 6mA MTase directly adjacent to it. In these cases (e.g., *Bacteroidota* bacterium isolate CP064983.1, *Moraxella ovis* strain CP011158.1), the putative DNA MTase has been inactivated by a frame shift. The PUA—wH—HNH and wH—HNH co-occurred in 47 cases with Eco57I-like MTases (of type IIG R-M-S fusion enzymes). These MTases are predicted to be 6mA MTases, with CTGAAG (site of methylation underlined) as the target sequence. In the case of the wH—GIY-YIG endonucleases, we found four instances of an EcoRI-like adenine MTase nearby. These MTases are expected to methylate GAATTC. Finally, for the PLD—wH endonucleases, we detected 17 cases of proximity to EcoEI-like ($GAGN_7ATGC$) or EcoR124I-like ($GAAN_6RTCG$) type I methyltransferases, also causing no conflict. Genetic conflict would not be expected in any of these cases if the new wH fusion proteins had specificity for 6mA in the GATC context.

Next, we looked for possible genetic conflicts on a genome-wide scale, assuming that the new wH fusion proteins were specific for m6A in a GATC context. Three types of such conflict are conceivable. First, a frequent adenine methyltransferase, such as M.EcoGII, may methylate adenine to m6A in GATC, among many other contexts. Second, a Dam-like methyltransferase may specifically modify GATC sequences. Finally, there are also methyltransferases that methylate target sequences that are longer than GATC but include the GATC site in the recognition sequence. We searched for cases of such potential conflict, scanning the entire genome, not just the genomic neighborhoods. Overall, less than 100 cases of potential conflict were identified (Supplementary Tables S1–S3) for over 1,000 wH fusion proteins. Most of the wH fusion proteins in potential conflict are stand-alone enzymes without an associated methyltransferase. Genomic conflict could be avoided if these proteins recognized modified DNA containing a mark other than m6A in the GATC context (i.e., either another methylation type or m6A in a different sequence context). Alternatively, conflict may be avoided or mitigated by tight expression control of the endonuclease or the methyltransferase.

Selection of wH fusion proteins for experimental characterization

Four types of endonuclease domains are noted in wH domain fusions: (1) DpnI-like PD-(E/D)XK endonuclease; (2) HNH endonuclease domain; (3) GIY-YIG endonuclease domain; and (4) PLD family endonuclease domain. An NTD-wH-NTPase fusion is usually paired with another endonuclease subunit, such as McrC-like catalytic subunit, which is not discussed in detail here. We have not studied evolutionary relationships within each endonuclease family since the endonuclease families have been the subject of numerous review articles and research papers (Mehta et al., 2004; Grazulis et al., 2005; Pingoud et al., 2005; Dunin-Horkawicz et al., 2006). For experimental characterization, we chose representatives of the PD-(D/E)XK—wH (FcyTI), PUA—wH—HNH (HhiV4I), wH—HNH (SruGXI), wH—GIY-YIG (Ahi29725I, Apa233), and PLD—wH (Aba4572I, CbaI) fusions for further experimental characterization. In the case of the PUA—wH—HNH architecture, many additional candidate MDREs were tested in *E. coli* cells only. An attempt to purify an NTD—wH—NTPase/McrC-like subunit fusion protein was unsuccessful. Therefore, we focused this study exclusively on single-chain proteins.

The wH fusion endonucleases exhibit Dam$^+$-dependent toxicity in *E. coli* cells

Endonuclease toxicity is a good proxy for restriction in bacterial cells (Heitman and Model, 1990; Fomenkov et al., 1994). If the putative MDREs were specific for Dam-methylated DNA, they should be more toxic to Dam$^+$ (C2566) than to Dam$^-$ (ER2948) *E. coli* cells. We tested this prediction with our IPTG-inducible expression constructs, both under basal (no IPTG) conditions and under induction conditions (0.5 mM IPTG). Since Dam$^-$ competent cells were roughly an order of magnitude less competent than Dam$^+$ cells, we avoided direct comparisons of transformation efficiency between Dam$^+$ and Dam$^-$ cells. Instead, we quantified the reduction in colony counts for transformations with expression plasmids compared to colony counts with an empty vector. With the exception of the Aba4572I expression construct, the plasmids for all other endonuclease-containing clones caused a reduction in colony count by two to three orders of magnitude compared to the empty vector control in Dam$^+$ cells under induced conditions (Figure 3). Aba4572I endonuclease may have strong non-specific endonuclease activity since it is toxic in a Dam-deficient strain under IPTG induction. The toxicity appeared to be less severe in Dam$^+$ cells. This result is not well understood. If the Aba4572I outlier is disregarded, the experimental results indicate that the wH fusion endonucleases display typical "restriction" on Dam$^+$ host but not Dam$^-$ cells. Whether this *in vivo* "restriction" was caused by tight binding to modified sites to inhibit replication or endonuclease cleavage of modified DNA remains to be investigated.

PD-(D/E)XK—wH endonucleases

As representatives of the PD-(D/E)XK—wH family, we selected Psp4BI and FcyTI (GenBank accession numbers WP_102090895 and WP_094411979). The two enzymes have 58.4 and 58.7% amino acid (aa) sequence identity to DpnI, respectively. Psp4BI was chosen because the source organism is psychrophilic, suggesting that the enzyme might be susceptible to heat inactivation, which would be desirable for biotechnological applications. The synthetic genes with *E. coli* optimized codons were cloned into pTXB1 in fusion with intein and CBD (chitin-binding domain) and expressed in the Dam-deficient T7 expression strain. The two enzymes were affinity purified on a chitin column and released from the column by DTT-triggered cleavage. The yield of Psp4BI was low due to poor expression of the fusion protein (Psp4BI-intein-CBD) (not shown); partially purified Psp4BI gave rise to a partial digestion pattern that was retained after 4 h at 25–37°C. The low activity could be caused by a low enzyme concentration or inhibition by some impurities in the preparation. By contrast, purified FcyTI was active on Dam$^+$ pBR322, pUC19 (HindIII-linearized), and phage λ DNA (Supplementary Figures S2A,B). FcyTI-specific activity was determined as approximately 32,000 U/mg protein in buffer 2.1. FcyTI could be inactivated by heating at 65°C for 30 min, which is a useful enzyme property (Supplementary Figure S3). FcyTI endonuclease was originally found in the genome of *Flavobacterium*

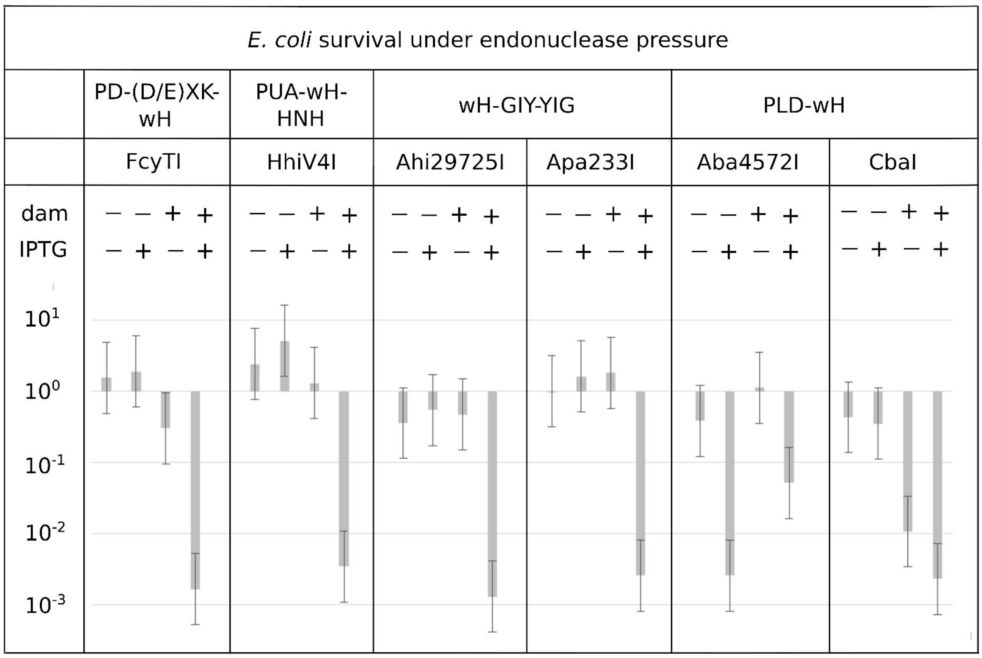

FIGURE 3
Toxicity of selected wH domain fusion endonucleases to a Dam⁺ but not a Dam⁻ host. Expression vectors containing open reading frames for putative MDREs or empty plasmid (50 ng) were transformed into Dam-positive C2566 (+) or Dam-negative ER2948 (–) *E. coli* cells, with (+) or without (–) IPTG induction. The reduction in colony counts for expression plasmid compared to the empty vector, plotted on the ordinate, is a measure of toxicity.

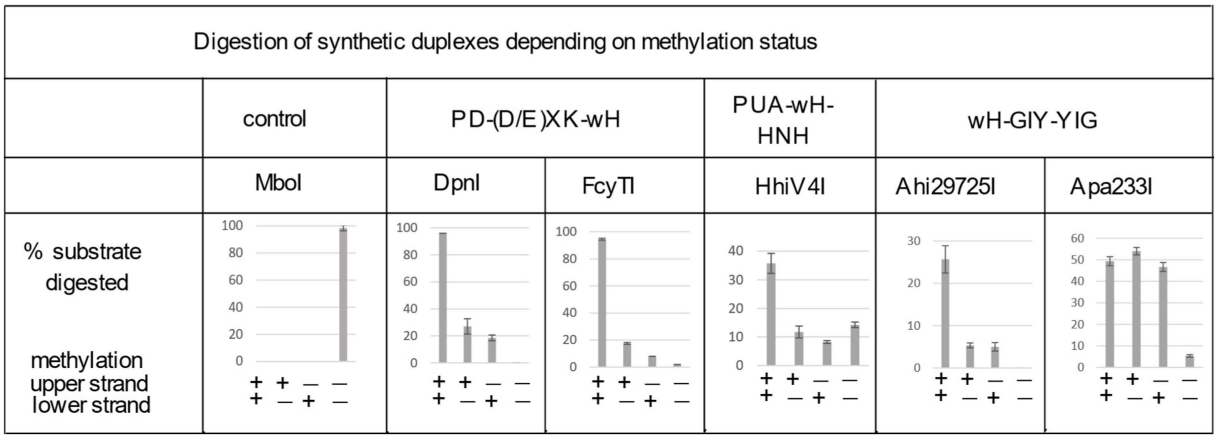

FIGURE 4
Digestion of synthetic DNA oligoduplexes depending on top and bottom strand methylation status in the Gm6ATC sequence context. 5 U of MboI (GATC) and 2 U of DpnI (Gm6ATC) were used as controls. FcyTI input was 0.1 µg protein (~3 U) (higher concentrations could obscure the digestion of hemi-modified duplex oligos). Duplex oligos and protein concentration in restriction digests: The duplex oligos concentration is approximately 18 nM (60mer, 21 ng in 30 µL reaction volume). The protein concentration is calculated below in the 30 µL reaction volume: (1) Ahi29725I protein (dimer) MW = 24.62 × 2 = 49.24 kDa, Ahi29725I, 1 µg = ~677 nM. (2) Apa233I protein (dimer) MW = 24.17 × 2 = 48.34 kDa, Apa233I, 1 µg = ~690 nM. (3) HhiV4I protein (dimer) MW = 44.67 × 2 = 89.34 kDa, HhiV4I, 1 µg = ~373 nM. (4) FcyTI protein (dimer) MW = 30.76 kDa × 2 = 61.52 kDa, FcyTI 0.1 µg = 54 nM.

cyanobacteriorum, which grows at 20–30°C. Due to its better biochemical properties, FcyTI was used for the *in vivo* toxicity study (Figure 3) and for the digestion of modified oligos (see Figure 4). The FcyTI expression plasmid showed over a 1,000-fold reduction in transformation efficiency in the Dam⁺ T7 strain compared to a Dam⁻ host (Figure 3). FcyTI could be over-expressed only in the Dam⁻ T7 expression strain. Run-off sequencing demonstrated that the enzyme was able to cleave within the Gm6A↓TC recognition sequence (Supplementary Figure S4). The purified 6× His-tagged FcyTI shown in Supplementary Figure S5 was used for the digestion of modified or hemi-modified oligoduplexes.

To compare the activity of the enzyme toward fully methylated, hemi-methylated, and non-methylated DNA, we digested synthetic DNA oligoduplexes and quantified substrate and product amounts after restriction digestion by capillary electrophoresis (CE) (Figure 4). The results showed that FcyTI was most active on fully methylated

DNA but also had partial activity on hemi-methylated DNA, similar to DpnI. No digestion product was detected for non-methylated DNA. MboI was used as a control and digested only unmodified GATC oligos. Fully and hemi-modified substrates were resistant to MboI restriction (Figure 4). The original CE digestion results are shown in raw data (PeakScan analysis of CE peaks). Consistent with the duplex oligos digestion, unmodified pUC19 (HindIII-linearized) or phage DNA were poorly cleaved by FcyTI (Supplementary Figure S2), although weak activity was observed on Dam$^-$ pUC19, probably due to the high enzyme concentration.

PUA-wH-HNH fusion endonuclease HhiV4I

6× His-tagged HhiV4I (see Supplementary Figure S5) was subjected to three-step chromatography (Ni-agarose column, DEAE column, and Heparin agarose column). Compared to the recently published paper on the same enzyme (Lu et al., 2023), two additional chromatography steps were used (DEAE and Heparin columns). Unfortunately, the Heparin agarose chromatography step was less efficient for purification than is typical for other DNA-binding proteins because HhiV4I was in the flow-through and did not bind to the Heparin column, as would be expected for a typical nucleic acid-binding protein. As a result, HhiV4I was not purified to homogeneity (Supplementary Figure S5). Mass spectrometry analysis of the contaminations identified, among other proteins, *E. coli* exonuclease (Exonuclease VII, Exo VII) as a minor contaminant (see raw data for the HhiV4I mass spectrometry study). Exo VII cleaves single-stranded DNA (ssDNA) from both the 5′→3′ and 3′→5′ directions. This enzyme is not active on linear or circular dsDNA. The contaminating exonuclease would not likely interfere with major cut site determination, but it may interfere with minor cut site(s) by removing a few nucleotides for the cleaved ends by HhiV4I if the overhang is single-stranded.

The partially purified 6× His-tagged HhiV4I was used for HhiV4I characterization. Consistent with the findings of Lu et al. (2023), we observed that HhiV4I was much more active in the presence of Mn^{2+} ions than in the presence of other divalent metal cations (Figure 5A). HhiV4I showed weak DNA-nicking activity in Mg^{2+} buffer.

In agreement with the toxicity experiments (Figure 3) and the results of Lu et al. (2023), we found that the enzyme had higher activity against Dam$^+$ than Dam$^-$ pBR322, pUC19, λ DNA, and synthetic duplex oligos. If HhiV4I cleaved at or near Dam$^+$ sites, its cleavage products should be of similar size as those of DpnI digestion, and discrete bands (as opposed to a smear on the gel) should be observed. In our experiments, we saw only a partial match of fragment sizes, likely due to incomplete digestion (Figure 6A) (see below for the two-site requirement for efficient cleavage). Dam$^+$ phage λ DNA was also only partially digested while Dam$^-$ λ DNA was not cut at all (λ DNA was partially methylated by the host Dam methylase during rapid phage replication in *E. coli*, unpublished observation) (Figure 6B). When Dam$^-$ λ DNA was methylated *in vitro* by Dam methylase or M.EcoGII, the DNA substrates now became cleavable by HhiV4I (Figure 6C), further demonstrating that GATC methylation is required for restriction. M.EcoGII-modified λ DNA appeared to be a slightly better substrate for HhiV4I restriction than Dam-modified λ DNA, indicating that the wH might not be strictly limited to the detection of 6mA in the GATC context.

We could digest non-methylated DNA with excess HhiV4I, suggesting that the dependence of the enzyme on adenine methylation was not absolute. Most restriction enzymes display star activity at high enzyme, high glycerol concentration, or low salt. This conclusion was confirmed with the digestion of synthetic DNA with a defined adenine methylation status. As expected, HhiV4I was most active on fully methylated DNA but had some activity on hemi- and non-methylated DNA (Figure 4). In agreement with the findings by Lu et al. (2023) we did not detect activity of HhiV4I toward PCR products, which contained 5mC or 5hmC instead of C, in conditions conducive to digestion of m6A containing DNA (Figure 5B). Since the PCR products contain 5mC and 5hmC in many different contexts, this result suggests that the enzyme has no activity against methylated or hydroxymethylated DNA, despite the presence of the PUA (SRA-like) domain. This was surprising because it had been shown previously that PUA superfamily REases VcaM4I, SRA-like domain-containing endonuclease TagI, and PvuRts1I restricted DNA containing modified cytosines (Janosi et al., 1994; Pastor et al., 2021). Possible activity against WT T4 [glucosylated(g)-5hmC] modified DNAs remains to be tested. HhiV4I shows no activity on dZ (modified adenine, 2-aminoadenine, or 2,6-diaminopurine)-modified PCR DNA (Figure 5B).

HhiV4I prefers to cut between two G6mATC sites with optimal spacers of 13–27 bp in Dam$^+$ pBR322. Shorter spacers of 8–11 bp or longer spacers >42 bp were cleaved more slowly. Run-off sequencing of Dam$^+$ HhiV4I DNA confirmed that the enzyme cleaved in the vicinity of but not within the G6mATC sequence, as previously reported (Supplementary Figure S6).

PUA—wH—HNH and wH—HNH endonucleases

In contrast to the prophage-encoded HhiV4I, most PUA—wH—HNH enzymes (375–496 aa long) and wH—HNH endonucleases (224–283 aa long) are bacterial/archaeal enzymes. For 15 of these enzymes and HhiV4I as a positive control, we attempted expression in the Dam$^-$ *E. coli* cells. Moreover, we analyzed the transformation efficiency into Dam$^+$ (C2566) and Dam$^-$ (ER2948) cells compared to the empty vector. Restriction activity was examined in the presence of IPTG induction to elevate the genome conflict (Table 1). Some restriction genes, such as HhiV4I and SruGXI, had a strong toxic effect, as detected by a 100–1,000-fold reduction in transformation efficiency in the Dam$^+$ host. Other ORFs caused an approximately 10-fold reduction in transformation efficiency, consistent with partial restriction (+/−). The transformation of the HhaN23I gene caused the formation of very small colonies in the presence or absence of IPTG, indicating partial restriction. A few ORF constructs showed no difference in transformation efficiency in the Dam$^+$ host, presumably as a result of poor expression or lack of activity (e.g., HboP9I). As a control, the pTXB1 empty vector could be readily transferred into C2566 (Dam$^+$) or ER2948 (Dam$^-$) cells in the presence of IPTG (Table 1; Figure 7).

Selected enzymes that appeared to be promising as Dam$^+$-dependent MDREs were partially purified, and their activity was tested on Dam$^+$ pBR322 or λ DNA. The partially purified HtuIII enzyme (GenBank accession number NC_013743, PUA—wH—HNH fusion) shows a low nicking activity in Mn^{2+} or Co^{2+} buffer (Supplementary Figure S7). DNA run-off sequencing of the partially nicked pBR322 indicated that the nick occurred upstream of the

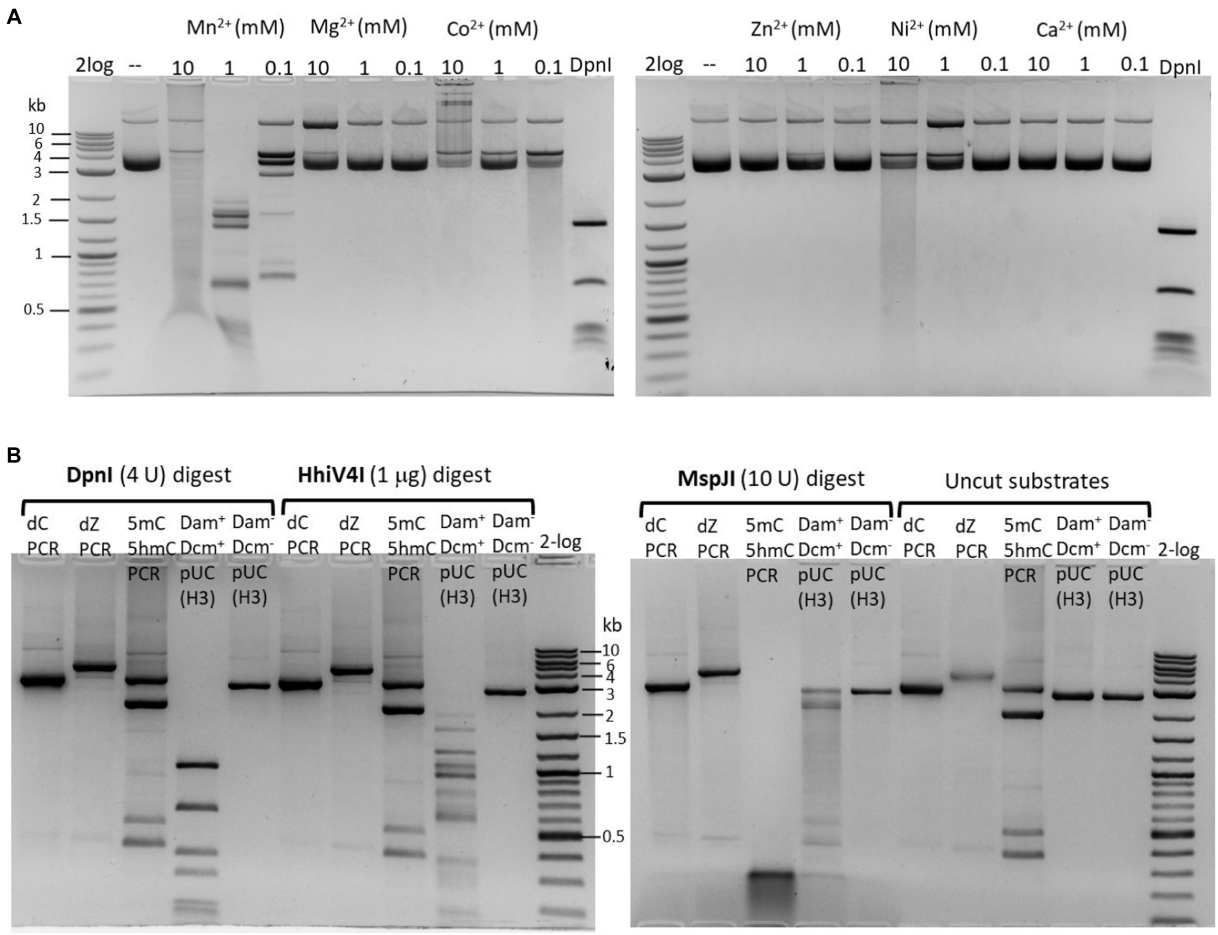

FIGURE 5
(A) Divalent cation requirement for HhiV4I restriction activity. Divalent cations were supplemented in a medium salt buffer (50 mM NaCl, 10 mM Tris–HCl, pH 7.5) at 10, 1, and 0.1 mM final concentration, respectively. HhiV4I (1 µg) was used in the digestion of pBR322 (Dam^+, 1 µg) at 37°C for 1 h. Mn^{2+} supports HhiV4I activity as the preferred cofactor. In 10 mM Mg^{2+} buffer, the enzyme showed partial nicking activity. In Co^{2+} and Ni^{2+} buffers, the enzyme showed weak but detectable activity. **(B)** HhiV4I incubation with 5mC–, 5hmC–, or dZ– (2-aminoadenine, 2,6-aminopurine) modified PCR DNAs. DpnI, MspJI, and HhiV4I digests were carried out in 1× NEB buffer 2.1, CutSmart buffer, and B2.1 plus 1 mM Mn^{2+}, respectively. Modified and unmodified DNA substrates: (1) dC regular PCR-unmodified (2.9 kb), (2) dZ PCR (2-aminoadenine modified, 4 kb), (3) a mixture of 5mC (2 kb) and 5hmC (2.9 kb) PCR products, and two minor PCR products (0.4–0.5 kb), (4) HindIII (H3)-prelinearized pUC19 (Dam^+ Dcm^+, 2.7 kb), (5) HindIII (H3)-prelinearized pUC19 (Dam^- Dcm^-, 2.7 kb). DpnI and HhiV4I digested linear Dam^+ pUC19 DNA. MspJI digested 5mC/5hmC modified PCR DNA and Dcm^+ linear pUC19 (MspJI site 5(h)mCNNR, Dcm-methylated sites C5mCWGG).

Gm6ATC site (top strand nicking only; ↓NGm6ATC-N14-Gm6ATC). The results of wH-HNH and PUA-wH-HNH endonuclease activities and *in vivo* toxicity are summarized in Table 1.

Analogous to HhiV4I, HtuIII also preferred Mn^{2+} or Co^{2+} for catalytic activity, suggesting that both enzymes have a unique metal ion binding site that is different from the typical HNH ββα-metal catalytic domain found in type II REases, homing endonucleases, Cas9, and non-specific endonucleases utilizing Mg^{2+} as a cofactor. The cofactor preferences are similar to the preferences of *E. coli* EcoKMcrA endonuclease and ScoMcrA (Liu et al., 2010).

wH—GIY-YIG endonucleases

Two wH—GIY-YIG fusion proteins, Ahi29725I (WP_035368356) and Apa233I (WP_026653965), were selected for characterization. The proteins occur naturally in *Acholeplasma hippikon* (ATCC29725 strain) and *Acholeplasma palma* (J233 strain), respectively. *Acholeplasma* are bacteria without cell walls in the Mollicutes class with small genomes (1.5–1.65 mbp). *Acholeplasma* species are found in animals, insects, and some plants in the environment. Some *Acholeplasma* species are pathogenic and can contaminate mammalian cell cultures. We expressed both proteins in Dam^- *E. coli* and purified the proteins by chromatography through chitin, DEAE, and Heparin columns. The analysis of the purified proteins on SDS-PAGE is shown in Supplementary Figure S5. Protein mass spectrometry analysis of the purified enzymes showed minimal exonuclease contamination (see raw data for protein composition analysis).

The divalent cation requirement for the Ahi29725I GIY-YIG endonuclease activity was assessed in a medium salt (50 mM NaCl) buffer (Supplementary Figure S8). Ahi29725I is active in Mg^{2+} (1–10 mM) and Mn^{2+} (0.1–10 mM) buffers and partially active in Co^{2+} (1–10 mM) and Ni^{2+} (1–10 mM) buffers in digestion of pBR322

10.3389/fmicb.2024.1286822

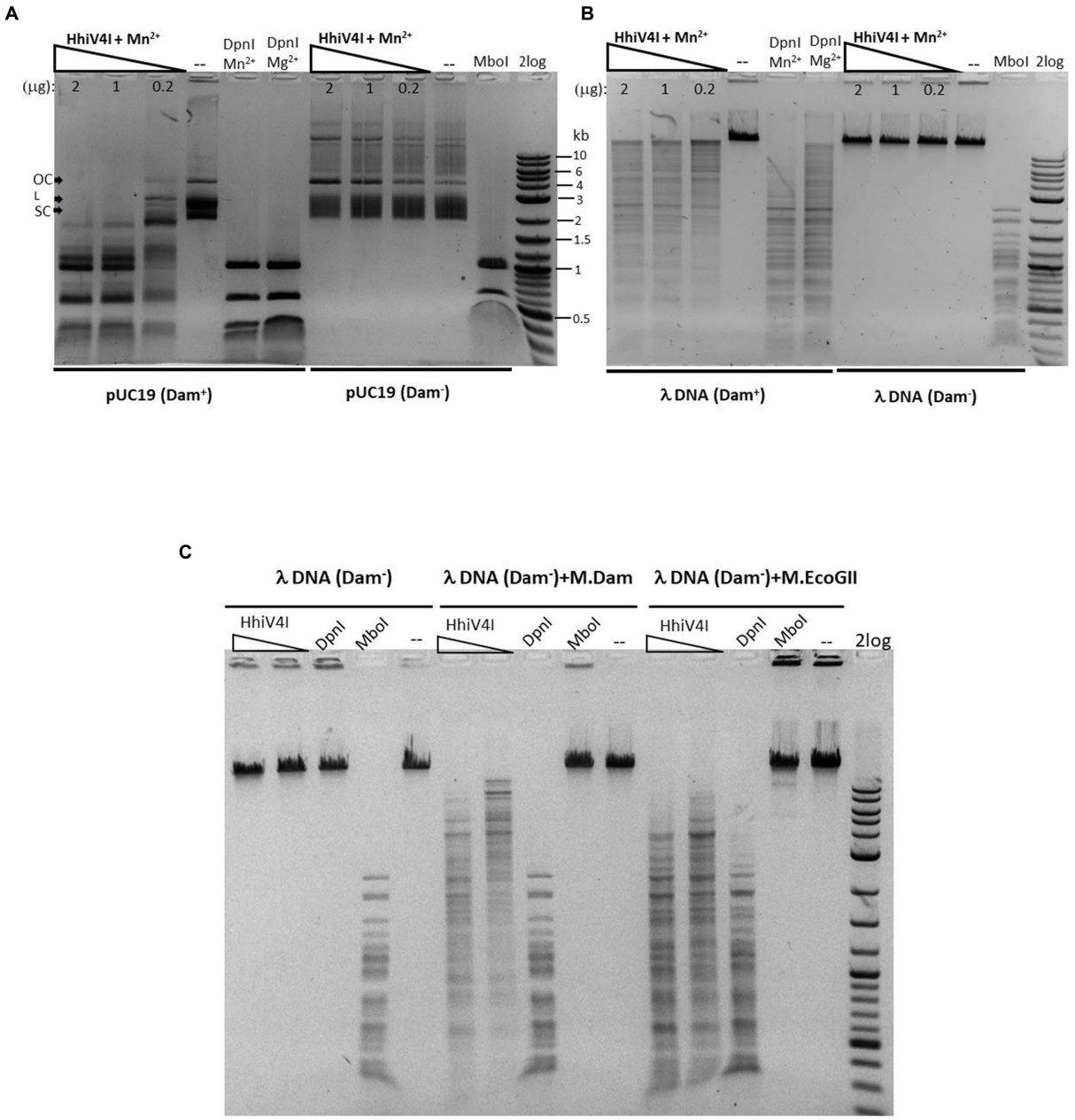

FIGURE 6
Methylation dependence of HhiV4I endonuclease. **(A)** Restriction digests of 1 µg of Dam$^+$ or Dam$^-$ pUC19 in Mn^{2+} buffer by HhiV4I (447.7, 223.9, and 44.8 nM, respectively, in 50 µL reaction volume). **(B)** Restriction digests of 1 µg of Dam$^+$ or Dam$^-$ phage λ DNA in Mn^{2+} buffer by HhiV4I (447.7, 223.9, and 44.8 nM, respectively, in 50 µL reaction volume). DpnI (2 U) cleaves Dam$^+$ DNA only. MboI (5 U) cleaves unmodified DNA only. **(C)** HhiV4I digestion (447.7 and 44.8 nM) of 1 µg of *in vitro* modified λ DNA (Dam$^-$) by Dam methylase or frequent adenine methylase M.EcoGII in Mn^{2+} buffer in 50 µL reaction volume. Following methylation reactions, the methylases were inactivated by heating at 65°C for 20 min. The methylated DNA was then diluted for restriction digestion in Mn^{2+} buffer.

(Dam$^+$). It has a nicking activity in Ca^{2+} (10 mM) buffer. To assess the 6mA-dependent restriction activity, Ahi29725I was also assayed on Dam$^-$ pBR322 in the presence of Mg^{2+}, Mn^{2+}, or Co^{2+} in restriction digests. The results showed that the enzyme digested Dam$^-$ DNA into a smearing pattern in Mn^{2+} buffer without discrete bands, probably as a result of loss of specificity (data not shown). It is known that type II REases and homing endonucleases (HEases) with the GIY-YIG endonuclease domain preferentially use Mg^{2+} divalent cation as a cofactor. The purified Apa233I showed a similar divalent cation preference as Ahi29725I since it is active in restriction buffers with Mg^{2+} or Mn^{2+} (data not shown).

The purified Ahi29725I enzyme was assayed on Dam$^+$ and Dam$^-$ λ DNA to test modification dependence (Figure 8). Dam$^-$ λ DNA was also methylated by Dam methylase (M.Dam) or EcoGII frequent adenine methylase (M.EcoGII) in the test tube and used for Ahi29725I digestions. The Ahi29725I endonuclease generated a partial digestion pattern on Dam$^+$ λ DNA. It showed no cleavage activity on Dam$^-$ λ DNA, indicating restriction activity dependent on Dam modification.

10.3389/fmicb.2024.1286822

TABLE 1 *In vivo* toxicity of PUA-wH-HNH and wH-HNH endonucleases.

GenBank accession number (protein)	*Halobacteria* (archaea) or bacterial strain	Enzyme name	Protein expression level	Toxicity *in vivo*	Activity *in vitro*
1. ARM71120.1	*Haloarcula hispanica* pleomorphic virus 4	HhiV4I (381 aa)	+++	+	+
2. WP_007980235.1	*Haladaptatus paucihalophilus* DX253	HpaD253I (234 aa)	+++	+	–
3. WP_009365977.1	*Halogranum salarium* B-1	Hsa1 (375 aa)	+++	+	–
4. WP_098725488.1	*Natrinema* sp. CBA1119	NspC1119I (480 aa)	+++	+	–
5. WP_049951637.1	*Halostagnicola larsenii* XH-48	HlaX48I (378 aa)	+++	+	–
6. WP_135828390.1	*Halorussus halobios* HD8-83	HhaH883I (224 aa)	+++	+	–
7. WP_103425724.1	*Salinigranum rubrum* GX10	SruGXI (232 aa)	+++	+	ND
8. WP_128477186.1	*Halorussus* sp. RC-68	HspR68I (283 aa)	++	+	ND
9. WP_012944509.1	*Haloterrigena turkmenica* DSM5511	HtuIII (496 aa)	++	+/–	+/–
10. WP_008893710.1	*Haloterrigena salina* JCM13891	Hsa13891I (496 aa)	+++	+/–	+/–
11. WP_126662294.1	*Haloterrigena salifodinae* ZY19	HsaZ19I (496 aa)	+++	+/–	–
12. WP_126597564.1	*Dictyobacter aurantiacus* S27 (G^+ bacterium)	DauS27I (372 aa)	–	+/–	–
13. WP_009486715.1	*Halobacterium* sp. DL1	HspD1I (230 aa)	–	+/–	–
14. WP_117591244.1	*Haloprofundus halophilus* NK23	HhaN23I (235 aa)	+++	+/– (small colonies)	ND
15. WP_006089674.1	*Natronorubrum tibetense* GA33	NtiG33I (492 aa)	+++	–	–
16. WP_159527272.1	*Halobacterrium bonnevillei* PCN9	HboP9I (230 aa)	+++	–	ND

Protein expression, *in vivo* toxicity, and *in vitro* activity of PUA—wH—HNH and wH—HNH endonucleases. The *in vivo* toxic effect of the restriction gene was measured by transformation into Dam^+ and Dam^- *E. coli* competent cells under IPTG induction. +, strong restriction (100–1,000-fold reduction in Dam^+ cells), +/–, mild restriction (~10-fold reduction in Dam^+ cells or formation of sick small colonies), –, no restriction. Protein expression level: +++, 2–10 mg protein per liter of IPTG-induced cells; ++, 1–1.5 mg/L; –, target protein not detected after chitin column purification or in IPTG-induced cell extract. Proteins with 375–496 aa residues are PUA—wH—HNH fusions; proteins below 300 aa residues are wH—HNH fusions. DauS27I is found in the G^+ bacterium *Dictyobacter aurantiacus* S27. The other enzymes are found in Archaea (*Halobacteria*). ND, not determined. Dam^+ pBR322 was used for *in vitro* cleavage assays.

When Dam^- λ DNA was methylated *in vitro* by Dam methylase or M.EcoGII, the modified substrates now became cleavable by Ahi29725I (Figure 8). In control digestion, MboI, DpnII, and Sau3AI are able to cleave Dam^- λ DNA, but DpnI cannot. Similarly, Ahi29725I and Apa233I endonucleases are also active on Dam^+ pBR322 and inactive on Dam^- pBR322 (Figure 9). However, high enzyme concentrations of Apa233I resulted in non-specific digestion (smearing) of Dam^- DNA. The finding was attributed to the non-specific activity on unmodified DNA, since most restriction enzymes display star activity at high enzyme, high glycerol concentration, or low salt.

The Ahi29725I and Apa233I digested pBR322 (Dam^+) DNA was subjected to run-off sequencing with primers annealing near the Gm6ATC sites. Cleavages occurred outside the recognition sequence, at a variable distance from the site of methylation (i.e., Ahi29725I cleaves G6mATC at N_{1-23}) (Supplementary Figures S9, S10). Ahi29725I and Apa233I have limited, if any, preference for cleavage at NN/RN and NN/GN sites, respectively (Supplementary Figure S11), which would have to be attributed to endonuclease sequence preferences.

To test whether Ahi29725I catalyzed DNA cleavage could be directed by adenine methylation in addition to the G6mATC sequence context, we digested M.EcoGII-modified pBR322 DNA (Dam^-) to see any enhancement of activity due to frequent adenine methylation. M.EcoGII is capable of methylating all adenines in DNA substrates except in polyA tracks (Murray et al., 2018). Ahi29725I activity was enhanced on M.EcoGII-methylated DNA

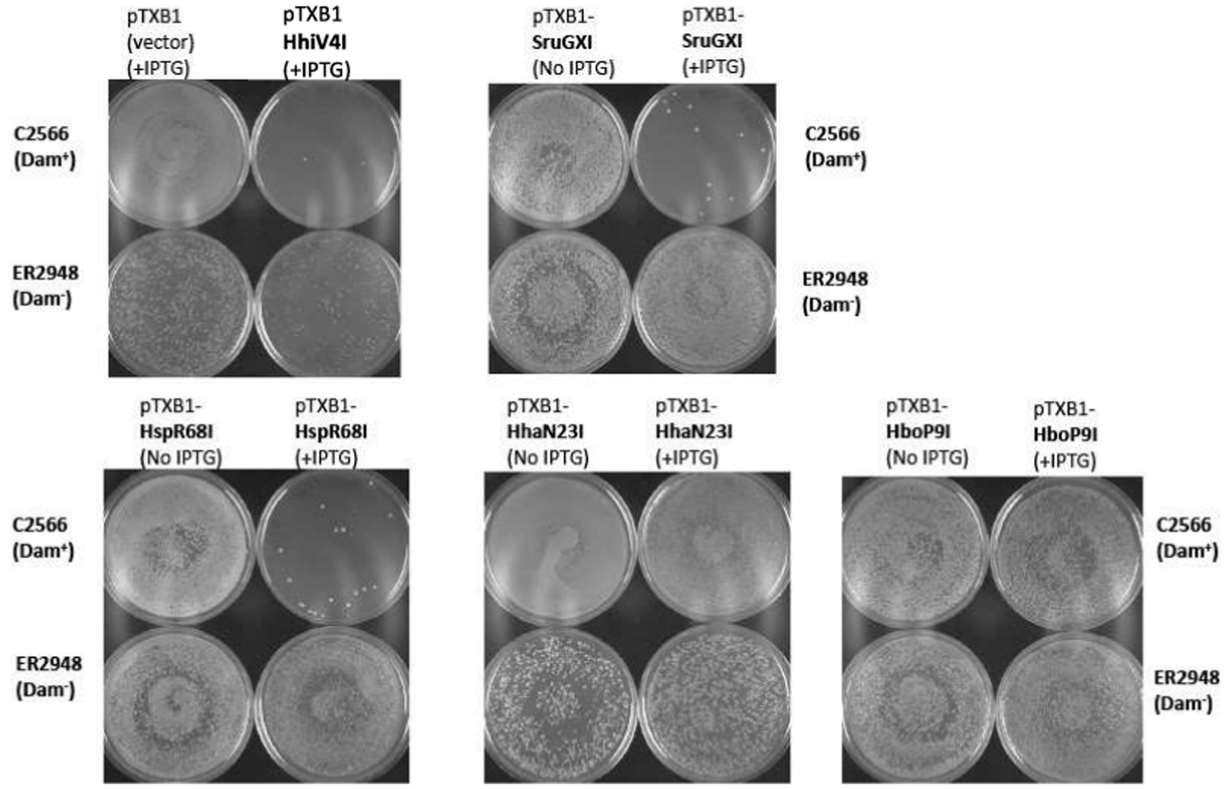

FIGURE 7
In vivo toxicity study: plasmid transfer into C2566 (Dam$^+$) and ER2948 (Dam$^-$) competent cells by transformation (~50 ng plasmid DNA). Three types of restriction phenotypes were observed: a strong reduction in transformation efficiency due to gene conflict (e.g., HhiV4I, SruGXI, and HspR68I); small colony formation in Dam$^+$ hosts presumably due to mild toxicity of the restriction gene (e.g., HhaN23I); and no noticeable change in transformation efficiency (e.g., HboP9I) compared to the empty vector control. The assay was done semi-qualitatively based on visualization of the transformation plates and not quantitatively since the number of colonies was not counted. Toxicity was more apparent with IPTG induction (0.5 mM IPTG in Amp plates). The overall transformation efficiency is lower in the Dam-deficient host.

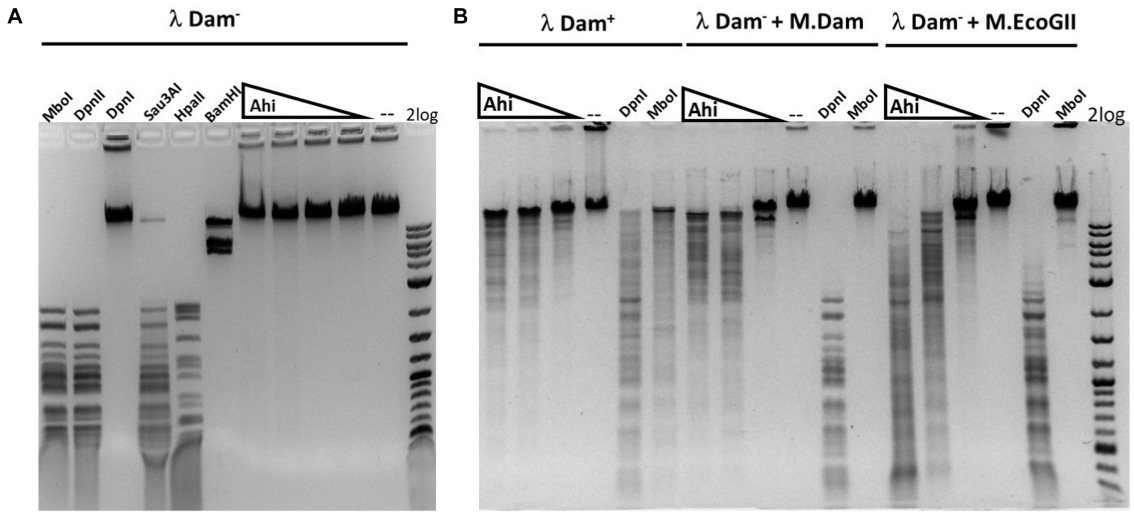

FIGURE 8
Restriction activity assay on Dam$^+$, Dam$^-$ λ DNA, Dam$^-$ λ DNA further modified by Dam methylase (M.Dam) or M.EcoGII. **(A)** Restriction digestion of Dam$^-$ λ DNA with Ahi29725I (Ahi) 10-fold serial dilution: 812.4, 81.2, and 8.1 nM (2, 0.2, and 0.02 μg) of protein incubated with 0.5 μg λ DNA in NEB B2.1 in 50 μL reaction volume at 37°C for 1 h. **(B)** 10-fold serial dilution of Ahi29725I (Ahi) in digestion of Dam$^+$ λ DNA, Dam$^-$ λ DNA methylated by Dam methylase or by M.EcoGII. λ DNA was only partially modified by *E. coli* host Dam methylase during rapid phage DNA replication. In control digests, MboI and DpnII cleave unmodified GATC sites only; DpnI cleaves Gm6ATC sites only; Sau3AI cleaves GATC sites regardless of m6A methylation. HpaII (CCGG) and BamHI-HF (GGATCC, not affected by Dam methylation) are additional controls.

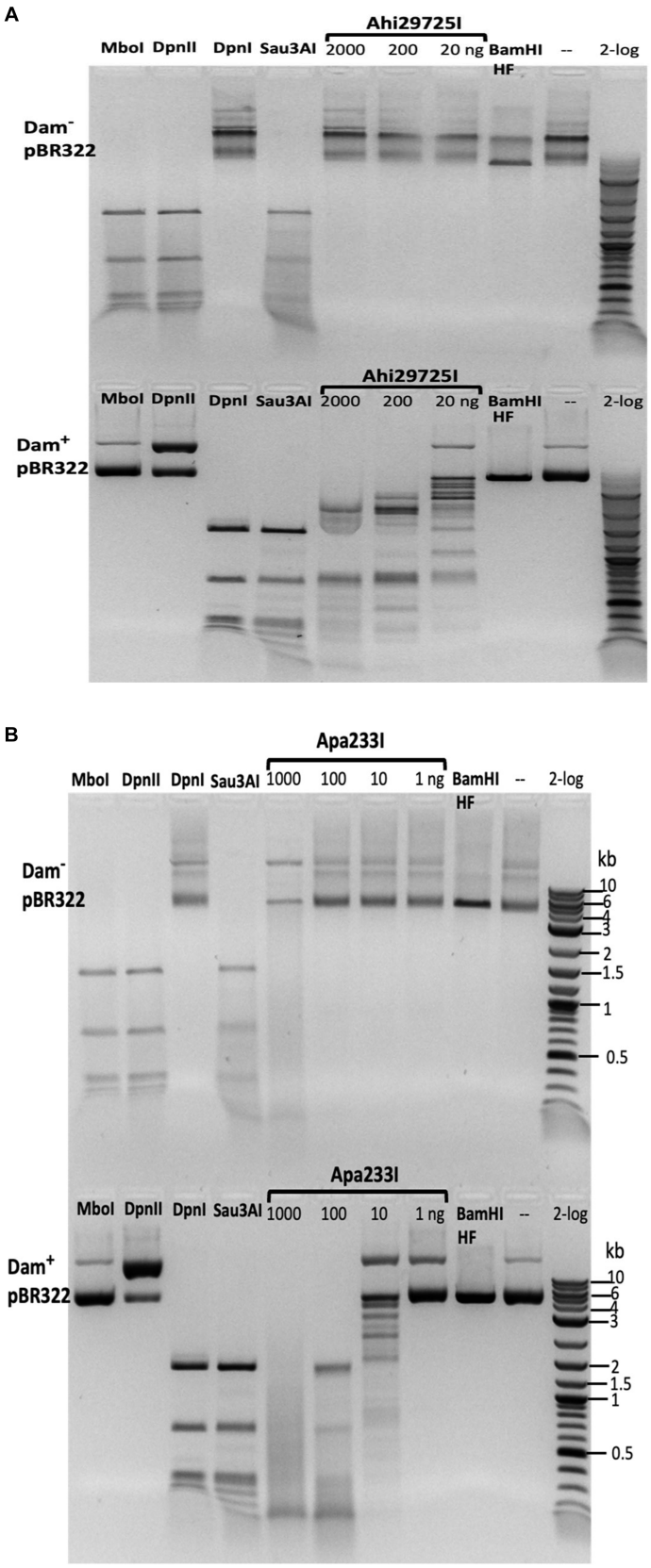

FIGURE 9
In vitro activity of Ahi29725I **(A)** and Apa233I **(B)** on Dam⁻ (top) and Dam⁺ (bottom) DNA. Ahi29725I (~812.4, 81.2, and 8.1 nM, respectively) or Apa233I (~413.7, 41.4, 4.1, and 0.4 nM, respectively) digestion of plasmid DNA (0.5 μg) was done in NEB B2.1 at 37°C for 1 h in a 50 μL reaction volume. In control digests, MboI and DpnII cleave unmodified GATC sites only; DpnI is unable to cut Dam⁻ DNA. Sau3AI cleaves GATC sites regardless of 6m A methylation. BamHI-HF (GGATCC) was used as an additional control. Apa233I showed non-specific endonuclease activity at high enzyme concentrations (~1 μg/0.4 μM). A smeared pattern was detected in both Dam⁺ and Dam⁻ pBR322.

substrate (Supplementary Figure S12). Three large fragments of Dam$^+$ DNA were further digested into smaller fragments after M.EcoGII methylation. However, it was not clear whether the enhanced activity was due to 6mA-dependent relaxed sequence recognition (e.g., cleavage near the Cm6ATC star site or Sm6ATS sites, S = G and C). The enhanced activity on M.EcoGII-modified DNA remains to be characterized in future by using defined modified oligos or restriction digestion/NGS sequencing mapping of M.EcoGII-modified λDNA.

In digestion of methylated duplex oligos with a single G6mATC site, it was noted that Ahi29725I preferentially cleaved fully methylated oligos (M+/M+) over hemi-modified substrates (M+/M− or M−/M+). However, Apa233I endonuclease was able to cut both fully modified and hemi-modified oligos (Figure 4). This discrepancy in methylation dependence between the two enzymes is unexplained.

If the *in vitro* divalent cation requirement of Ahi29725I was relevant in cells, invading 6mA-modified DNA would be digested by Mg^{2+}-bound Ahi29725I. By contrast, activation of the non-specific endonuclease activity of Ahi29725I with Mn^{2+}, Co^{2+}, or Ni^{2+} in the active site could lead to cleavage of both cellular and invading DNA regardless of modifications, triggering cell death and preventing phage release.

PLD—wH endonucleases

We identified 27 predicted PLD—wH fusion endonucleases in bacterial genomes. Two putative restriction genes from *Anaerolineaceae bacterium* (Aba4572I) and *Chloroflexi bacterium* (CbaI) were cloned into the pTXB1 expression vector. However, upon IPTG induction, no over-expressed proteins were detected. In the gene neighborhood analysis, the Aba45721 ORF resides in a genomic region of (1) DNA MTase (predicted specificity CGATCG, amino-MTase), (2) PLD-wH endonuclease, and (3) and (4) hypothetical proteins. If the CGATCG site is methylated to become CGm6ATCG, it would be a substrate for the PLD—wH endonuclease, which could potentially result in self-restriction. The CbaI enzyme is located in a region with (1) Leu-tRNA ligase, (2) restriction endonuclease, (3) PLD—wH endonuclease, (4) hypothetical protein, and (5) dimethyl-menaquinone MTase. Since the transformation of Aba4572I and CbaI was less toxic in Dam$^-$ cells in a non-induced condition (see Figure 3), the lack of expression in Dam$^-$ cells is surprising and requires further investigation. Expression of two more PLD-wH fusion proteins containing the conserved catalytic residues HxDx(4)K and HxEx(4)K in the PLD endonuclease domain in *E. coli* Dam$^-$ cells was not successful due to toxicity. More work is necessary to explain the reasons for the poor expression of PLD-wH fusion endonucleases in *E. coli*.

NTD—wH—NTPase fusions

The N-terminal domain-wH-NTPase fusions are usually paired with another catalytic subunit, such as an McrC (Ross et al., 1989) McrC-like protein with a PD-(D/E)XK endonuclease motif (Pieper and Pingoud, 2002). This arrangement is reminiscent of McrBC, a type IV restriction system acting on modified cytosines (Stewart et al., 2000). We only made one unsuccessful attempt to purify a putative heterodimeric NTD—wH—NTPase/McrC complex. Therefore, we have not studied the possible activity of these enzymes toward 6mA-containing DNA or their more general activity in the restriction of modified DNA. More work is needed to characterize this group of putative type IV restriction systems with wH-NTPase fusion.

Discussion

wH domain as a sensor of fully methylated ApT in a dsDNA context

The wH domain was first associated with adenine methylation because of its presence in the C-terminal region of *E. coli* and phage T4 Dam methyltransferases (Teichmann et al., 2012), and later because of its presence in the adenine methylation-dependent DpnI restriction endonuclease. A subsequent study on DpnI showed that the wH domain binds dsDNA at fully methylated ApT sites without base flipping. The two methyl groups, which are in close proximity, are bound in a single pocket of the wH domain of DpnI (Mierzejewska et al., 2014). The study on DpnI also showed that the specificity of the domain for the flanking sequence was somewhat relaxed with respect to the Dam Gm6ATC consensus and the Sm6ATS sites (where S stands for G or C) (Siwek et al., 2012). In this study, we show that the properties of an adenine methylation reader carry over to many fusions with HNH, GIY-YIG, and likely also PLD endonuclease domains. If methylation is seen in the ApT context, there is a preference for fully methylated sites except for Apa233I (see Figures 3, 4). Our study shows that in all tested fusion proteins, the wH domain can operate as an adenine methylation reader for the Gm6ATC context. A study on the wH—GIY-YIG endonucleases further indicates that additional cleavage sites are likely created when DNA is hypermethylated by M.EcoGII (see Supplementary Figure S12). Hence, the wH domains of the wH—GIY-YIG endonucleases also suggest that the Gm6ATC preference may be relaxed, but the star binding sites remain to be characterized (star sites are usually defined as DNA sequences with one base off from the cognate site; if there are two bases off, these sequences are usually called non-cognate sites; Pingoud et al., 2016).

The identification of prokaryotic winged helix domains as sensors of adenine methylation contrasts with the role of some eukaryotic winged helix domains as sensors of non-methylated CpG (Stielow et al., 2021; Becht et al., 2023; Weber et al., 2023). The superposition of the winged helix domains of prokaryotic DpnI (Mierzejewska et al., 2014) and eukaryotic KAT6A (also called histone acetyltransferase KAT6A, lysine acetyltransferase 6A, zinc finger protein 220, and MYST-3) (Weber et al., 2023) shows that the dsDNA molecules are bound to opposite faces of the wH domain (Figure 10), indicating that the two DNA-binding modes have likely evolved independently for needs that are characteristic for prokaryotes (sensing of Dam methylation) and eukaryotes (sensing of the absence of CpG methylation).

Cooperation with endonuclease domains

For most NTP-independent MDREs, there is a clear division of labor between the modification reader and endonuclease domains. The

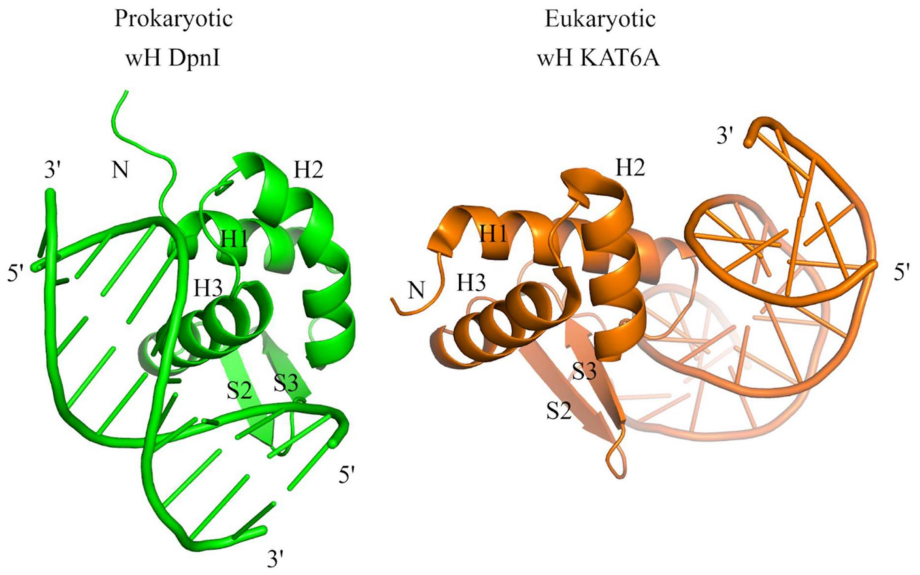

FIGURE 10
Structural comparison of the wH domains of DpnI (G6mATC) with bacterial origin and mammalian KAT6A protein that recognizes unmodified CpG sites. The docking of wH domains on dsDNA shows a major difference in recognition. The wH domain-containing protein KAT6A is a histone lysine acetyltransferase that acetylates lysine residues in histones H3 and H4 (*in vitro*) (Weber et al., 2023).

former recruits the enzyme to modification sites, and the latter cleaves the DNA at a distance from the recognition site, which is likely defined by the length of the linker that connects the two domains. The nuclease domain has generally low or only a very relaxed sequence specificity, and it is likely not modification-specific. How the modification sensor domain keeps the activity of the nuclease domain in check is not well understood. In some cases, it can be shown that the linker has an inhibitory role for the endonuclease that is only relieved once modified DNA is bound to the reader domain and the complex reorganizes structurally (Pastor et al., 2021). The PUA—wH—HNH (HhiV4I) and wH—GIY-YIG (Ahi29725I and Apa233I) that were tested by run-off sequencing are consistent with this expectation. As for the cytosine modification-specific MDREs, cleavage occurred mostly at a distance from the recognition sequence except for the BisI family REases (e.g., Eco15I and NhoI) that cut within the recognition sequence GCNGC with 2–4 modified cytosines (Xu et al., 2016).

Among the wH fusion endonucleases, DpnI and its isochizomers are the exception to the rule that cleavage occurs always outside of and not within the recognition sequence. Mechanically, DpnI DNA cleavage within the recognition sequence is a consequence of the fact that the endonuclease domain has separate sequence and modification specificities (Siwek et al., 2012). In this scenario, the role of the wH domain is similar to the role of the extra specificity domain in type IIE restriction endonucleases (Roberts et al., 2003), except that the target sequence contains a modified base. Type IIE restriction endonucleases are one of the subfamilies that require a pair of target sequences in order to show activity (Senesac and Allen, 1995; Colandene and Topal, 1998). Run-off sequencing shows that FcyTI and Psp4BI cleave inside the recognition sequence, like DpnI, pointing to the separate sequence and modification specificity of the catalytic PD-(D/E)XK domain. Given the high sequence conservation of the PD-(D/E)XK—wH endonuclease family, it is likely that this property is general for the entire family.

Sensors/readers for N6-methyladenine in DNA

Despite the many roles of DNA adenine methylation in prokaryotes (and eukaryotic organelles), the repertoire of reader domains for m6A in DNA is still surprisingly limited (Iyer et al., 2016). Perhaps the best-known adenine methylation sensors are the YTH (Liao et al., 2018) domains, which belong to (or are related to) the PUA superfamily domains. The PUA superfamily domains are believed to flip the modified 2′-deoxynucleotide out of duplex DNA (Shao et al., 2014) or to bind a single nucleotide in RNA in the reader pocket (Li et al., 2014; Xu et al., 2014). Therefore, at least when acting in isolation, they can be considered as sensors of a single modified adenine. Consistent with this role, most YTH domains sense adenine methylation in RNA rather than DNA. However, some YTH and ASCH domains in prokaryotes are considered as DNA adenine methylation sensors (Iyer et al., 2016). For the ASCH domains, this remains to be experimentally shown, since currently only a 4mC reader role is experimentally supported (Stanislauskiene et al., 2020). Apart from the YTH and ASCH domains, the HARE-HTH and RAMA (Restriction enzyme and Adenine Methylation Associated) domains have also been suggested to serve as readers of adenine methylation in DNA (Teichmann et al., 2012). The HARE-HTH domains are related to winged helix domains, which would be consistent with a role as adenine methylation sensors. However, they have an extra helix inserted into the HTH motif of the winged helix domain, and recent analysis suggests that they are more likely to sense cytosine modifications (Aravind and Iyer, 2012). Unlike

the HARE-HTH domains, the RAMA domains are unrelated to the wH domain in fold (Yang et al., 2023). For the RAMA domain-containing MPND protein, there is some biochemical evidence for adenine methylation sensing (Kweon et al., 2019). However, a preference for adenine-methylated DNA could not be experimentally confirmed (Yang et al., 2023). We noticed the occurrence of RAMA—Mrr catalytic domain (PD-QXK)—NTPase (three-domain fusion) and GIY-YIG—RAMA (two-domain fusion) in prokaryotes, which might indicate that the RAMA domain is utilized similarly to the wH domain in these fusions. Future studies will be focused on the characterization of YTH-NTPase, YTH-HNH, and RAMA fusion endonucleases as 6mA readers/sensors in type IV restriction systems.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

IH: Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – review & editing. DH: Data curation, Formal analysis, Investigation, Methodology, Writing – review & editing. WY: Data curation, Formal analysis, Investigation, Methodology, Software, Writing – review & editing. TV: Data curation, Investigation, Writing – review & editing. AH: Data curation, Investigation, Writing – review & editing. TL: Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – review & editing. LE: Data curation, Formal analysis, Investigation, Methodology, Software, Writing – review & editing. MB: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing. S-yX: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Part of this study was supported by the Polish NAWA (PPI/APM/2018/1/00034/U/001) and NCN grants (2018/30/Q/NZ2/00669). This study was also supported by New England Biolabs, Inc. (NEB). This publication cost was paid for by NEB.

Acknowledgments

We thank Andy Gardner, Tom Evans, Rich Roberts, Jim Ellard, and Salvatore Russello for their continued support. We are grateful to Lise Raleigh for providing the Dam-deficient T7 expression strain ER2948. We thank Honorata Czapinska and Lise Raleigh for their critical reading of the manuscript.

Conflict of interest

WY, DH, TV, AH, LE, and S-yX (retired) are employees of New England Biolabs, Inc., a company providing molecular biology reagents to the research and diagnostic community. IH and TL were short-term visiting scientists at New England Biolabs, Inc.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2024.1286822/full#supplementary-material

References

Aravind, L., and Iyer, L. M. (2012). The HARE-HTH and associated domains: novel modules in the coordination of epigenetic DNA and protein modifications. *Cell Cycle* 11, 119–131. doi: 10.4161/cc.11.1.18475

Au, K. G., Welsh, K., and Modrich, P. (1992). Initiation of methyl-directed mismatch repair. *J. Biol. Chem.* 267, 12142–12148. doi: 10.1016/S0021-9258(19)49816-5

Becht, D. C., Klein, B. J., Kanai, A., Jang, S. M., Cox, K. L., Zhou, B. R., et al. (2023). MORF and MOZ acetyltransferases target unmethylated CpG islands through the winged helix domain. *Nat. Commun.* 14:697. doi: 10.1038/s41467-023-36368-5

Bertonati, C., Punta, M., Fischer, M., Yachdav, G., Forouhar, F., Zhou, W., et al. (2009). Structural genomics reveals EVE as a new ASCH/PUA-related domain. *Proteins* 75, 760–773. doi: 10.1002/prot.22287

Borgaro, J. G., and Zhu, Z. (2013). Characterization of the 5-hydroxymethylcytosine-specific DNA restriction endonucleases. *Nucleic Acids Res.* 41, 4198–4206. doi: 10.1093/nar/gkt102

Boye, E., and Lobner-Olesen, A. (1990). The role of dam methyltransferase in the control of DNA replication in *E. coli*. *Cell* 62, 981–989. doi: 10.1016/0092-8674(90)90272-G

Brennan, R. G. (1993). The winged-helix DNA-binding motif: another helix-turn-helix takeoff. *Cell* 74, 773–776. doi: 10.1016/0092-8674(93)90456-Z

Bujnicki, J. M., and Rychlewski, L. (2001). Grouping together highly diverged PD-(D/E)XK nucleases and identification of novel superfamily members using structure-guided alignment of sequence profiles. *J. Mol. Microbiol. Biotechnol.* 3, 69–72. doi: 10.1016/s0378-1119(01)00405-x

Cerrudo, C. S., Ghiringhelli, P. D., and Gomez, D. E. (2014). Protein universe containing a PUA RNA-binding domain. *FEBS J.* 281, 74–87. doi: 10.1111/febs.12602

Chan, S. H., Bao, Y., Ciszak, E., Laget, S., and Xu, S. Y. (2007). Catalytic domain of restriction endonuclease BmrI as a cleavage module for engineering endonucleases

with novel substrate specificities. *Nucleic Acids Res.* 35, 6238–6248. doi: 10.1093/nar/gkm665

Cohen-Karni, D., Xu, D., Apone, L., Fomenkov, A., Sun, Z., Davis, P. J., et al. (2011). The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. *Proc. Natl. Acad. Sci. USA* 108, 11040–11045. doi: 10.1073/pnas.1018448108

Colandene, J. D., and Topal, M. D. (1998). The domain organization of NaeI endonuclease: separation of binding and catalysis. *Proc. Natl. Acad. Sci. USA* 95, 3531–3536. doi: 10.1073/pnas.95.7.3531

Czapinska, H., Kowalska, M., Zagorskaite, E., Manakova, E., Slyvka, A., Xu, S. Y., et al. (2018). Activity and structure of EcoKMcrA. *Nucleic Acids Res.* 46, 9829–9841. doi: 10.1093/nar/gky731

Dunin-Horkawicz, S., Feder, M., and Bujnicki, J. M. (2006). Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics* 7:98. doi: 10.1186/1471-2164-7-98

Flick, K. E., Jurica, M. S., Monnat, R. J. Jr., and Stoddard, B. L. (1998). DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature* 394, 96–101. doi: 10.1038/27952

Fomenkov, A., Xiao, J. P., Dila, D., Raleigh, E., and Xu, S. Y. (1994). The 'endo-blue method' for direct cloning of restriction endonuclease genes in *E. coli*. *Nucleic Acids Res.* 22, 2399–2403. doi: 10.1093/nar/22.12.2399

Frickey, T., and Lupas, A. (2004). CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 20, 3702–3704. doi: 10.1093/bioinformatics/bth444

Gajiwala, K. S., and Burley, S. K. (2000). Winged helix proteins. *Curr. Opin. Struct. Biol.* 10, 110–116. doi: 10.1016/S0959-440X(99)00057-3

Gajiwala, K. S., Chen, H., Cornille, F., Roques, B. P., Reith, W., Mach, B., et al. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* 403, 916–921. doi: 10.1038/35002634

Grazulis, S., Manakova, E., Roessle, M., Bochtler, M., Tamulaitiene, G., Huber, R., et al. (2005). Structure of the metal-independent restriction enzyme BfiI reveals fusion of a specific DNA-binding domain with a nonspecific nuclease. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15797–15802. doi: 10.1073/pnas.0507949102

Heitman, J., and Model, P. (1990). Substrate recognition by the EcoRI endonuclease. *Proteins* 7, 185–197. doi: 10.1002/prot.340070207

Horton, J. R., Wang, H., Mabuchi, M. Y., Zhang, X., Roberts, R. J., Zheng, Y., et al. (2014). Modification-dependent restriction endonuclease, MspJI, flips 5-methylcytosine out of the DNA helix. *Nucleic Acids Res.* 42, 12092–12101. doi: 10.1093/nar/gku871

Hosford, C. J., Bui, A. Q., and Chappie, J. S. (2020). The structure of the *Thermococcus gammatolerans* McrB N-terminal domain reveals a new mode of substrate recognition and specificity among McrB homologs. *J. Biol. Chem.* 295, 743–756. doi: 10.1016/S0021-9258(17)49932-7

Iyer, L. M., Burroughs, A. M., and Aravind, L. (2006). The ASCH superfamily: novel domains with a fold related to the PUA domain and a potential role in RNA metabolism. *Bioinformatics* 22, 257–263. doi: 10.1093/bioinformatics/bti767

Iyer, L. M., Zhang, D., and Aravind, L. (2016). Adenine methylation in eukaryotes: apprehending the complex evolutionary history and functional potential of an epigenetic modification. *Bioessays* 38, 27–40. doi: 10.1002/bies.201500104

Jablonska, J., Matelska, D., Steczkiewicz, K., and Ginalski, K. (2017). Systematic classification of the his-me finger superfamily. *Nucleic Acids Res.* 45, 11479–11494. doi: 10.1093/nar/gkx924

Janosi, L., Yonemitsu, H., Hong, H., and Kaji, A. (1994). Molecular cloning and expression of a novel hydroxymethylcytosine-specific restriction enzyme (PvuRts1I) modulated by glucosylation of DNA. *J. Mol. Biol.* 242, 45–61. doi: 10.1006/jmbi.1994.1556

Josephs, E. A., Zheng, T., and Marszalek, P. E. (2015). Atomic force microscopy captures the initiation of methyl-directed DNA mismatch repair. *DNA Repair* 35, 71–84. doi: 10.1016/j.dnarep.2015.08.006

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637. doi: 10.1002/bip.360221211

Kaminska, K. H., Kawai, M., Boniecki, M., Kobayashi, I., and Bujnicki, J. M. (2008). Type II restriction endonuclease R.Hpy188I belongs to the GIY-YIG nuclease superfamily, but exhibits an unusual active site. *BMC Struct. Biol.* 8:48. doi: 10.1186/1472-6807-8-48

Kazrani, A. A., Kowalska, M., Czapinska, H., and Bochtler, M. (2014). Crystal structure of the 5hmC specific endonuclease PvuRts1I. *Nucleic Acids Res.* 42, 5929–5936. doi: 10.1093/nar/gku186

Kosinski, J., Feder, M., and Bujnicki, J. M. (2005). The PD-(D/E)XK superfamily revisited: identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function. *BMC Bioinformatics* 6:172. doi: 10.1186/1471-2105-6-172

Kweon, S. M., Chen, Y., Moon, E., Kvederaviciute, K., Klimasauskas, S., and Feldman, D. E. (2019). An adversarial DNA N(6)-Methyladenine-sensor network preserves Polycomb silencing. *Mol. Cell* 74, 1138–1147.e6. doi: 10.1016/j.molcel.2019.03.018

Lai, E., Clark, K. L., Burley, S. K., and Darnell, J. E. Jr. (1993). Hepatocyte nuclear factor 3/fork head or "winged helix" proteins: a family of transcription factors of diverse biologic function. *Proc. Natl. Acad. Sci. U. S. A.* 90, 10421–10423. doi: 10.1073/pnas.90.22.10421

Li, F., Zhao, D., Wu, J., and Shi, Y. (2014). Structure of the YTH domain of human YTHDF2 in complex with an m(6)a mononucleotide reveals an aromatic cage for m(6)a recognition. *Cell Res.* 24, 1490–1492. doi: 10.1038/cr.2014.153

Liao, S., Sun, H., Xu, C., and Domain, Y. T. H. (2018). A family of N(6)-methyladenosine (m(6)a) readers. *Genomics Proteomics Bioinformatics* 16, 99–107. doi: 10.1016/j.gpb.2018.04.002

Liu, G., Ou, H. Y., Wang, T., Li, L., Tan, H., Zhou, X., et al. (2010). Cleavage of phosphorothioated DNA and methylated DNA by the type IV restriction endonuclease ScoMcrA. *PLoS Genet.* 6:e1001253. doi: 10.1371/journal.pgen.1001253

Lu, X., Huang, F., Cheng, R., and Zhu, B. (2023). A unique m6A-dependent restriction endonuclease from an archaeal virus. *Microbiol. Spectr.* 11:e0426222. doi: 10.1128/spectrum.04262-22

Lutz, T., Flodman, K., Copelas, A., Czapinska, H., Mabuchi, M., Fomenkov, A., et al. (2019). A protein architecture guided screen for modification dependent restriction endonucleases. *Nucleic Acids Res.* 47, 9761–9776. doi: 10.1093/nar/gkz755

Makarova, K. S., Wolf, Y. I., Snir, S., and Koonin, E. V. (2011). Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.* 193, 6039–6056. doi: 10.1128/JB.05535-11

Marinus, M. G., and Casadesus, J. (2009). Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS Microbiol. Rev.* 33, 488–503. doi: 10.1111/j.1574-6976.2008.00159.x

Mehta, P., Katta, K., and Krishnaswamy, S. (2004). HNH family subclassification leads to identification of commonality in the his-me endonuclease superfamily. *Protein Sci.* 13, 295–300. doi: 10.1110/ps.03115604

Mierzejewska, K., Siwek, W., Czapinska, H., Kaus-Drobek, M., Radlinska, M., Skowronek, K., et al. (2014). Structural basis of the methylation specificity of R.DpnI. *Nucleic Acids Res.* 42, 8745–8754. doi: 10.1093/nar/gku546

Miyazono, K., Furuta, Y., Watanabe-Matsui, M., Miyakawa, T., Ito, T., Kobayashi, I., et al. (2014). A sequence-specific DNA glycosylase mediates restriction-modification in Pyrococcus abyssi. *Nat. Commun.* 5:3178. doi: 10.1038/ncomms4178

Murray, I. A., Morgan, R. D., Luyten, Y., Fomenkov, A., Correa, I. R. Jr., Dai, N., et al. (2018). The non-specific adenine DNA methyltransferase M.EcoGII. *Nucleic Acids Res.* 46, 840–848. doi: 10.1093/nar/gkx1191

Orlowski, J., and Bujnicki, J. M. (2008). Structural and evolutionary classification of type II restriction enzymes based on theoretical and experimental analyses. *Nucleic Acids Res.* 36, 3552–3569. doi: 10.1093/nar/gkn175

Pastor, M., Czapinska, H., Helbrecht, I., Krakowska, K., Lutz, T., Xu, S. Y., et al. (2021). Crystal structures of the EVE-HNH endonuclease VcaM4I in the presence and absence of DNA. *Nucleic Acids Res.* 49, 1708–1723. doi: 10.1093/nar/gkaa1218

Pieper, U., and Pingoud, A. (2002). A mutational analysis of the PD…D/EXK motif suggests that McrC harbors the catalytic center for DNA cleavage by the GTP-dependent restriction enzyme McrBC from *Escherichia coli*. *Biochemistry* 41, 5236–5244. doi: 10.1021/bi0156862

Pingoud, A., Fuxreiter, M., Pingoud, V., and Wende, W. (2005). Type II restriction endonucleases: structure and mechanism. *Cell. Mol. Life Sci.* 62, 685–707. doi: 10.1007/s00018-004-4513-1

Pingoud, A., and Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* 29, 3705–3727. doi: 10.1093/nar/29.18.3705

Pingoud, A., Wilson, G. G., and Wende, W. (2016). Type II restriction endonucleases - a historical perspective and more. *Nucleic Acids Res.* 44:8011. doi: 10.1093/nar/gkw513

Pommer, A. J., Cal, S., Keeble, A. H., Walker, D., Evans, S. J., Kuhlmann, U. C., et al. (2001). Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. *J. Mol. Biol.* 314, 735–749. doi: 10.1006/jmbi.2001.5189

Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., et al. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* 31, 1805–1812. doi: 10.1093/nar/gkg274

Roberts, R. J., and Cheng, X. (1998). Base flipping. *Annu. Rev. Biochem.* 67, 181–198. doi: 10.1146/annurev.biochem.67.1.181

Ross, T. K., Achberger, E. C., and Braymer, H. D. (1989). Identification of a second polypeptide required for McrB restriction of 5-methylcytosine-containing DNA in *Escherichia coli* K12. *Mol. Gen. Genet.* 216, 402–407. doi: 10.1007/BF00334382

Sasnauskas, G., Zakrys, L., Zaremba, M., Cosstick, R., Gaynor, J. W., Halford, S. E., et al. (2010). A novel mechanism for the scission of double-stranded DNA: BfiI cuts both 3′-5′ and 5′-3′ strands by rotating a single active site. *Nucleic Acids Res.* 38, 2399–2410. doi: 10.1093/nar/gkp1194

Schwartz, T., Behlke, J., Lowenhaupt, K., Heinemann, U., and Rich, A. (2001). Structure of the DLM-1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins. *Nat. Struct. Biol.* 8, 761–765. doi: 10.1038/nsb0901-761

Senesac, J. H., and Allen, J. R. (1995). Oligonucleotide activation of the type IIe restriction enzyme NaeI for digestion of refractory sites. *BioTechniques* 19, 990–993.

Shao, C., Wang, C., and Zang, J. (2014). Structural basis for the substrate selectivity of PvuRts1I, a 5-hydroxymethylcytosine DNA restriction endonuclease. *Acta Crystallogr. D Biol. Crystallogr.* 70, 2477–2486. doi: 10.1107/S139900471401606X

Siwek, W., Czapinska, H., Bochtler, M., Bujnicki, J. M., and Skowronek, K. (2012). Crystal structure and mechanism of action of the N6-methyladenine-dependent type IIM restriction endonuclease R.DpnI. *Nucleic Acids Res.* 40, 7563–7572. doi: 10.1093/nar/gks428

Slyvka, A., Zagorskaite, E., Czapinska, H., Sasnauskas, G., and Bochtler, M. (2019). Crystal structure of the EcoKMcrA N-terminal domain (NEco): recognition of modified cytosine bases without flipping. *Nucleic Acids Res.* 47, 11943–11955. doi: 10.1093/nar/gkz1017

Sokolowska, M., Czapinska, H., and Bochtler, M. (2009). Crystal structure of the beta beta alpha-me type II restriction endonuclease Hpy99I with target DNA. *Nucleic Acids Res.* 37, 3799–3810. doi: 10.1093/nar/gkp228

Sokolowska, M., Czapinska, H., and Bochtler, M. (2011). Hpy188I-DNA pre- and post-cleavage complexes--snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Res.* 39, 1554–1564. doi: 10.1093/nar/gkq821

Stanislauskiene, R., Laurynenas, A., Rutkiene, R., Aucynaite, A., Tauraite, D., Meskiene, R., et al. (2020). YqfB protein from *Escherichia coli*: an atypical amidohydrolase active towards N(4)-acylcytosine derivatives. *Sci. Rep.* 10:788. doi: 10.1038/s41598-020-57664-w

Stewart, F. J., Panne, D., Bickle, T. A., and Raleigh, E. A. (2000). Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *J. Mol. Biol.* 298, 611–622. doi: 10.1006/jmbi.2000.3697

Stielow, B., Zhou, Y., Cao, Y., Simon, C., Pogoda, H. M., Jiang, J., et al. (2021). The SAM domain-containing protein 1 (SAMD1) acts as a repressive chromatin regulator at unmethylated CpG islands. *Sci. Adv.* 7:eabf2229. doi: 10.1126/sciadv.abf2229

Sukackaite, R., Grazulis, S., Tamulaitis, G., and Siksnys, V. (2012). The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. *Nucleic Acids Res.* 40, 7552–7562. doi: 10.1093/nar/gks332

Tang, Q., Rigby, R. E., Young, G. R., Hvidt, A. K., Davis, T., Tan, T. K., et al. (2021). Adenosine-to-inosine editing of endogenous Z-form RNA by the deaminase ADAR1 prevents spontaneous MAVS-dependent type I interferon responses. *Immunity* 54, 1961–1975.e5. doi: 10.1016/j.immuni.2021.08.011

Teichmann, M., Dumay-Odelot, H., and Fribourg, S. (2012). Structural and functional aspects of winged-helix domains at the core of transcription initiation complexes. *Transcription* 3, 2–7. doi: 10.4161/trns.3.1.18917

Wah, D. A., Hirsch, J. A., Dorner, L. F., Schildkraut, I., and Aggarwal, A. K. (1997). Structure of the multimodular endonuclease FokI bound to DNA. *Nature* 388, 97–100. doi: 10.1038/40446

Weber, L. M., Jia, Y., Stielow, B., Gisselbrecht, S. S., Cao, Y., Ren, Y., et al. (2023). The histone acetyltransferase KAT6A is recruited to unmethylated CpG islands via a DNA binding winged helix domain. *Nucleic Acids Res.* 51, 574–594. doi: 10.1093/nar/gkac1188

Wolberger, C., and Campbell, R. (2000). New perch for the winged helix. *Nat. Struct. Biol.* 7, 261–262. doi: 10.1038/74004

Wu, C. C., Lin, J. L. J., and Yuan, H. S. (2020). Structures, mechanisms, and functions of his-me finger nucleases. *Trends Biochem. Sci.* 45, 935–946. doi: 10.1016/j.tibs.2020.07.002

Xu, S. Y., Klein, P., Degtyarev, S., and Roberts, R. J. (2016). Expression and purification of the modification-dependent restriction enzyme BisI and its homologous enzymes. *Sci. Rep.* 6:28579. doi: 10.1038/srep28579

Xu, D., Shao, J., Song, H., and Wang, J. (2020). The YTH Domain family of N6-Methyladenosine "readers" in the diagnosis and prognosis of colonic adenocarcinoma. *Biomed. Res. Int.* 2020:9502560. doi: 10.1155/2020/9502560

Xu, C., Wang, X., Liu, K., Roundtree, I. A., Tempel, W., Li, Y., et al. (2014). Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat. Chem. Biol.* 10, 927–929. doi: 10.1038/nchembio.1654

Yang, M., Li, X., Tian, Z., Ma, L., Ma, J., Liu, Y., et al. Structures of MPND reveal the molecular recognition of nucleosomes. *Int. J. Mol. Sci.* 24:3368. doi: 10.3390/ijms24043368

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 11 April 2024
DOI 10.3389/fmicb.2024.1388522

OPEN ACCESS

EDITED BY
Axel Cloeckaert,
Institut National de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), France

REVIEWED BY
Sung-Jae Lee,
Kyung Hee University, Republic of Korea
Jozsef Soki,
University of Szeged, Hungary
Hirokazu Suzuki,
Tottori University, Japan

*CORRESPONDENCE:
Milton H. Saier Jr.
✉ msaier@ucsd.edu
Zhongge Zhang
✉ zzhongge@ucsd.edu

RECEIVED 19 February 2024
ACCEPTED 22 March 2024
PUBLISHED 11 April 2024

CITATION
Kopkowski PW, Zhang Z and Saier MH Jr (2024) The effect of DNA-binding proteins on insertion sequence element transposition upstream of the *bgl* operon in *Escherichia coli*.
Front. Microbiol. 15:1388522.
doi: 10.3389/fmicb.2024.1388522

The effect of DNA-binding proteins on insertion sequence element transposition upstream of the *bgl* operon in *Escherichia coli*

Peter W. Kopkowski, Zhongge Zhang* and Milton H. Saier Jr.*

Department of Molecular Biology, School of Biological Sciences, University of California, San Diego, La Jolla, CA, United States

The *bglGFB* operon in *Escherichia coli* K-12 strain BW25113, encoding the proteins necessary for the uptake and metabolism of β-glucosides, is normally not expressed. Insertion of either IS1 or IS5 upstream of the *bgl* promoter activates expression of the operon only when the cell is starving in the presence of a β-glucoside, drastically increasing transcription and allowing the cell to survive and grow using this carbon source. Details surrounding the exact mechanism and regulation of the IS insertional event remain unclear. In this work, the role of several DNA-binding proteins in how they affect the rate of insertion upstream of *bgl* are examined via mutation assays and protocols measuring transcription. Both Crp and IHF exert a positive effect on insertional Bgl^+ mutations when present, active, and functional in the cell. Our results characterize IHF's effect in conjunction with other mutations, show that IHF's effect on IS insertion into *bgl* also affects other operons, and indicate that it may exert its effect by binding to and altering the DNA conformation of IS1 and IS5 in their native locations, rather than by directly influencing transposase gene expression. In contrast, the cAMP-CRP complex acts directly upon the *bgl* operon by binding upstream of the promoter, presumably altering local DNA into a conformation that enhances IS insertion.

KEYWORDS

IS element, insertional mutation, adaptive mutation, Crp, IHF, DNA-binding protein

1 Introduction

Since their discovery, transposable elements ("jumping genes") have been studied in prokaryotic and eukaryotic models. In both types of organisms, they are best known for their ability to insert into variable locations within an organism's genome (McClintock, 1950). In some cases, transposition takes place upstream of or within a structural gene, which may cause a change in protein expression for the former and loss of gene function for the latter (Whiteway et al., 1998; Zhang and Saier, 2009a,b).

One of the best characterized examples of this phenomenon is IS (Insertion Sequence) element insertion into the *bglGFB* operon, which is not expressed in wild-type (WT) *Escherichia coli* K-12 strain BW25113. Both the pathogenic and non-pathogenic forms of *E. coli* contain this operon, and to our knowledge it is never expressed in WT. Binding of the global histone-like nucleoid structuring (H-NS) protein at two sites on either side of the

otherwise active *bgl* promoter is the most important factor in silencing transcription of the *bgl* operon by a strong repression mechanism (Schnetz, 1995; Dole et al., 2004; Lam et al., 2022). Data published by other groups indicate that the upstream and downstream H-NS binding sites exhibit synergy with each other (Nagarajavel et al., 2007). Building upon this discovery, recent data from our group suggest the formation of a repression loop structure that blocks access of RNA polymerase to the *bgl* promoter (Lam et al., 2022; Tran et al., 2022).

The first gene in the operon, *bglG*, contains the downstream H-NS binding site within its coding region and is itself flanked by two Rho-independent terminators, limiting the amount of RNA transcript made of *bglG* as well as the two other downstream genes, *bglF* and *bglB* (Figure 1; Mahadevan and Wright, 1987; Schnetz and Rak, 1988). BglG is a homodimer that binds to its own transcript to prevent early termination, allowing transcription to continue and thereby promoting expression of the entire operon (Amster-Choder and Wright, 1992). BglG also has other functions including the positive regulation of insertional and non-insertional Bgl^+ mutations, although how it accomplishes these functions has yet to be elucidated (Zhang et al., 2022). The *bglF* gene immediately follows *bglG*'s downstream terminator and encodes a membrane-integrated protein responsible for the uptake and concomitant phosphorylation of β-glucosides via a phosphotransferase (PTS)-dependent mechanism (Fox and Wilson, 1968). Thus, BglF passes a phosphoryl group from HPr or a BglG monomer to the incoming β-glucoside concomitant with transport (Chen et al., 1997), marking it for hydrolysis of the aglycone from the glucose-phosphate moiety by BglB, the operon's third and final gene product (Prasad et al., 1973), and this sugar-P feeds directly into glycolysis. Since phosphorylated BglG is in equilibrium with phosphorylated BglF, transfer of the phosphoryl group from BglG to BglF causes the dephosphorylation of BglG, allowing it to dimerize into its active anti-termination configuration. BglG can be phosphorylated on two specific histidyl residues, one that promotes antitermination and the other which prevents antitermination (Görke and Rak, 1999; Rothe et al., 2012).

While repression of *bgl* operon expression by H-NS is strong, preventing almost 100% of the maximal transcription rate, transposition of an IS element may occur upstream of the *bgl* promoter (Reynolds et al., 1981). The vast majority of IS elements that insert into this area are either IS1 or IS5, and both elements insert in either orientation (forward or backward) within an area spanning ~200 bp (Reynolds et al., 1981). This insertional event eliminates repression of *bgl* operon expression by H-NS (Lopilato and Wright, 1990; Singh et al., 1995), increases *bgl* operon expression several hundred-fold (Lam et al., 2022), and allows cells to grow, divide and form colonies using β-glucosides as their sole carbon source (Prasad and Schaefler, 1974). This state corresponds to a "Bgl^+" phenotype.

Two details about this insertional event remain of great interest. First, Bgl^+ mutations of any class, insertional or otherwise, only occur when *E. coli* cells are starving in the presence of a β-glucoside, which can be freely taken in and used after activation of the *bgl* operon via mutation or IS insertion. Second, the rate of insertion into *bgl* under these circumstances is much higher than the random mutation rate for *E. coli* (Hall, 1998).

Since mutations are widely considered by the scientific community to be random events rather than environmentally directed, the above facts indicate that the *bgl* operon is operating under an additional, fundamentally novel layer of regulation that is not well understood. Several paradigms like *bgl* exist, where insertional events occurring under specific circumstances of cellular stress lead to a phenotype which relieves that stress. These operons deserve further study so that the mechanisms of their regulation may be incorporated into today's accepted models of mutation and evolution.

As noted above, *bgl* is one of several operons in *E. coli* that are preferential sites for insertion of IS elements. These sites are commonly found in the operon's promoter region, have high Gibbs free energy signatures, and exhibit an increase in IS insertion frequency when the bacterial cell is experiencing a specific type of stress related to the operon's function (Humayun et al., 2017). These so-called superhelical stress-induced duplex destabilization (SIDD) sites are under study by our group and others to determine whether insertion into a SIDD site is directed by environmental conditions, as well as to uncover whether SIDDS evolved specifically to allow IS-mediated operon activation (Humayun et al., 2017).

Since the hydrogen bonds between individual base pairs are relatively weak in a SIDD site, the opening of the DNA may be what allows IS insertion to occur there with increased frequency. This would support previous unpublished findings that H-NS lowers the rate of insertion into the *bgl* operon (Z.Z., unpublished data). When H-NS binds at or near the SIDD site, it may prevent it from opening to a conformation that allows for easy insertion of an IS element. It is therefore important to identify what roles, if any, that DNA-binding proteins have upon transposition into the sequences on or near their binding sites. In pursuit of this goal, we studied several DNA-binding proteins to better understand their roles in regulating the insertion of IS elements and the phenomena of "adaptive" or "directed" mutations in general. Adding to the body of knowledge of how transposons and their movements are regulated is vital for proper understanding of the "mobilomes" that are present in almost every life form on Earth.

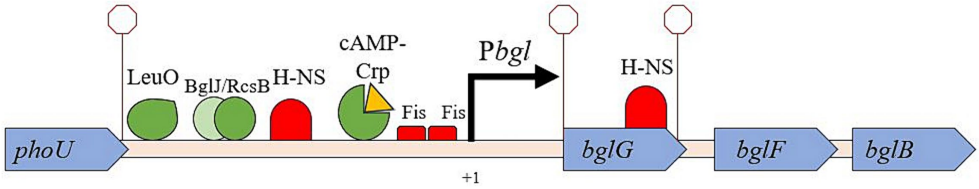

FIGURE 1
The *bgl* operon with relevant DNA-binding protein sites and structural features. The *phoU*/*bglG* intergenic region is where IS insertion occurs, upstream of the promoter. Several DNA binding proteins are known to bind to the *bglGFB* regulatory region and influence transcription. A major focus of this work is to determine whether protein binding in this region affects the rate of insertion of IS1 and/or IS5.

Integration Host Factor (IHF) is a heterodimeric global histone-like DNA-binding protein involved in several cellular functions including transcriptional regulation, DNA recombination, and chromosome compaction (Goosen and van de Putte, 1995; Engelhorn and Geiselmann, 1998). After binding to the minor groove of the double helix, it significantly bends the DNA at least 160° per dimer (Sugimura and Crothers, 2006). Under starvation conditions, IHF concentrations rise 4–8-fold in *E. coli* (Bushman et al., 1985; Ditto et al., 1994) and has been implicated in the induction of proteins related to carbon starvation (Sclavi et al., 2007). Also of interest is IHF's role in the transposition of Mu prophage in *Pseudomonas putida*. *In vitro* studies of IHF binding to the Mu promoter suggest that IHF assists in formation of a transposase complex that may facilitate excision via supercoiling relief (Surette et al., 1989; Allison and Chaconas, 1992). *In vivo*, IHF is not required for Mu transposition, but has a dual effect on Mu transposase transcription. IHF binds to the end of the Mu transposase promoter region and activated transcription of the transposase while simultaneously relieving repression by H-NS at a downstream site (van Ulsen et al., 1996). Multiple groups have shown that a Gly to Glu mutation at the 62nd amino acid in IhfA causes loss of IHF's ability to bind to its DNA consensus sequence but does not prevent dimerization with IhfB (Granston and Nash, 1993; Hales et al., 1994).

Crp is a DNA-binding protein that relies upon cyclic AMP (cAMP) to become activated (Kolb et al., 1993). It is a global transcriptional regulator that affects expression of almost 200 genes in *E. coli* (Fic et al., 2009). Most applicable to this work, cAMP-Crp is generally involved in the positive regulation of genes concerned with the metabolism of carbon sources (Deutscher, 2008; Görke and Stülke, 2008) and is a necessary factor in the activation of *bgl*. Even if repression by H-NS is relieved, *bgl* remains transcriptionally inactive and unable to grow using only β-glucosides, unless activated Crp is present (Mukerji and Mahadevan, 1997; Gulati and Mahadevan, 2000).

In this work, we provide evidence that two DNA-binding proteins affect the rate of IS insertion into *bgl*: the cAMP-Crp complex and Integration Host Factor (IHF). Both proteins are positive regulators of insertion; that is, their presences and abundances maintain or increase the frequencies of IS-related Bgl$^+$ mutations observed in WT cells. The effect of IHF is not specific to *bgl* but applies to several other operons into which IS elements insert in response to environmental stresses. Our results suggest that binding of IHF to known sites in IS1 and IS5 does not influence transposase gene transcription, and is therefore likely to influence DNA conformation, transpososome formation and/or energetics at the site of transposition.

2 Results

2.1 Deletion of *ihfA* decreases the Bgl$^+$ mutation rate

The first phase of our research was to determine which genes, if any, demonstrated a clear effect on the frequency of IS insertion into *bgl* upon deletion. Figures 2A,B show data gathered from several single deletion strains over a 10-day period. The full list of deletion strains tested is tabulated in Supplementary Table S1. Genes were selected for study based on the DNA-binding abilities of their gene products and were loosely separated into two groups: the global histone-like DNA-binding proteins and operon-specific DNA-binding proteins with known binding sites within the *bgl* control region. Colony PCR was performed on the mutant colonies of each strain, using primers that flank the P*bgl* regulatory region into which IS1 and IS5 are known to insert (Supplementary Table S2). Of the deletion strains tested, Δ*ihfA* was selected for further study, since it caused the greatest change in the frequency of Bgl$^+$ mutant appearances compared to WT (Figure 2A). Further confirmatory Bgl$^+$ mutation assays were performed on WT and Δ*ihfA* to precisely ascertain the difference in mutant incidence between them. These results again showed an approximately eight-fold drop in total Bgl$^+$ mutants in Δ*ihfA*, and a more than 11-fold drop in insertional mutants (Figures 2C,D). Deletion of *ihfB*, encoding IHF's other subunit IhfB, showed the same effect as Δ*ihfA* (Figures 2C,D). To determine whether the Δ*ihfA* mutation prevented cell multiplication but not necessarily insertion, IS5 was inserted upstream of *bgl* in its usual position in the Δ*ihfA* mutant. This strain grew as quickly on salicin media as WT Bgl$^+$ cells. This indicates that Δ*ihfA* acts specifically to lower the IS insertion frequency rather than hinder the growth of Bgl$^+$ mutants after insertion takes place. These observations led us to consider the mechanism of IHF's role in IS insertion, as well as its specificity for the *bgl* operon.

2.2 Further characterization of an *ihfA* deletion mutant

The Δ*ihfA* background was transferred to the backgrounds P*tet*-G and Iq-G, which constitutively express *bglG* at the intS locus, leaving the native *bgl* operon intact. Our objective was to observe how IHF's observed effect would interact with increased BglG levels, which has been previously shown to increase the rate of Bgl$^+$ insertional and non-insertional mutations (Zhang et al., 2022). The results presented in Figures 3A,B show that the stimulatory effect of increased *bglG* expression on insertion frequency counteracts Δ*ihfA*'s negative effect on the same. As with the WT background, deletion of *ihfA* in either background led to a greater than 10-fold drop in insertional mutants. In the case of P*tet*-G, the total mutants fell by four-fold when *ihfA* was deleted, and a more than 10-fold decrease was observed among insertional mutants. These data further solidify BglG's effect on insertional mutation rate and confirm our results on *ihfA* deletion, which appears to mitigate both the insertional and non-insertional mutant incidences in the P*tet*-G strain, but not the Iq-G strain.

2.3 Characterization of P*tet*-driven *ihfA/B* expression

Since the IHF protein has a significantly positive effect on IS insertion upstream of the *bgl* promoter, we decided to determine if changing the amounts of both IhfA and IhfB subunits would have a different effect. PK01, a strain expressing both the *ihfA* and *ihfB* genes in the chromosome using the constitutive P*tet* promoter, was therefore constructed and subjected to a mutation assay as previously described. Interestingly, a decrease in both total and insertional mutants was observed after 10 days (Supplementary Figures S1A,B). We consider several possible reasons for why this occurred (see Discussion).

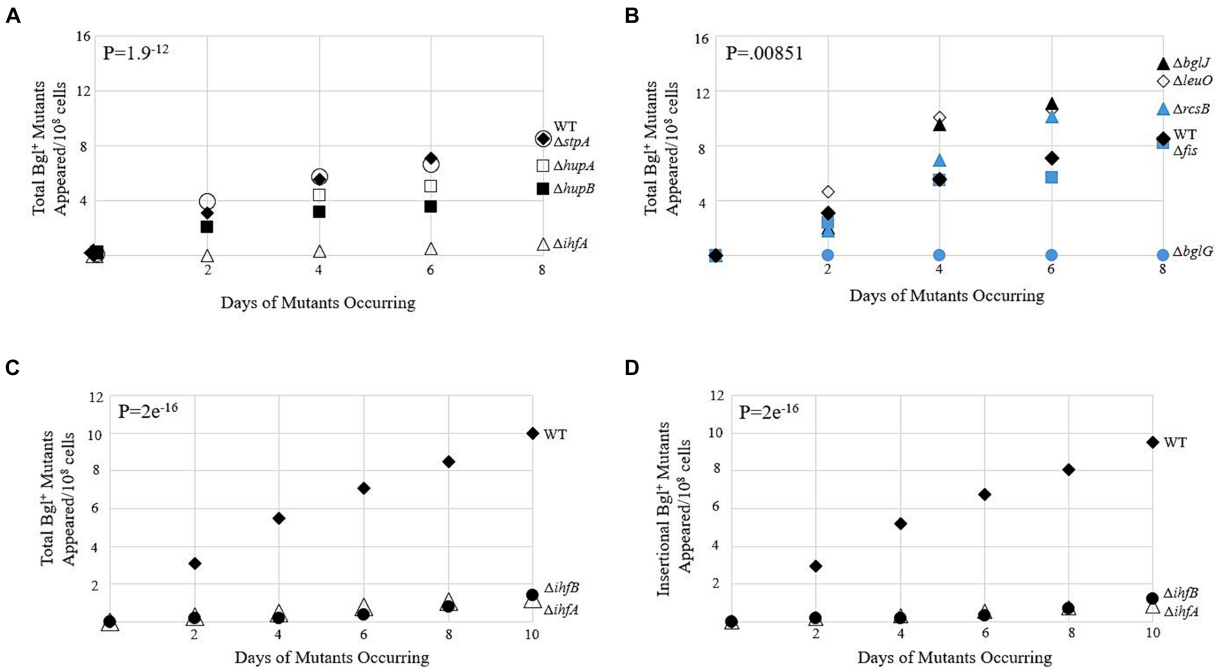

FIGURE 2
Bgl$^+$ mutation frequencies due to deletion of genes encoding DNA-binding proteins. **(A,B)** Appearances of total Bgl$^+$ mutants over time. Mutation assays were carried out on M9 + 0.5% salicin plates as described in Materials and Methods. Strains are loosely grouped in A and B based on their global nature **(A)** or more specific binding to *bgl*. **(B,C)** Effects of deleting *ihfA* and *ihfB* on appearance of total Bgl$^+$ mutants. **(D)** Effects of deleting *ihfA* and *ihfB* on appearance of insertional Bgl$^+$ mutants. Colony PCR using primers flanking the region of *bgl* insertion was performed to differentiate between insertional and non-insertional colonies. WT = ◆; Δ*ihfA* = △; Δ*bglG* = ●; Δ*stpA* = ○; Δ*hupA* = □; Δ*hupB* = ■; Δ*bglJ* = ▲; Δ*leuO* = ◇; Δ*rcsB* = ▲; Δ*fis* = ■; Δ*ihfB* = ●.

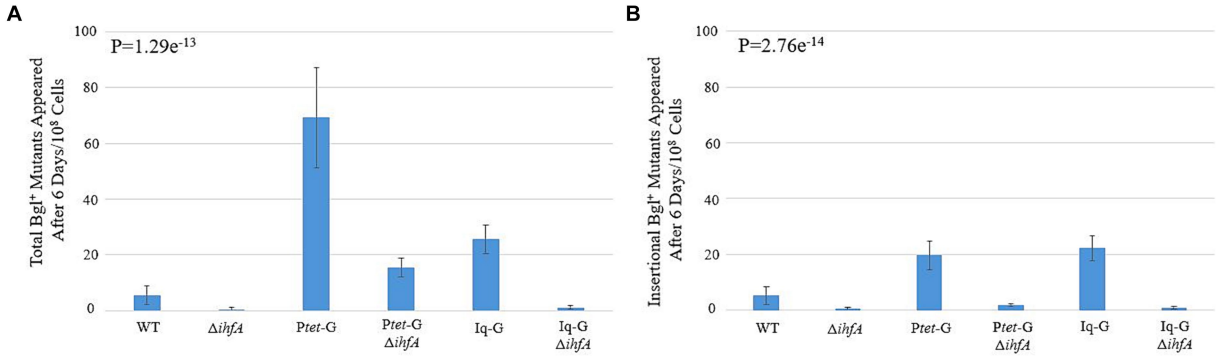

FIGURE 3
Effect of *ihfA* deletion combined with increased *bglG* expression on Bgl$^+$ mutations. **(A)** Effect of deleting *ihfA* on total Bgl$^+$ mutations in the cells over expressing *bglG*. **(B)** Effect of deleting *ihfA* on insertional Bgl$^+$ mutations in the cells over expressing *bglG*. Bgl$^+$ mutation assays were performed as previously described. Two *bglG* overexpression strains, P*tet*-G (P*tet* driving *bglG* at the *intS* locus) and Iq-G (*lacIq* driving *bglG* at the *intS* locus), were used together with or without the *ihfA* deletion. This experiment serves to determine if the *ihfA* deletion has similar inhibitory impact on IS1/IS5 insertions in the cells with higher levels of BglG. Our previous study showed that more BglG produced leads to greater insertional frequencies (Zhang et al., 2022). Colony PCR was used to differentiate insertional and non-insertional Bgl$^+$ mutants.

PK01_R was constructed, containing the *tetR* gene. *tetR* encodes the P*tet* repressor protein TetR, which is most effective in the absence of its inducer, any one of several tetracycline derivatives. Our goal was to see if different levels of P*tet* induction would reveal a similar trend as that already observed with normal PK01. Supplementary Figures S1C,D show the results of a mutation assay as described, but with the addition of several different concentrations of the *Tet* inducer anhydrotetracycline (aTC). As expected, the increase of IhfA and IhfB led to more Bgl$^+$ mutations. The number of mutants that appeared was like the previous assay at levels of maximal aTC induction (~30 μM), supporting the results of Supplementary Figure S1B.

To create a strain expressing both IHF subunits at even higher levels, a pZA31 plasmid containing P*tet*-driven *ihfA* and *ihfB* was constructed and then transformed into wild type BW25113 and its counterpart containing the *Tet* repressor gene *tetR*. These new strains, called PK02 and PK02_R, respectively, were compared to each other and an isogenic control (WT Rf) in a mutation assay. All three strains behaved very similarly (Figures 4A,B), suggesting that the increased

transcription of both IHF subunits did little to influence the insertion rate. To confirm this, PK02_R was plated on M9 salicin media containing differential levels of the Tet promoter's inducer aTC (Figures 4C,D). The amount of inducer present did not appear to cause a corresponding change in total or insertional Bgl$^+$ mutants, as consistent with the findings in Figures 4A,B.

2.4 Changing IHF levels has no significant effect on transcription of *bgl*

Our next goal was to establish a mechanism by which IHF exerts its effect on the expression of *bgl*. If Δ*ihfA*'s effect on the frequency of Bgl$^+$ insertional mutations is due to the inability of IHF binding to the *bgl* upstream region, then LacZ assays could show a difference in transcriptional activity when IHF is present versus when it is absent. The constructs we devised for determining this are shown in Supplementary Figure S2. The native *bgl* operon was left intact to maintain normal induction by β-glucosides. A second, altered *bgl* promoter and its upstream region were positioned in front of the native *lacZ* gene in place of its usual promoter. This new construct, P*bgl-Z*, can measure the promoter activity of this second *bgl* operon. A similar construct, P*bgl-G-Z*, contained the *bglG* gene between P*bgl* and *lacZ* for the purpose of determining operon activity. Figures 5A,B show the *lacZ* activity results, which demonstrated a ~ 2-fold change in promoter activity but no significant change in operon activity between WT and Δ*ihfA*. We repeated these experiments using a BglB assay, which functions similarly in practice to the LacZ assay. However, this assay uses PNP-Glucoside (PNPG) as its substrate and measures activity of the *bgl* operon directly, rather than having to rely on additional constructs and the *lacZ* gene. The results of the BglB assay confirmed those of the LacZ experiments, showing no increase or decrease in *bgl* activity (Figure 5C). We suspected that any potential change may not be detectable due to the natural low levels of *bgl* activity; to investigate this possibility, the ΔP*bgl*-G strains were constructed. ΔP*bgl*-G strains are similar to the strains originally used for BglB but exhibit a 200-fold increase in operonic activity due to the removal of the *bglG* gene, both of its flanking terminator sequences, and the downstream H-NS binding site. To see if ΔP*bgl*-G would lead to observable differences between strains, BglB assays were run again using several previously used strains now in a ΔP*bgl*-G background. No large change was observed across the board, leading us to reject the notion that specific binding of IHF significantly affects transcription of *bgl* (Figure 5D).

2.5 The effects of *ihfA* deletion on expression of several IS-activated operons

Since our LacZ data did not reveal a substantial change in transcriptional activity within the context of *bgl*, we hypothesized that Δ*ihfA* does not affect *bgl* directly but may exert its effect on insertion upstream of *bgl* by a nonspecific means. If IHF, being nonspecific, is important for IS insertion in general, then it may be involved in the upstream process of IS excision from other locations in the genome. In this situation, other operons into which IS1 and IS5 insert would also experience a change in insertion frequency upon deletion of *ihfA*. To explore this possibility, a Δ*ihfA* strain was tested alongside an isogenic WT strain in mutation assays corresponding to several other operons where preferential IS insertion into their respective SIDD sites have been previously documented (McCalla et al., 1978; Chen et al., 1989; Whiteway et al., 1998; Zhang et al., 2010, 2017; Humayun et al., 2017).

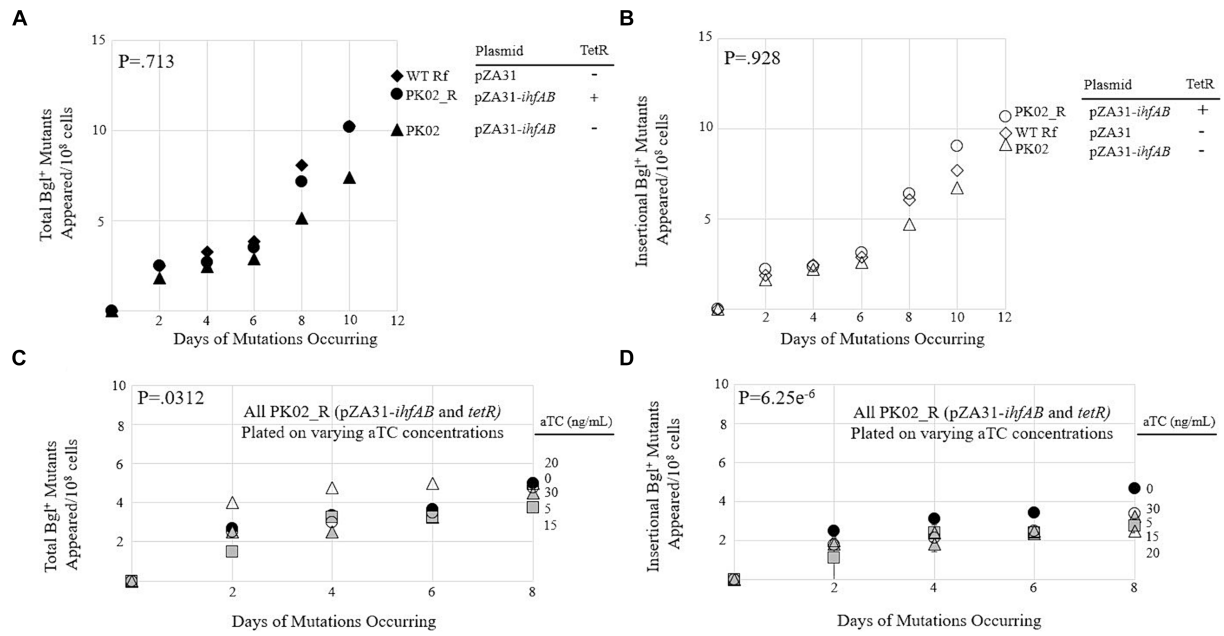

FIGURE 4
Overexpression of *ihfA* and *ihfB* on a plasmid slightly lowers the IS insertional rate into *bgl*. **(A)** Effect of IhfA and IhfB overproduction on total Bgl$^+$ mutations. **(B)** Effect of IhfA and IhfB overproduction on insertional Bgl$^+$ mutations. **(C)** Effect of titrating *ihfA* and *ihfB* expressions on a plasmid on total Bgl$^+$ mutations. **(D)** Effect of titrating *ihfA* and *ihfB* expressions on a plasmid on insertional Bgl$^+$ mutations. For **(C,D)**, mutation assays were performed as previously described on M9 salicin plates containing aTC at 0 to 30 ng/mL over a 10-day period. 0aTC = ●; 5aTC = ▲; 15aTC = ■; 20aTC = Δ; 30aTC = ○.

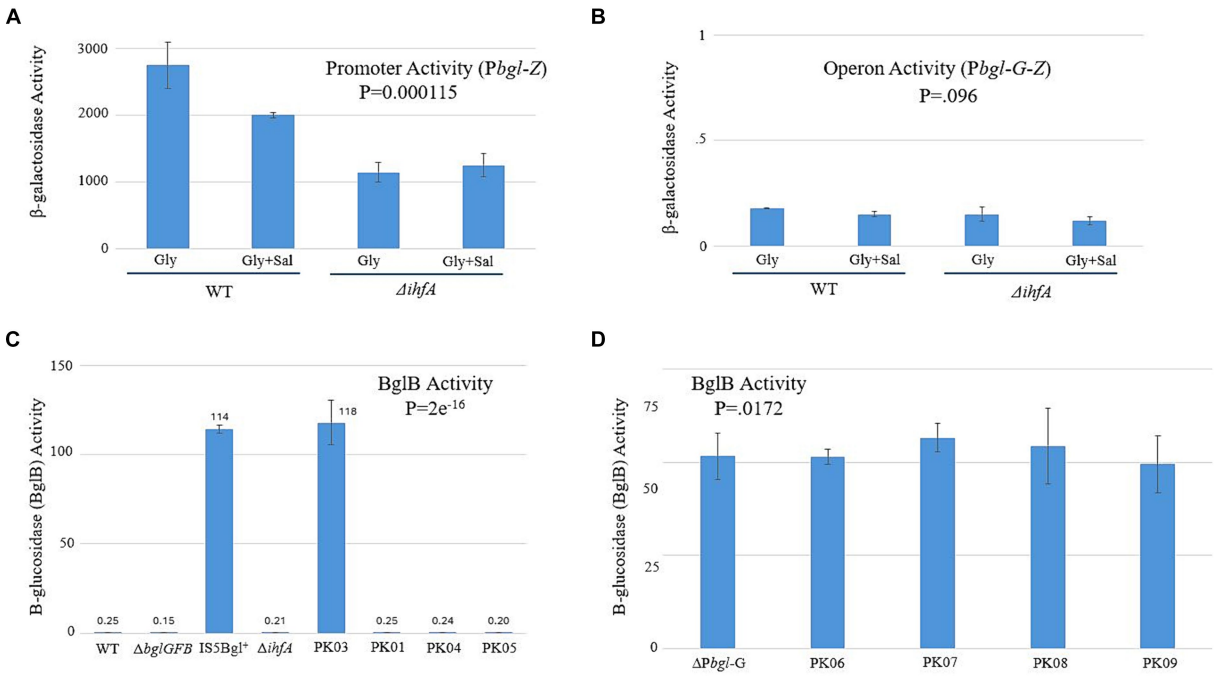

FIGURE 5
Changing IHF levels does not have a significant effect on *bgl* operon transcriptional activity. LacZ and BglB assays were carried out as previously described (Materials and Methods). **(A)** The *bglGFB* promoter (P*bgl*) activities determined using the LacZ assay. **(B)** The *bglGFB* operon (P*bgl*-*bglG*) activities determined using LacZ assay. **(C)** The *bglGFB* operon activities were determined via the BglB assay. PK01 = P*tet*-*ihfA*_P*tet*-*ihfB*; PK03 = Δ*ihfA* IS+; PK04 = Δ*ihfA* in G50; PK05 = P*tet*-*ihfA*_P*tet*-*ihfB* in G50. **(D)** The *bglGFB* operon activities in strain ΔP*bgl*-G determined using the BglB assay. Strain ΔP*bgl*-G was deleted for the *bglG* gene plus its flanking terminators and is Bgl$^+$. PK06 = Δ*ihfA* in ΔP*bgl*-G; PK07 = Δ*fis* in ΔP*bgl*-G; PK08 = Δ*ihfA* and Δ*fis* in ΔP*bgl*-G; PK09 = P*tet*-*ihfA*_P*tet*-*ihfB* in ΔP*bgl*-G.

Agar plates were made with the intention of promoting insertion into the *glpFK*, *fucAO*, and *flhDC* operons, and were inoculated with a number of cells specific to the paradigm tested. The results for these three operons are presented in Figures 6A–C, and all three showed a significant decrease in insertional frequency, just as observed for *bgl*.

As when testing *bgl*, these experiments compared Δ*ihfA* to the WT strain. However, the parameters of each experiment differed depending on the specific operon being tested. Interestingly, all the operons examined showed a significant decrease in the insertional mutant appearance rate of the Δ*ihfA* counterpart, suggesting that the dependency of this process on IHF affects several operons into which insertions occur, and consequently, it may affect an early step (IS excision or transposition) of the IS1 or IS5 element itself.

2.6 Examining the mechanism of the non-specific effect of IHF on IS1/IS5 transposition via loss of *ihfA*'s DNA-binding function and LacZ measurements of IS1/IS5 promoters

Since our view of IHF's effect proved to be relevant to other operons, the question remained: What is the mechanism of the IHF effect on transposition? First, we tested whether the DNA-binding ability of IHF was the cause of the observed multi-operon effect. A mutation assay was performed comparing WT and Δ*ihfA* to two other strains. These strains have P*tet*-driven *ihfA* at the *intS* locus, but here the *ihfA* gene product contains a G-E substitution at the 62nd position (referred to as G62E). The product of this *ihfA* mutant loses its DNA-binding function but is still able to dimerize with IhfB (Granston and Nash, 1993; Hales et al., 1994). The *ihfA*_G62E construct was placed into WT as well as the deletion strain for *ihfA* at its native position (Δ*ihfA*), yielding PK10 and PK11. If binding to DNA was important for IHF's effect, then PK11 would show similar insertion rates as Δ*ihfA*, while PK10 would have binding and nonbinding IhfA and therefore would not experience as drastic of a change. Figures 7A,B show the results of a *bgl* mutation assay, conducted as previously described. The strains acted according to expectations, with PK10 having fewer mutants than WT, but more than its Δ*ihfA* counterpart PK11.

To confirm these results, the G62E constructs were also used in mutation assays for the *glpFK* and *fucAO* operons as previously described; the results of these are in Figures 7C–F. The G62E strains act similarly with respect to other operons as with *bgl*; in all operons tested, the G62E mutant generated around the same number of mutations as the WT if the native *ihfA* gene was intact. However, the G62E mutation in the Δ*ihfA* deletion mutant still caused a drastic drop to the same level of Δ*ihfA* or lower.

2.7 IHF has no observable effect on transcription of the transposase gene encoded within IS1 or IS5

Research conducted by other laboratory groups had demonstrated that IHF binding sites are present at both ends of IS1 and on one end of IS5 (hereafter referred to as IS5C; IS5A is the end without an IHF binding site) (Gamas et al., 1987; Prentki et al., 1987; Muramatsu et al., 1988).

10.3389/fmicb.2024.1388522

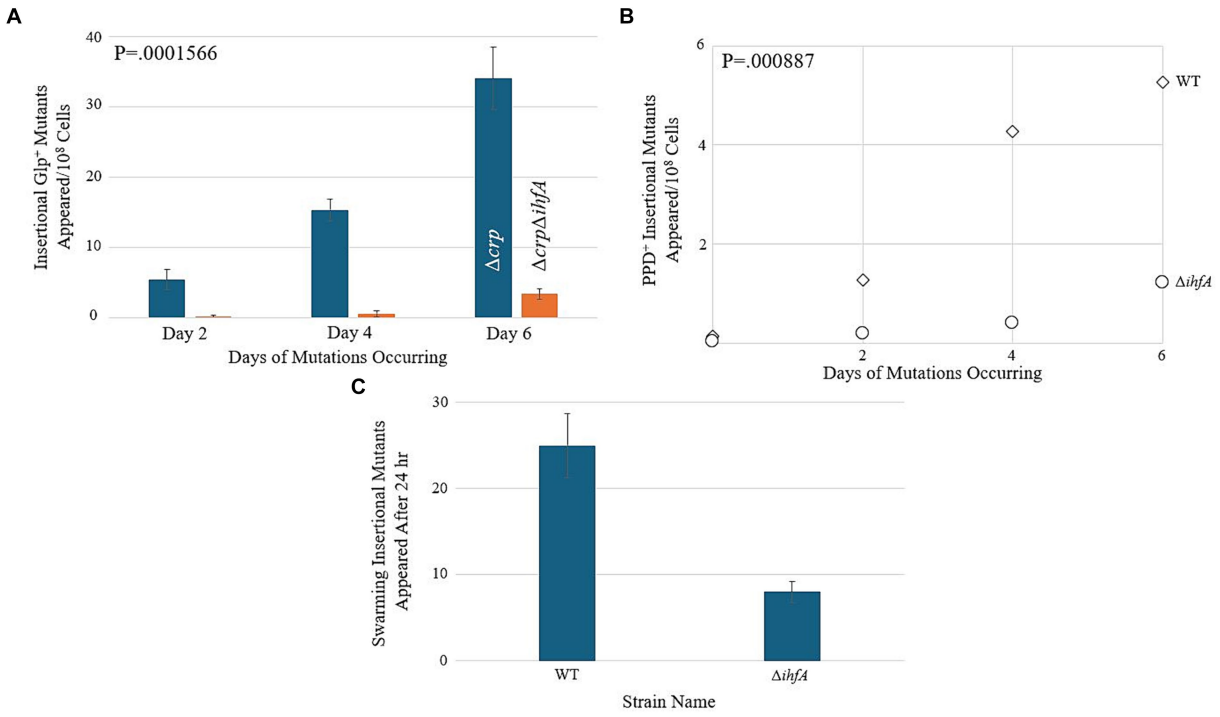

FIGURE 6
Deletion of *ihfA* affects IS element insertion into other operons. **(A)** Effect of *ihfA* deletion on Glp$^+$ mutations. A Δ*crp* mutant and a Δ*crp* Δ*ihfA* double mutant were subjected to the Glp$^+$ mutation assay on M9 + 0.5% glycerol media as previously described. **(B)** Effect of *ihfA* deletion on PPD$^+$ mutations. WT and Δ*ihfA* strains were subjected to the PPD$^+$ mutation assay on M9 + 1% propanediol media. **(C)** Effect of *ihfA* deletion on swarming mutations. WT and Δ*ihfA* were subjected to the swarming mutation assay on 0.3% LB agar media.

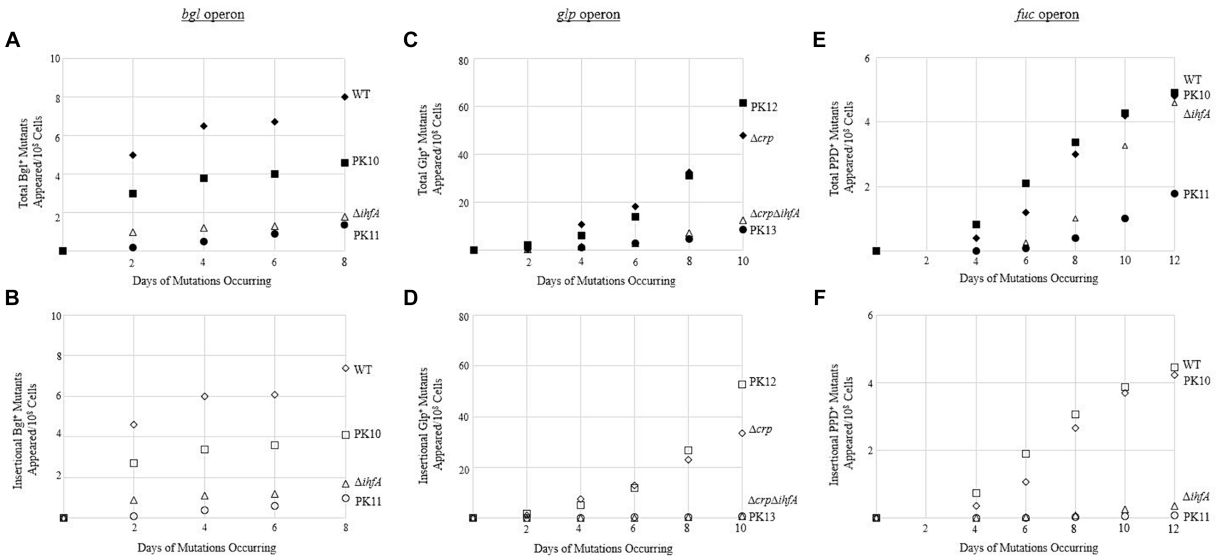

FIGURE 7
Binding of IHF is necessary to effectively exert its positive effect on IS insertion upstream of the *bgl* operon. **(A,B)** Negative effect of expressing *ihfA*_G62E on total Bgl$^+$ mutations **(A)** or on insertional Bgl$^+$ mutations **(B)**. **(C,D)** Negative effect of expressing *ihfA*_G62E on total Glp$^+$ mutations **(C)** or on insertional Glp$^+$ mutations **(D)**. **(E,F)** Negative effect of expressing *ihfA*_G62E on total PPD$^+$ mutations **(E)** or on insertional PPD$^+$ mutations **(F)**. IhfA with the G62E mutation loses the DNA binding property but is still able to dimerize with IhfB. The dimers consisting of IhfA_G62E and IhfB are incapable of DNA binding. WT and Δ*crp* = ◆ total and ◇ insertional; PK10 and PK12 = ■ total and □ insertional; Δ*ihfA* and Δ*crp*Δ*ihfA* = ▲ total and Δ insertional; PK11 and PK13 = ● total and ○ insertional.

We therefore decided to test whether the presence and/or binding of IHF causes a change in the level of transcription of either or both IS elements. This was accomplished by placing the promoter regions for IS1 and both ends of IS5 (containing the Inverted Repeats to which IHF binds) directly before the native *lacZ* gene (Figure 8A). This construct was placed in the WT, Δ*ihfA*, P*tet*-driven *ihfA* at the *intS* locus, and P*tet*-driven *ihfA* G62E

10.3389/fmicb.2024.1388522

at the *intS* locus (containing the G-E substitution at the 62nd position that abolishes DNA-binding ability of IHF). The results for each reporter are presented in Figures 8B–D. Overall, no significant change was observed, regardless of the genetic background. Since IHF is known to bind to either end of IS1 and to IS5C, these results strongly suggest that IHF exerts its effect via direct binding to IS1 and IS5 without altering the expression level of either IS element.

2.8 cAMP-Crp is a positive regulator of IS insertion when present upstream of P*bgl*

To test whether Crp (when activated by cAMP) affects the rate of insertional mutations in the upstream region, two types of mutation assays were conducted. Simply testing a Δ*crp* mutant is not an option in this case, as Crp is required for expression of the *bgl* operon. If insertion takes place in a *crp* deletion mutant, we would be unable to observe it since the colony would not grow. Therefore, we first decided to perform mutation assays using WT on M9 salicin plates with minimal cAMP and compare mutation rates to those growing on plates with an excess of cAMP. The results of this experiment are shown in Figure 9A. They demonstrate no change in insertional mutants as the amount of extracellular cAMP increases.

Next, we compared these WT results to those of a *cyaA* deletion mutant. *cyaA* encodes adenylate cyclase, and without it, the cells are unable to produce cAMP and must rely on extracellular cAMP provided in the growth media. In Figure 9A, we observe an increase in insertional mutants as the level of extracellular cAMP increases. Interestingly, the number of total colonies was like those observed for the WT strain, suggesting that the lack of *cyaA* has a significant effect on IS insertion specifically. The next set of experiments used *cpdA* to control the levels of intracellular cAMP. *cpdA* encodes a phosphodiesterase which degrades cAMP to 5'-AMP, and in its absence, cAMP levels increase. Similarly, if *cpdA* is expressed at higher levels, the amount of cAMP, and therefore the amount of active Crp, should decrease. A *cpdA* deletion mutant was constructed and was given an "empty" pZA31 plasmid for isogeneity, producing Δ*cpdA* Rf. A pZA31 plasmid containing P*tet-cpdA* was electroporated into WT cells, yielding PK20. All three strains were subjected to the *bgl* mutation assay as previously described. The results are presented in Figures 9B,C. The Δ*cpdA* mutant showed a more than two-fold increase in insertional mutants compared to the WT strain, and the *cpdA* overexpression strain PK20 showed a similar number of total colonies as WT, but with only half as many insertional mutants. Together, these data suggest that Crp, when activated by cAMP, is a positive regulator of insertion upstream of the *bgl* operon. In the absence of cAMP via deletion of *cyaA* or via increased degradation of cAMP, the insertion rate decreased 2–3-fold.

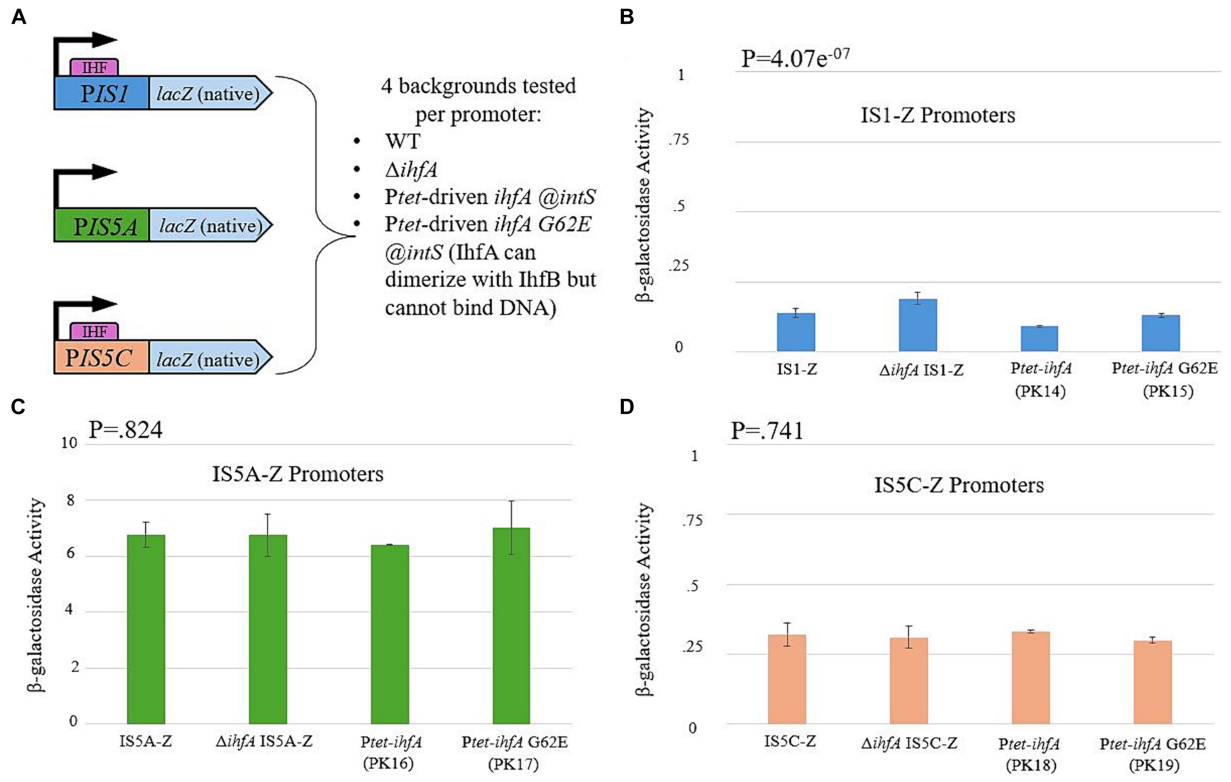

FIGURE 8
IHF does not influence transcriptional activity of IS1 or IS5. **(A)** A graph showing the construction of each LacZ reporter, which uses the promoter region of the 5' end of IS1 and both ends of IS5 (PIS5A and PIS5C). **(B)** The IS1 transposase promoter (PIS1) activity in various genetic backgrounds. **(C)** The *ins5A* transposase promoter (PIS5A) activity in various genetic backgrounds. **(D)** The *ins5BC* operon promoter (PIS5C) activity in various genetic backgrounds.

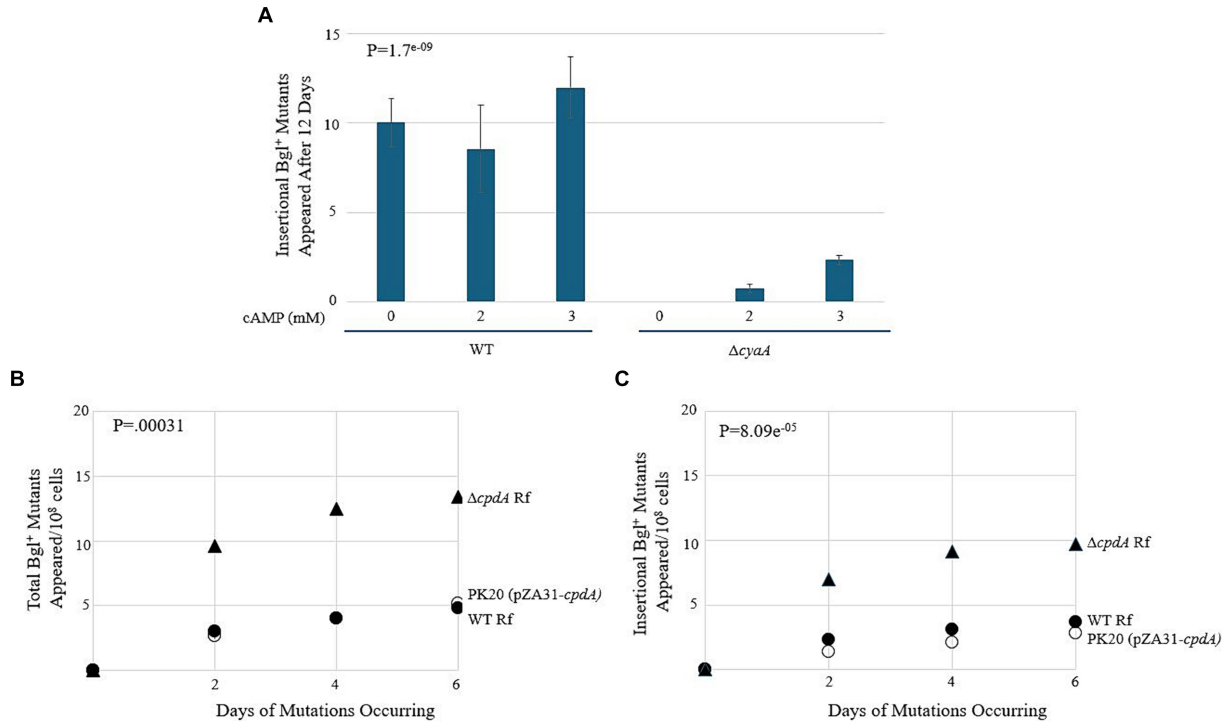

FIGURE 9
cAMP-Crp is a positive regulator of IS insertion into *bgl*. **(A)** Effect of adding external cAMP on insertional Bgl⁺ mutations in wild type cells and Δ*cyaA* cells. cAMP was added to M9 + Salicin agar from 0 mM to 3 mM. **(B,C)** Effects of *cpdA* deletion and overexpression on total Bgl⁺ mutations **(B)** and insertional Bgl⁺ mutations **(C)**. *cpdA* encodes a phosphodiesterase which degrades cAMP to 5′-AMP, and in its absence, cAMP levels increase. For overexpression, the strong constitutive *tet* promoter, P*tet*, was used to drive *cpdA* on the plasmid pZA31. WT Rf = ●; Δ*cpdA* Rf = ▲; PK20 = ○.

2.9 *ihfA* deletion or *ihfA*/*B* constitutive expression with differential cAMP-Crp levels

To observe how loss of *ihfA* affects mutation frequencies along with differing cAMP-Crp complex levels, the Δ*cpdA* Rf and PK20 backgrounds were transferred to the Δ*ihfA* strain. The results of this set of mutation essays are shown in Figures 10A,B. In the absence of IHF, increased Crp-cAMP by deleting *cpdA* led to a more than 2-fold elevation of Bgl⁺ mutations. Similarly, decreased Crp-cAMP by overexpressing *cpdA* led to a 3-fold reduction of Bgl⁺ mutations in the absence of IHF. Combining both IHF and Crp, these two proteins affect Bgl⁺ mutations by up to 33-fold (Figures 9B, 10A).

3 Discussion

3.1 The effect of PK01, causing a decrease in the IS insertion rate

PK01, which replaced the native promoters of *ihfA* and *ihfB* with the constitutive P*tet* promoter, had the surprising effect of lowering the insertion rate into *bgl* by nearly two-fold compared to WT. Additionally, the titratable strain PK02_R, which grows similarly to WT, grew only around half as many colonies when saturated with its inducer aTC. We here provide a few possible explanations for these phenomena. The first and most likely possibility is that the native promoters for *ihfA* and *ihfB* are already strong. Since IHF is a global protein with several functions in the cell and with genes in locations adjacent to other important genes, promoters for both subunits may be very strong (Pozdeev et al., 2022). It may be that the promoters driving these subunits are already operating at a higher rate of expression than P*tet* itself. Therefore, in an effort to measure the mutational response of a cell by increasing expression, it is possible that replacement by P*tet*, while strong, could actually have expressed *ihfA* and *ihfB* at a lower rate than the native promoter. Our data on the titratable strains PK01_R and PK02_R are in support of this hypothesis, because even at maximal levels of inductions, the mutation rate remains lower than WT.

Another possibility is that this strain is in fact expressing IHF at levels higher than WT, but that this has an overall negative effect on mutation frequency. For example, these increased levels of expression may be toxic to the cell, or the increased binding of IHF to the multitude of its recognized binding sites may have pleiotropic effects that lead to lower growth rates, and consequently less insertion.

A final possibility is related to the fact that IHF levels increase greatly when the cell is starving (Bushman et al., 1985; Ditto et al., 1994). It may be that this increase is regulated by transcription factors affecting the native *ihfA*/*ihfB* promoters under starvation conditions such as during the mutation assay. If we change the promoters to P*tet*, we may well have inadvertently removed the cell's normal ability to respond to starvation by upregulating *ihfA* and *ihfB* expression, which would also bring IHF levels to a lower-than-expected value.

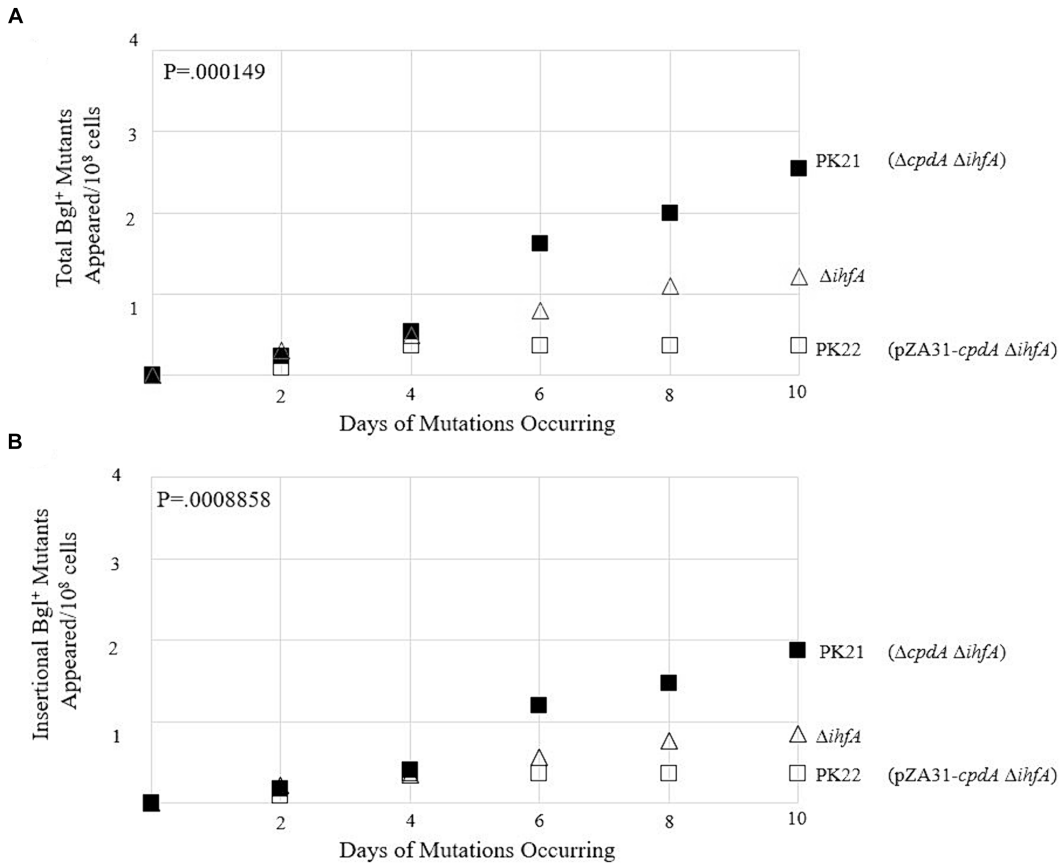

FIGURE 10
Effects of varying levels of cAMP on Bgl$^+$ mutations in Δ*ihfA* cells. **(A)** Effects of *cpdA* deletion and overexpression on total Bgl$^+$ mutations in Δ*ihfA* cells. **(B)** Effects of *cpdA* deletion and overexpression on insertional Bgl$^+$ mutations in Δ*ihfA* cells. To elevate cellular cAMP levels, *cpdA* was deleted. To decrease cellular cAMP levels, *cpdA* was driven by P*tet* on plasmid pZA31.

3.2 Effect of *cyaA* deletion on the IS1/5 insertion rate

cAMP-activated Crp, which is known to bind in *bgl*'s promoter region and positively regulate transcription, was also shown to positively regulate the rate of IS element transposition upstream of *bgl*. Results of a mutation assay comparing WT with an adenylate cyclase deletion mutant (Δ*cyaA*) showed a decrease in the incidence of insertional colonies, although the overall Bgl$^+$ colony counts remained comparable. Similarly, in backgrounds overexpressing or deleting the cAMP-degrading phosphodiesterase gene, *cpdA*, the rate of insertion decreased or increased, respectively. Thus, these results reveal a new role for Crp within the context of *bgl* as a regulator of transposition as well as transcription. Interestingly, previous data from our group have suggested that in the *glpFK* operon, cAMP-Crp has a negative effect on upstream IS insertion. It appears that while Crp maintains its status as a positive transcriptional regulator in the vast majority of operons which it influences, its role with respect to mutational changes in the *bgl* operon may be responsive to the sequences and/or DNA structures surrounding its binding site.

Crp has a known binding site in *bgl*'s promoter region, and this site is close to or overlapping the projected SIDD site into which IS elements insert (Humayun et al., 2017). This novel effect of cAMP-activated Crp presented here likely results from the increased energetic favorability for insertion to take place due to its binding. However, advances in computational programs that analyze energetics would be helpful to confirm this, as the current programs do not account for the binding of proteins to the DNA.

3.3 Strength of the IS1 promoter region

The IS1-*lacZ* reporter activity in section 2.7 proved to be low, even when compared to the two orientations of the IS5 reporter. After a thorough examination of the sequence of the construct, no mutation was found, leading us to consider other possibilities as to why IS1 exerted such a low operon transcriptional activity compared to IS5. We submit that either the growth conditions were not sufficient to increase IS1's promoter to higher levels, or the region upstream of the IS1-*lacZ* gene construct lacks a site for a postulated positive transcription factor for the IS1 promoter. If so, it must be present in the other native locations in the genome where IS1 exists. Both considerations may independently or together explain why IS1's reporter showed such low activity.

3.4 Effect of IHF via DNA conformation of the IS element to allow for efficient transposase binding

As stated in the Results section, both ends of IS1 and at least one end of IS5 contain IHF binding sites (Gamas et al., 1987; Prentki et al., 1987; Muramatsu et al., 1988). However, LacZ assays measuring the activities of the IS1 and IS5 promoters showed no clear change when the level of IHF binding was altered. Despite this, our other data indicate a non-specific effect of IHF on transposition of IS1 and IS5. Thus, the effect exerted by IHF on these IS elements must be of a nature other than transcriptional. Both IS1 and IS5 have DDE catalytic regions (Ohta et al., 2002; Curcio and Derbyshire, 2003), and may therefore act similarly for the purposes of replicative excision from the *E. coli* genome. Our IHF mutation assay data supports this; if IHF only affected one element's transposition and not the other, then only a 2-fold change at best would be expected, and all the insertional colonies would contain the same element, instead of the roughly equal levels of IS1 and IS5 insertion that has been classically observed (Schnetz and Rak, 1988; P.K., unpublished data). Work on other transposons with DDE activity has suggested that IHF plays an important, but not essential, role in the efficient transposition of Tn10 and the Mu bacteriophage (Kobryn et al., 2002; Swinger and Rice, 2004; Gueguen et al., 2005). The role of IHF involves assisting in the conformation of the transposon via binding so that the transposase protein can effectively bind to the transposon ends and then dimerize (or tetramerize in the case of Mu). It has been shown that when IHF is not present, DNA supercoiling has a compensatory effect, allowing for some transposition to occur even when an important conformational factor is not present (Chalmers et al., 1998). Interestingly, Tn10 *in vitro* experiments mimicking *in vivo* supercoiling levels in the absence of IHF observed a ~ 10-fold decrease in transposition frequency (called the "basal level" of 27 excision), which is a similar decrease to what we observed in the *ihfA* deletion mutant (Δ*ihfA*) for the Bgl$^+$ mutation assay (Chalmers et al., 1998; Figure 2A). We utilized *in vivo* models to support past findings in relation to other transposable elements and draw further similarities between IS elements and other transposons in the DDE transposase family.

4 Conclusion

In this work, we demonstrate that IHF and Crp are positive regulators of IS1 and IS5 insertion. Without both functional IHF subunits and the ability to bind DNA, a strong decrease in total Bgl$^+$ mutations is observed, and an even larger decrease in insertional Bgl$^+$ mutations. No matter the amount of antiterminator protein BglG, this strong decrease is still observed. This effect extends to other operons where IS1 and IS5 insertion takes place. We show that the effect of IHF is not transcriptional in nature - that is, binding of IHF to the IS element causes no significant change in transposase gene expression. Therefore, we suggest that IHF's role may instead be related to assisting DNA conformation or transpososome formation at the site of the IS element. We show that increasing the amount of extracellular cAMP increases insertional Bgl$^+$ mutations. Additionally, by increasing intracellular cAMP levels via *cpdA* deletion, a more than 2-fold increase in insertional Bgl$^+$ mutations is observed. IHF and Crp's effects are independent of each other, and combining their effects causes an up to 33-fold change in insertional mutations.

5 Materials and methods

5.1 Construction of strains used

5.1.1 Construction of deletion mutants

CGSC strains JW3964-1, JW0430-3, JW1702-1, and JW3000-1 (*E. coli* Genetic Stock Center, Yale Univ.) carry the deletion mutations of *hupA*, *hupB*, *ihfA*, and *cpdA*, respectively. For each of these mutants, a kanamycin resistance (*km^r*) gene was substituted for the target gene. These mutations were individually transferred to strain BW25113 (wild type/WT) (Datsenko and Wanner, 2000) by P1 transduction, and the *km^r* gene was subsequently flipped out by pCP20 (Datsenko and Wanner, 2000), yielding deletion mutant strains Δ*hupA*, Δ*hupB*, Δ*ihfA*, and Δ*cpdA*, respectively (Supplementary Table S1). Using P1 transduction, the *ihfA* mutation was transferred into the *crp* deletion strain Δ*crp* (Zhang and Saier, 2009b), yielding the Δ*crp*Δ*ihfA* double mutant. The same *ihfA* mutation was transferred into strain Δ*cpdA*, yielding the Δ*cpdA*Δ*ihfA* double mutant.

5.1.2 Construction of *cpdA* overexpression plasmid

The *cpdA* gene was amplified from BW25113 chromosomal DNA using oligos cpdA-Kpn-F and cpdA-Bam-R (Supplementary Table S2). The PCR products were gel purified, digested with KpnI/BamHI and then ligated into the same sites of pZA31P*tet* (Lutz and Bujard, 1997) yielding pZA31-*cpdA*. pZA31-*cpdA* was transformed into strains BW23113 and Δ*ihfA*, yielding PK20 and PK22, respectively (Supplementary Table S1). Plasmid pZA31 P*tet*-Rf (Levine et al., 2007) carries a random fragment (RF) and was used as a negative control (WT Rf).

5.1.3 Construction of P*tet* driving *ihfA*, *ihfB*, and *ihfA*G62E on the chromosome

Using plasmid pKDT:P*tet* (Klumpp et al., 2009) as a template, the cassette "*km^r*:*rrnB*T:P*tet*," containing the *km^r* gene, the *rrnB* terminator (*rrnB*T) and the P*tet* promoter, was amplified using the primer pair Ptet-ihfA-P1 and Ptet-ihfA-P2 (Supplementary Table S2). Using lambda-red system (Datsenko and Wanner, 2000), the PCR products were integrated into the chromosome of BW25113 to replace the "CCT" nucleotides immediately upstream of the *ihfA* translational start point. Chromosomal integration was confirmed, first by colony PCR, and subsequently by DNA sequencing. This yielded strain P*tet-ihfA*, in which the strong *tet* promoter drives expression of the *ihfA* gene. Similarly, the "*km^r*:*rrnB*T:P*tet*" cassette amplified by Ptet-ihfB-P1 and Ptet-ihfB-P2 (Supplementary Table S2) from pKDT:P*tet* was substituted for the *ihfB* upstream promoter region (−46 to −1 relative to the *ihfB* translational start point), yielding strain P*tet-ihfB*.

To make an *ihfA*/*ihfB* double overexpression strain, the *km^r* gene was first flipped out from strain P*tet-ihfB* by pCP20. The cassette of "*km^r*:*rrnB*T:P*tet*" driving *ihfA* gene from strain P*tet-ihfA* was then transferred to Km-sensitive P*tet-ihfB* by P1- transduction, yielding strain PK01 (Supplementary Table S1), in which both *ihfA* and *ihfB* are simultaneously driven by P*tet*. To titrate expression of *ihfA* and

ihfB, the transcription unit including a constitutively expressed *tetR* gene and a spectinomycin resistance (*sp*^{r}) marker was transferred to PK03 by P1-transduction as mentioned above, yielding strain PK01_R.

To further increase *ihfA* expression, a second copy of P*tet* driving *ihfA* was inserted to another chromosomal location. To do so, the *km*^{r}:*rrnB*T:P*tet-ihfA* expression cassette was amplified from the genomic DNA of strain P*tet-ihfA* using primers intS1-P1 and ihfA2-P2. The products were integrated into the *intS* site to replace the region between −229 and + 1,101 (relative to the *intS* translational initiation site). The chromosomal integration was confirmed by colony PCR and subsequently by DNA sequencing, yielding strain P*tet-ihfA @intS*.

The Glycine residue at position 62 is required for IhfA to bind to the DNA (Granston and Nash, 1993; Hales et al., 1994). To reduce or abolish IhfA's DNA-binding ability, this residue was changed to a glutamate residue using fusion PCR. The first part (5′ region) of *ihfA* was amplified from BW25113 gnomic DNA using ihfA-Kpn-F and ihfA.G62E-R (carrying the G62E alteration). The second part (3′ region) was amplified using ihfA.G62E-F (carrying the G62E alteration and overlapping ihfA.G62E-R) and ihfA-Bam-R. Both products were gel purified and fused together using primers ihfA-Kpn-F and ihfA-Bam-R. The fused products (that is, ihfA.G62E) were gel purified, digested with KpnI and BamHI, and then ligated into the same sites of pKDT_P*tet*, yielding pKDT_P*tet-ihfA*.G62E. The cassette "*km*^{r}:*rrnB*T:P*tet-ihfA*.G62E" was amplified using intS-P1 and ihfA-P2, gel purified and then integrated into the *intS* site as for P*tet-ihfA @intS*. The chromosomal integration was confirmed by colony PCR and subsequently by DNA sequencing, yielding strain PK10, which still maintains the native *ihfA* but constitutively expresses the modified *ihfA*.G62E at the *intS* locus. This cassette was transferred to an *ihfA* deletion background by P1 transduction, yielding PK11.

To over express *ihfA* and *ihfB* simultaneously, both structural genes were first PCR amplified individually from *E. coli* genomic DNA. After being gel purified, these two DNA fragments were fused together by PCR. There is a 20 bp intergenic region with nucleotide sequences "tctgattAGAGGAaacagct" between these two genes. The capitalized nucleotides refer to the RBS site for *ihfB*, which is the same as one for *ihfA* located within P*tet*. The fusion *ihfA*/*ihfB* products were digested with KpnI and BamHI and then ligated into pZA31P*tet* digested with the same enzymes, yielding pZA31-*ihfAB*, in which *ihfA* and *ihfB* are driven by P*tet* and both genes have the same RBS site.

5.2 Construction of the IS1 and IS5 promoter *lacZ* reporters

The promoter region (−55 to +30 relative to the *insA* translation initiation site) driving the IS1 transposase gene *insAB1* was amplified using IS1p-Xho-F and IS1pBam-R from BW25113 genomic DNA. This region contains all of the upstream region (including the left-hand IR) and the first 10 residues of *insA* (plus a stop codon TAA). The amplified products were digested with XhoI and BamHI and cloned into the same sites of the default integration vector, pKDT (Klumpp et al., 2009), yielding pKDT-PIS1. The region carrying the *km*^{r}, *rrnB*T and PIS1 (*km*^{r}:*rrnB*T:PIS1) was PCR amplified using oligos IS1- Z-P1 and IS1-Z-P2 (Supplementary Table S2) and then integrated into the chromosomal default strain EQ42 (Klumpp et al., 2009) to replace the *lacI* gene and the *lacZ* promoter. The resultant strain carries the *km*^{r}:*rrnB*T:PIS1 cassette followed by *lacZ*'s ribosomal binding site (RBS) and the *lacZ* structural gene within the *lac* locus. After being confirmed by PCR and sequencing, the promoter reporter, PIS1, driving *lacZ* expression (that is, PIS1-*lacZ*) was transferred into BW25113 and various genetic backgrounds by P1 transduction. This yielded the IS1 promoter reporter strains IS1-Z, Δ*ihfA*_IS1-Z, PK14, and PK15, respectively.

IS5 carries three open reading frames (*ins5A*, *ins5B*, and *ins5C*). *ins5A* encodes the main transposase, and it is transcribed from its own promoter located close to the left-hand IR while the divergent *ins5B* and *ins5C* genes may form an operon that is driven by another promoter located close to the right-hand IR (Sawers, 2005). The *ins5A* promoter region (−68 to +30 relative to the *ins5A* translational initiation site) and the *ins5CB* promoter region (−207 to +30 relative to the *ins5C* translational initiation site) were individually cloned into pKDT, yielding pKDT_PIS5A and pKDT_PIS5C, respectively. As for PIS1 described above, PIS5A and PIS5CB each carries the first 10 amino acids from the target gene followed by a stop codon TAA. These promoters were integrated into the *lac* site on BW25113 chromosome as for PIS1-Z, yielding strains IS5A-Z and IS5C-Z, respectively. Using P1 transductions, these new IS5 promoter *lacZ* reporters were transferred to other genetic backgrounds by P1 transduction. This yielded the IS5 transposase promoter (PIS5A) reporter strains Δ*ihfA*_IS5A-Z, PK16, and PK17, as well as the IS5C promoter reporter strains Δ*ihfA*_IS5C-Z, PK18, and PK19.

5.3 β-galactosidase (LacZ) assays

Escherichia coli reporter strains were cultured in 4 mL of LB contained in glass test tubes (1.5 cm in diameter × 15 cm in length) with shaking at 37°C for 8 h. An amount of 30 μL of LB cultures were used to inoculate 3 mL of M63 minimal media in smaller glass tubes (1.2 cm × 12 cm), and the tubes were shaken at 37°C overnight. The carbon sources were 0.5% glycerol, 0.5% salicin, or both. The tubes were rotated at 250 rpm and 37°C, and cell densities (OD_{600}) were measured with a Bio-Rad spectrophotometer. During the exponential growth phase, four samples were collected in the range of OD_{600} from 0.1 to 1. The samples (roughly 0.3 mL for promoter reporter strains, 0.9 mL for the IS1/5 promoters), and 0.6 mL for operon reporter strains were immediately frozen at −20°C prior to β-galactosidase assays. To measure β-galactosidase activities in *bgl* promoter reporter strains, 0.8 mL of Z-buffer containing β-mercaptoethanol (2.7 μL/mL) and sodium dodecyl sulfate (SDS) (0.005%) was mixed with 0.2 mL of sample and 25 μL of $CHCl_3$ in test tubes. Alternatively, for *bgl* operon reporter strains, 0.5 mL of Z-buffer was mixed with 0.5 mL of the sample. The tubes were vortexed twice (each time for 10 s at a constant speed) and incubated in a 37°C water bath until temperature equilibration. A 0.2 mL aliquot of O-nitrophenyl galactoside (ONPG) substrate (4 mg/mL) was then added to each test tube. When a yellow color developed, the reaction was stopped by adding 0.5 mL of 1 M Na_2CO_3 followed by vortexing. Reaction mixtures were centrifuged (15,000 rpm, 3 min), and the absorbance values of the supernatants were measured at 420 nm and 550 nm. A control tube was run in parallel using M63 salts instead of the test sample. β-galactosidase activity (Miller units) = $[(OD_{420} - 1.75 \times OD_{550})/(\text{sample volume in mL} \times \text{time in min})] \times 1{,}000$ (Miller, 1972). For a given test strain, the slope of OD_{600} values versus β-galactosidase activities was referred to as the promoter activity or the operon activity.

To test the effect of IHF on *bglGFB* transcription, the *bgl* promoter reporter (P*bgl*-Z) and the *bgl* operon reporter (P*bgl*-G-Z) (Tran et al., 2022) were individually transferred to an *ihfA* deletion background. This yielded strains Δ*ihfA*_P*bgl*-Z and Δ*ihfA*_P*bgl*-G-Z.

5.4 β-glucosidase (BglB) assay

β-glucosidase assay was performed as described in our previous study (Lam et al., 2022). Briefly, 20 μL of LB culture was transferred to M63 minimal medium with 0.66% casamino acids (CAA). After overnight growth at 37°C with shaking, the culture was diluted into 6 mL of M63 + 0.66% CAA +0.5% salicin minimal medium at a starting OD_{600} of 0.025. The cells were grown at 37°C with shaking. Five samples with 0.8 mL each were collected during the late exponential growth phase when the *bgl* operon was fully induced with an OD_{600} of above 1.5. The samples were centrifuged at a speed of 5,500 rpm for 2.5 min and the cells were suspended with 1 mL of Z-buffer with 50 μg/mL chloramphenicol.

To measure β-glucosidase (BglB) activities, test samples were warmed in a 37°C water bath, and 200 μL of *p*-nitrophenyl-β-D-glucoside (PNPG, 8 mg/mL) was added to the cell suspension in Z-buffer. After a visible yellow color appeared, the reaction was terminated by adding 0.5 mL of 1 M Na_2CO_3 and subsequently vortexing. The reaction mixture was centrifuged, and the absorbance of the reaction mixture was measured at 420 nm and 550 nm. The BglB activity of each sample was calculated using the equation: β-glucosidase (BglB) activity = [1,000 × ($OD_{420\,nm}$ − 1.75 × $OD_{550\,nm}$) × Dilution factor]/[$OD_{600\,nm}$ × Time of reaction (min) × Volume of sample (mL)]. The activity of the strain was determined by averaging the BglB activities of the samples measured.

β-glucosidase (BglB) assays were conducted in wild type and Δ*ihfA* backgrounds. To see the IHF effect on the cells with higher operon expression, strains G50 (deleted for the two terminators flanking *bglG*) and ΔP*bgl*-G (deleted for P*bgl* and *bglG* together with two terminators; Lam et al., 2022) was used.

5.5 Bgl$^+$ mutation assays

Bgl$^+$ mutation assays were performed on minimal M9 agar plates with 0.5% of a β-glucoside (salicin) as the sole carbon source. Strains to be tested (from single fresh colonies) were cultured in LB liquid medium for approximately 7 h at 37°C, washed twice with carbon source-free M9 salts (M9) and applied onto plates (2×10^7 cells/plate). The plates were then incubated in a 30°C incubator and were examined every 2 days for the appearance of Bgl$^+$ colonies, with each colony representing a new Bgl$^+$ mutation. On these β-glucoside minimal agar plates, any colonies appearing by day 2 were considered to be from Bgl$^+$ cells initially applied onto the plates. They were therefore subtracted from the subsequent measurements. IS1 and IS5 have been previously shown to insert into the same upstream *bgl* promoter region. Colony PCR using primers flanking this region was used to differentiate between insertional and non-insertional mutants. Mutation frequency was determined in the manner described in Cairns and Foster (1991) and Zhang and Saier (2009b).

For the mutation assays which used differential levels of aTC, the growth media was prepared with different amounts of aTC to obtain several ng/mL concentrations as outlined in Results 2.3.

To determine the effects of other carbon sources on Bgl$^+$ mutations, mutation assays were performed on minimal M9 agar plates with 0.25% glycerol or 1% propanediol as the sole carbon source. To determine the total populations, the cells were washed off the minimal M9 agar plates at relevant time points, serially diluted, and plated onto LB agar plates. To determine Bgl$^+$ populations, appropriate dilutions were applied on M9 + salicin agar plates. The frequencies of Bgl$^+$ mutations were determined as described above for Bgl$^+$ on M9 + salicin agar plates.

5.6 Glp$^+$ mutation assays

Glp$^+$ mutation assays were conducted on glycerol M9 minimal agar plates as described in Zhang and Saier (2009b). Strains Δ*crp* and Δ*crp*Δ*ihfA* were used for the mutation assays.

5.7 Swarming mutation assays

Using the wild-type and Δ*ihfA* strains, the swarming mutation assays for the appearance of hyper-swarming mutants (outgrowing subpopulations from the inoculated cells) were carried out using the method of Barker et al. (2004). Briefly, overnight cell cultures in LB media were washed once with M9 salts and diluted to an OD_{600} of 1.0 prior to use. Two microliters of the cell suspensions were streaked across the centers of LB semisolid (0.3% agar) plates (diameter = 9 cm) using a plastic transfer loop. The plates were incubated at 30°C. The swarming mutants, represented by outgrowths of motile subpopulations from the streaked cells, were counted. The mutation frequency was normalized as outgrowths (mutations) per 9-cm cell streak. Insertional mutants were verified via colony PCR.

5.8 Propanediol growth mutation assay

The assay for PPD$^+$ mutations was conducted by applying cell suspensions from fresh overnight cultures onto propanediol (1%) M9 minimal agar plates (~10^8 cells/plate). The wild-type and other strains were tested over a period of several days as in the Bgl$^+$ mutation assay. Total populations and mutation frequency were determined as described above under "Bgl$^+$ mutation assays."

5.9 Statistical analysis

The program R was used for all statistical analyses. *p*-values were generated to determine statistical significance, with values under 0.05 treated as statistically significant. The strains tested were plated along with known controls for each experimental run. Results shown are from a minimum of 3 separate runs/strain, with each run consisting of several plates. The data reported on the figures are expressed as the mean of a minimum of three runs for each strain at each time point.

LacZ activity assay results are expressed as the slope of a line generated from 4 time points/run and a minimum of 2 runs (at least 8 data points total). BglB assay results are expressed as the mean of 12 data points from a minimum of 3 separate runs.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

PK: Data curation, Investigation, Methodology, Writing – original draft, Writing – review & editing. ZZ: Conceptualization, Data curation, Methodology, Supervision, Writing – original draft, Writing – review & editing. MS: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was funded by NIH grant GM077402 and private contributions, both to MS.

Acknowledgments

We thank Arturo Medrano-Soto, Dennis Tran, Katie Lam, and Harry Zhou for their expertise and support in the preparation of this manuscript. We also thank Jialu Huo for his help with statistical analysis.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2024.1388522/full#supplementary-material

References

Allison, R. G., and Chaconas, G. (1992). Role of the a protein-binding sites in the in vitro transposition of mu DNA. A complex circuit of interactions involving the mu ends and the transpositional enhancer. *J. Biol. Chem.* 267, 19963–19970. doi: 10.1016/S0021-9258(19)88651-9

Amster-Choder, O., and Wright, A. (1992). Modulation of the dimerization of a transcriptional antiterminator protein by phosphorylation. *Science* 257, 1395–1398. doi: 10.1126/science.1382312

Barker, C. S., Prüss, B. M., and Matsumura, P. (2004). Increased motility of *Escherichia coli* by insertion sequence element integration into the regulatory region of the flhD operon. *J. Bacteriol.* 186, 7529–7537. doi: 10.1128/JB.186.22.7529-7537.2004

Bushman, W., Thompson, J. F., Vargas, L., and Landy, A. (1985). Control of directionality in lambda site specific recombination. *Science* 230, 906–911. doi: 10.1126/science.2932798

Cairns, J., and Foster, P. L. (1991). Adaptive reversion of a frameshift mutation in *Escherichia coli*. *Genetics* 128, 695–701. doi: 10.1093/genetics/128.4.695

Chalmers, R., Guhathakurta, A., Benjamin, H., and Kleckner, N. (1998). IHF modulation of Tn10 transposition: sensory transduction of supercoiling status via a proposed protein/DNA molecular spring. *Cell* 93, 897–908. doi: 10.1016/S0092-8674(00)81449-X

Chen, Q., Arents, J. C., Bader, R., Postma, P. W., and Amster-Choder, O. (1997). BglF, the sensor of the *E. coli* bgl system, uses the same site to phosphorylate both a sugar and a regulatory protein. *EMBO J.* 16, 4617–4627. doi: 10.1093/emboj/16.15.4617

Chen, Y. M., Lu, Z., and Lin, E. C. (1989). Constitutive activation of the fucAO operon and silencing of the divergently transcribed fucPIK operon by an IS5 element in *Escherichia coli* mutants selected for growth on L-1,2-propanediol. *J. Bacteriol.* 171, 6097–6105. doi: 10.1128/jb.171.11.6097-6105.1989

Curcio, M. J., and Derbyshire, K. M. (2003). The outs and ins of transposition: from mu to kangaroo. *Nat. Rev. Mol. Cell Biol.* 4, 865–877. doi: 10.1038/nrm1241

Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* 97, 6640–6645. doi: 10.1073/pnas.120163297

Deutscher, J. (2008). The mechanisms of carbon catabolite repression in bacteria. *Curr. Opin. Microbiol.* 11, 87–93. doi: 10.1016/j.mib.2008.02.007

Ditto, M. D., Roberts, D., and Weisberg, R. A. (1994). Growth phase variation of integration host factor level in *Escherichia coli*. *J. Bacteriol.* 176, 3738–3748. doi: 10.1128/jb.176.12.3738-3748.1994

Dole, S., Nagarajavel, V., and Schnetz, K. (2004). The histone-like nucleoid structuring protein H-NS represses the *Escherichia coli* bgl operon downstream of the promoter. *Mol. Microbiol.* 52, 589–600. doi: 10.1111/j.1365-2958.2004.04001.x

Engelhorn, M., and Geiselmann, J. (1998). Maximal transcriptional activation by the IHF protein of *Escherichia coli* depends on optimal DNA bending by the activator. *Mol. Microbiol.* 30, 431–441. doi: 10.1046/j.1365-2958.1998.01078.x

Fic, E., Bonarek, P., Gorecki, A., Kedracka-Krok, S., Mikolajczak, J., Polit, A., et al. (2009). cAMP receptor protein from *escherichia coli* as a model of signal transduction in proteins--a review. *J. Mol. Microbiol. Biotechnol.* 17, 1–11. doi: 10.1159/000178014

Fox, C. F., and Wilson, G. (1968). The role of a phosphoenolpyruvate-dependent kinase system in beta-glucoside catabolism in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 59, 988–995. doi: 10.1073/pnas.59.3.988

Gamas, P., Chandler, M. G., Prentki, P., and Galas, D. J. (1987). *Escherichia coli* integration host factor binds specifically to the ends of the insertion sequence IS1 and to its major insertion hot-spot in pBR322. *J. Mol. Biol.* 195, 261–272. doi: 10.1016/0022-2836(87)90648-6

Goosen, N., and van de Putte, P. (1995). The regulation of transcription initiation by integration host factor. *Mol. Microbiol.* 16, 1–7. doi: 10.1111/j.1365-2958.1995.tb02386.x

Görke, B., and Rak, B. (1999). Catabolite control of *Escherichia coli* regulatory protein BglG activity by antagonistically acting phosphorylations. *EMBO J.* 18, 3370–3379. doi: 10.1093/emboj/18.12.3370

Görke, B., and Stülke, J. (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat. Rev. Microbiol.* 6, 613–624. doi: 10.1038/nrmicro1932

Granston, A. E., and Nash, H. A. (1993). Characterization of a set of integration host factor mutants deficient for DNA binding. *J. Mol. Biol.* 234, 45–59. doi: 10.1006/jmbi.1993.1562

Gueguen, E., Rousseau, P., Duval-Valentin, G., and Chandler, M. (2005). The transpososome: control of transposition at the level of catalysis. *Trends Microbiol.* 13, 543–549. doi: 10.1016/j.tim.2005.09.002

Gulati, A., and Mahadevan, S. (2000). Mechanism of catabolite repression in the bgl operon of *Escherichia coli*: involvement of the anti-terminator BglG, CRP-cAMP and EIIAGlc in mediating glucose effect downstream of transcription initiation. *Genes Cells* 5, 239–250. doi: 10.1046/j.1365-2443.2000.00322.x

Hales, L. M., Gumport, R. I., and Gardner, J. F. (1994). Mutants of *Escherichia coli* integration host factor: DNA-binding and recombination properties. *Biochimie* 76, 1030–1040. doi: 10.1016/0300-9084(94)90027-2

Hall, B. G. (1998). Activation of the bgl operon by adaptive mutation. *Mol. Biol. Evol.* 15, 1–5. doi: 10.1093/oxfordjournals.molbev.a025842

Humayun, M. Z., Zhang, Z., Butcher, A. M., Moshayedi, A., and Saier, M. H. (2017). Hopping into a hot seat: role of DNA structural features on IS5-mediated gene activation and inactivation under stress. *PLoS One* 12:e0180156. doi: 10.1371/journal.pone.0180156

Klumpp, S., Zhang, Z., and Hwa, T. (2009). Growth rate-dependent global effects on gene expression in bacteria. *Cell* 139, 1366–1375. doi: 10.1016/j.cell.2009.12.001

Kobryn, K., Watson, M. A., Allison, R. G., and Chaconas, G. (2002). The mu three-site synapse: a strained assembly platform in which delivery of the L1 transposase binding site triggers catalytic commitment. *Mol. Cell* 10, 659–669. doi: 10.1016/S1097-2765(02)00596-8

Kolb, A., Busby, S., Buc, H., Garges, S., and Adhya, S. (1993). Transcriptional regulation by cAMP and its receptor protein. *Annu. Rev. Biochem.* 62, 749–797. doi: 10.1146/annurev.bi.62.070193.003533

Lam, K. J. K., Zhang, Z., and Saier, M. H. (2022). Histone-like nucleoid structuring (H-NS) protein silences the beta-glucoside (bgl) utilization operon in *Escherichia coli* by forming a DNA loop. *Comput. Struct. Biotechnol. J.* 20, 6287–6301. doi: 10.1016/j.csbj.2022.11.027

Levine, E., Zhang, Z., Kuhlman, T., and Hwa, T. (2007). Quantitative characteristics of gene regulation by small RNA. *PLoS Biol.* 5:e229. doi: 10.1371/journal.pbio.0050229

Lopilato, J., and Wright, A. (1990). "Mechanisms of activation of the cryptic bgl operon of *Escherichia coli* K-12" in *The bacterial chromosome*. eds. K. Drilca and M. Riley (Washington DC: American Society for Microbiology), 435–444.

Lutz, R., and Bujard, H. (1997). Independent and tight regulation of transcriptional units in *Escherichia coli* via the lac R/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* 25, 1203–1210. doi: 10.1093/nar/25.6.1203

Mahadevan, S., and Wright, A. (1987). A bacterial gene involved in transcription antitermination: regulation at a rho-independent terminator in the bgl operon of *E. coli*. *Cell* 50, 485–494. doi: 10.1016/0092-8674(87)90502-2

McCalla, D. R., Kaiser, C., and Green, M. H. (1978). Genetics of nitrofurazone resistance in *Escherichia coli*. *J. Bacteriol.* 133, 10–16. doi: 10.1128/jb.133.1.10-16.1978

McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA* 36, 344–355. doi: 10.1073/pnas.36.6.344

Miller, F. (1972). Glycopeptides of human immunoglobulins. 3. The use and preparation of specific glycosidases. *Immunochemistry* 9, 217–228. doi: 10.1016/0019-2791(72)90087-0

Mukerji, M., and Mahadevan, S. (1997). Characterization of the negative elements involved in silencing the bgl operon of *Escherichia coli*: possible roles for DNA gyrase, H-NS, and CRP-cAMP in regulation. *Mol. Microbiol.* 24, 617–627. doi: 10.1046/j.1365-2958.1997.3621725.x

Muramatsu, S., Kato, M., Kohara, Y., and Mizuno, T. (1988). Insertion sequence IS5 contains a sharply curved DNA structure at its terminus. *Mol. Gen. Genet. MGG* 214, 433–438. doi: 10.1007/BF00330477

Nagarajavel, V., Madhusudan, S., Dole, S., Rahmouni, A. R., and Schnetz, K. (2007). Repression by binding of H-NS within the transcription unit. *J. Biol. Chem.* 282, 23622–23630. doi: 10.1074/jbc.M702753200

Ohta, S., Tsuchida, K., Choi, S., Sekine, Y., Shiga, Y., and Ohtsubo, E. (2002). Presence of a characteristic D-D-E motif in IS1 transposase. *J. Bacteriol.* 184, 6146–6154. doi: 10.1128/JB.184.22.6146-6154.2002

Pozdeev, G., Beckett, M. C., Mogre, A., Thomson, N. R., and Dorman, C. J. (2022). Reciprocally rewiring and repositioning the integration host factor (IHF) subunit genes in *Salmonella enterica* serovar typhimurium: impacts on physiology and virulence. *Microb. Genom.* 8:000768. doi: 10.1099/mgen.0.000768

Prasad, I., and Schaefler, S. (1974). Regulation of the beta-glucoside system in Escherchia coli K-12. *J. Bacteriol.* 120, 638–650. doi: 10.1128/jb.120.2.638-650.1974

Prasad, I., Young, B., and Schaefler, S. (1973). Genetic determination of the constitutive biosynthesis of phospho-β-glucosidase a in *Escherichia coli* K-12. *J. Bacteriol.* 114, 909–915. doi: 10.1128/jb.114.3.909-915.1973

Prentki, P., Chandler, M., and Galas, D. J. (1987). *Escherichia coli* integration host factor bends the DNA at the ends of IS1 and in an insertion hotspot with multiple IHF binding sites. *EMBO J.* 6, 2479–2487. doi: 10.1002/j.1460-2075.1987.tb02529.x

Reynolds, A. E., Felton, J., and Wright, A. (1981). Insertion of DNA activates the cryptic bgl operon in *E. coli* K12. *Nature* 293, 625–629. doi: 10.1038/293625a0

Rothe, F. M., Bahr, T., Stülke, J., Rak, B., and Görke, B. (2012). Activation of *Escherichia coli* antiterminator BglG requires its phosphorylation. *Proc. Natl. Acad. Sci. USA* 109, 15906–15911. doi: 10.1073/pnas.1210443109

Sawers, R. G. (2005). Transcript analysis of *Escherichia coli* K-12 insertion element IS5. *FEMS Microbiol. Lett.* 244, 397–401. doi: 10.1016/j.femsle.2005.02.019

Schnetz, K. (1995). Silencing of *Escherichia coli* bgl promoter by flanking sequence elements. *EMBO J.* 14, 2545–2550. doi: 10.1002/j.1460-2075.1995.tb07252.x

Schnetz, K., and Rak, B. (1988). Regulation of the bgl operon of *Escherichia coli* by transcriptional antitermination. *EMBO J.* 7, 3271–3277. doi: 10.1002/j.1460-2075.1988.tb03194.x

Sclavi, B., Beatty, C. M., Thach, D. S., Fredericks, C. E., Buckle, M., and Wolfe, A. J. (2007). The multiple roles of CRP at the complex acs promoter depend on activation region 2 and IHF. *Mol. Microbiol.* 65, 425–440. doi: 10.1111/j.1365-2958.2007.05797.x

Singh, J., Mukerji, M., and Mahadevan, S. (1995). Transcriptional activation of the *Escherichia coli* bgl operon: negative regulation by DNA structural elements near the promoter. *Mol. Microbiol.* 17, 1085–1092. doi: 10.1111/j.1365-2958.1995.mmi_17061085.x

Sugimura, S., and Crothers, D. M. (2006). Stepwise binding and bending of DNA by *Escherichia coli* integration host factor. *Proc. Natl. Acad. Sci. USA* 103, 18510–18514. doi: 10.1073/pnas.0608337103

Surette, M. G., Lavoie, B. D., and Chaconas, G. (1989). Action at a distance in mu DNA transposition: an enhancer-like element is the site of action of supercoiling relief activity by integration host factor (IHF). *EMBO J.* 8, 3483–3489. doi: 10.1002/j.1460-2075.1989.tb08513.x

Swinger, K. K., and Rice, P. A. (2004). IHF and HU: flexible architects of bent DNA. *Curr. Opin. Struct. Biol.* 14, 28–35. doi: 10.1016/j.sbi.2003.12.003

Tran, D., Zhang, Z., Lam, K. J. K., and Saier, M. H. (2022). Effects of global and specific DNA-binding proteins on transcriptional regulation of the *E. coli* bgl operon. *Int. J. Mol. Sci.* 23:10343. doi: 10.3390/ijms231810343

van Ulsen, P., Hillebrand, M., Zulianello, L., van de Putte, P., and Goosen, N. (1996). Integration host factor alleviates the H-NS-mediated repression of the early promoter of bacteriophage mu. *Mol. Microbiol.* 21, 567–578. doi: 10.1111/j.1365-2958.1996.tb02565.x

Whiteway, J., Koziarz, P., Veall, J., Sandhu, N., Kumar, P., Hoecher, B., et al. (1998). Oxygen-insensitive nitroreductases: analysis of the roles of nfsA and nfsB in development of resistance to 5-nitrofuran derivatives in *Escherichia coli*. *J. Bacteriol.* 180, 5529–5539. doi: 10.1128/JB.180.21.5529-5539.1998

Zhang, Z., Kukita, C., Humayun, M. Z., and Saier, M. H. (2017). Environment-directed activation of the *Escherichia coli*flhdc operon by transposons. *Microbiology* 163, 554–569. doi: 10.1099/mic.0.000426

Zhang, Z., and Saier, M. H. (2009a). A novel mechanism of transposon-mediated gene activation. *PLoS Genet.* 5:e1000689. doi: 10.1371/journal.pgen.1000689

Zhang, Z., and Saier, M. H. (2009b). A mechanism of transposon-mediated directed mutation. *Mol. Microbiol.* 74, 29–43. doi: 10.1111/j.1365-2958.2009.06831.x

Zhang, Z., Yen, M. R., and Saier, M. H. (2010). Precise excision of IS5 from the intergenic region between the fucPIK and the fucAO operons and mutational control of fucPIK operon expression in *Escherichia coli*. *J. Bacteriol.* 192, 2013–2019. doi: 10.1128/JB.01085-09

Zhang, Z., Zhou, K., Tran, D., and Saier, M. (2022). Insertion sequence (IS) element-mediated activating mutations of the cryptic aromatic β-glucoside utilization (Bgl GFB) operon are promoted by the anti-terminator protein (BglG) in *Escherichia coli*. *Int. J. Mol. Sci.* 23:1505. doi: 10.3390/ijms23031505

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 06 June 2024
DOI 10.3389/fmicb.2024.1423762

OPEN ACCESS

EDITED BY
Axel Cloeckaert,
Institut National de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), France

REVIEWED BY
Karl Pedersen,
Aarhus University, Denmark
Magnus Rasmussen,
Lund University, Sweden

*CORRESPONDENCE
Oddvar Oppegaard
✉ O.Oppegaard@uib.no

RECEIVED 26 April 2024
ACCEPTED 15 May 2024
PUBLISHED 06 June 2024

CITATION
Glambek M, Skrede S, Sivertsen A, Kittang BR, Kaci A, Jonassen CM, Jørgensen HJ, Norwegian Study Group on *Streptococcus dysgalactiae* and Oppegaard O (2024) Antimicrobial resistance patterns in *Streptococcus dysgalactiae* in a One Health perspective.
Front. Microbiol. 15:1423762.
doi: 10.3389/fmicb.2024.1423762

Antimicrobial resistance patterns in *Streptococcus dysgalactiae* in a One Health perspective

Marte Glambek[1], Steinar Skrede[1,2], Audun Sivertsen[3], Bård Reiakvam Kittang[2,4], Alba Kaci[5], Christine Monceyron Jonassen[5,6], Hannah Joan Jørgensen[7], Norwegian Study Group on *Streptococcus dysgalactiae* and Oddvar Oppegaard[1,2]*

[1]Department of Medicine, Haukeland University Hospital, Bergen, Norway, [2]Department of Clinical Medicine 2, Department of Clinical Science, University of Bergen, Bergen, Norway, [3]Department of Microbiology, Haukeland University Hospital, Bergen, Norway, [4]Department of Internal Medicine, Haraldsplass Deaconess Hospital (HDS), Bergen, Norway, [5]Center for Laboratory Medicine, Østfold Hospital, Grålum, Norway, [6]Department of Virology, Division of Infection Control and Environmental Health, Norwegian Institute of Public Health, Oslo, Norway, [7]Norwegian Veterinary Institute, Ås, Norway

Background: *Streptococcus dysgalactiae* (SD) is an important pathogen in humans as well as in a broad range of animal species. Escalating rates of antibiotic resistance in SD has been reported in both human and veterinary clinical practice, but the dissemination of resistance determinants has so far never been examined in a One Health Perspective. We wanted to explore the occurrence of zoonotic transmission of SD and the potential for exchange of resistance traits between SD from different host populations.

Methods: We compared whole genome sequences and phenotypical antimicrobial susceptibility of 407 SD isolates, comprising all isolates obtained from human bloodstream infections in 2018 ($n = 274$) and available isolates associated with animal infections from the years 2018 and 2019 ($n = 133$) in Norway.

Results: Antimicrobial resistance genes were detected in 70 (26%), 9 (25%) and 2 (2%) of the isolates derived from humans, companion animals and livestock, respectively. Notably, distinct host associated genotypic resistomes were observed. The *erm*(A) gene was the dominant cause of erythromycin resistance in human associated isolates, whereas only *erm*(B) and *lsa*(C) were identified in SD isolates from animals. Moreover, the tetracycline resistance gene *tet*(O) was located on different mobile genetic elements in SD from humans and animals. Evidence of niche specialization was also evident in the phylogenetic analysis, as the isolates could be almost perfectly delineated in accordance with host species. Nevertheless, near identical mobile genetic elements were observed in four isolates from different host species including one human, implying potential transmission of antibiotic resistance between different environments.

Conclusion: We found a phylogenetic delineation of SD strains in line with host adapted populations and niche specialization. Direct transmission of strains or genetic elements carrying resistance genes between SD from different ecological niches appears to be rare in our geographical region.

KEYWORDS

***Streptococcus dysgalactiae*, antibiotic resistance, One Health, whole genomic sequencing, horizontal genetic transfer, animal, human, Norway**

Introduction

Streptococcus dysgalactiae (SD) causes a broad spectrum of human infections ranging from asymptomatic carriage via non-invasive soft tissue infections to life threatening conditions like necrotizing fasciitis and toxic shock syndrome (Brandt and Spellerberg, 2009). In the past decades there has been a significant increase in invasive infections in humans caused by SD in several geographical regions, and SD is currently among the most common pathogens detected in bloodstream infections in some countries (Couture-Cossette et al., 2018; UK Health Security Agency, 2022; Nevanlinna et al., 2023; Oppegaard et al., 2023).

SD is not a strict human pathogen, but capable of infecting a broad range of host species. It is recognized as a major cause of bovine mastitis, arthritis in swine and ovine lambs and necrotic ulcers in fish in aquaculture (Abdelsalam et al., 2010; Moreno et al., 2016; Zhang et al., 2018; Smistad et al., 2021). Moreover, SD is associated with a variety of infections in dogs, cats, and horses, underpinning the ecological versatility of this pathogen (Acke et al., 2015). At the same time, a phylogenetic diversity is evident within the SD taxon, and likely extends beyond the current delineation into the subspecies *Streptococcus dysgalactiae* subsp. *dysgalactiae* (SDSD) and *Streptococcus dysgalactiae* subsp. *equisimilis* (SDSE) (Ciszewski et al., 2016; Alves-Barroco et al., 2022). SDSD is, by definition, restricted to α-hemolytic group C strains, predominantly infecting cattle and sheep, whereas SDSE comprises all β-hemolytic strains (Vieira et al., 1998).

Penicillin remains a cornerstone in the treatment of infections caused by SD, and resistance to this antibiotic is exceedingly rare. Regarding second line alternatives, however, the situation is more alarming. Rising rates of macrolide, lincosamide and streptogramin (MLS) resistance have been noted, and in the United Kingdom MLS resistance is approaching 40% in SD isolates collected from humans (UK Health Security Agency, 2022). Even higher rates have been reported in bovine associated SD in China, as well as SD isolated from swine in South America (Moreno et al., 2016; Zhang et al., 2018). Moreover, increasing numbers of tetracycline resistant pyogenic streptococci are observed, and, as for the MLS resistance, this trend seems independent of host species (Abdelsalam et al., 2010; Ciszewski et al., 2016; Garch et al., 2020).

The main drivers for increasing antibiotic resistance are selection of resistant microbes by use and overuse of antibiotics, the possibility of horizontal genetic transfer of resistance traits between bacteria, and the spread of resistant bacteria between different geographic and ecological environments, both locally and globally. The complex ecological processes call for collaboration of multiple science fields in a "One health perspective" to approach and overcome emerging antibiotic resistance (McEwen and Collignon, 2018). However, our current knowledge on antimicrobial resistance in SD is predominantly based on studies limited to distinct host reservoirs, and data on dissemination of resistance determinants between ecological niches is scarce. Hence, there is a need for comparative studies on contemporary and spatially related isolates from different host species to investigate possible pathways for spreading of resistance traits.

We sought to explore antimicrobial resistance of SD in a One Health perspective and have examined clinically relevant SD isolates collected from various host species within a confined temporal and geographical setting. By dissecting phenotypic susceptibility patterns and whole genome sequences, we compared the resistomes associated with the different host specific reservoirs and attempted to elucidate possible routes for spread of resistance traits.

Methods

Bacterial isolates

All SD isolates identified in human blood cultures in Norway during 2018 were collected as part of the Norwegian surveillance program for antimicrobial resistance (NORM, 2019). SD isolates from bovine sources were randomly selected among isolates from bovine mastitis by TINE SA mastitis laboratory (Molde, Norway) in 2018. Additionally, all SD isolates from clinical samples from animals submitted to the Norwegian Veterinary Institute in 2018 and 2019, were included. Only one isolate per person and one isolate per animal flock or herd was included.

Species identification in the primary laboratories was based on colony morphology (hemolytic reaction on 5% sheep blood agar and colony size >0.5 mm after 24 h of incubation), serogroup specificity using rapid Lancefield agglutination test, and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). All isolates identified as SD were submitted to either Haukeland University Hospital, Bergen, or Østfold Hospital Trust, Grålum, for susceptibility testing and genomic characterization.

Susceptibility testing

All isolates were examined for susceptibility to benzylpenicillin, tetracycline, erythromycin, clindamycin, and trimethoprim-sulphamethoxazole according to the NORM protocol (NORM, 2019). Briefly, isolates were plated on Mueller-Hinton agar supplemented with defibrinated horse blood and β-NAD, and minimum inhibitory concentrations (MIC) were determined using MIC gradient strips. The Kirby-Bauer double disc diffusion method was used to assign the constitutive macrolide-lincosamide-streptogramin (cMLS), the inducible MLS (iMLS) and the macrolide (M) resistance phenotypes. Clinical breakpoints (version 14.0) set by the European Committee for Antimicrobial Susceptibility Testing (EUCAST) were used for interpretation of susceptibility.

Whole genome sequencing

Whole Genome Sequencing of 119 SD of human origin as well as all 133 SD from animals was performed at Haukeland University Hospital on an Illumina 4,000 HiSeq system to produce 150 bp paired end reads, as previously described (Oppegaard et al., 2017). The remaining 155 human isolates were sequenced at Østfold Hospital Trust by the Ion Torrent technology on an Ion S5XL system as previously described (Kaci et al., 2023).

In silico analysis

For data generated on the Illumina HiSeq system, reads were trimmed with Trimmomatic v0.39 (Bolger et al., 2014). For Ion Torrent generated data, reads were processed with the incorporated

S5 software plug-ins. All trimmed reads from the sequenced isolates were *de novo* assembled by SPAdes v5.14 (Bankevich et al., 2012). Genome annotation was accomplished using RAST v1.073 (Aziz et al., 2008). Species identity was confirmed by 16S rDNA analysis. A core genome single-nucleotide polymorphism phylogeny was generated by CSI Phylogeny at the Center for Genomic Epidemiology[1] using default settings and the SDSE type strain NCTC13762 as a reference. The resulting maximum likelihood phylogenetic tree was visualized and annotated using the Interactive Tree of Life platform, iTol v6 (Letunic and Bork, 2021).

Multilocus sequence typing (MLST) of the isolates was performed using the MLST 2.0 software available at the CGE webpage. The *emm*-database at the Centers for Disease Control and Prevention webpage was used to determine the *emm*-types.[2] A minimum spanning phylogenetic tree using MLST types was constructed using the Phyloviz online tool,[3] using triple-locus variant limitation for clustering.

RESfinder was used to screen for the presence of resistance genes (Florensa et al., 2022). Geneious Prime v 2022.2 was used to inspect the genetic context of the resistance genes, and screen for known mobilization genes from streptococcal mobile genetic elements using a database adapted form CONJdb.[4] BLASTn was used to search for closest matches to putative mobile elements.

There are no validated methods for genotypic distinction between the two subspecies SDSE and SDSD. In accordance with the phenotypic definition proposed by Vieira et al. (1998), we defined SDSD *in silico* as genomes harboring the Lancefield group C-antigen operon, lacking the streptolysin S operon (corresponding to an α- or nonhemolytic reaction on blood agar), and lacking the streptokinase gene (inferring that streptokinase activity on human plasminogen does not occur). All other genomes were classified as SDSE.

Statistical analysis

Differences in resistance rates between the host associated SD populations were tested for statistical significance by Fisher's exact test. A two-sided p-value <0.05 was considered significant. Due to the low number of isolates available from some host associated populations, the data were pooled into isolates derived from humans, companion animals (dog and horse) and livestock (cow, sheep, swine) for statistical analysis.

Results

A total of 407 SD isolates were included in the study (Supplementary Table S1). Of these, 274 were isolated from human blood cultures in 2018, constituting all SD isolates registered in the national surveillance program this year (NORM, 2019). Among the 133 animal associated isolates, 97 originated from livestock, including cattle ($n = 74$), sheep ($n = 11$) and swine ($n = 12$). The remaining 36 isolates were from dogs ($n = 20$) and horses ($n = 16$).

Whole genome sequencing and phylogenetic analyses

The draft genomes of the 407 SD isolates had an average assembly length of 2.10 Mb, GC content of 39.3%, 2,100 protein encoding genes, and a coverage of approximately 220x. Based on whole genome sequencing analysis, 83 out of 85 isolates from bovine and ovine sources were classified as SDSD, whereas all other isolates belonged to the subspecies *equisimilis*. This phylogenetic delineation was in line with the phenotypic species identification.

Phylogenetic analysis delineated the isolates largely in accordance with host species (Figure 1). A notable exception was the distinct cluster corresponding to SDSD, where the isolates derived from bovine and ovine hosts were phylogenetically inseparable. The large clade of human associated isolates was clearly demarcated from the SD isolated from different animals. Nevertheless, indications of SD isolates crossing species barriers were observed. One isolate obtained from a human blood culture clustered phylogenetically with dog associated isolates. Inversely, five isolates obtained from dogs were found scattered in the cluster of isolates from humans, comprising 25% of all dog associated isolates. The minimum spanning tree based on MLST types was congruent with the single nucleotide polymorphism phylogenetic tree (Supplementary Figure S1).

Phenotypic susceptibility testing

The phenotypic antimicrobial resistance rates were relatively similar among isolates derived from humans, companion animals and livestock (Figure 2A). The only significant difference was a higher tetracycline resistance rate in SD from livestock (41%) compared to SD from companion animals (19%, $p = 0.02$). Resistance to erythromycin or clindamycin (MLS-resistance) was detected in 33 (12%), 3 (8%) and 8 (8%) of the isolates obtained from humans, companion animals and livestock, respectively. Of note, MLS-resistance in companion animals was derived from isolates from dogs only, and all MLS-resistance among livestock was detected in bovine associated isolates. All isolates in the study were susceptible to benzylpenicillin and trimethoprim-sulfamethoxazole.

Resistome analysis

Whole genome sequences of all the included bacterial strains were screened for antimicrobial resistance genes (Table 1). In total, resistance genes were detected in 70 (26%), 9 (25%), and 2 (2%) of the isolates derived from humans, companion animals and livestock, respectively. We almost exclusively detected genes encoding resistance to either tetracyclines or MLS-antibiotics, and *tet*(M), *tet*(O) and *erm*(A) were the most abundant. However, distinct host associated resistance gene profiles were noted; *erm*(A) accounted for 82% of the MLS resistance in human associated isolates, whereas only *erm*(B) and *lsa*(C) were identified in MLS resistant SD from companion animals. Moreover, the rates of genotypic resistance to

1 https://www.genomicepidemiology.org/
2 www.cdc.gov/streplab
3 online.phyloviz.net
4 conjdb.web.pasteur.fr

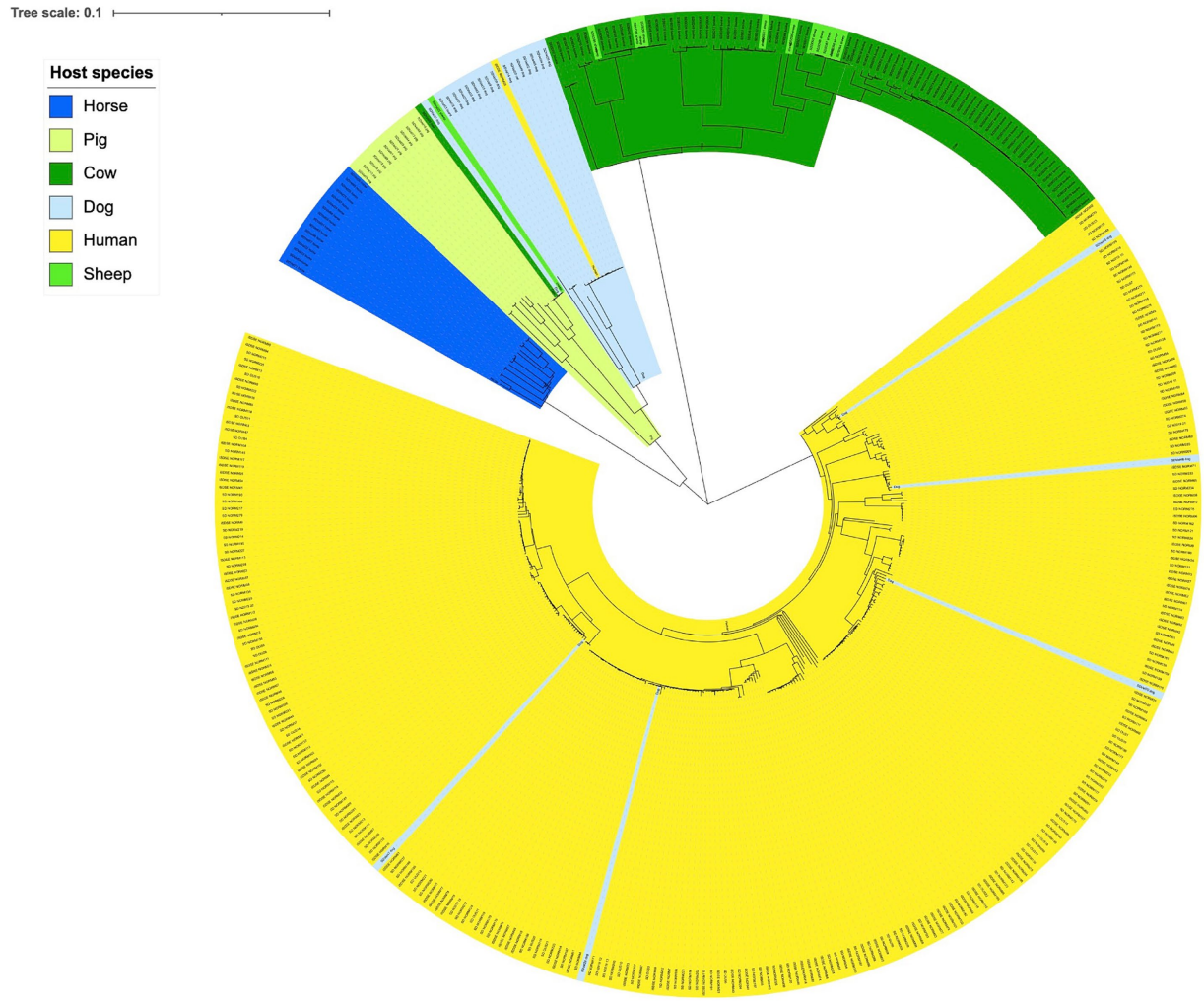

FIGURE 1
Genetic relationship between *Streptococcus dysgalactiae* isolates from different host species. Scale indicates substitutions per site. The phylogenetic tree is constructed based on a core genome single nucleotide polymorphism alignment using a maximum likelihood method.

macrolides, clindamycin and tetracycline were all significantly lower among isolates procured from livestock than from humans and companion animals (Figure 2B).

The predicted genotypic resistance rates were substantially lower than the observed phenotypic resistance rates, particularly for tetracycline. Exploring this discrepancy, we found that the overall distribution of the tetracycline MIC values in our SD population appeared to be trimodal, and only the cluster with the highest MIC values correlated with isolates harboring genes encoding resistance to tetracycline (Figure 3A). The central cluster was intersected by the EUCAST tetracycline breakpoint, but the isolates lacked identifiable validated resistance genes. This resistance gene negative, low-grade resistant population included 39 of 40 livestock associated SD isolates with reduced susceptibility to tetracycline, but also comprised a distinct phylogenetic cluster of human associated strains (Figure 3B).

Incongruence between phenotypic and genotypic resistance traits was also observed for other antimicrobial agents. All eight erythromycin resistant livestock associated isolates had MIC values just above the susceptibility breakpoint, and none of them harbored identifiable resistance genes. An identical pattern was observed for 7 of 33 human associated SD isolates displaying reduced susceptibility to erythromycin.

Inversely, a few isolates had identifiable resistance genes but displayed phenotypical susceptibility, including three isolates harboring *lsa*(C)-genes, and one strain with a truncated *tet*(M)-gene.

Mobile genetic elements and resistance genes

Analyses of flanking sequences of the detected resistance genes revealed a location on mobile genetic elements (MGEs) in almost all cases (Table 2). The one exception to this was a *tet*(M) gene located on a contig with a flanking sequence too short to determine the location with certainty. Integrative conjugative elements (ICEs) were the predominant form of MGEs detected, but the *erm*(T) gene was carried on a small *p5580*-like plasmid, and one dog isolate harbored a bacteriophage carrying an *erm*(B) gene. The major vector for MLS resistance was MGEs belonging to the ICE*Sp2905* family. These ICEs harbored 85% of the MLS resistance genes, including *lsa*(C) genes

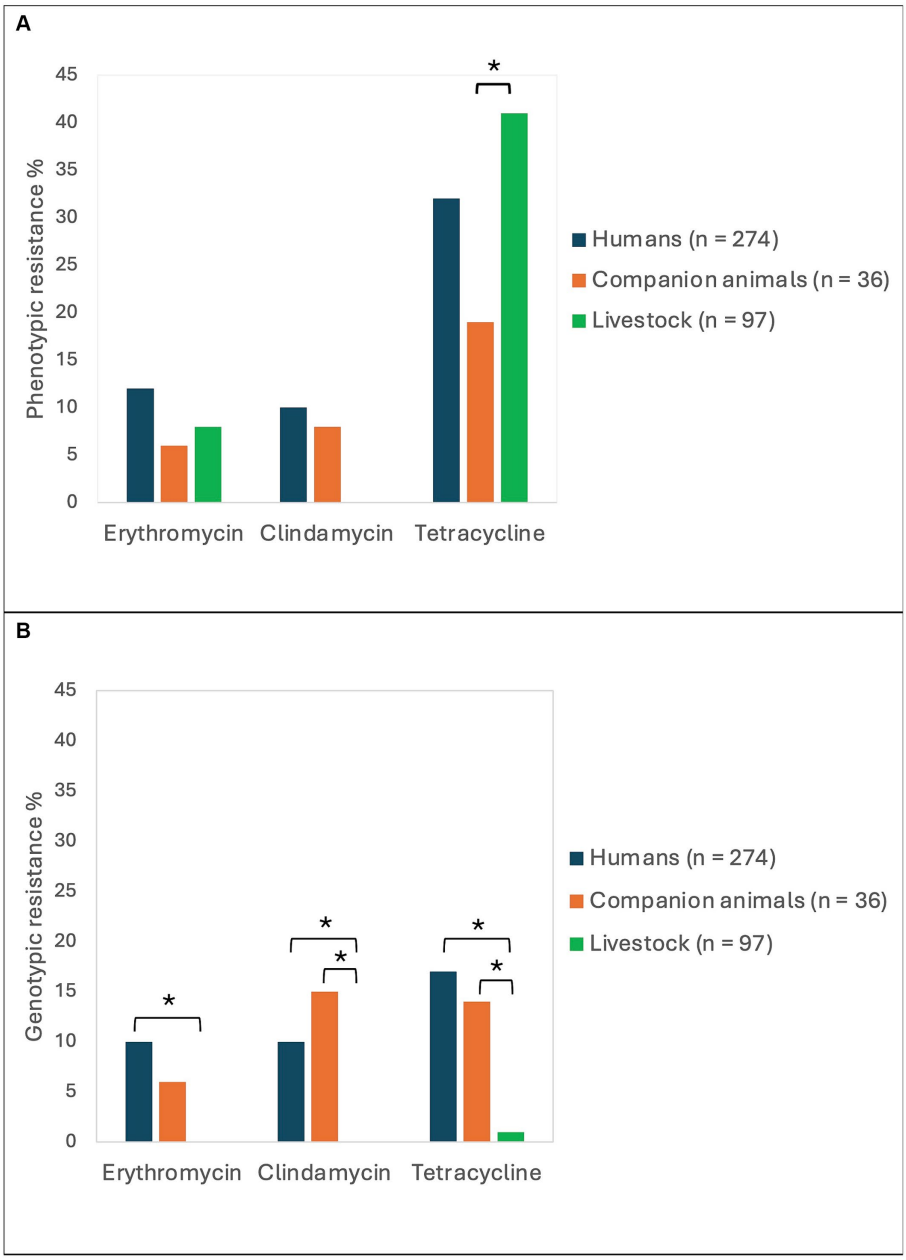

FIGURE 2
Antimicrobial resistance rates in *Streptococcus dysgalactiae* isolates from different host groups. The figure displays the prevalence of phenotypic **(A)** and genotypic **(B)** resistance to erythromycin, clindamycin, and tetracycline among *Streptococcus dysgalactiae* isolates procured from humans, companion animals and livestock animals. Asterisks mark significant differences in antimicrobial resistance levels.

in both human and animal associated isolates. Nevertheless, ICE*Sp2905* elements in strains from different host sources displayed less than 95% sequence similarity based on core ICE conjugation genes.

On a similar note, the *tet*(O) genes, giving resistance to tetracycline, were located on ICEs belonging to the ICE*Sp2905* and ICE*Sa2603* family, but were distinctly associated with SD from human and animal sources, respectively. The location of *tet*(M) was more diverse, but the two major vectors in human associated SD isolates were the ICEs *Tn916* and *Tn6944*.

The element *Tn5801* was the most common harboring *tet*(M) in SD isolated from animal hosts. *Tn5801* is divided into type A and B, where the type B variant is lacking two genes at the beginning of the element (León-Sampedro et al., 2016). Two isolates from horses and one from a dog carried the type A variant of *Tn5801*, whereas one human associated isolate harbored the type B. Notably, the *Tn5801* element in the dog associated isolate SDVet48 clustered phylogenetically to the horse associated elements (Figure 4), even though the bacterial isolate itself phylogenetically resided among human isolates. By BLASTn search, both variants of the *Tn5801* elements were detected in a range of streptococcal, enterococcal and staphylococcal strains, isolated from both human and animal sources. Several of these showed more than 99% overall sequence homology to the type A and B *Tn5801* elements detected in our isolates.

TABLE 1 Genotypic resistance found in isolates of *Streptococcus dysgalactiae* of different host origin.

	Resistance gene	Humans (*n* = 274)	Companion animals (*n* = 36)	Livestock (*n* = 97)	Total
Tetracycline	*tet*(M)	29 (11%)	4 (11%)	0	33
	tet(O)	16 (6%)	1 (3%)	1 (1%)	18
	tet(T)	1 (< 1%)	0	0	1
	tet(W)	1 (< 1%)	0	0	1
MLS	*erm*(A)	23 (8%)	0	0	23
	erm(B)	2 (1%)	2 (6%)	0	4
	erm(T)	1 (<1%)	0	0	1
	lsa(C)	1 (<1%)	3 (8%)	0	4
	mef(A)	1 (<1%)	0	0	1
Other	*ant6Ia*	0	1 (3%)	2 (2%)	3
	dfrF	1 (< 1%)	0	0	1
	cat	1 (< 1%)	0	0	1

MLS, macrolide, lincosamide and streptogramin antibiotics. Numbers in parenthesis represent percentage of total number of isolates in the host category group.

Discussion

To our best knowledge, this is the first comparative analysis of antimicrobial resistance patterns in contemporary *Streptococcus dysgalactiae* (SD) isolates in a One Health perspective. Overall, we found that direct transmission of bacterial strains or genetic content between different ecological niches appears to be infrequent. This is underpinned by the observed host specialization among the SD isolates and predominantly a difference in the genetic recipe for resistance, inferring different origins of the expressed traits.

A delineation of SD populations in accordance with host species has also been reported previously (Porcellato et al., 2021; Alves-Barroco et al., 2022). Porcellato et al. found that SD isolates harbored several host specific virulence factors and appeared to have evolved through genetic exchange with other bacterial species residing within their ecological niche (Porcellato et al., 2021). Segregation into host adapted phylogenetic lineages is also seen in studies of *Streptococcus agalactiae* and *Staphylococcus aureus*, and such niche specialization potentially represents a barrier to cross-species transmission (Maeda et al., 2020; Matuszewska et al., 2020). Supporting this, epidemiological collections of whole genomes of SD isolates from Japan, Canada and Denmark all reveal a very low frequency of animal associated MLST-profiles among SD isolated from humans, suggesting that zoonotic transmission of this pathogen is rare (Lother et al., 2017; Rebelo et al., 2022; Shinohara et al., 2023).

Despite the apparent transmission barrier conferred by niche adaptation, we found evidence of some level of cross-species exchange of SD isolates, predominantly between humans and companion animals. This is not surprising considering the close contact that often exists between humans and their dogs, and to a certain degree also between humans and horses. Similarly, Pinho et al. found that two SD isolates from dog and horse, respectively, were clustered together with human isolates in a phylogenetic study based on MLST analysis (Pinho et al., 2016). An SD isolate with the same MLST sequence type as the dog isolate to which it is referred was sampled from a boy living in the same household as the dog, reinforcing the hypothesis of a cross-species transmission event (Schrieber et al., 2014). In Singapore, an SD isolate phylogenetically resembling piscine associated SD strains was identified in a skin infection in a fish handler (Koh et al., 2009). Taken together, these findings indicate that cross-species transmission does occur, but predominantly in situations with prolonged or extensive exposure, such as between humans and their companion animals. Moreover, transmission events in a human to animal direction appear to be more common than the opposite, but the numbers are too small to draw firm conclusions.

We found highly diverging resistomes in SD isolates of human and animal origin, including both resistance genes and their associated MGEs, inferring limited genetic exchange between the host associated populations. Nevertheless, we observed an almost identical resistance element, *Tn5801*, in two SD isolates derived from a dog and a horse, respectively (Figure 4). Moreover, a highly similar element from a human associated *Streptococcus mitis* isolate was deposited in GenBank, strongly supporting the presence of genetic transfer between different ecological niches. However, the resistance MGEs detected in human associated SD isolates predominantly displayed similarities to MGEs derived from other β-hemolytic streptococcal isolates from humans, including *S. agalactiae* and *S. pyogenes*. Thus, the bulk of conjugative transfer and transduction of resistance determinants likely occurs within the boundaries of the ecological niche.

Reports on antimicrobial resistance in SD are quite sparse and limited to SD infecting humans and cattle. Studies on antimicrobial susceptibility in SD from the past decade demonstrate MLS resistance rates varying from 1 to 48% among bovine associated SD (Zhang et al., 2018; Duse et al., 2021) and from 17 to 42% among invasive, human associated strains (Park et al., 2019; Rojo-Bezares et al., 2021). Regarding tetracycline resistance, available data from the same period has shown resistance rates varying in the ranges 33–100% and 30–56% for bovine and human associated isolates, respectively (Traverso et al., 2016; Zhang et al., 2018; Park et al., 2019; Shen et al., 2021). Compared to these numbers, our findings indicate a relatively low frequency of resistant SD strains of both human and animal origin in Norway, possibly reflecting the strict policy regarding the use of antibiotics in both human and veterinary medicine in our country. European

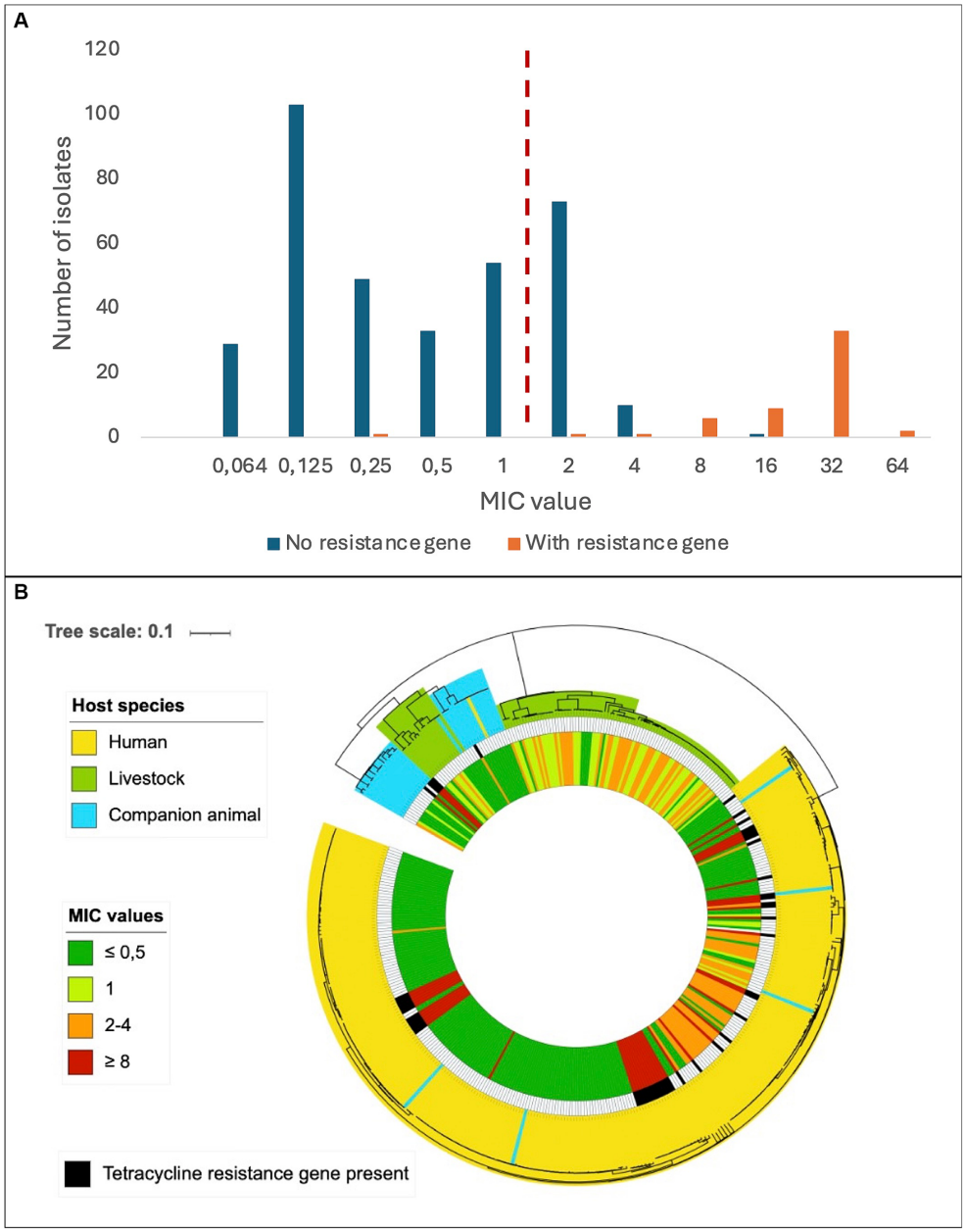

FIGURE 3
Distribution of tetracycline resistance in *Streptococcus dysgalactiae* isolates. **(A)** Minimum inhibitory concentrations of tetracycline among *Streptococcus dysgalactiae* isolates with or without a known tetracycline resistance gene. The dashed, red line represents the current EUCAST susceptibility breakpoint. **(B)** Phylogenetic tree of *Streptococcus dysgalactiae* isolates indicating host species (outer circle), presence of known tetracycline resistance gene (middle circle) and minimum inhibitory concentration of tetracycline (inner circle). Scale indicates substitutions per site.

surveillance data about veterinary antimicrobial consumption obtained from the European Medicines Agency shows an at least tenfold higher consumption of all antibiotics relevant to this study in most European countries compared to Norway (EMA, 2024).

We observed a substantial incongruence between phenotypic and genotypic susceptibility rates for tetracycline, particularly in bovine associated isolates. In January 2023, EUCAST lowered the MIC breakpoint for tetracycline resistance in SD by merging the "I" (susceptible, increased exposure) category into the "R" (resistant) group. This change had a great impact on our results, doubling the number of strains entering the resistant category relative to earlier versions of the EUCAST clinical breakpoint table. Notably, most of these low-grade resistant strains lacked identifiable validated resistance genes. A MIC distribution intersected by the current EUCAST breakpoint was also reported in a recent Scandinavian study examining oxytetracycline-susceptibility among 231 SD isolates of bovine origin (Jensen et al., 2024). They found a uniform distribution with a proposed tentative epidemiological cut off (TECOFF) of 8 mg/L, which is three dilution steps above the breakpoint. Retrospectively applying the current breakpoint to previously published reports on tetracycline susceptibility in bovine associated SD, a tetracycline MIC distribution encircling the novel

TABLE 2 Antimicrobial resistance genes and their associated mobile genetic elements among human (H) and animal (A) associated isolates of *Streptococcus dysgalactiae*.

Resistance			Mobile genetic element											
Phenotype	Gene	*N*, total	*Sp2905*		*Sa2603*		*Tn916*		*Tn6944*		*Tn5801*		Other	
			H	A	H	A	H	A	H	A	H	A	H	A
Tetracycline resistance	*tet*(M)	33					13	1	14		1	3	1[a]	
	tet(O)	18	15		1	2								
	tet(T)	1	1											
	tet(W)	1	1											
MLS resistance	*erm*(A)	23	23											
	erm(B)	4				1	1						1[b]	1[c]
	erm(T)	1											1[d]	
	lsa(C)	4	1	3										
	mef(A)	1	1											

Data presented as number of isolates harboring the resistance gene. MLS; macrolide, lincosamide and streptogramin antibiotics. [a]Unknown location. [b]Tn6218. [c]Bacteriophage. [d]Plasmid.

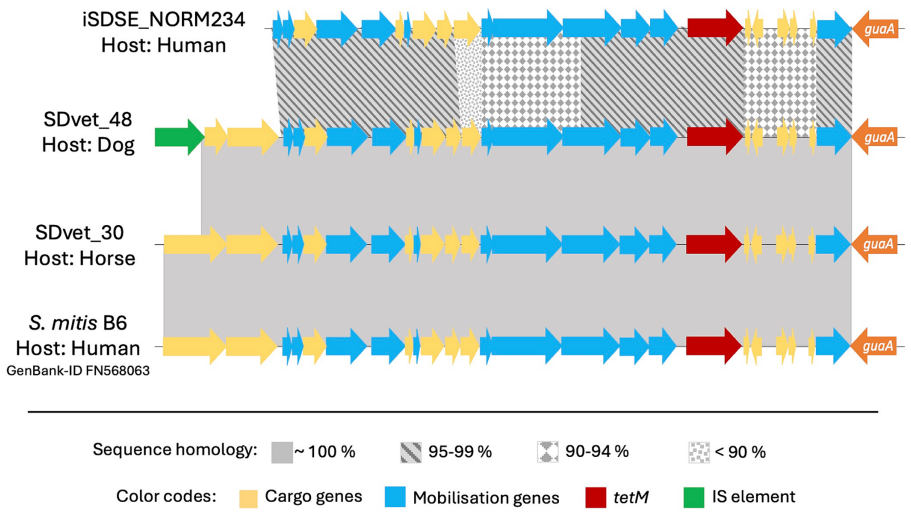

FIGURE 4
Comparative analyses of the *tet(M)* carrying mobile genetic element *Tn5801* in different bacteria. *Tn5801* was in all cases studied found integrated immediately downstream to the chromosomally located gene *guaA*. Among the *Streptococcus dysgalactiae* isolates included in this study, *Tn5801* was present in four: one from a human, one from a dog and two from horses. The *Tn5801* in the animal isolates had an inter-sequence homology close to 100%, while the human sequence was more divergent. This finding was not in line with the phylogenetic relationship of the bacterial isolates themselves, of which the isolate from the dog was more closely linked to the human isolate than to the horse isolates. By BLAST search we found *Tn5801* also to be present in other species than *S. dysgalactiae*, here represented by an element located in an invasive human isolate of *S. mitis*, which interestingly is a closer match to the mobile genetic element in *S. dysgalactiae* of animal than of human origin.

breakpoint was observed also in studies from Canada, New Zealand, and Europe, suggesting that this is a widespread feature in this species (McDougall et al., 2014; Cameron et al., 2016; Garch et al., 2020). In the present study a low-grade tetracycline resistant subpopulation was also evident among SD isolates of human origin (Figure 3B), indicating that this phenomenon is not limited to isolates of bovine origin, nor related to the distinction between the two subspecies SDSE and SDSD.

Breakpoints intersecting defined bacterial populations is generally avoided by EUCAST, as inherent analytic variations in the susceptibility testing makes the susceptible/resistant categorization unreliable. The observed high proportion of SD isolates displaying tetracycline MIC values encircling the breakpoint could either reflect a breakpoint poorly adapted to the SD wild type or be the result of a so far unrecognized mechanism of low-grade resistance. The trimodal distribution of the low-grade tetracycline resistant strains in our material could infer the latter, and the genetic basis for this phenomenon should be subjected to scrutiny.

In our study, we have included SD from a relatively large selection of host species, which entail a multifaceted base for comparative studies. The collection of strains from all hosts from within the same

temporal and geographical delimitated setting is also a strength of the study, enabling a real time comparison of SD in a One Health perspective.

A limitation of the study is the low prevalence of antimicrobial resistance in Norway, potentially underestimating the extent of cross-species transmission. Antimicrobial resistance in streptococci in a One Health perspective should therefore be explored also in regions with higher rates of antimicrobial resistance. Moreover, analyzing pooled resistance rates for companion animals and livestock does not reflect the phylogenetic diversity within the SD taxon, and could potentially obscure significant differences between these ecological niches. Another potential limitation is the use of MIC gradient strips instead of disc diffusion or broth dilution. Nevertheless, a large proportion of our isolates was also examined by disc diffusion, and the results were congruent (Supplementary Table S1). A potential confounder is the delimitation of included human associated strains to exclusively bloodborne isolates, whereas the animal associated strains predominantly are from non-invasive infections. Nevertheless, national surveillance data on antimicrobial susceptibility in non-invasive human associated isolates of SD from 2018 does not reveal major differences between invasive and non-invasive strains (NORM, 2019). Lastly, we only examined isolates from a confined temporal context, and dissemination of resistance traits over time could not be evaluated. Longitudinal collection of isolates from asymptomatic carriers in contemporary and spatially related animal and human populations would be interesting. However, the execution of such an investigation probably would entail ethical challenges regarding sampling procedure on healthy animals.

In conclusion, we found a phylogenetic delineation of SD strains in line with host adapted populations and niche specialization. Moreover, the resistome differed significantly between SD in these host associated groups both regarding the repertoire of circulating resistance genes and their associated mobile gene elements. Our findings indicate that direct transmission events of strains or genetic elements carrying resistance genes between SD from different ecological niches are rare in our geographic region.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found here: BioProject, accession number PRJEB74563, available at NCBI (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJEB74563).

Ethics statement

The studies involving humans were approved by Regional Committee for Medical Research Ethics in Western Norway (2021/63132). The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from a by-product of routine care or industry. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements. Ethical approval was not required for the study involving animals in accordance with the local legislation and institutional requirements because samples from animals were collected from sick animals for diagnostic purposes by veterinarians in clinical practice, which does not require ethical approval.

Author contributions

MG: Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. SS: Conceptualization, Funding acquisition, Writing – review & editing. AS: Writing – review & editing. BK: Conceptualization, Project administration, Writing – review & editing. AK: Writing – review & editing. CJ: Writing – review & editing. HJ: Writing – review & editing. OO: Conceptualization, Data curation, Investigation, Methodology, Project administration, Supervision, Writing – review & editing.

Members of the Norwegian Study Group on *Streptococcus dysgalactiae*

Aasmund Fostervold, Department of Clinical Microbiology, Stavanger University Hospital, Stavanger, Norway; Aleksandra Jakovljev, Department of Microbiology, St. Olav's University Hospital, Trondheim, Norway; Nadine Durema Pullar, Department of Clinical Microbiology, Vestre Viken Hospital Trust, Bærum, Norway; Åshild Marvik, Department of Microbiology, Vestfold Hospital Trust, Tønsberg, Norway; Einar Nilsen, Department of Microbiology, Møre and Romsdal Hospital Trust, Ålesund, Norway; Fredrik Müller, Department of Microbiology, Oslo University Hospital, Oslo, Norway; Ghantous Milad Chedid, Laboratory for Clinical Microbiology, Fonna Hospital Trust, Haugesund, Norway; Gunnar Skov Simonsen, Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, Norway; Elisabeth Sirnes, Department of Microbiology, Division of Medicine, District General Hospital of Førde, Førde, Norway; Roar Magne Bævre-Jensen, Department of Clinical Microbiology, Vestre Viken Hospital Trust, Drammen, Norway; Sandra Åsheim, Unit for Clinical Microbiology, Norland Hospital Trust, Bodø, Norway; Ståle Tofteland, Department of Clinical Microbiology, Sørlandet Hospital Trust, Agder, Norway; Rolf-Arne Sandnes, Department of Clinical Microbiology, Innlandet Hospital Trust, Lillehammer, Norway; Truls Michael Leegaard, Department of Microbiology and Infection Control, Akershus University Hospital, Lørenskog, Norway.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by grants from the Norwegian surveillance program for antimicrobial resistance (NORM) (Grant number: 19/08, to OO and

19/16 to CJ) and Center for Laboratory Medicine at Østfold Hospital Trust. The genomic Core Facility (GCF) at the University of Bergen, which is part of the NorSeq consortium, provided services on the whole genome sequencing in this study; GCF is supported by major grants from the Research Council of Norway (grant no. 245979/F50) and Trond Mohn Foundation (grant no. BFS2017TMT04/BFS2017TMT08). This study was received also financial contributions from the Western Norway Regional Health Authority (grant no. 912231). SS received grants from the Norwegian Research Council under the frames of NordForsk (Project: 90456, PerAID) and ERA PerMed (Project: 2018-151, PerMIT).

Acknowledgments

We thank the staff at Microbiological Department at Haukeland University Hospital for excellent technical assistance and for providing access to their laboratory facilities. We acknowledge Haima Mylvaganam for her significant contribution to streptococcal research at Haukeland University Hospital during the past two decades, and for inspiration to continued scientific exploration.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2024.1423762/full#supplementary-material

References

Abdelsalam, M., Chen, S., and Yoshida, T. (2010). Phenotypic and genetic characterizations of *Streptococcus dysgalactiae* strains isolated from fish collected in Japan and other Asian countries. *FEMS Microbiol. Lett.* 302, 32–38. doi: 10.1111/j.1574-6968.2009.01828.x

Acke, E., Midwinter, A., Lawrence, K., Gordon, S., Moore, S., Rasiah, I., et al. (2015). Prevalence of *Streptococcus dysgalactiae* subsp. equisimilis and *S. equi* subsp. zooepidemicus in a sample of healthy dogs, cats and horses. *N. Zealand Vet. J.* 63, 265–271. doi: 10.1080/00480169.2015.1016133

Alves-Barroco, C., Brito, P. H., Santos-Sanches, I., and Fernandes, A. R. (2022). Phylogenetic analysis and accessory genome diversity reveal insight into the evolutionary history of *Streptococcus dysgalactiae*. *Front. Microbiol.* 13:952110. doi: 10.3389/fmicb.2022.952110

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Brandt, C. M., and Spellerberg, B. (2009). Human infections due to *Streptococcus dysgalactiae* subspecies equisimilis. *Clin. Infect. Dis.* 49, 766–772. doi: 10.1086/605085

Cameron, M., Saab, M., Heider, L., McClure, J. T., Rodriguez-Lecompte, J. C., and Sanchez, J. (2016). Antimicrobial susceptibility patterns of environmental streptococci recovered from bovine Milk samples in the maritime provinces of Canada. *Front. Vet. Sci.* 3:79. doi: 10.3389/fvets.2016.00079

Ciszewski, M., Zegarski, K., and Szewczyk, E. M. (2016). *Streptococcus dysgalactiae* subsp. equisimilis isolated from infections in dogs and humans: are current subspecies identification criteria accurate? *Curr. Microbiol.* 73, 684–688. doi: 10.1007/s00284-016-1113-x

Couture-Cossette, A., Carignan, A., Mercier, A., Desruisseaux, C., Valiquette, L., and Pépin, J. (2018). Secular trends in incidence of invasive beta-hemolytic streptococci and efficacy of adjunctive therapy in Quebec, Canada, 1996-2016. *PLoS One* 13:e0206289. doi: 10.1371/journal.pone.0206289

Duse, A., Persson-Waller, K., and Pedersen, K. (2021). Microbial Aetiology, antibiotic susceptibility and pathogen-specific risk factors for udder pathogens from clinical mastitis in dairy cows. *Animals* 11:2113. doi: 10.3390/ani11072113

EMA (2024). *European surveillance of veterinary antmicrobial consumption (ESVAC): 2009–2023*. Available at: https://www.ema.europa.eu/en/veterinary-regulatory-overview/antimicrobial-resistance-veterinary-medicine/european-surveillance-veterinary-antimicrobial-consumption-esvac-2009-2023 (Accessed April 26, 2024).

Florensa, A. F., Kaas, R. S., Clausen, P. T. L. C., Aytan-Aktug, D., and Aarestrup, F. M. (2022). ResFinder – an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microb. Genom.* 8:000748. doi: 10.1099/mgen.0.000748

Garch, F. E., Youala, M., Simjee, S., Moyaert, H., Klee, R., Truszkowska, B., et al. (2020). Antimicrobial susceptibility of nine udder pathogens recovered from bovine clinical mastitis milk in Europe 2015–2016: VetPath results. *Vet. Microbiol.* 245:108644. doi: 10.1016/j.vetmic.2020.108644

Jensen, V. F., Damborg, P., Norström, M., Nonnemann, B., Slettemeås, J. S., Smistad, M., et al. (2024). Estimation of epidemiological cut-off values for eight antibiotics used for treatment of bovine mastitis caused by Streptococcus uberis and *Streptococcus dysgalactiae* subsp. dysgalactiae. *Vet. Microbiol.* 290:109994. doi: 10.1016/j.vetmic.2024.109994

Kaci, A., Jonassen, C. M., Skrede, S., Sivertsen, A., Steinbakk, M., Norwegian Study Group on Streptococcus dysgalactiaeet al. (2023). Genomic epidemiology of *Streptococcus dysgalactiae* subsp. equisimilis strains causing invasive disease in Norway during 2018. *Front. Microbiol.* 14:1171913. doi: 10.3389/fmicb.2023.1171913

Koh, T. H., Sng, L.-H., Yuen, S. M., Thomas, C. K., Tan, P. L., Tan, S. H., et al. (2009). Streptococcal cellulitis following preparation of fresh raw seafood. *Zoonoses Public Heal.* 56, 206–208. doi: 10.1111/j.1863-2378.2008.01213.x

León-Sampedro, R., Novais, C., Peixe, L., Baquero, F., and Coque, T. M. (2016). Diversity and evolution of the Tn 5801-tet (M)-like integrative and conjugative elements among Enterococcus, Streptococcus, and Staphylococcus. *Antimicrob. Agents Chemother.* 60, 1736–1746. doi: 10.1128/aac.01864-15

Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Lother, S. A., Demczuk, W., Martin, I., Mulvey, M. R., Dufault, B., Lagacé-Wiens, P., et al. (2017). Clonal clusters and virulence factors of group C and G Streptococcus causing severe infections, Manitoba, Canada, 2012–2014. *Emerg. Infect. Dis.* 23, 1079–1088. doi: 10.3201/eid2307.161259

Maeda, T., Tsuyuki, Y., Fujita, T., Fukushima, Y., Goto, M., Yoshida, H., et al. (2020, 2019). Comparison of *Streptococcus agalactiae* isolates from humans and companion animals reveals genotypic and phenotypic differences. *Jpn. J. Infect. Dis.* 73:441. doi: 10.7883/yoken.jjid.2019.441

Matuszewska, M., Murray, G. G. R., Harrison, E. M., Holmes, M. A., and Weinert, L. A. (2020). The evolutionary genomics of host specificity in *Staphylococcus aureus*. *Trends Microbiol.* 28, 465–477. doi: 10.1016/j.tim.2019.12.007

McDougall, S., Hussein, H., and Petrovski, K. (2014). Antimicrobial resistance in *Staphylococcus aureus*, Streptococcus uberis and *Streptococcus dysgalactiae* from dairy cows with mastitis. *N. Zealand Vet. J.* 62, 68–76. doi: 10.1080/00480169.2013.843135

McEwen, S. A., and Collignon, P. J. (2018). Antimicrobial resistance: a one health perspective. *Microbiol. Spectr.* 6. doi: 10.1128/microbiolspec.arba-0009-2017

Moreno, L. Z., da Costa, B. L. P., Matajira, C. E. C., Gomes, V. T. M., Mesquita, R. E., Silva, A. P. S., et al. (2016). Molecular and antimicrobial susceptibility profiling of *Streptococcus dysgalactiae* isolated from swine. *Diagn. Microbiol. Infect. Dis.* 86, 178–180. doi: 10.1016/j.diagmicrobio.2016.07.020

Nevanlinna, V., Huttunen, R., Aittoniemi, J., Luukkaala, T., and Rantala, S. (2023). Incidence, seasonal pattern, and clinical manifestations of *Streptococcus dysgalactiae* subspecies equisimilis bacteremia; a population-based study. *Eur. J. Clin. Microbiol. Infect. Dis.* 42, 819–825. doi: 10.1007/s10096-023-04607-8

NORM (2019). NORM/NORM-VET 2018. Usage of Antimicrobial Agents and Occurrence of Antimicrobial Resistance in Norway. Tromsø/Oslo 2019. Available at: https://www.unn.no/4a79d7/siteassets/documents/kompetansetjenester--sentre-og-fagrad/norm---norsk-overvakingssystem-for-antibiotikaresistens-hos-mikrober/rapporter/norm-norm-vet-2018.pdf

Oppegaard, O., Glambek, M., Skutlaberg, D. H., Skrede, S., Sivertsen, A., and Kittang, B. R. (2023). *Streptococcus dysgalactiae* bloodstream infections, Norway, 1999–2021. *Emerg. Infect. Dis.* 29, 260–267. doi: 10.3201/eid2902.221218

Oppegaard, O., Mylvaganam, H., Skrede, S., Lindemann, P. C., and Kittang, B. R. (2017). Emergence of a *Streptococcus dysgalactiae* subspecies equisimilis st G62647-lineage associated with severe clinical manifestations. *Sci. Rep.* 7:7589. doi: 10.1038/s41598-017-08162-z

Park, J. H., Jung, J., Kim, M. J., Sung, H., Kim, M.-N., Chong, Y. P., et al. (2019). Incidence, clinical characteristics, and outcomes of *Streptococcus dysgalactiae* subspecies equisimilis bacteremia in a tertiary hospital: comparison with *S. agalactiae* bacteremia. *Eur. J. Clin. Microbiol. Infect. Dis.* 38, 2253–2258. doi: 10.1007/s10096-019-03667-z

Pinho, M. D., Erol, E., Ribeiro-Gonçalves, B., Mendes, C. I., Carriço, J. A., Matos, S. C., et al. (2016). Beta-hemolytic *Streptococcus dysgalactiae* strains isolated from horses are a genetically distinct population within the *Streptococcus dysgalactiae* taxon. *Sci. Rep.* 6:31736. doi: 10.1038/srep31736

Porcellato, D., Smistad, M., Skeie, S. B., Jørgensen, H. J., Austbø, L., and Oppegaard, O. (2021). Whole genome sequencing reveals possible host species adaptation of Streptococcusdysgalactiae. *Sci. Rep.* 11:17350. doi: 10.1038/s41598-021-96710-z

Rebelo, A. R., Ibfelt, T., Bortolaia, V., Leekitcharoenphon, P., Hansen, D. S., Nielsen, H. L., et al. (2022). One day in Denmark: Nationwide point-prevalence survey of human bacterial isolates and comparison of classical and whole-genome sequence-based species identification methods. *PLoS One* 17:e0261999. doi: 10.1371/journal.pone.0261999

Rojo-Bezares, B., Toca, L., Azcona-Gutiérrez, J. M., Ortega-Unanue, N., Toledano, P., and Sáenz, Y. (2021). *Streptococcus dysgalactiae* subsp. equisimilis from invasive and non-invasive infections in Spain: combining epidemiology, molecular characterization, and genetic diversity. *Eur. J. Clin. Microbiol. Infect. Dis.* 40, 1013–1021. doi: 10.1007/s10096-020-04119-9

Schrieber, L., Towers, R., Muscatello, G., and Speare, R. (2014). Transmission of *Streptococcus dysgalactiae* subsp. equisimilis between child and dog in an aboriginal Australian community. *Zoonoses Public Heal.* 61, 145–148. doi: 10.1111/zph.12057

Shen, J., Wu, X., Yang, Y., Lv, Y., Li, X., Ding, X., et al. (2021). Antimicrobial resistance and virulence factor of *Streptococcus dysgalactiae* isolated from clinical bovine mastitis cases in Northwest China. *Infect. Drug Resist.* 14, 3519–3530. doi: 10.2147/idr.s327924

Shinohara, K., Murase, K., Tsuchido, Y., Noguchi, T., Yukawa, S., Yamamoto, M., et al. (2023). Clonal expansion of multidrug-resistant *Streptococcus dysgalactiae* subspecies equisimilis causing bacteremia, Japan, 2005–2021. *Emerg. Infect. Dis.* 29, 528–539. doi: 10.3201/eid2903.221060

Smistad, M., Tollersrud, T. S., Austbø, L., Porcellato, D., Wolff, C., Asal, B., et al. (2021). Molecular detection and genotype characterization of *Streptococcus dysgalactiae* from sheep flocks with outbreaks of infectious arthritis. *Vet. Microbiol.* 262:109221. doi: 10.1016/j.vetmic.2021.109221

Traverso, F., Blanco, A., Villalón, P., Beratz, N., Nieto, J. A. S., Lopardo, H., et al. (2016). Molecular characterization of invasive *Streptococcus dysgalactiae* subsp. equisimilis. Multicenter study: Argentina 2011–2012. *Rev. Argent. Microbiol.* 48, 279–289. doi: 10.1016/j.ram.2016.07.001

UK Health Security Agency (2022). Laboratory surveillance of pyogenic and non-pyogenic streptococcal bacteraemia in England: 2021 update. *Health Protect. Rep.* 16:13.

Vieira, V. V., Teixeira, L. M., Zahner, V., Momen, H., Facklam, R. R., Steigerwalt, A. G., et al. (1998). Genetic relationships among the different phenotypes of *Streptococcus dysgalactiae* strains. *Int J Syst Evol Micr* 48, 1231–1243. doi: 10.1099/00207713-48-4-1231

Zhang, S., Piepers, S., Shan, R., Cai, L., Mao, S., Zou, J., et al. (2018). Phenotypic and genotypic characterization of antimicrobial resistance profiles in *Streptococcus dysgalactiae* isolated from bovine clinical mastitis in 5 provinces of China. *J. Dairy Sci.* 101, 3344–3355. doi: 10.3168/jds.2017-14031

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 20 June 2024
DOI 10.3389/fmicb.2024.1416665

OPEN ACCESS

EDITED BY
Axel Cloeckaert,
Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE), France

REVIEWED BY
Khald Blau,
University of Applied Sciences Emden/Leer, Germany
Nozomu Obana,
University of Tsukuba, Japan

*CORRESPONDENCE
Shan Goh
✉ s.goh5@herts.ac.uk

†These authors share first authorship

‡These authors have contributed equally to this work and share last authorship

§PRESENT ADDRESS
Amer Nubgan,
Department of Biology,
College of Science, University of Baghdad,
Baghdad, Iraq

RECEIVED 12 April 2024
ACCEPTED 04 June 2024
PUBLISHED 20 June 2024

CITATION
Hussain H, Nubgan A, Rodríguez C, Imwattana K, Knight DR, Parthala V, Mullany P and Goh S (2024) Removal of mobile genetic elements from the genome of *Clostridioides difficile* and the implications for the organism's biology.
Front. Microbiol. 15:1416665.
doi: 10.3389/fmicb.2024.1416665

Removal of mobile genetic elements from the genome of *Clostridioides difficile* and the implications for the organism's biology

Haitham Hussain[1†], Amer Nubgan[2†§], César Rodríguez[3], Korakrit Imwattana[4,5], Daniel R. Knight[4,6], Valerija Parthala[2], Peter Mullany[1‡] and Shan Goh[2*‡]

[1]Department of Microbial Diseases, Eastman Dental Institute, University College London, London, United Kingdom, [2]Department of Clinical, Pharmaceutical and Biological Sciences, University of Hertfordshire, Hatfield, United Kingdom, [3]Facultad de Microbiología and Centro de Investigación en Enfermedades Tropicales (CIET), Universidad de Costa Rica, San José, Costa Rica, [4]School of Biomedical Sciences, The University of Western Australia, Perth, WA, Australia, [5]Department of Microbiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Salaya, Thailand, [6]Department of Microbiology, PathWest Laboratory Medicine WA, Queen Elizabeth II Medical Centre, Nedlands, WA, Australia

Clostridioides difficile is an emerging pathogen of One Health significance. Its highly variable genome contains mobile genetic elements (MGEs) such as transposons and prophages that influence its biology. Systematic deletion of each genetic element is required to determine their precise role in *C. difficile* biology and contribution to the wider mobilome. Here, Tn*5397* (21 kb) and ϕ027 (56 kb) were deleted from *C. difficile* 630 and R20291, respectively, using allele replacement facilitated by CRISPR-Cas9. The 630 Tn*5397* deletant transferred PaLoc at the same frequency (1×10^{-7}) as 630 harboring Tn*5397*, indicating that Tn*5397* alone did not mediate conjugative transfer of PaLoc. The R20291 ϕ027 deletant was sensitive to ϕ027 infection, and contained two unexpected features, a 2.7 kb remnant of the mutagenesis plasmid, and a putative catalase gene adjacent to the deleted prophage was also deleted. Growth kinetics of R20291 ϕ027 deletant was similar to wild type (WT) in rich medium but marginally reduced compared with WT in minimal medium. This work indicates the commonly used pMTL8000 plasmid series works well for CRISPR-Cas9-mediated gene deletion, resulting in the largest deleted locus (56.8 kb) described in *C. difficile*. Removal of MGEs was achieved by targeting conjugative/integrative regions to promote excision and permanent loss. The deletants created will be useful strains for investigating Tn*5397* or ϕ027 prophage contribution to host virulence, fitness, and physiology, and a platform for other mutagenesis studies aimed at functional gene analysis without native transposon or phage interference in *C. difficile* 630 and R20291.

KEYWORDS

***C. difficile*, prophage deletion, transposon deletion, CRISPR-Cas9, site-specific recombinase**

1 Introduction

Clostridioides difficile, also known as *Clostridium difficile* (Lawson et al., 2016), is a Gram-positive, anaerobic, endospore-forming bacterium that causes gastrointestinal illness. It is a leading cause of antibiotic-associated diarrhea (He et al., 2013; Smits et al., 2016), and recurrent *C. difficile* infection (CDI) is difficult to treat with antibiotics alone (van Prehn et al., 2021). It is an important nosocomial and community-acquired pathogen worldwide (Magill et al., 2014; Collins et al., 2020; Finn et al., 2021; Viprey et al., 2022). Sources of infection include community spaces, environmental water and soil, animals, and the food chain (Candel-Pérez et al., 2019; Knight and Riley, 2019; Jo et al., 2022). Genetically similar strains from pigs and humans have been reported, indicating possible zoonotic or anthropogenic transmission (Knetsch et al., 2014; Moloney et al., 2021).

Clostridioides difficile has a highly variable genome with up to 30% being made up of mobile genetic elements (MGEs) (Sebaihia et al., 2006). These are very common in bacteria and can loosely be defined as any genetic element that can mediate its own transfer from one part of the genome to another. *C. difficile* contains a range of MGEs from the very simple, such as insertion sequences (IS) to complex integrative conjugative elements (ICE, also sometimes referred to as conjugative transposons) and integrated phage genomes called prophage (reviewed in Roberts et al., 2014). These MGEs can have a profound effect on the biology of *C. difficile*. For example, ICE often encode resistance to antibiotics; e.g., Tn*916,* and Tn*5397* (tetracycline resistance). A large ICE, 023_CTnT found in *C. difficile* clade 3 strains contains genes encoding a sortase, putative sortase substrates, lantibiotic ABC transporters and a putative siderophore biosynthetic cluster (Shaw et al., 2019). Similar genes are found throughout the gut microbiome indicating that ICE have a role in allowing organisms to adapt to their local environment and can transfer through the gut microbiome. *C. difficile* also contains integrative mobilizable elements such as Tn*4453*a/b (chloramphenicol resistance), and Tn*5398* (erythromycin resistance), which spread via conjugation, between and beyond *C. difficile* (Roberts et al., 2014). Furthermore, bioinformatic analysis of *C. difficile* ICE show that they contain different sigma factors implying that they can have a global role in gene expression in the organism (Brouwer et al., 2011). Prophage and ICE are modular MGEs and both typically enter the host bacterial genome via the activity of site-specific recombinases (Johnson and Grossman, 2015). These belong to two different families namely serine or tyrosine. The amino acid named referring to that responsible for cutting DNA at the active site (Johnson and Grossman, 2015). Comparison of ICE and phage show further relationships indicating that they can exchange modules and have an intertwined evolutionary history (Johnson and Grossman, 2015).

Lysogeny is frequently observed in *C. difficile* (Sebaihia et al., 2006; Ramirez-Vargas et al., 2018), with prophages most commonly belonging to the order Siphoviridae and Myoviridae, and most commonly identified with ϕC2 (Goh et al., 2005), ϕMMP04 (Meessen-Pinard et al., 2012), ϕCD119 (Govind et al., 2006), ϕCDHM1 (Hargreaves et al., 2014), ϕCD38-2 (Fortier and Moineau, 2007), and ϕCD27 (Mayer et al., 2008), ranging in size from 31 to 56 kb with a GC content similar to that of the *C. difficile* genome (28–30%) (Knight et al., 2015). *C. difficile* prophages can influence host toxin regulation (Goh et al., 2005; Govind et al., 2009, 2011; Sekulovic et al., 2011; Riedel et al., 2017), quorum sensing (Hargreaves et al., 2014), biofilm formation (Slater et al., 2019) and fitness including transduction (Goh et al., 2013), phage immunity (Boudry et al., 2015; Sekulovic et al., 2015; Li et al., 2020), and plasmid/prophage maintenance (Peltier et al., 2020). Some of these studies were carried out by infecting *C. difficile* with a phage of interest and examining changes to the transcriptome or selected phenotype. However due to the presence of prophages in the studied strains, it can be difficult to attribute changes solely to the infecting phage.

To prove the role of ICE and prophage in *C. difficile* biology it is necessary to make clean scarless deletions of these large genetic elements. The best understood *C. difficile* ICE is Tn*5397,* which encodes resistance to tetracycline and is capable of broad host range transfer within several Gram-positive organisms (Wang et al., 2006). Tn*5379* translocates between strains by excising and forming a circular molecule, which is then capable of conjugal transfer to a suitable recipient or reintegration into the host genome (Supplementary Figure S1). In *C. difficile* 630 this element integrates into the genome close to the region of the chromosome which encodes the major virulence factors of the organism toxins A and B, termed the PaLoc (Brouwer et al., 2013). The later element can transfer at low frequency to non-toxigenic *C. difficile* strains converting them to toxin producers. The PaLoc does not have any genes which are obviously involved in its own transfer, so it was proposed that one of the *C. difficile* ICE mediated its transfer via a mechanism like Hfr in *E. coli* (Brouwer et al., 2013).

Two studies have deleted a prophage from *C. difficile* 630, lysogenic for two inducible prophages, ϕCD630-1 and ϕCD630-2 (Sebaihia et al., 2006; Goh et al., 2007). Hong *et al* deleted ϕCD630-2 using CRISPR-Cpf1 (Hong et al., 2018), and Peltier *et al* deleted ϕCD630-2 using a toxin-antitoxin system to select for double crossover mutants (Peltier et al., 2020). R20291 is a hypervirulent epidemic *C. difficile* strain that is well-characterized, and its genome was predicted to contain a complete prophage genome (Stabler et al., 2009), ϕ027, shown to spontaneously excise from the bacterial chromosome and circularize to exist extrachromosomally (Sekulovic and Fortier, 2015). ϕ027 has not been characterized as a functional phage, perhaps due to the lack of a suitable indicator strain for phage infection.

C. difficile possesses the type I-B CRISPR-Cas system utilizing several Cas proteins (Boudry et al., 2015; Andersen et al., 2016), which is different to the commonly used type II system utilizing a Cas9 or Cas12a protein. The use of heterologously expressed type II systems in *C. difficile* could avoid native type I-B CRISPR-Cas systems interference, which was successfully re-programmed for gene deletion in both *C. difficile* 630Δ*erm* and R20291 (Maikova et al., 2019). Indeed, native CRISPR-Cas systems of *C. difficile* 630 and R20291 (Boudry et al., 2015) did not interfere with several reports of successful gene deletants selected using Cas9 or Cas12a (McAllister et al., 2017; Hong et al., 2018; Wang et al., 2018; Ingle et al., 2019) expressed from pMTL8000 plasmids (Heap et al., 2009). In this study, we used a similar strategy of CRISPR-Cas9 to select deletants of Tn*5397* from *C. difficile* 630 and ϕ027 prophage from R20291(Stabler et al., 2009), allowing their contribution to *C. difficile* biology to be determined.

2 Materials and methods

2.1 Bacterial strains and growth conditions

The *C. difficile* and *Escherichia coli* strains used in this study are listed in Table 1. All bacterial strains were stored at −80°C in their

TABLE 1 Bacterial strains and plasmids used in this study.

Organism/plasmid	Relevant features[1]	Source and reference
C. difficile R20291	Ribotype (RT) 027, MLST sequence type (ST) 1 (76) (Bletz et al., 2018), toxinotype III, lysogen of ϕ027 prophage (Stabler et al., 2009)	Brendan Wren, LSHTM CRG2021 lineage (Monteford et al., 2021)
C. difficile NCTC11207	RT 001, ST3, toxinotype 0, susceptible to ϕ027 infection	Melinda Mayer, Quadram Institute. Genome sequence determined in this study. BioProject PRJNA993731, Accession number CP129979.
C. difficile CD37	RT 009, ST03, non-toxigenic, Tet^S Erm^S Rif^R	Smith et al. (1981) and Gawlik et al. (2015)
C. difficile 630	RT 012, ST54, toxinotype 0, Tet^R Erm^R Rif^S	Sebaihia et al. (2006), Monot et al. (2011), and Riedel et al. (2015)
630Δ*erm tcdB*::*erm*(B)	*C. difficile* 630Δ*erm*::Δ*tcdB,* contains wild type Tn*5397*, Tet^R Erm^R Rif^S	Kuehne et al. (2010)
630Δerm::Δ*Conj*	630Δ*erm*::Δ*tcdB* containing a 5 kb deletion of the Tn*5397* conjugation region, Tet^R Erm^R Rif^S	This study
630Δerm::ΔTn*5397*	630Δ*erm*::Δ*Conj* that had lost Tn*5397*, Tet^S Erm^R Rif^S	This study
630Δerm::*tcdB*$^-$ Δ*Conj*	Transconjugant from the mating of *630Δerm*::ΔTn*5397* and CD37. Tet^S Erm^R Rif^R	This study
68P10-23/68P10-30	*C. difficile* R20291Δϕ027, Tm^S prophage deletants	This study
NEB® 10-beta	High efficiency competent *E. coli* DH10β from NEB. Δ(*ara-leu*) 7697 *ara*D139 *fhuA* Δ*lacX74 gal*K16 *gal*E15 *e14-* ϕ80*dlacZ* M15 *rec*A1 *rel*A1 *end*A1 *nup*G *rps*L (StrR) rph spoT1 Δ(*mrr-hsd*RMS-*mcr*BC)	New England Biolabs, UK
NEB® 5-alpha	Competent *E. coli* DH5α derivative from NEB. *fhuA2*Δ(*argF-lacZ*)U169 *phoA glnV44* Φ80Δ(*lacZ*)M15 *gyrA96 rec*A1 *rel*A1	New England Biolabs, UK
E. coli CA434	HB101 carrying the IncP conjugative plasmid R702	Williams et al. (1990) and Purdy et al. (2002)
pMTL83151	replicon of pCB102, *catP*, *colE1*, *traJ*	Heap et al. (2009)
pPM100	pMTL83151 modified to include P_{xyl}/tetO, Cas9, P_{tetM}, gRNA scaffold, *lacZ*. This is the basic modular vector that can be used to manipulate *C. difficile* (Figure 1).	This study
pPM101	pPM100 containing sequences encoding gRNA targeting region B (Figures 1, 2).	This study
pPM102	pPM100 containing sequences encoding gRNA targeting region C (Figures 1, 2).	This study
pPM103	Contains LHA and RHA1 (Figure 2) cloned into pPM101	This study
pPM104	Contains RHA1 and RHA2 (Figure 2) cloned into pPM102	This study
pAN721	gRNA_1040 targeting the coding strand nt 76..95 of ϕ027 integrase gene (CDR20291_1415) with a PAM of tgg, cloned into pPM100	This study
pAN821	1 kb (HA1) cassette of homology arms consisting of 500 bp of sense sequences flanking the ϕ027 integrase gene (CDR20291_1415) of the circularized phage genome, cloned into pAN721	This study

[1]Antibiotics are abbreviated as follows: erythromycin (Erm), thiamphenicol (Tm), tetracycline (Tet).

respective medium [brain heart infusion broth (BHIB, Neogen, UK) for *C. difficile* and Luria-Bertani (LB, Neogen, UK) for *E. coli*] with 20% (v/v) glycerol. *C. difficile* agar cultures were freshly prepared weekly from −80°C stocks on Brazier's agar (Neogen, UK) supplemented with 1% defibrinated horse blood (Thermo Scientific, UK), 250 μg/mL cycloserine and 8 μg /mL cefoxitin (Merck, UK) incubated anaerobically (Don Whitley DG250: 10% H_2, 5% CO_2, 85% N_2) at 37°C for 2–3 days. *C. difficile* broth cultures were prepared from agar cultures either in BHI, BHI supplemented with 5 g/L yeast extract (Oxoid, UK) and 0.1% L-cysteine (Merck, UK) (BHIS), or BHIS supplemented with the following antibiotics/inducer when appropriate: thiamphenicol (Tm, 15 μg/mL, Merck, UK), D-cycloserine (250 μg/mL) and kanamycin (50 μg/mL, Merck UK), and incubated 16–18 h or BHIS agar supplemented with the following antibiotics/inducer when appropriate: 7% defibrinated horse blood, Tm (15 μg/mL), D-cycloserine (250 μg/mL) and kanamycin (50 μg/mL), and incubated 2–3 days. Log phase cultures were prepared from 1 mL of 16–18 h cultures in 10 mL pre-reduce BHIB incubated anaerobically for 4 h at 37°C. *E. coli* NEB® 5 -alpha or NEB® 10-beta (New England Biolabs or NEB, UK) was used as the general host for plasmid construction and gene cloning. *E. coli* CA434 was used as the donor for conjugation with *C. difficile*. Transformation of *E. coli* was carried out by heat-shock at 42°C for either 45 s (*E. coli* CA434) or 30 s (*E. coli* NEB® 5 -alpha or NEB® 10-beta), and transformants were selected on LB agar plates (Difco, UK) supplemented with 25 μg/mL chloramphenicol (Biological Life Science USA), and grown in LB broth (Neogen, UK) with 12.5 μg/mL chloramphenicol.

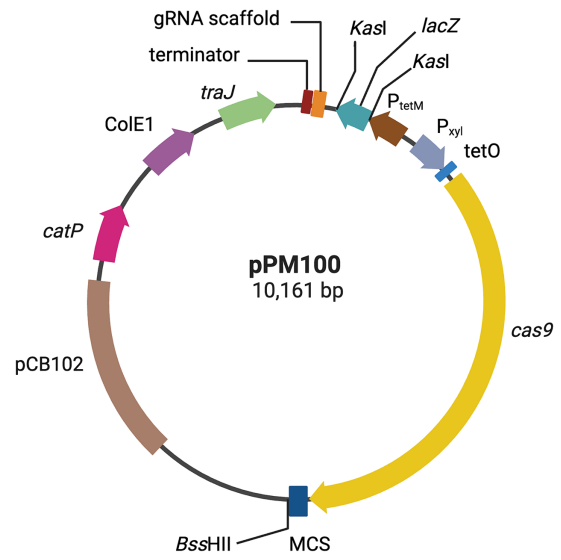

FIGURE 1
CRISPR-Cas9 genome editing vector. Plasmid map representation of the CRISPR-Cas9 vector developed in this study, pPM100. The *S. pyogenes cas9*, is under the control of the Pxyl-tetO promoter; a gRNA scaffold, consisting of a gRNA handle under the control of *C. difficile* P_{tetM} promoter, and a multiple cloning site (MCS) for cloning of an editing template, containing upstream and downstream chromosomal regions flanking a deletion target site. The editing regions are cloned into the MCS (see text for more details). Created with biorender.com.

Growth curves of *C. difficile* R20291 and R20291Δϕ027 were generated from $OD_{600\,nm}$ readings over 18 h in 96-well plates. Microtiter plates, sealing films, BHIB and *C. difficile* minimal medium (CDMM) (Cartman and Minton, 2010) were pre-reduced before use. Bacterial cultures in BHIB of 18 h were anaerobically diluted in growth media to $OD_{600\,nm}$ of 0.1, and 200 µL volumes distributed into triplicate wells (3 technical repeats). Growth media alone were similarly distributed, serving as blanks and negative controls. The 96-well microtiter plate was sealed anaerobically, then transferred to a microtiter plate reader set at 37°C and kinetic measurements taken for 18 h every 15 min after 5 s of agitation. The experiment was repeated four times from which average $OD_{600\,nm}$ values and standard deviation were calculated in Microsoft Excel and plotted in Prism 10 (GraphPad).

2.2 Phage induction, propagation, and purification

To induce ϕ027, known to exist within *C. difficile* R20291 (Stabler et al., 2009), a 16–18 h culture of *C. difficile* R20291 in 10 mL BHIB was treated with 3 µg/mL of mitomycin C (Merck, UK) for 6 h at 37°C. The induced culture was centrifuged at 4500 x g for 15 min and the supernatant was filtered through a sterile 0.45 µm membrane filter (Fisher Scientific, UK). Plaque assays were carried out in anaerobe basal agar (Oxoid, UK) with 600 µL of 4 h log phase cultures of NCTC11207 in BHIB and 100 µL of R20291 filtrate as previously described (Goh et al., 2005). A no-phage control was included with every plaque assay to ensure no spontaneously induced prophages from NCTC11207 were co-cultivated with ϕ027. Two rounds of single plaque propagation were carried out on NCTC11207, followed by whole plate assays for phage propagation to obtain crude phage suspensions as previously described (Goh et al., 2005). Crude phage suspensions were treated with DNase I (2 U/µL, Merck, UK) and RNase A (10 µg/mL, Merck, UK), precipitated by 1 M NaCl and 10% w/v PEG 8000, and recovered with chloroform to yield semi-pure suspensions. These were then either purified through a pre-formed CsCl density gradient of 1.3 g/mL, 1.5 g/mL, and 1.7 g/mL at 60,000 x *g* for 2 h at 4°C (Sorvall WX 80+ Ultracentrifuge, AH 650 swing out rotor) and dialyzed as previously described (Sambrook and Russell, 2001a) to yield purified suspensions, or concentrated by ultrafiltration through Amicon Ultra-15 3 kDa MWCO Centrifugal Filter Devices (Merck, Germany) spun at 4,000 × *g* for 60 min at room temperature.

2.3 Bacterial, phage, and plasmid DNA extraction

Five milliliters of a 16–18 h *C. difficile* broth culture was pelleted and frozen at −20°C before genomic DNA was extracted using the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich, UK). Phage DNA was extracted from semi-purified or purified dialyzed phage suspensions using either phenol:chloroform:isoamyl alcohol (Invitrogen, UK) (Sambrook and Russell, 2001b) or Phage DNA Isolation kit (Norgen Biotek, Canada). Plasmid DNA was extracted from 1 mL of 16–18 h *E. coli* broth cultures grown aerobically with agitation at 200 rpm using either the plasmid miniprep kit (Qiagen, UK) or Monarch Plasmid Miniprep kit (NEB UK).

2.4 PCR of phage, and phage and bacterial attachment sites in *Clostridioides difficile* R20291

Primer sets LCF889/LCF 890 and LCF 887/LCF 888 from Sekulovic and Fortier (2015) were used to confirm phage attachment site for ϕ027 in R20291 (Supplementary Table S1). Six primer sets specific for ϕ027 *orf 1415, 1416, 1417, 1418, 1419,* and *1464a* were used to confirm the presence of the prophage in R20291 (Supplementary Table S1). OneTaq DNA polymerase and reaction buffer (NEB UK) were used for PCR reactions according to cycling conditions recommended by the manufacturer.

2.5 Construction of plasmids for ICE and prophage deletion

A modular vector pPM100, i.e., where desired DNA modules can be inserted in a single step, was constructed (Table 1 and Figure 1). The starting point was the *E. coli–C. difficile* shuttle vector pMTL83151, which contains the origin of replication from plasmid pCB102 (this replicon is unstable in *C. difficile* allowing it to be used as conditional lethal vector), along with the *catP* selective marker, the ColEI replicon and the *mob* region from RK2 (Heap et al., 2009). The Cas9 gene cassette of *Streptococcus pyogenes* from pCas9 (Addgene, UK) and the inducible promotor $P_{xyl/tetO}$ from pRPF185 (Fagan and Fairweather, 2011) were used as templates for PCR utilizing Pxyl/tetO-F with Pxyl/tetO-R and Cas9-F with Cas9-R primer pair, respectively. The two fragments were fused

10.3389/fmicb.2024.1416665

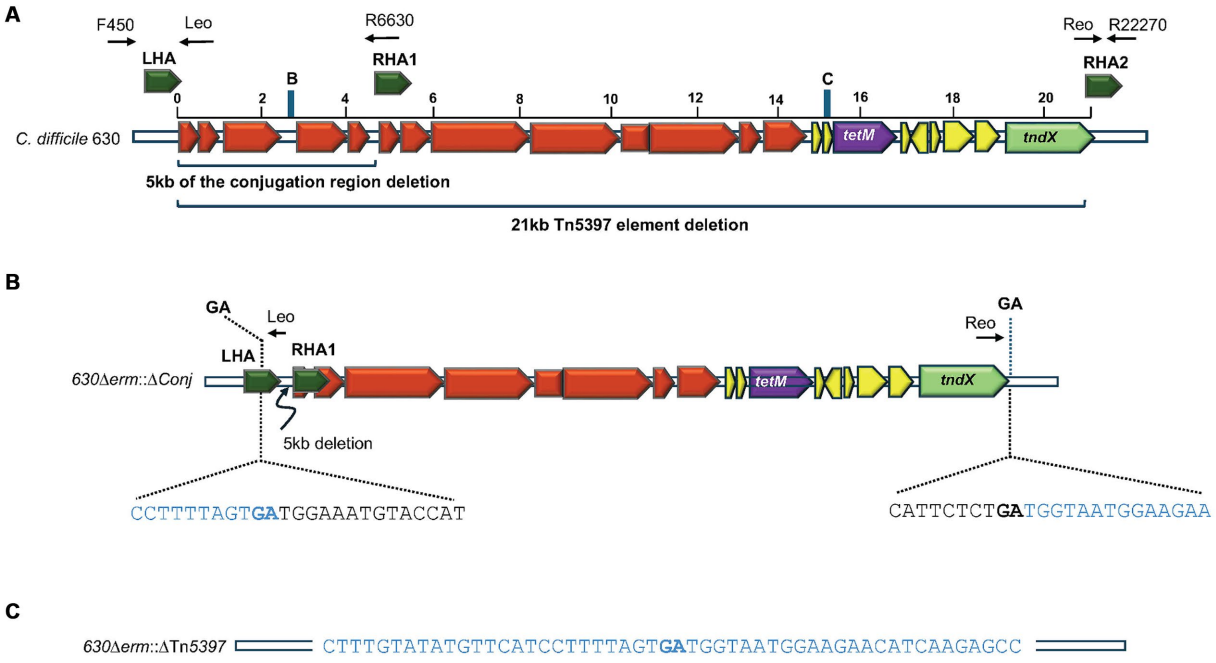

FIGURE 2
Deletion of Tn*5397* from *C. difficile* 630. **(A)** Schematic representation of *C. difficile* Tn*5397* conjugative transposon sequence. Black arrows above the figure represent binding sites of primers used in PCR analysis, the sequences of these are shown in Supplementary Table S1. The conjugation region genes are shown in red, the *tetM* gene is shown in purple, regulatory genes are shown in yellow, excision/insertion (*tndX*) shown in green. Brackets indicate the 5 kb conjugation region deleted from Tn*5397,* and the whole 21 kb Tn*5397.* Homology arms flanking regions for deletion are in deep green. Blue bars represent the positions of 20-nt sgRNA retargeting sequences within the region targeted for deletion from Tn*5397* labelled B and C, corresponding to the sequences in Supplementary Table S2. **(B)** *630Δerm*::Δ*Conj* carrying the 5 kb deletion and sequence at the ends of the remaining element. Bacterial sequences are in blue, Tn*5397* sequences are in black, GA in bold is the junction between the element. **(C)** Regenerated DNA sequences of *630Δerm*::ΔTn*5397* after deletion of whole element.

using splicing by overlap extension (SOE) PCR utilizing Pxyl/tetO-F and Cas9-R primer pair. The Pxyl/tetO-Cas9 fragment was cloned into pMTL83151 upstream of the Fdx terminator between *Xma*I and *Sal*I restriction enzyme (Thermo Scientific, UK) sites resulting in plasmid pPM100. For deleting Tn*5397* from *C. difficile* 630, the sgRNA encoding fragment was synthesized by Thermo Fisher Scientific and consisted of: the strong Tn*916* derived promoter (P_{tetM}) (Su et al., 1992) and a 20 bp gRNA targeting sequence that was selected using an algorithm for scoring and ranking potential target sites with the Benchling CRISPR design tool[1] (Supplementary Table S2). The sgRNA fragment was annealed by heating for 5 min then cloned into pMTL83151-CRISPR-Cas9 (pPM100) upstream of CD0164 terminator between *Xma*I and *Not*I sites. The editing regions were amplified by PCR using two pairs of primers (Supplementary Table S1) to produce fragments homologous to sequences targeted for recombination, and they were cloned next to the multiple cloning site (Figures 1, 2). Individual editing fragments were then fused together by SOE PCR at the *Bss*HII site resulting in plasmid (pPM103 and pPM104) respectively (Table 1 and Figure 2).

For deleting ϕ027 from *C. difficile* R20291, a gRNA sequence targeting the region of interest identified by Benchling was chosen to which pairs of self-complementary oligos (Supplementary Tables S1, S2) were annealed at 50 pmol/μL in annealing buffer (10 mM Tris pH 8, 50 mM NaCl, 1 mM EDTA pH 8), phosphorylated with 15 U T4 polynucleotide kinase (NEB, UK), ligated to *Kas*I-linearized (10 U of SspD1, Thermo Scientific, UK) and gel-extracted plasmid (Monarch kit, NEB, UK) with 5 U of T4 DNA ligase (Thermo Scientific, UK), then transformed by chemically competent NEB® 5-alpha cells. Cloned gRNA was confirmed by PCR using OneTaq DNA polymerase (NEB, UK) and Sanger sequencing (Source Biosciences Ltd., UK) with primers pMTL83151bb_99 and PtetM_191 (Supplementary Table S1). NEBuilder® HiFi DNA Assembly Tool designed primers for amplification and assembly of homology arms (Supplementary Table S1). Phusion High-Fidelity DNA Polymerase (Thermo Scientific, UK) amplified 0.5 kb sequences flanking the ϕ027 integrase gene with primer pairs int_RLA_rev/phi027_1415_LHA_R and int_RLA_fwd/ phi027_1415_RHA_F. Amplicons were cloned with NEBuilder® HiFi DNA Assembly Master Mix (NEB) into *Bss*HII linearized and gel-extracted pAN721. Assembled plasmids were transformed by heat-shock to NEB® 10-beta to generate pAN821. Cloned inserts were Sanger sequenced (Source Biosciences Ltd., UK) with primer pairs pHHCas9_HACS_F/R, and 14152HA_pwalk1/2 (Supplementary Table S1).

2.6 Mating experiments

2.6.1 Filter-mating experiments between *Clostridium difficile* 630 and CD37

To test for PaLoc and Tn*5397* transfer, methods described in Brouwer et al. (2013) were used with the following modifications.

1 https://www.benchling.com/

C. difficile was grown in BHIB for 18 h anaerobically, then subcultured to fresh broth at 37°C until mid-exponential phase ($OD_{600\,nm}$ of 0.45). Cultures of *C. difficile* 630 (Tet^R Erm^R Rif^S) or 630Δ*erm tcdB*::*erm*(B) (Tet^R Erm^R Rif^S) donors and CD37 (Tet^S Erm^S Rif^R) recipient were mixed, and 200 µL was spread on nitrocellulose 0.45 µm pore-size filters on BHI agar and incubated for 18 h at 37°C in an anaerobic environment. The filters were removed from the agar plates and placed in 20 mL bottles and vigorously washed with 1 mL BHIB. Aliquots (100 µL) were spread on BHI agar supplemented with either 10 µg/mL tetracycline, 100 µg/mL erythromycin, 15 µg/mL thiamphenicol, or (25 µg/mL) rifampicin and incubated anaerobically for 48 h. Putative transconjugants were subcultured on fresh selective plates and incubated for a further 48 h. Selection of transfer of Tn*5397* from 630Δ*erm tcdB*::*erm*(B) to CD37 was made on plates containing rifampicin and tetracycline. Transfer of the PaLoc from 630Δ*erm tcdB*::*erm*(B) and strains that had lost Tn*5397,* or contained a deletion of part of the conjugation region, was made on plates containing erythromycin and rifampicin.

2.6.2 Transfer of plasmids from *Escherichia coli* to *Clostridium difficile*

Single colonies of *E. coli* CA434 containing mutagenic plasmids were grown anaerobically overnight at 37°C in pre-reduced BHIS with 12.5 µg/mL chloramphenicol, 1 mL was pelleted and washed in pre-reduced BHIS. Two hundred microliters of overnight *C. difficile* culture was heated to 52°C for 5 min, cooled for 2 min, (the heating step was only required when R20291 was the recipient) then mixed with *E. coli* donor cell pellets and incubated for 8 h as described previously (Kirk and Fagan, 2016). The mating mixture was spotted onto BHIS agar, grown anaerobically for 8 h, harvested in 1 mL pre-reduced phosphate-buffered saline (PBS) and plated onto BHIS with 250 µg/mL cycloserine, 50 µg/mL kanamycin, and 15 µg/mL Tm (CKTm). After 24–48 h of growth, colonies were picked and transfer of mutagenic plasmid pPM103, pPM104 or pAN821 was confirmed by PCR (see Supplementary Table S1 and the results section for more details). *C. difficile* transconjugants were re-streaked onto BHIS agar with appropriate antibiotics. After 2 days, a single colony was inoculated into pre-reduced 10 mL BHIS supplemented with appropriate antibiotics and grown overnight for gDNA extraction for further PCR confirmation. Conjugation frequency was calculated against either donor or recipient in mating mixtures. Donor and recipient cultures were serially diluted 10-fold in pre-reduced PBS and plated onto LB and BHI plates, respectively. Colonies were counted after 24 h anaerobic incubation for *E. coli* and 48 h incubation for *C. difficile*. The conjugation frequency was calculated as colony forming units (CFU) of transconjugants/CFU of donor or recipients.

2.7 Induction of CRISPR-Cas9 system and MGE deletion

Clostridioides difficile containing mutagenic plasmid on BHIS CKTm and 7% defibrinated horse blood plates were grown in 10 mL BHIB with Tm (15 µg/mL) for 16–18 h, serially diluted 10-fold in pre-reduced sterile 1 x PBS, spread-plated onto dried and pre-reduced BHIS + anhydrotetracycline (aTC, Merck, UK, 30 ng/mL) + Tm (15 µg/mL) plates and grown for 2 days. Five to 10 colonies were screened by colony PCR for ϕ027 prophage or Tn*5397* deletion. For prophage deletion primer pairs phiR2_1415_F/phiR2_1,415_R, phi027_1464a_F/phi027_1464a_R, LCF887/LCF889, LCF888/LCF890 and catP_2/3 were used (Supplementary Table S1). To confirm Tn*5397* conjugation region deletion, primers Tn*5397*(F450) and Tn*5397*(R6630), and Tn*5397*(R450) and Tn*5397*(R22270) were used to determine if the whole element was absent (Supplementary Table S1).

2.8 Plasmid curing

To cure the plasmid from strains with mutant or deleted target MGE (ϕ027 or Tn*5397*), a single colony of the mutant with the desired deletion (i.e., deletant) was subcultured in BHIB with no antibiotics. After 18 h, 100 µL of the culture was used to inoculate 10 mL of BHIB. This subculture was repeated daily for up to 10 days. Ten-fold serial dilutions of deletant culture in pre-reduced 1 x PBS were made after each subculture for Tn*5397* deletants, but only at the end of 10 days for ϕ027 deletants, and 100 µL of the 10^{-5} dilution was spread onto BHI agar plates without antibiotic. Replica plating was performed on agar supplemented with 15 µg/mL Tm, tetracycline (Tet) or erythromycin (Erm) to identify colonies that have lost the mutagenesis plasmid (i.e., either pPM103, pPM104 or pAN821). Colonies that had lost antibiotic resistance at the least number of subcultures were isolated for further study to avoid excessive subculture. Loss of the plasmid was confirmed by PCR using plasmid-specific primers (Cas9-F and Cas9-R, or catP_2/3) (Supplementary Table S1). For Tm sensitive (Tm^S) prophage deletants putatively cured of plasmid, gDNA was extracted and checked with 1464a_F/R, NF1643/44, and catP_2/3 primers, then sequenced by Illumina sequencing. To screen for further loss of a truncated Cas9 plasmid remnant in sequenced Tm^S prophage deletant strains, single colonies were picked for colony PCR using primers 68P1023_LF/LR, specific for the left junction of the integrated plasmid remnant (Supplementary Table S1). Putatively negative colonies were grown in BHIB for gDNA extraction and confirmation by PCR with the same primers in addition to 68P1023_RF/RR, and pHHCas9_3F/end.

2.9 Genomic DNA library preparation for Nanopore sequencing

For long-read Oxford Nanopore Technology (ONT, UK) sequencing of NCTC11207, 5 mL of 18 h culture was pelleted and frozen at −20°C before DNA extraction using the Qiagen MagAttract High Molecular Weight DNA Kit (Qiagen, UK). DNA quality and quantity were assessed using NanoDrop, Qubit (Thermo Fisher Scientific, UK), and Agilent TapeStation instruments (Agilent, UK). ONT sequencing libraries were prepared by multiplexing DNA from *C. difficile* isolates per flow cell using a Nanopore protocol for native barcoding of genomic DNA (version NBE_9065_v109_revAC_14Aug2019). This firstly involved DNA repair and end-prep carried out with NEBNext FFPE DNA Repair Mix (M6630, NEB, UK), NEBNext Ultra II End repair/dA-tailing module (E7546, NEB, UK), and AMPure XP beads (Beckman Coulter, UK). Secondly, native barcode ligation was carried out using the Native Barcoding Expansion kit (EXP-NBD104; ONT, UK) and NEB Blunt/TA Ligase Master mix (M0367, NEB, UK). Thirdly, adapter ligation using Ligation Sequencing Kit (SQK-LSK109, ONT, UK), NEBNext Quick T4 DNA Ligase (E6057, NEB UK), NEBNext Quick Ligation Reaction

Buffer (B6058, NEB, UK). Sequencing libraries were loaded onto a R9 generation flow cell (FLO-MIN106) and sequenced in MinION Mk1C (ONT, UK), stopping after 25 h.

2.10 Nanopore sequence analysis of NCTC11207

Before assembly, long read sequences were filtered using Filtlong v0.2.0, keeping the minimum length of 1,000 bp and 90% of best quality sequences[2]. Genome assembly was performed using Flye assembler v2.9 (Kolmogorov et al., 2019) and Trycycler v0.5.0 (Wick et al., 2021), with a standard protocol recommended by the developers (Wick et al., 2021). The final assembly graph was visualized and polished with Bandage v0.8.1 (Wick et al., 2015). Multi-Locus Sequence Typing was performed *in silico* using BIGSdb v1.32.0 hosted at PubMLST[3] and the scheme of Griffiths et al. (2010). Prophage screening of the NCTC11207 chromosome was performed using PHASTER (Arndt et al., 2016). The final circular genome was annotated using the NCBI Prokaryotic Genomes Annotation Pipeline (PGAP) (Tatusova et al., 2016) and is now available in GenBank under BioProject PRJNA993731 (accession number CP129979). Similarity of the ϕ027 genome to genomes of NCTC11207 prophages 1 and 2 was determined using VIRIDIC (Moraru et al., 2020).

2.11 Illumina sequencing of ϕ027 deletants and deletion/insertion confirmations

Clostridioides difficile gDNA was extracted using either GenElute Bacterial kit (Merck, UK) or Qiagen MagAttract High Molecular Weight DNA Kit (Qiagen, UK), checked for quality, and paired-end sequenced at Microbes NG (UK, 2 × 250 bp, 30 x coverage) or SeqCenter (USA, 2 × 151 bp, 30 x coverage). Trimmed Illumina reads facilitated by these vendors were mapped against the genome of *C. difficile* R20291 (accession number NZ_CP029423.1). Unmapped reads were afterwards *de novo* assembled using Unicycler v0.4.8 and mapped against pAN821 using the BWA-MEM algorithm (arXiv:1303.3997v2). To detect ORF 1465 in deletant and WT by PCR, primers LF1 and RR1 (Supplementary Table S1) were used on two batches of genomic DNA prepared from WT and deletant cultures as described in 2.3. OneTaq DNA polymerase and reaction buffer (NEB UK) were used for PCR reactions according to cycling conditions recommended by the manufacturer.

3 Results

3.1 Construction of a modular vector for gene knock out in *C. difficile*

A simple modular vector, pPM100, was constructed for generating CRISPR-directed mutations in *C. difficile* (see Materials and Methods and Figure 1). This plasmid is unstable in *C. difficile* and therefore an ideal delivery vector. Furthermore, all the modules on this vector can be easily replaced or modified making it a useful and efficient tool for relatively easy manipulation of *C. difficile*.

3.2 Editing regions designed to delete the conjugation region of Tn*5397* resulted in a mixture of clones some of which had lost just the conjugation region and some the whole of Tn*5397*

To investigate the role of Tn*5397* in genome mobility and to generate a Tet sensitive (Tet^S) derivative of 630Δ*erm tcdB*::*erm*(B) [this has a Clostron insertion conferring Erm resistance (Erm^R) in the *tcdB* gene to allow for the selection of PaLoc transfer (Kuehne et al., 2011; Brouwer et al., 2013)], we initially wanted to precisely delete the whole of Tn*5397*. This was attempted by generating a CRISPR-Cas9 vector (pPM103) with gRNA encoding sequences targeting region C of Tn*5397* (Figure 2A and Supplementary Table S2) and editing regions flanking the insertion site of Tn*5397* in 630Δ*erm tcdB*::*erm*(B) (LHA and RHA2, Figures 1, 2A). This plasmid was conjugated into 630Δ*erm tcdB*::*erm*(B) and the resulting four Tm-resistant (Tm^R) transconjugants were subject to Cas9 induction and then screened by PCR for loss of Tn*5397* using primers flanking Tn*5397*. All four transconjugants still had Tn*5397*. It was assumed that the reason for this failure was that the region we were trying to delete is too large. Therefore, it was decided to delete part of the conjugation region. To do this, the CRISPR-Cas9 vector (pPM104) containing a gRNA encoding region targeting region B at 2500 bp on Tn*5397* and an editing region consisting of LHA and RHA1 was used (Figure 2A and Supplementary Table S2). Plasmids were transferred by conjugation to 630Δ*erm tcdB*::*erm*(B), and five Tm^R colonies arose. These were subject to PCR with primers F450 and R6630 and these yielded a product of 1.5 kb (no product was obtained with strains carrying wild-type Tn*5397* presumably because the product was too large) (Figure 2A). DNA sequence analysis of this product confirmed that a precise 5 kb deletion of part of the conjugation region had occurred (Figure 2B and Supplementary Figure S2).

One of the strains, *630Δerm*::Δ*Conj,* carrying the 5 kb deletion was selected for further study. It was grown for 18 h in drug free broth then plated onto drug free media, 600 colonies were screened for loss of resistance to Tet, and 2 of 600 were sensitive, hence the mutation efficiency was 0.3%.

PCR analysis of the two tetracycline-sensitive mutants using primers flanking Tn*5397* (F450 and R22270 in Figure 2A) gave a product of 1.2 kb (no product was obtained in strains carrying wild-type Tn*5397* or those carrying the 5 kb deletion). One of these strains was selected for further study and designated *630Δerm*::ΔTn*5397*. DNA sequencing of the PCR product showed that the target site of Tn*5397* had been regenerated (Figure 2C) and that the whole of the transposon had been lost. This implies that deletion of part of the conjugation region destabilizes Tn*5397* so that it can still excise from the host chromosome and circularize but presumably due to the large deletion some circular molecules are lost. This idea was supported by the fact that we could detect the presence of a circular form of the element using primers Tn*5397* (Leo) and Tn*5397* (Reo) in the *630Δerm*::Δ*Conj* mutants that contained the 5 kb deletion. These

2 https://github.com/rrwick/Filtlong

3 https://pubmlst.org/

primers read out from the ends of Tn*5397* and will only form a product when the ends are ligated together in a circular form of the element (Supplementary Figure S1 and Figure 2A). Diagrams showing these events in wild-type Tn*5397* have been previously published and are summarized in Supplementary Figure S1 (Wang et al., 2000; Wang and Mullany, 2000; Brouwer et al., 2011). No product was obtained from the tetracycline sensitive strains.

3.3 Strains that have lost Tn*5397* can still transfer the PaLoc at the same frequency as WT

Tn*5397* is the nearest MGE to the PaLoc in the 630Δ*erm tcdB*::*erm*(B) chromosome (Brouwer et al., 2013) and it was proposed that this element might be responsible for its mobilization. However, mutants that contain a deletion of the conjugation region (*630*Δ*erm*, Δ*Conj*) and those that have lost Tn*5397* completely (*630*Δ*erm*, ΔTn*5397*) both transfer the PaLoc at the same frequency of around 1×10^{-7} of erm^R transconjugants per donor (encoded by the Clostron inserted in the *tcdB* gene) as the WT contains an intact Tn*5397*. This shows that a genetic element other than Tn*5397* is responsible for PaLoc transfer, although we cannot completely rule out a role for this element.

3.4 Φ027 is a functional phage integrated in R20291

ϕ027 was first identified as a putative prophage integrated into the chromosome of *C. difficile* R20291 (GenBank accession numbers FN545816.1 and CP029423.1) (Stabler et al., 2009). This phage and its bacterial attachment sites, *attP* and *attB,* were later identified by PCR (Sekulovic and Fortier, 2015), as there was a population of spontaneously excised and re-circularized ϕ027 genomes in DNA preparations of the lysogen. In this study, we firstly confirmed R20291 was PCR positive for 6 predicted genes of ϕ027 (Supplementary Figure S3). Then we found that ϕ027 in *C. difficile* R20291 of CRG2021 lineage (i.e., closest to the original R20291 clinical isolate and less amenable to conjugation) (Monteford et al., 2021) is a functional and inducible phage that can be propagated on *C. difficile* NCTC11207 (GenBank accession CP129979), obtaining yields of 10^8–10^9 plaque forming units (pfu)/mL. NCTC11207 is a ribotype (RT) 001 strain (Table 1), which was sequenced here and predicted to contain two intact prophages whose features are summarized in Supplementary Table S3. To ensure that NCTC11207 prophages were not co-propagated with ϕ027, a phage buffer (i.e., no phage) control was included with every batch of propagated phage to ensure no plaques were derived from spontaneously induced NCTC11207 prophages. ϕ027 virion DNA was extracted and used for PCR to confirm the *attP* sequence 5′ tattacaacttaagtaaata 3′, is as previously found in circularized ϕ027 prophage DNA within R20291 (Sekulovic and Fortier, 2015). The linear phage DNA annotation is re-arranged to convention in Supplementary Figure S4 and Supplementary Table S4. We also obtained similar PCR results when using R20291 bacterial genomic DNA, confirming previous observations that ϕ027 spontaneously excises, and exists extra-chromosomally and as an integrated prophage located at nt. 1670843…1,726,837 encompassing CDS CDR20291_1415 to CDR20291_1464a (Genbank accession number FN545816.1 and Supplementary Figure S4; Sekulovic and Fortier, 2015).

As ϕ027 prophage spontaneously excises and circularizes within host cells, we hypothesized that removal of the integrase gene (CDR20291_1415) from the circular form would lead to loss of the phage as the circular form would not be able to reintegrate (Figure 3).

The conjugation frequencies of pPM100 (the plasmid backbone), pAN721 (targeting integrase and not containing homology arms), and pAN821 (targeting integrase and containing the 1 kb homology arms for integrase deletion) are shown in Supplementary Table S5. Conjugation frequencies of cells harboring pPM100 and pAN721 were comparable to those of cells harboring pAN821, indicating *cas9* expression is likely to be repressed in the absence of the inducer and did not affect cell viability.

3.5 Generation of *Clostridioides difficile* R20291 ΔΦ027

Four of 8 transconjugants containing pAN821 with intact sgRNA, *cas9* and deletion cassette (Supplementary Figure S5) were devoid of prophage after aTC induction of Cas9 (Figure 4), could not be induced by mitomycin C to form plaques, and were susceptible to reinfection by ϕ027. The mutation efficiency was 50%. Curing of pAN821 from prophage deletants was attempted by passaging in non-selective BHIB for 10 days, then screening colonies for loss of susceptibility to Tm. Cultures containing control plasmids pPM100 (the plasmid backbone) and pAN721 (targeting integrase and not containing homology arms) were also screened for plasmid curing in the same way (Table 2). Just 1.5% (2/137) of colonies from the culture containing pAN821 had lost plasmid-encoded resistance after 10 passages. In contrast, plasmids lacking homology arms were all rapidly cured in this time from R20291 (Table 2). It is possible the homology arms allowed pAN821 to survive by recombining with the host genome. The actual mechanism for this requires further investigation. Susceptibility of two cured prophage deletants to ϕ027 infection was confirmed (Figure 5).

3.6 Whole-genome sequencing confirmed Φ027 deletion

Two Tm^S prophage deletants, 68P10-23 and 68P10-30, were subject to Illumina sequencing and had identical sequences, having the phage attachment site but neither the ϕ027 prophage genome nor the entire pAN821. Also ORF 1465 (678 bp), which was downstream of the *attR* site, hence predicted to be a bacterial gene (Figure 3), was deleted. Essentially a 56.8 kb locus containing the prophage and ORF 1465, was removed at the expected locations of nt. 1670843..1726837 (Figures 6A,B). However, 48–52% of the sequence reads contained a 2.7 kb remnant of pAN821 where the prophage was previously integrated (Figure 6C). The 2.7 kb remnant of pAN821 aligned with 1767 of 4,107 bases of the 3′ end of *cas9*, a 262 nt downstream intergenic region, 500 nt of the RHA, and 204 nt of the LHA (Figure 6D). Its presence was confirmed by PCR in all of 254 single colonies of 68P10-23 tested, and all of 160 single colonies of 68P10-30 tested (results not shown). To see if ORF 1465 could be absent from the WT genome naturally due to prophage excision, PCR and Sanger sequencing was

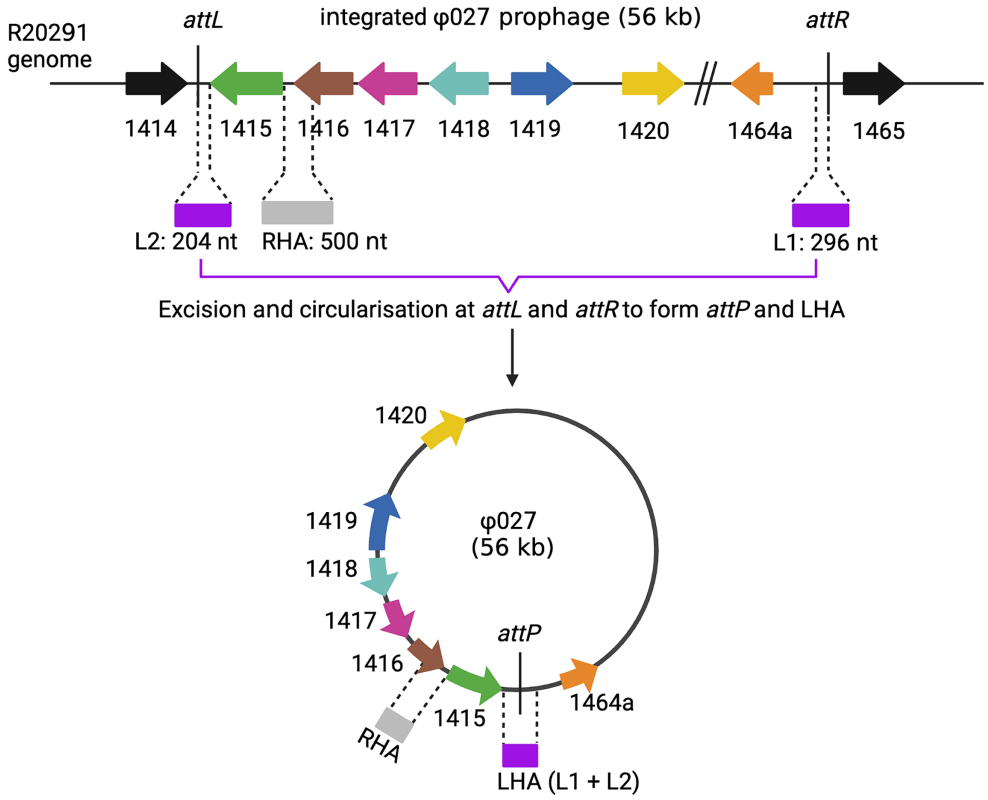

FIGURE 3
Homology arms for prophage deletion. The top line shows the integrated form of ϕ027, the junction between phage and bacterial DNA is shown at *attL* and *attR*. The phage can excise and circularize by site-specific recombination between *attL* and *attR* to generate a circular molecule with *attP* at the joint of the circular form. The location of homology arms that were used in mutagenesis experiments are shown. Note that LHA is only present at the joint of the circular form as it is composed of *attP*. Created with biorender.com.

carried out to show that deletion of ORF 1465 was likely a consequence of the mutagenic plasmid (Supplementary Figure S6). There was no other genomic difference between the prophage deletants and the WT. Batch culture growth curves of R20291Δϕ027 (68P10-23) was similar to WT in rich medium but marginally reduced compared with WT in minimal medium (Figure 7).

4 Discussion

Large MGEs such as prophages and ICE use integrase enzymes to facilitate their entry and exit from the host chromosome. There are two general families, i.e., the tyrosine and the serine recombinases. The former requires an accessory protein, Xis to excise the MGE while the serine recombinases can mediate both integration and excision (Stark, 2014). In this work, we showed that a prophage could be deleted from the *C. difficile* genome by targeting the integrase and that the ICE Tn*5397* can be cleanly removed by targeting the conjugation region. These observations show that it should be possible to specifically remove any of the large integrated MGEs from the *C. difficile* genome and determine their contribution to the organism's biology. In this work, we ruled out a direct role for Tn*5397* in the transfer of the PaLoc. However, *C. difficile* does contain many different ICE and phages which have the potential to mediate chromosomal transfer, and systematic deletion of each of these is required to determine their precise role in the organism's biology and their contribution to the wider mobilome.

Tn*5397* and the closely related genetic element Tn*916* are both very stable in bacterial genomes. The rate of loss of Tn*5397* being much less than the 2 in 600 observed in this work; we have tested 3,000 colonies containing wild-type Tn*5397* with no loss of the element (unpublished data). Therefore, it is likely that the deletion of part of the conjugation region destablises the element. There have only been a small number of studies examining gene regulation in the Tn*916/5397* family in genetic elements and these have all been done in Tn*916* (Scornec et al., 2017). This work has shown that transcription is tightly regulated and expression of the conjugation region requires transcription initiating at the strong *tet*(m) promoter progressing over the joint of the circular form into the conjugation region. Our mutant that lacks part of the conjugation region still contains the ends of the element on which TndX can act (explaining why circular forms of the element are still detected). It is possible that the deletion of part of the conjugation region results in premature transcriptional termination and that not enough TndX is produced to allow reintegration, hence the element is lost. However more work is required to determine the exact mechanism for loss of Tn*5397*.

Our hypothesis for transfer of the PaLoc is that integrated origins of transfers (*oriT*) result in the mobilization of the chromosome from donor to recipient (Brouwer et al., 2013). As Tn*5397* was the nearest *oriT* to the PaLoc this seemed like a good candidate for mobilizing the

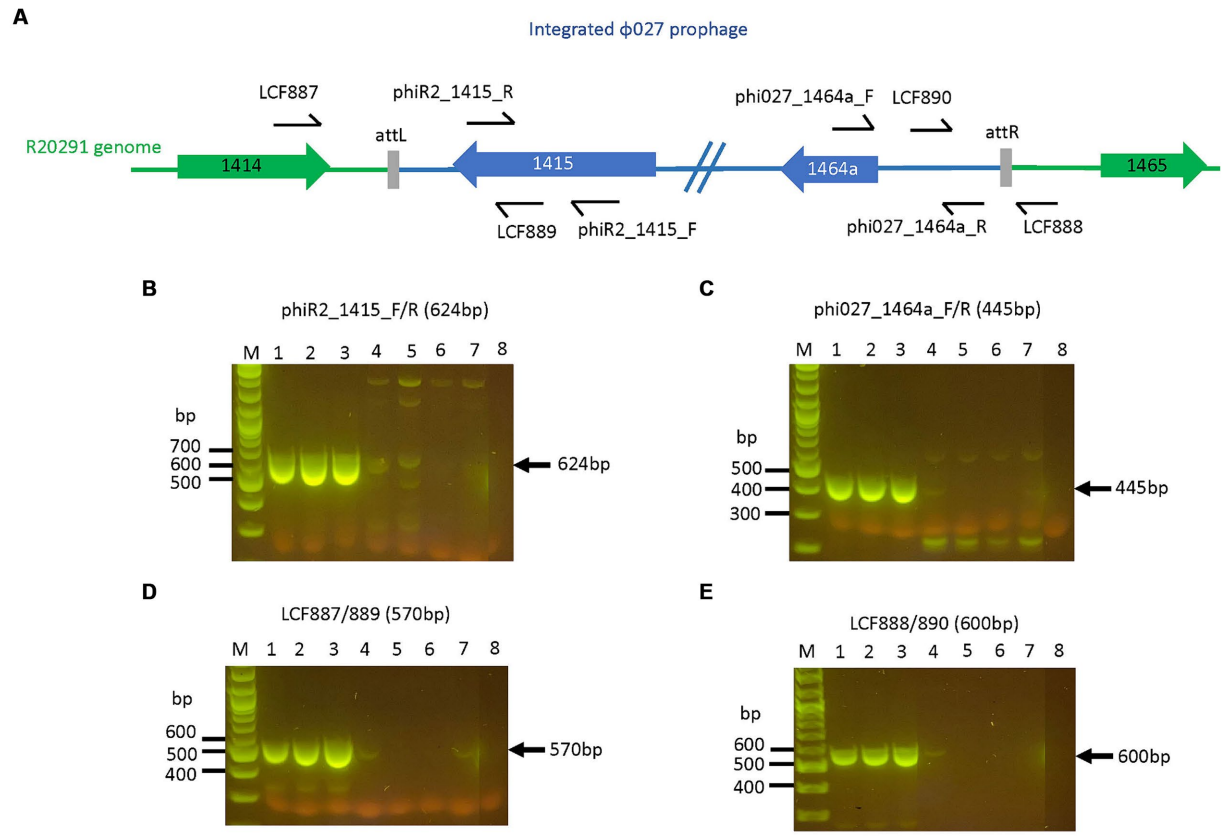

FIGURE 4
R20291 prophage deletants devoid of prophage features. Primers targeting selected features of the prophage are shown in **(A)**. The loss of **(B)** prophage *orf 1415* (integrase), **(C)** *orf 1464a*, **(D)** *attL*, and **(E)** *attR* were observed in R20291 deletants. Lane 1: WT R20291, 2: transconjugant of pPM100, 3: transconjugant of pAN721, 4–7: transconjugants of pAN821, 8: No template control. M: 1 kb Plus DNA ladder (NEB, UK).

TABLE 2 Colonies screened for loss of mutagenesis plasmid.

Transconjugants	Growth on replica plates		Tm^S	*catP* positive
	BHI	BHITm[1]		
pPM100	105	0	105	0
pAN721	130	0	130	0
pAN821	137	135	2	135

[1]Tm is thiamphenicol.

PaLoc. However, as the PaLoc still transferred from strains lacking Tn*5397*, this element is obviously not required for PaLoc transfer. The *C. difficile* genome does contain a number of integrated *oriTs* that could transfer the PaLoc, or it may transfer by a completely different mechanism, for example cell fusion to form a zygote. Further work is required to determine the exact mechanism of PaLoc transfer.

Clean deletion of large DNA fragments (up to 8,000 bp) in Clostridia using CRISPR-Cas9 is challenging (Wang et al., 2016, 2018). The potential causes include: (i) the Cas9 protein is toxic to the host; (ii) the Cas9-carrying plasmid is often large and therefore potentially unstable; (iii) homologous arms present on the mutagenesis plasmid enable repeated re-integration after double crossover events. Here, we demonstrate the successful deletion of the ~56 kb ϕ027 prophage in R20291 using the pMTL83151 backbone modified with CRISPR-Cas9. Four previous studies which described gene deletions in *C. difficile* 630 or R20291 by CRISPR-Cas 9 or Cas12a were built on pMTL84151 (McAllister et al., 2017), pMTL82151 (Hong et al., 2018; Wang et al., 2018), and pMTL83151 (Ingle et al., 2019). The main difference between these plasmids is the replicon, with pMTL82151 having a replicon from pBP1, pMTL83151 a replicon from pCB102, and pMTL84151 a replicon from pCD6 (Heap et al., 2009). Compared to other reports, our mutation efficiency of 0.3% for deleting Tn*5397* (21 kb) in *C. difficile* 630 was very low. This could be because of the size of the deletion and the limits of the Cas9 nuclease in *C. difficile* 630, since other studies which used Cas9 for selecting deletions at a high efficiency targeted sequences up to 3.6 kb. In *C. difficile* 630, Cas9 on pMTL82151 selected *spo0A* deletants (825 bp) at 100% efficiency (Wang et al., 2018), and Cas9 on pMTL83151 allowed selection of *pyrE* (234 bp), and *ermB1* and *ermB2* (3.6 kb) deletants at 89 and 96% efficiency, respectively (Ingle et al., 2019). Interestingly, Hong et al. was unable to obtain transconjugants when they attempted to use Cas9 on pMTL82151 for deletion of ϕCD630-2 (49 kb) in *C. difficile* 630. However, they succeeded using Cas12a (Cpf1) nuclease to select for deleted prophage ϕCD630-2 (49.2 kb) at mutation efficiencies of 37.5–58.3% (Hong et al., 2018). They also deleted *fur* (390 bp), *cwp66* (1.8 kb), *tetM* (1.9 kb), *ermB1* and *ermB2* (3.2 kb), and *tcdA* (8.1 kb) at mutation efficiencies of 25–100% (Hong et al., 2018). Our 50% mutation efficiency of deleting ϕ027 (56 kb) prior to plasmid curing

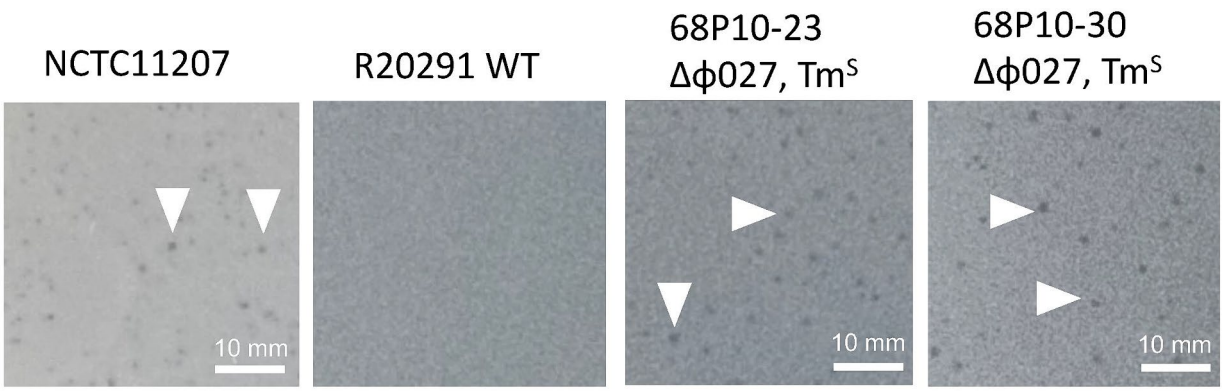

FIGURE 5
Phage susceptibility of R20291 prophage deletants. Susceptibility to ϕ027 infection was determined by plaque assays of Tm^S prophage deletants (68P10-23 and 68P10-30) that had lost plasmid-encoded resistance, the indicator strain NCTC11207 which is susceptible to ϕ027, and the lysogen WT R20291, which is resistant to ϕ027. White arrowheads indicate plaques, which vary in size and clarity.

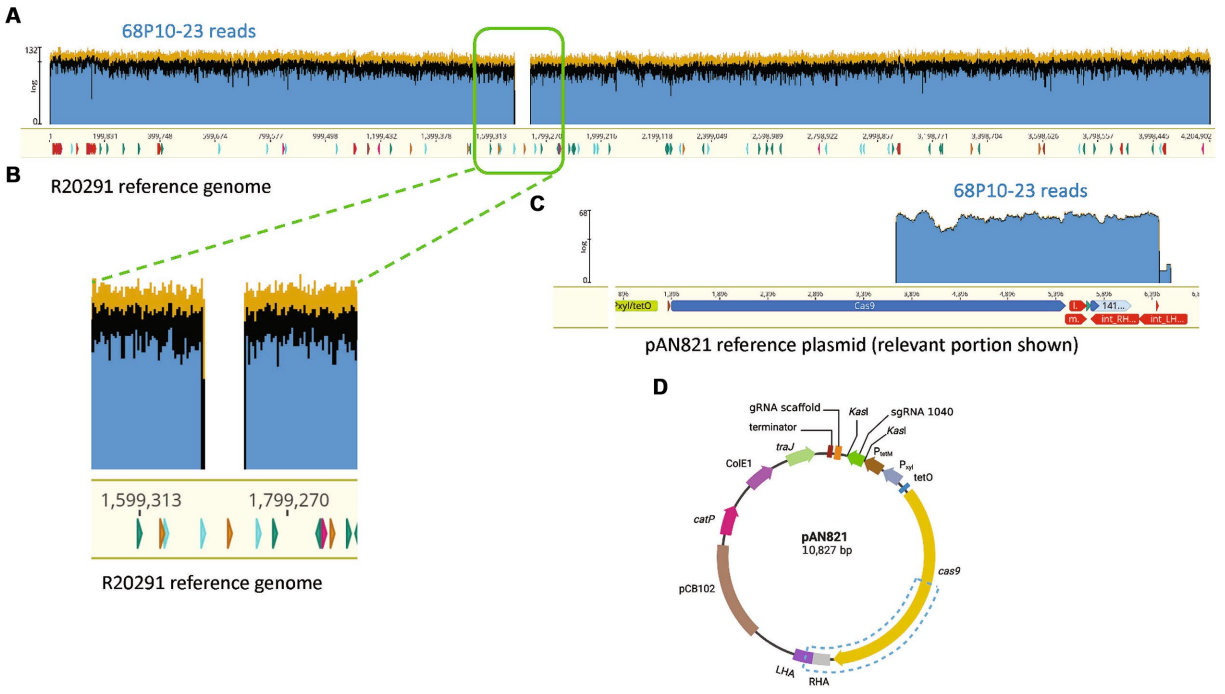

FIGURE 6
R20291Δϕ027 (68P10-23) contained a fragment of the mutagenesis plasmid. **(A)** Sequence reads of 68P10-23 (top blue bar) mapped to WT R20291 genome (bottom yellow bar with nucleotide numbers) showed missing reads as a gap, which belonged to the prophage, indicating its deletion. **(B)** Magnified section of where prophage sequence reads were missing from the WT genome. **(C)** A population of reads not mapping to R20291 was aligned to a 2.7 kb fragment of the pAN821 plasmid (bottom yellow bar) encompassing part of *cas9* and downstream sequences. **(D)** Blue dashed box shows the location of the 2.7 kb plasmid portion remaining in the prophage deletant. Created with biorender.com.

in R20291 was comparable to McAllister *et al* (McAllister et al., 2017). They deleted *pyrE* (585 bp) and *selD* (951 bp) at 50 and 20%, respectively (McAllister et al., 2017). However, we were unable to completely cure the mutagenesis plasmid. It is worth noting that (Maikova et al., 2019) re-programmed the endogenous Cas I-B system in R20291 and achieved 90% mutation efficiency for deleting a 261 bp gene. This strategy could be useful to avoid toxic effects of Cas9.

In this work, the pAN821 mutagenesis plasmid deleted the ϕ027 prophage and a downstream predicted bacterial gene (CDR20291_1,465) from R20291. However, the plasmid was not completely cured; a truncated *cas9* and the RHA from the plasmid remained stably integrated in a population of bacterial cells. This likely occurred from an imprecise double crossover event and was detected by whole genome sequencing, although previous studies in *C. difficile* did not report this phenomenon, perhaps because it is undetectable by standard PCR assays for loss of gene targets. Primers flanking the *attL* and *attR* sites (LCF887/888, Figure 4) were not used to check for prophage deletion to avoid false positive results of prophage deletion,

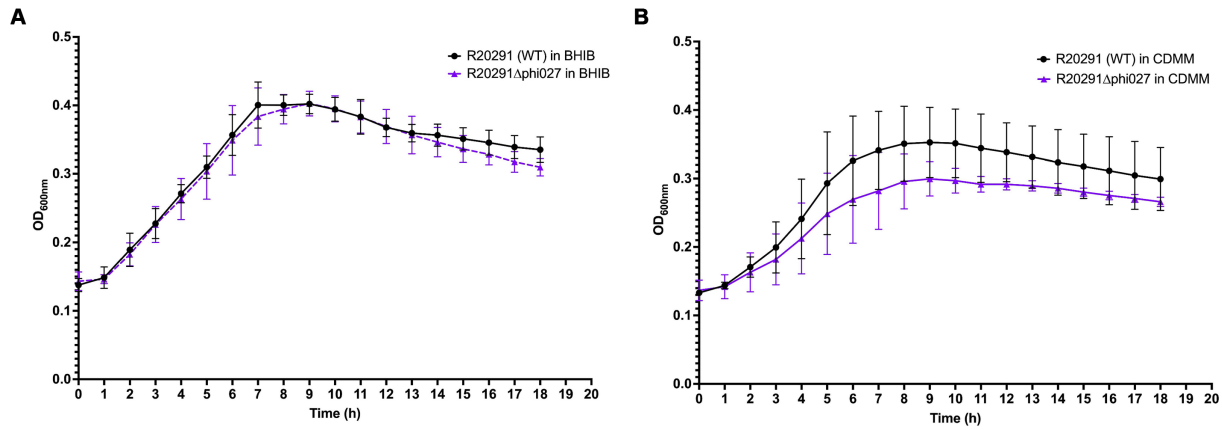

FIGURE 7
Growth of R20291 WT and Δϕ027 in **(A)** BHIB and **(B)** CDMM ($n = 4$).

since the prophage could spontaneously exist extrachromosomally. A study in *C. beijerinckii* reported difficulties in plasmid curing, which was overcome with the inclusion of CRISPR-Cas9 self-targeting of the mutagenesis plasmid (Wang et al., 2016). This could be explored in future. It was not possible to quantify the subset of cells based on the number of sequence reads due to amplification bias in sequencing. However, cells with the plasmid remnant appear to be the dominant cell type based on PCR screening of single colonies. The truncated *cas9* translates to amino acid (aa) residues 781 to 1,368, which consists of the HNH, RuvCIII, Topo-homology, and PI domains that function in nuclease and PAM recognition, i. e. the "nuclease lobe" (Jinek et al., 2014; Nishimasu et al., 2014). However, without the other Cas9 protein domains that form the "recognition lobe" for facilitating guide RNA binding to DNA, this truncated *cas9* will likely be inactive if translated in the prophage deletant (Jinek et al., 2014; Nishimasu et al., 2014). Interestingly, Ingle et al. used the same backbone pMTL83151 for CRISPR-Cas9 deletion in *C. difficile* 630 and found truncation of Cas9 (Ingle et al., 2019). However, their truncated Cas9 was missing 87 aa from the N-terminus (Ingle et al., 2019) while the *cas9* remaining in our ϕ027 deletant would be missing approximately 780 aa (or 2,340 bases) from the N-terminus (or 5′ end of the gene) as mentioned above, if it was translated.

The *C. difficile* gene deleted adjacent to the prophage CDR20291_1465 is homologous to a putative manganese-containing catalase found in the *Bacillus subtilis* spore coat protein CotJC. In R20291, two other genes encode CotJC1 (CotCB) and CotJC2 (CotD), which have 70 and 50% amino acid sequence similarity, respectively, to CotJC (Permpoonpattana et al., 2011). Insertional mutations of *cotCB* and *cotD* in *C. difficile* 630 resulted in a reduction of catalase activity, but otherwise no significant defect to spore coat formation (Permpoonpattana et al., 2013). This suggests functional redundancy of CDR20291_1465 in R20291, the deletion of which would be unlikely to affect spore coat formation.

We were able to assay for ϕ027 plaque formation and hence propagate the phage using NCTC11207, although it was of a different ribotype to R20291 and contained two predicted prophage genomes. This indicates the potential of ϕ027 to lysogenize isolates other than RT 027, in which it is commonly found (He et al., 2013). Sensitivity of R20291 to ϕ027 was restored after prophage deletion, indicating in R20291 superinfection exclusion provided by the lysogenic prophage was the main mechanism of superinfection immunity. For instance, ϕ027 prophage encoded an Abi-like protein with similarity to *abiD* from *Lactococcus lactis*. Abi proteins protect uninfected bacterial cells from phage infection by infected cell suicide, hence aborting further phage infection (Lopatina et al., 2020). CRISPR-Cas is another system which provides immunity against phage. CRISPR arrays found in R20291 did not target ϕ027, though it is noteworthy that ϕ027 carried two CRISPR arrays which very likely conferred immunity to 12 phages (Boudry et al., 2015). Toxin-antitoxin systems, which is another possible phage defense system, has not been predicted in R20291. Since ϕ027 harbored two phage-defense systems, its deletion from R20291 may increase bacterial susceptibility to phage infection. The prophage deletant did not differ significantly in growth compared to WT in rich or minimal medium, although its growth in minimal medium was consistently lower than WT under nutrient limiting conditions (i.e., in the late stationary and death phase of growth curves) and could indicate the prophage was involved in regulation of genes for survival under those conditions. Our prophage deletion approach resulted in an unexpected genetic feature in the prophage deletant that could affect bacterial behavior. This could be determined by re-lysogenizing the prophage deletant with ϕ027 and comparing it to WT. Nevertheless, we anticipate the R20291 prophage deletant to be a useful strain for investigating prophage contribution to host virulence, fitness, and physiology, and a platform for other mutagenesis studies aimed at functional gene analysis without native phage interference.

In conclusion, we have shown that it is feasible to make a clean deletion of the ICE Tn*5397*. A phage genome could also be precisely deleted from the host chromosome. However, it was also observed that a fragment of the vector used for generating phage deletion could not be completely removed from the host cells. This is probably due to continual recombination between the host genome and the vector DNA. We have also observed this type of interaction with other host vector systems (Hussain et al. unpublished). Our previous work has also shown that some vectors can transfer between *E. coli* and *C. difficile* without the requirement for an obvious *oriT* and that transfer is sensitive to DNase (Khodadoost et al., 2017). Therefore, it is recommended that researchers undertake whole genome sequence

analysis after mutant construction to determine the exact genotype of their mutant as this could impact how downstream physiological experiments are interpreted. Furthermore, it is important that further work is done to get a deeper understanding of the mechanism of transfer of MGEs between *C. difficile* strains.

Data availability statement

The datasets generated for this study can be found in BioProject, accession number: PRJNA993731 and GenBank, accession number: CP129979.1. Materials generated in this study are available upon request from the corresponding author.

Author contributions

HH: Data curation, Investigation, Methodology, Visualization, Writing – review & editing, Writing – original draft. AN: Writing – original draft, Writing – review & editing, Data curation, Investigation, Methodology, Visualization. CR: Methodology, Resources, Visualization, Writing – review & editing, Data curation, Formal analysis. KI: Formal Analysis, Methodology, Writing – review & editing. DK: Data curation, Resources, Writing – review & editing. VP: Investigation, Visualization, Writing – review & editing. PM: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. SG: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. KI was supported by the Mahidol Scholarship. SG is supported by an MRC Equipment award (MC_PC_ MR/X012190/1). This work was supported by an Emerging Leaders Grant from WA Department of Health (WANMA2021) awarded to DK.

Acknowledgments

This research used the facilities and services of the Pawsey Supercomputing Centre, Perth, Western Australia.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2024.1416665/full#supplementary-material

References

Andersen, J. M., Shoup, M., Robinson, C., Britton, R., Olsen, K. E. P., and Barrangou, R. (2016). CRISPR diversity and microevolution in *Clostridium difficile*. *Genome Biol. Evol.* 8, 2841–2855. doi: 10.1093/gbe/evw203

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387

Bletz, S., Janezic, S., Harmsen, D., Rupnik, M., and Mellmann, A. (2018). Defining and evaluating a Core genome multilocus sequence typing scheme for genome-wide typing of *Clostridium difficile*. *J. Clin. Microbiol.* 56, e01987–17. doi: 10.1128/JCM.01987-17

Boudry, P., Semenova, E., Monot, M., Datsenko, K. A., Lopatina, A., Sekulovic, O., et al. (2015). Function of the CRISPR-Cas system of the human pathogen *Clostridium difficile*. *MBio* 6, e01112–e01115. doi: 10.1128/mBio.01112-15

Brouwer, M. S., Roberts, A. P., Hussain, H., Williams, R. J., Allan, E., and Mullany, P. (2013). Horizontal gene transfer converts non-toxigenic *Clostridium difficile* strains into toxin producers. *Nat. Commun.* 4:2601. doi: 10.1038/ncomms3601

Brouwer, M. S., Warburton, P. J., Roberts, A. P., Mullany, P., and Allan, E. (2011). Genetic organisation, mobility and predicted functions of genes on integrated, mobile genetic elements in sequenced strains of *Clostridium difficile*. *PLoS One* 6:e23014. doi: 10.1371/journal.pone.0023014

Candel-Pérez, C., Ros-Berruezo, G., and Martínez-Graciá, C. (2019). A review of *Clostridioides [Clostridium] difficile* occurrence through the food chain. *Food Microbiol.* 77, 118–129. doi: 10.1016/j.fm.2018.08.012

Cartman, S. T., and Minton, N. P. (2010). A mariner-based transposon system for *in vivo* random mutagenesis of *Clostridium difficile*. *Appl. Environ. Microbiol.* 76, 1103–1109. doi: 10.1128/AEM.02525-09

Collins, D. A., Sohn, K. M., Wu, Y., Ouchi, K., Ishii, Y., Elliott, B., et al. (2020). *Clostridioides difficile* infection in the Asia-Pacific region. *Emerg. Microbes Infect.* 9, 42–52. doi: 10.1080/22221751.2019.1702480

Fagan, R. P., and Fairweather, N. F. (2011). *Clostridium difficile* has two parallel and essential sec secretion systems. *J. Biol. Chem.* 286, 27483–27493. doi: 10.1074/jbc.M111.263889

Finn, E., Andersson, F. L., and Madin-Warburton, M. (2021). Burden of *Clostridioides difficile* infection (CDI) - a systematic review of the epidemiology of primary and recurrent CDI. *BMC Infect. Dis.* 21:456. doi: 10.1186/s12879-021-06147-y

Fortier, L. C., and Moineau, S. (2007). Morphological and genetic diversity of temperate phages in *Clostridium difficile*. *Appl. Environ. Microbiol.* 73, 7358–7366. doi: 10.1128/AEM.00582-07

Gawlik, D., Slickers, P., Engelmann, I., Müller, E., Lück, C., Friedrichs, A., et al. (2015). DNA-microarray-based genotyping of *Clostridium difficile*. *BMC Microbiol.* 15:158. doi: 10.1186/s12866-015-0489-2

Goh, S., Chang, B., and Riley, T. (2005). Effect of phage infection on toxin production by *Clostridium difficile*. *J. Med. Microbiol.* 54, 129–135. doi: 10.1099/jmm.0.45821-0

Goh, S., Hussain, H., Chang, B. J., Emmett, W., Riley, T. V., and Mullany, P. (2013). Phage phiC2 mediates transduction of Tn*6215*, encoding erythromycin resistance,

between *Clostridium difficile* strains. *MBio* 4, e00840–e00813. doi: 10.1128/mBio.00840-13

Goh, S., Ong, P. F., Song, K. P., Riley, T. V., and Chang, B. J. (2007). The complete genome sequence of *Clostridium difficile* phage phiC2 and comparisons to phiCD119 and inducible prophages of CD630. *Microbiology* 153, 676–685. doi: 10.1099/mic.0.2006/002436-0

Goh, S., Riley, T. V., and Chang, B. J. (2005). Isolation and characterization of temperate bacteriophages of *Clostridium difficile*. *Appl. Environ. Microbiol.* 71, 1079–1083. doi: 10.1128/AEM.71.2.1079-1083.2005

Govind, R., Fralick, J. A., and Rolfe, R. D. (2006). Genomic organization and molecular characterization of *Clostridium difficile* bacteriophage PhiCD119. *J. Bacteriol.* 188, 2568–2577. doi: 10.1128/JB.188.7.2568-2577.2006

Govind, R., Fralick, J. A., and Rolfe, R. D. (2011). *In vivo* lysogenization of a *Clostridium difficile* bacteriophage ΦCD119. *Anaerobe* 17, 125–129. doi: 10.1016/j.anaerobe.2011.05.012

Govind, R., Vediyappan, G., Rolfe, R. D., Dupuy, B., and Fralick, J. A. (2009). Bacteriophage-mediated toxin gene regulation in *Clostridium difficile*. *J. Virol.* 83, 12037–12045. doi: 10.1128/JVI.01256-09

Griffiths, D., Fawley, W., Kachrimanidou, M., Bowden, R., Crook, D. W., Fung, R., et al. (2010). Multilocus sequence typing of *Clostridium difficile*. *J. Clin. Microbiol.* 48, 770–778. doi: 10.1128/JCM.01796-09

Hargreaves, K. R., Kropinski, A. M., and Clokie, M. R. J. (2014). What does the talking?: quorum sensing Signalling genes discovered in a bacteriophage genome. *PLoS One* 9:e85131. doi: 10.1371/journal.pone.0085131

He, M., Miyajima, F., Roberts, P., Ellison, L., Pickard, D. J., Martin, M. J., et al. (2013). Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nat. Genet.* 45, 109–113. doi: 10.1038/ng.2478

Heap, J. T., Pennington, O. J., Cartman, S. T., and Minton, N. P. (2009). A modular system for *Clostridium* shuttle plasmids. *J. Microbiol. Methods* 78, 79–85. doi: 10.1016/j.mimet.2009.05.004

Hong, W., Zhang, J., Cui, G., Wang, L., and Wang, Y. (2018). Multiplexed CRISPR-Cpf1-mediated genome editing in *Clostridium difficile* toward the understanding of pathogenesis of *C. difficile* infection. *ACS Synth. Biol.* 7, 1588–1600. doi: 10.1021/acssynbio.8b00087

Ingle, P., Groothuis, D., Rowe, P., Huang, H., Cockayne, A., Kuehne, S. A., et al. (2019). Generation of a fully erythromycin-sensitive strain of *Clostridioides difficile* using a novel CRISPR-Cas9 genome editing system. *Sci. Rep.* 9:8123. doi: 10.1038/s41598-019-44458-y

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., et al. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343:1247997. doi: 10.1126/science.1247997

Jo, J., Gonzales-Luna, A. J., Lancaster, C. K., McPherson, J. K., Begum, K., Jahangir Alam, M., et al. (2022). Multi-country surveillance of *Clostridioides difficile* demonstrates high prevalence of spores in non-healthcare environmental settings. *Anaerobe* 75:102543. doi: 10.1016/j.anaerobe.2022.102543

Johnson, C. M., and Grossman, A. D. (2015). Integrative and conjugative elements (ICEs): what they do and how they work. *Annu. Rev. Genet.* 49, 577–601. doi: 10.1146/annurev-genet-112414-055018

Khodadoost, L., Hussain, H., and Mullany, P. (2017). Plasmids can transfer to *Clostridium difficile* CD37 and 630Δ*erm* both by a DNase resistant conjugation-like mechanism and a DNase sensitive mechanism. *FEMS Microbiol. Lett.* 364. doi: 10.1093/femsle/fnx208

Kirk, J. A., and Fagan, R. P. (2016). Heat shock increases conjugation efficiency in *Clostridium difficile*. *Anaerobe* 42, 1–5. doi: 10.1016/j.anaerobe.2016.06.009

Knetsch, C. W., Connor, T. R., Mutreja, A., van Dorp, S. M., Sanders, I. M., Browne, H. P., et al. (2014). Whole genome sequencing reveals potential spread of *Clostridium difficile* between humans and farm animals in the Netherlands, 2002 to 2011. *Euro Surv.* 19:20954. doi: 10.2807/1560-7917.ES2014.19.45.20954

Knight, D. R., Elliott, B., Chang, B. J., Perkins, T. T., and Riley, T. V. (2015). Diversity and evolution in the genome of *Clostridium difficile*. *Clin. Microbiol. Rev.* 28, 721–741. doi: 10.1128/CMR.00127-14

Knight, D. R., and Riley, T. V. (2019). Genomic delineation of zoonotic origins of *Clostridium difficile*. *Front. Public Health* 7:164. doi: 10.3389/fpubh.2019.00164

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546. doi: 10.1038/s41587-019-0072-8

Kuehne, S. A., Cartman, S. T., Heap, J. T., Kelly, M. L., Cockayne, A., and Minton, N. P. (2010). The role of toxin a and toxin B in *Clostridium difficile* infection. *Nature* 467, 711–713. doi: 10.1038/nature09397

Kuehne, S. A., Cartman, S. T., and Minton, N. P. (2011). Both, toxin a and toxin B, are important in *Clostridium difficile* infection. *Gut Microbes* 2, 252–255. doi: 10.4161/gmic.2.4.16109

Lawson, P. A., Citron, D. M., Tyrrell, K. L., and Finegold, S. M. (2016). Reclassification of *Clostridium difficile* as *Clostridioides difficile* (hall and O'Toole 1935) Prévot 1938. *Anaerobe* 40, 95–99. doi: 10.1016/j.anaerobe.2016.06.008

Li, T., Zhang, Y., Dong, K., Kuo, C. J., Li, C., Zhu, Y. Q., et al. (2020). Isolation and characterization of the novel phage JD032 and global transcriptomic response during JD032 infection of *Clostridioides difficile* Ribotype 078. *mSystems* 5, e00017–20. doi: 10.1128/mSystems.00017-20

Lopatina, A., Tal, N., and Sorek, R. (2020). Abortive infection: bacterial suicide as an antiviral immune strategy. *Ann. Rev. Virol.* 7, 371–384. doi: 10.1146/annurev-virology-011620-040628

Magill, S. S., Edwards, J. R., Bamberg, W., Beldavs, Z. G., Dumyati, G., Kainer, M. A., et al. (2014). Multistate point-prevalence survey of health care-associated infections. *N. Engl. J. Med.* 370, 1198–1208. doi: 10.1056/NEJMoa1306801

Maikova, A., Kreis, V., Boutserin, A., Severinov, K., and Soutourina, O. (2019). Using an endogenous CRISPR-Cas system for genome editing in the human pathogen *Clostridium difficile*. *Appl. Environ. Microbiol.* 85, e01416–19. doi: 10.1128/AEM.01416-19

Mayer, M. J., Narbad, A., and Gasson, M. J. (2008). Molecular characterization of a *Clostridium difficile* bacteriophage and its cloned biologically active endolysin. *J. Bacteriol.* 190, 6734–6740. doi: 10.1128/JB.00686-08

McAllister, K. N., Bouillaut, L., Kahn, J. N., Self, W. T., and Sorg, J. A. (2017). Using CRISPR-Cas9-mediated genome editing to generate *C. difficile* mutants defective in selenoproteins synthesis. *Sci. Rep.* 7. doi: 10.1038/s41598-017-15236-5

Meessen-Pinard, M., Sekulovic, O., and Fortier, L. C. (2012). Evidence of *in vivo* prophage induction during *Clostridium difficile* infection. *Appl. Environ. Microbiol.* 78, 7662–7670. doi: 10.1128/AEM.02275-12

Moloney, G., Eyre, D. W., Mac Aogáin, M., McElroy, M. C., Vaughan, A., Peto, T. E. A., et al. (2021). Human and porcine transmission of *Clostridioides difficile* Ribotype 078. *Europe Emerg. Infect. Diseases* 27, 2294–2300. doi: 10.3201/eid2709.203468

Monot, M., Boursaux-Eude, C., Thibonnier, M., Vallenet, D., Moszer, I., Medigue, C., et al. (2011). Reannotation of the genome sequence of *Clostridium difficile* strain 630. *J. Med. Microbiol.* 60, 1193–1199. doi: 10.1099/jmm.0.030452-0

Monteford, J., Bilverstone, T. W., Ingle, P., Philip, S., Kuehne, S. A., Minton, N. P., et al. (2021). What's a SNP between friends: the lineage of *Clostridioides difficile* R20291 can effect research outcomes. *Anaerobe* 71:102422.

Moraru, C., Varsani, A., and Kropinski, A. M. (2020). VIRIDIC—A novel tool to calculate the Intergenomic similarities of prokaryote-infecting viruses. *Viruses* 12:1268. doi: 10.3390/v12111268

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., et al. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156, 935–949. doi: 10.1016/j.cell.2014.02.001

Peltier, J., Hamiot, A., Garneau, J. R., Boudry, P., Maikova, A., Hajnsdorf, E., et al. (2020). Type I toxin-antitoxin systems contribute to the maintenance of mobile genetic elements in *Clostridioides difficile*. *Commun. Biol.* 3:718. doi: 10.1038/s42003-020-01448-5

Permpoonpattana, P., Phetcharaburanin, J., Mikelsone, A., Dembek, M., Tan, S., Brisson, M. C., et al. (2013). Functional characterization of *Clostridium difficile* spore coat proteins. *J. Bacteriol.* 195, 1492–1503. doi: 10.1128/JB.02104-12

Permpoonpattana, P., Tolls, E. H., Nadem, R., Tan, S., Brisson, A., and Cutting, S. M. (2011). Surface layers of *Clostridium difficile* endospores. *J. Bacteriol.* 193, 6461–6470. doi: 10.1128/JB.05182-11

Purdy, D., O'Keeffe, T. A., Elmore, M., Herbert, M., McLeod, A., Bokori-Brown, M., et al. (2002). Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Mol. Microbiol.* 46, e00276–15. doi: 10.1046/j.1365-2958.2002.03134.x

Ramirez-Vargas, G., Goh, S., and Rodriguez, C. (2018). The novel phages phiCD5763 and phiCD2955 represent two groups of big Plasmidial Siphoviridae phages of *Clostridium difficile*. *Front. Microbiol.* 9:26. doi: 10.3389/fmicb.2018.00026

Riedel, T., Bunk, B., Thürmer, A., Spröer, C., Brzuszkiewicz, E., Abt, B., et al. (2015). Genome resequencing of the virulent and multidrug-resistant reference strain *Clostridium difficile* 630. *Genome Announc.* 3, e00276–15. doi: 10.1128/genomeA.00276-15

Riedel, T., Wittmann, J., Bunk, B., Schober, I., Sproer, C., Gronow, S., et al. (2017). A *Clostridioides difficile* bacteriophage genome encodes functional binary toxin-associated genes. *J. Biotechnol.* 250, 23–28. doi: 10.1016/j.jbiotec.2017.02.017

Roberts, A. P., Allan, E., and Mullany, P. (2014). The impact of horizontal gene transfer on the biology of *Clostridium difficile*. *Adv. Microb. Physiol.* 65, 63–82. doi: 10.1016/bs.ampbs.2014.08.002

Sambrook, J., and Russell, D. W. (2001a). *Protocol 8. Purification of bacteriophage lambda particles by isopycnic centrifugation through CsCl gradients. Molecular cloning: A laboratory manual. 3rd* Edn. New Yor: Cold Spring Harbor Laboratory Press.

Sambrook, J., and Russell, D. W. (2001b). *Protocol 11. Extraction of bacteriophage lambda DNA from large-scale cultures using proteinase K and SDS. Molecular cloning: A laboratory manual. 3rd* Edn. New York: Cold Spring Harbor Laboratory Press.

Scornec, H., Bellanger, X., Guilloteau, H., Groshenry, G., and Merlin, C. (2017). Inducibility of Tn*916* conjugative transfer in *Enterococcus faecalis* by subinhibitory concentrations of ribosome-targeting antibiotics. *J. Antimicrob. Chemother.* 72, 2722–2728. doi: 10.1093/jac/dkx202

Sebaihia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R., et al. (2006). The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat. Genet.* 38, 779–786. doi: 10.1038/ng1830

Sekulovic, O., and Fortier, L. C. (2015). Global transcriptional response of *Clostridium difficile* carrying the phiCD38-2 prophage. *Appl. Environ. Microbiol.* 81, 1364–1374. doi: 10.1128/AEM.03656-14

Sekulovic, O., Meessen-Pinard, M., and Fortier, L. C. (2011). Prophage-stimulated toxin production in *Clostridium difficile* NAP1/027 lysogens. *J. Bacteriol.* 193, 2726–2734. doi: 10.1128/JB.00787-10

Sekulovic, O., Ospina Bedoya, M., Fivian-Hughes, A. S., Fairweather, N. F., and Fortier, L. C. (2015). The *Clostridium difficile* cell wall protein CwpV confers phase-variable phage resistance. *Mol. Microbiol.* 98, 329–342. doi: 10.1111/mmi.13121

Shaw, H. A., Khodadoost, L., Preston, M. D., Corver, J., Mullany, P., and Wren, B. W. (2019). *Clostridium difficile* clade 3 (RT023) have a modified cell surface and contain a large transposable island with novel cargo. *Sci. Rep.* 9:15330. doi: 10.1038/s41598-019-51628-5

Slater, R. T., Frost, L. R., Jossi, S. E., Millard, A. D., and Unnikrishnan, M. (2019). *Clostridioides difficile* LuxS mediates inter-bacterial interactions within biofilms. *Sci. Rep.* 9:9903. doi: 10.1038/s41598-019-46143-6

Smith, C. J., Markowitz, S. M., and Macrina, F. L. (1981). Transferable tetracycline resistance in *Clostridium difficile*. *Antimicrob. Agents Chemother.* 19, 997–1003. doi: 10.1128/AAC.19.6.997

Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H., and Kuijper, E. J. (2016). *Clostridium difficile* infection. *Nat. Rev. Dis. Primers* 2:16020. doi: 10.1038/nrdp.2016.20

Stabler, R. A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., et al. (2009). Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biol.* 10:R102. doi: 10.1186/gb-2009-10-9-r102

Stark, W. M. (2014). The serine recombinases. *Microbiol Spectr* 2:6. doi: 10.1128/microbiolspec.MDNA3-0046-2014

Su, Y. A., He, P., and Clewell, D. B. (1992). Characterization of the *tet(M)* determinant of Tn*916*: evidence for regulation by transcription attenuation. *Antimicrob. Agents Chemother.* 36, 769–778. doi: 10.1128/AAC.36.4.769

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44, 6614–6624. doi: 10.1093/nar/gkw569

van Prehn, J., Reigadas, E., Vogelzang, E. H., Bouza, E., Hristea, A., Guery, B., et al. (2021). European Society of Clinical Microbiology and Infectious Diseases: 2021 update on the treatment guidance document for *Clostridioides difficile* infection in adults. *Clin. Microbiol. Infect.* 27, S1–s21. doi: 10.1016/j.cmi.2021.09.038

Viprey, V. F., Davis, G. L., Benson, A. D., Ewin, D., Spittal, W., Vernon, J. J., et al. (2022). A point-prevalence study on community and inpatient *Clostridioides difficile* infections (CDI): results from combatting bacterial resistance in Europe CDI (COMBACTE-CDI), July to November 2018. *Euro Surv.* 27, pii=2100704. doi: 10.2807/1560-7917.ES.2022.27.26.2100704

Wang, S., Hong, W., Dong, S., Zhang, Z. T., Zhang, J., Wang, L., et al. (2018). Genome engineering of *Clostridium difficile* using the CRISPR-Cas9 system. *Clin. Microbiol. Infect.* 24, 1095–1099. doi: 10.1016/j.cmi.2018.03.026

Wang, H., and Mullany, P. (2000). The large resolvase TndX is required and sufficient for integration and excision of derivatives of the novel conjugative transposon Tn*5397*. *J. Bacteriol.* 182, 6577–6583. doi: 10.1128/JB.182.23.6577-6583.2000

Wang, H., Roberts, A. P., Lyras, D., Rood, J. I., Wilks, M., and Mullany, P. (2000). Characterization of the ends and target sites of the novel conjugative transposon Tn*5397* from *Clostridium difficile*: excision and circularization is mediated by the large resolvase. *TndX. J Bacteriol.* 182, 3775–3783. doi: 10.1128/JB.182.13.3775-3783.2000

Wang, H., Smith, M. C., and Mullany, P. (2006). The conjugative transposon Tn*5397* has a strong preference for integration into its *Clostridium difficile* target site. *J. Bacteriol.* 188, 4871–4878. doi: 10.1128/JB.00210-06

Wang, Y., Zhang, Z.-T., Seo, S.-O., Lynn, P., Lu, T., Jin, Y.-S., et al. (2016). Bacterial genome editing with CRISPR-Cas9: deletion, integration, single nucleotide modification, and desirable "clean" mutant selection in *Clostridium beijerinckii* as an example. *ACS Synth. Biol.* 5, 721–732. doi: 10.1021/acssynbio.6b00060

Wick, R. R., Judd, L. M., Cerdeira, L. T., Hawkey, J., Méric, G., Vezina, B., et al. (2021). Trycycler: consensus long-read assemblies for bacterial genomes. *Genome Biol.* 22:266. doi: 10.1186/s13059-021-02483-z

Wick, R. R., Schultz, M. B., Zobel, J., and Holt, K. E. (2015). Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* 31, 3350–3352. doi: 10.1093/bioinformatics/btv383

Williams, D. R., Young, D. I., and Young, M. (1990). Conjugative plasmid transfer from *Escherichia coli* to *Clostridium acetobutylicum*. *Microbiology* 136, 819–826.

Frontiers in Microbiology

Explores the habitable world and the potential of microbial life

The largest and most cited microbiology journal which advances our understanding of the role microbes play in addressing global challenges such as healthcare, food security, and climate change.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact