frontiers | Research Topics

Bacterial pathogens and virulence factor genes: diversity and evolution

Edited by
Renmao "Tim" Tian, Qun Gao, Xuanyu Tao and Xiangyu Fan

Published in
Frontiers in Microbiology

ISSN 1664-8714
ISBN 978-2-8325-6658-9
DOI 10.3389/978-2-8325-6658-9

Generative AI statement
Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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

Bacterial pathogens and virulence factor genes: diversity and evolution

Topic editors

Renmao "Tim" Tian — Illinois Institute of Technology, United States
Qun Gao — Beijing Normal University, China
Xuanyu Tao — University of Oklahoma, United States
Xiangyu Fan — University of Jinan, China

Citation

Tian, R. T., Gao, Q., Tao, X., Fan, X., eds. (2025). *Bacterial pathogens and virulence factor genes: diversity and evolution*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6658-9

Table of contents

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 15 April 2024
DOI 10.3389/fmicb.2024.1384166

OPEN ACCESS

EDITED BY
Renmao Tian,
Illinois Institute of Technology, United States

REVIEWED BY
Hanyan Li,
University of Oklahoma, United States
Mengyuan Ji,
University of Padua, Italy
Xuanyu Tao,
University of Oklahoma, United States
Dongyu Wang,
University of Oklahoma Norman,
United States, in collaboration with
reviewer XT

*CORRESPONDENCE
Hong Zhang
✉ 17755132396@163.com

RECEIVED 08 February 2024
ACCEPTED 29 March 2024
PUBLISHED 15 April 2024

CITATION
Zhang D, Li X, Wang Y, Zhao Y and Zhang H (2024) The clinical importance of metagenomic next-generation sequencing in detecting disease-causing microorganisms in cases of sepsis acquired in the community or hospital setting. *Front. Microbiol.* 15:1384166. doi: 10.3389/fmicb.2024.1384166

The clinical importance of metagenomic next-generation sequencing in detecting disease-causing microorganisms in cases of sepsis acquired in the community or hospital setting

Dan Zhang, Xingxing Li, Yu Wang, Yong Zhao and Hong Zhang*

Department of Emergency Medicine, The First Affiliated Hospital of Anhui Medical University, Anhui, Hefei, China

Objectives: Although metagenomic next-generation sequencing (mNGS) is commonly used for diagnosing infectious diseases, clinicians face limited options due to the high costs that are not covered by basic medical insurance. The goal of this research is to challenge this bias through a thorough examination and evaluation of the clinical importance of mNGS in precisely identifying pathogenic microorganisms in cases of sepsis acquired in the community or in hospitals.

Methods: A retrospective observational study took place at a tertiary teaching hospital in China from January to December 2021. Data on 308 sepsis patients were collected, and the performance of etiological examination was compared between mNGS and traditional culture method.

Results: Two hundred twenty-nine cases were observed in the community-acquired sepsis (CAS) group and 79 cases in the hospital-acquired sepsis (HAS) group. In comparison with conventional culture, mNGS showed a significantly higher rate of positivity in both the CAS group (88.21% vs. 25.76%, adj.$P < 0.001$) and the HAS group (87.34% vs. 44.30%, adj.$P < 0.001$), particularly across various infection sites and specimens, which were not influenced by factors like antibiotic exposure or the timing and frequency of mNGS technology. Sepsis pathogens detected by mNGS were broad, especially viruses, *Mycobacterium tuberculosis*, and atypical pathogens, with mixed pathogens being common, particularly bacterial-viral co-detection. Based on the optimization of antimicrobial therapy using mNGS, 58 patients underwent antibiotic de-escalation, two patients were switched to antiviral therapy, and 14 patients initiated treatment for tuberculosis, resulting in a reduction in antibiotic overuse but without significant impact on sepsis prognosis. The HAS group exhibited a critical condition, poor prognosis, high medical expenses, and variations in etiology, yet the mNGS results did not result in increased medical costs for either group.

Conclusions: mNGS demonstrates efficacy in identifying multiple pathogens responsible for sepsis, with mixed pathogens of bacteria and viruses being prevalent. Variability in microbiological profiles among different infection setting

underscores the importance of clinical vigilance. Therefore, the adoption of mNGS for microbiological diagnosis of sepsis warrants acknowledgment and promotion.

KEYWORDS

community-acquired sepsis, hospital-acquired sepsis, mNGS, microorganisms, optimizing antimicrobial therapy, medical expenses

Introduction

Sepsis, a prevalent and severe clinical condition characterized by organ dysfunction resulting from immune response dysfunction caused by infection, has garnered significant attention worldwide due to its substantial health risks and financial burden (Singer et al., 2016; Rhodes et al., 2017; Buchman et al., 2020). In 2017, the World Health Organization (WHO) introduced a resolution stressing the importance of improving the recognition, diagnosis, treatment, and prevention of sepsis due to the pressing need for emergency care (WHO, 2017). The management of sepsis treatment is complicated due to the wide range of pathogenic infections, the challenge of promptly identifying pathogens, and the need for precise treatment in the early stages. While culture has long been viewed as the benchmark for diagnosing sepsis, it does have some drawbacks. These include the time-consuming nature of the process, which can take 2–7 days, a low rate of positive results, susceptibility to contamination, a limited spectrum of pathogen detection, and vulnerability to the influence of antibiotics. These constraints may lead to delayed treatment with antibacterial drugs, excessive use of broad-spectrum antibiotics, a rise in resistance to microbial drugs, and increased medical costs (Miao et al., 2018). Hence, the identification and prompt detection of pathogens without the aforementioned issues necessitate the exploration of a technology. Researchers have shown growing interest in metagenomic next-generation sequencing (mNGS) technology in this particular situation.

mNGS employs high-throughput gene sequencing technology to simultaneously detect the deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) of all microorganisms present in clinical samples, enabling the determination of potential pathogenic microorganism types through database comparison and bioinformatics analysis (Sharon and Banfield, 2013; Chiu and Miller, 2019). It has the characteristics of no culture, no dependence on specific sequence amplification, no bias, less time consumption, high sensitivity, not affected by a variety of bacterial species and antibiotic treatment, which is widely used in infectious diseases (Ishihara et al., 2020; Gu et al., 2021; Miller and Chiu, 2022; Yang et al., 2022). Prior research have shown that mNGS results obtain higher positive rate and clinical coincidence rate in sepsis pathogen detection, which is helpful for medical decision-making, optimizing antibiotic management and improving prognosis (Wang et al., 2023; Zuo et al., 2023). Clinicians frequently do not prioritize utilizing mNGS due to its expensive nature and limited coverage by health insurance.

Research found that community-acquired sepsis (CAS) makes up around 70% of sepsis cases (Reinhart et al., 2017), with distinct pathogen characteristics, treatment strategies, and survival rates among CAS and hospital-acquired sepsis (HAS) (Tonai et al., 2022; Kim et al., 2023). While timely administration of effective antibiotics is crucial in reducing sepsis mortality, the significance of pathogen culture should not be overlooked (Niederman et al., 2021). Given the limited exploration of the disparities between CAS and HAS in existing literature, further investigation into the utility of mNGS in these contexts is warranted. Therefore, the aim of this research was to evaluate the medical importance of mNGS in detecting pathogens and offering treatment recommendations for sepsis acquired in the community or in a hospital setting.

Materials and methods

Study participants and groups

A retrospective observational study on sepsis was conducted at a single center, using analytical cross-sectional cohort methods. Between January 2021 and December 2021, the First Affiliated Hospital of Anhui Medical University enrolled 308 participants. Patients were selected if they met specific criteria, including the third international consensus diagnostic criteria for sepsis and septic shock (Sepsis-3) from 2016, with a Sequential Organ Failure Assessment (SOFA) score of $\geq$2, being over 18 years old, providing consent for mNGS examination, having complete clinical and laboratory data available, and undergoing both mNGS and conventional culture methods for examination. Participants who did not meet the study criteria were excluded. This included individuals under 18 years old, those who refused mNGS examination, pregnant or lactating individuals, and those with incomplete or insufficient clinical data.

Patients were categorized into two groups, namely, the CAS group and the HAS group, based on the location of onset as documented in the clinical electronic medical record system. The CAS group consisted of patients who developed sepsis due to infections acquired prior to hospitalization, as well as those who exhibited symptoms within 48 h of hospitalization during the incubation period. In contrast, sepsis caused by the aforementioned infections was classified as HAS if it met the following criteria: (1) Infections without an indeterminate incubation period beyond 48 h of admission; (2) Infections with a clearly defined incubation period that exceeded the average incubation period upon admission; (3) Infections directly associated with the most recent hospitalization; (4) Occurrence of new infections in other anatomical sites based on the original infection; (5) Identification of new pathogens in addition to known pathogens from the original infection. Based on the findings from mNGS, sepsis was divided into two categories: those with positive mNGS results and those

with negative mNGS results. Utilize a comprehensive approach that incorporates mNGS and traditional culture findings, clinical presentation, inflammatory markers, and imaging studies to tailor and refine therapeutic interventions.

Clinical information collection

Data collected from the electronic medical record system includes a range of clinical details. These details encompass demographic information, medical history, infection location, SOFA score, acute physiology and chronic health evaluation II (APACHE II) score, length of hospital stay, intensive care unit (ICU) admission, ICU length of stay, treatments (like vasoactive drugs, mechanical ventilation, and renal replacement therapy), and mortality rate. Furthermore, the laboratory results included various blood parameters such as white blood cell (WBC) count, neutrophils, lymphocytes, neutrophil to lymphocyte ratio (NLR), red blood cell (RBC) count, hemoglobin, and platelets. Additionally, levels of total bilirubin (TBIL), alanine aminotransferase (ALT), aspartate aminotransferase (AST), prothrombin time (PT), activated partial thrombin time (APTT), prothrombin activity (PTA), fibrinogen, d-dimer (D-D), albumin, blood urea nitrogen (BUN), serum creatinine (Scr), lactic acid (Lac), and inflammatory markers like C-reactive protein (CRP), and procalcitonin (PCT) were recorded. The factors to be considered in this study also include the administration of antibiotics, the detecting timing and frequency of the mNGS technique, adjustment of antimicrobial drugs, and various medical costs. These costs consist of total hospitalization fees, average daily hospitalization fees, diagnosis costs (including laboratory diagnosis costs and clinical diagnosis project fees), integrated medical service costs (comprising medical service fees, treatment operation fees, nursing fees, and operation fees), as well as treatment costs (including western medicine fees, antibacterial drug fees, Chinese patent medicine fees, and blood fees), and consumables expenses.

Microbiological analyses

Samples from various infection sites of sepsis, including blood, sputum, bronchoalveolar lavage fluid (BALF), urine, cerebrospinal fluid (CSF), pleural effusion, ascites, pus, tissue, hydropericardium, and bone marrow, were collected in accordance with National Clinical Laboratory Procedures (Shang, 2015). These samples were expeditiously transferred to the microbiology laboratory and mNGS Laboratory at the First Affiliated Hospital of Anhui Medical University for Standard Microbial Culture (Shang, 2015) and mNGS detection procedures (Lu and Wang, 2020). Various pathogens were identified through microbial cultivation and automatic analysis in the microbiology laboratory. Blood culture was emphasized, with each group necessitating two culture bottles, one aerobic and one anaerobic. In cases where the infection site was ambiguous or the specimen was unattainable, a blood specimen was chosen, and the culture protocol was adjusted according to the suspected infection site. Performing all culture types for each patient was deemed unnecessary.

Metagenomic next-generation sequencing experiments and data analysis

In accordance with the manufacturer's guidelines, samples were processed to extract and purify DNA utilizing the QIAamp DNA Micro Kit (QIAGEN, Hilden, Germany). Subsequent DNA library construction was completed using the Qiagen library construction kit (QIAseq Ultralow Input library kit). Quality assessment of the library was performed using the Qubit 3.0 Fluorometer (Invitrogen, Q33216) and Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, USA). Following this, sequencing was conducted on the Illumina Nextseq 550 sequencing platform (Illumina, San Diego, USA) with SE75bp sequencing strategy. The data underwent quality filtering to remove adapters, low-quality, low-complexity, and short sequences, followed by the utilization of Scalable Nucleotide Alignment Program (SNAP; v2.0.1) software to eliminate human-derived sequences aligned with the human reference database (hg38). Subsequently, the non-human data were classified by simultaneous alignment to the reference microbial sequences from bacteria, viruses, fungi, which were obtained from the NCBI Nucleotide database (https://ftp.ncbi.nlm.nih.gov/genomes/) (Chiu and Miller, 2019). Sequence alignment was performed by BLASTN (v2.11.0+) with "megablast" option, and only reads uniquely aligning to microbial taxa were tallied. The final microbial identification results for the samples were then determined. Using peripheral blood samples from healthy donors as negative controls and sterile deionized water as non-template controls. Reads per million mapped reads (RPM) was defined as the number of reads of target pathogen per million of total filtered reads. The identification of positive criteria is not reliant on any singular indicator, including but not limited to the number of identified sequences for particular microorganisms, the ratio of normalized RPM, or the genome coverage of detected species. The formula used to determine the normalized RPM of the pathogen is expressed as follows: RPM of the pathogen = (number of reads mapped to the pathogen $\times 10^6$)/(total number of mapped reads from the given library) (Li et al., 2023b). For bacteria other than *Mycobacterium tuberculosis*, fungi other than *Cryptococcus*, and parasites, identification was based on sequencing coverage ranking within the top 10 of all detected pathogens and the absence in the negative control (NTC), or a sample/NTC RPM ratio exceeding 10. Conversely, for viruses, *M. tuberculosis*, and Cryptococci, identification relied on the presence of at least one specific sequence not found in the NTC, or a sample/NTC RPM ratio >5 (Zhang et al., 2023).

Statistical analysis

The data were analyzed and graphed using GraphPad-Prism 9. Continuous variables were depicted as mean ± standard deviation (SD) for normally distributed data and as median (25th percentile, 75th percentile) for non-normally distributed data. Inter-group comparisons for continuous variables were conducted using independent sample *t*-tests or non-parametric tests. Categorical variables were expressed as numerical values and percentages, and evaluated through chi-square or Fisher's exact test. *P*-values

10.3389/fmicb.2024.1384166

TABLE 1 The baseline characteristics of participants.

Clinical feature	CAS group	HAS group	*P*-value	adj. *P*
Male *N* (%)	62.01% (142/229)	67.09% (53/79)	0.4192	0.6627
Age (years)	58 (48–69.50)	57 (45–68)	0.7876	0.7876
SOFA score	5 (3–9)	8 (5–12)	**<0.0001**	**0.0004**
APACHE II score	14 (9–21)	18 (12–23)	**0.0065**	**0.0321**
Length of admission (day)	22 (13–35.50)	32 (25–42)	**<0.0001**	**0.0003**
ICU admission (%)	43.23% (99/229)	67.09% (53/79)	**0.0003**	**0.0018**
ICU length of stay (day)	16 (9–24)	19 (13–31)	**0.0413**	0.1552
Past history				
Hypertension (%)	37.99% (87/229)	39.24% (31/79)	0.8439	0.9672
Diabetes (%)	16.16% (37/229)	20.25% (16/79)	0.4056	0.9390
Cardiopathy (%)	9.61% (22/229)	21.52% (17/79)	**0.006**	0.0584
Neurogenic disease	10.48% (24/229)	20.25% (16/79)	**0.0259**	0.2104
Chronic kidney dysfunction (%)	10.04% (23/229)	6.33% (5/79)	0.3735	0.9390
Chronic obstructive pulmonary disease (%)	9.17% (21/229)	7.59% (6/79)	0.819	0.9672
Immune-related diseases (%)	28.82% (66/229)	32.91% (26/79)	0.5687	0.9390
Smoking	14.41% (33/229)	8.86% (7/79)	0.2472	0.8630
Drinking	10.48% (24/229)	6.33% (5/79)	0.3726	0.9390
No underlying diseases	26.20% (60/229)	16.46% (13/79)	0.0919	0.5375
Treatments				
Vasoactive drug therapy (%)	39.30% (90/229)	68.35% (54/79)	**<0.0001**	**0.0004**
Mechanical ventilation (%)	37.12% (85/229)	62.03% (49/79)	**0.0001**	**0.0005**
Duration of mechanical ventilation (day)	11 (7–20)	15 (10.50–22.50)	0.0729	0.2031
Renal replacement therapy (%)	15.72% (36/229)	24.05% (19/79)	0.1242	0.2330
Duration of renal replacement therapy (hour)	87 (40.63–205.2)	64 (40–128.5)	0.363	0.3630
Glucocorticoid therapy (%)	63.32% (145/229)	81.01% (64/79)	**0.0034**	**0.0135**
Blood products therapy (%)	68.56% (157/229)	97.47% (77/79)	**<0.0001**	**0.0001**
Case fatality rates	26.64% (61/229)	36.71% (29/79)	0.114	0.3045

CAS, community-acquired sepsis; HAS, hospital-acquired sepsis; SOFA, sequential organ failure assessment; APACHE II, acute physiology and chronic health evaluation II; ICU, intensive care unit; adj.*P*, Adjusted *P*-value. The bold values indicate statistical significance.

were adjusted using the Holm-Sidak method. G*Power software was used to calculate sample size and statistical testing power. A significance level of adjusted *P* (adj.*P*) value < 0.05 was utilized to determine statistical significance.

Results

Baseline characteristics of study participants

Table 1 shows the distribution of baseline characteristics of 308 participants between CAS and HAS, including clinical features, past history, treatments, and mortality rate. Of the 308 participants, 74.35% (229/308) belonged to CAS, while 25.65% (79/308) belonged to HAS. The median days of admission were 22 for CAS and 32 for HAS, with a significant difference (adj.$P = 0.0003$). The median age for both groups was similar, with 58 years for CAS and 57 years for HAS (adj.$P = 0.7876$). The medical histories of the two cohorts of patients encompass a range of conditions such as hypertension, diabetes, heart disease, neurogenic disease, chronic renal insufficiency, chronic obstructive pulmonary disease, and immune-related diseases, etc. Analysis of Table 1 reveals no statistically significant difference in medical history between the two groups (adj.$P > 0.05$). Additionally, the severity of the patient's condition was assessed using SOFA scores and APACHE II scores, which indicated that HAS had a higher severity. Furthermore, HAS had a higher ICU admission rate, longer hospitalization time, received more vasoactive drugs and mechanical ventilation, and required more glucocorticoid therapy, and blood products therapy. HAS had a higher likelihood of developing multiple organ failure and a slightly increased mortality rate. However, there was no statistical difference in mortality rate compared to CAS.

Analysis of laboratory data among the CAS and HAS groups

When examining the correlation between laboratory results in CAS and HAS groups during sepsis diagnosis, it was discovered that the HAS group had lower levels of RBC, hemoglobin, and PTA compared to the CAS group (all adj.$P < 0.05$). Furthermore, the levels of BUN and PT were elevated in the HAS group compared to the CAS group, with statistically significant differences observed (all adj.$P < 0.05$). The results suggest that people in the HAS category had a higher likelihood of experiencing anemia, impaired blood clotting, and damage to kidney function (see Supplementary Table 1).

Comparison clinical diagnostic outcome of mNGS and traditional culture

The study found that the detection rates of mNGS and traditional culture were 87.99% (271/308) and 30.52% (94/308) in every instance. There were 387 mNGS specimens and 552 culture specimens collected for mNGS and culture detection, with positive rates of 85.27% (330/387) and 22.28% (123/552), respectively. Certainly, the mNGS positive rate was increased by almost 60% in both cases and samples compared to traditional culture, a statistically significant difference shown in the Chi-squared test of positive rate (adj.$P < 0.001$). Moreover, within the CAS cohort, the detection rate of mNGS (202/229, 88.21%) was approximately 63% greater than that of conventional culture (59/229, 25.76%), showing a statistically significant disparity (adj.$P < 0.001$). Similarly, the HAS group also showed comparable results [(69/79, 87.34%) vs. (35/79, 44.30%), adj.$P < 0.001$; Figure 1A]. These findings suggest that the detection rates of mNGS were similar in the CAS and HAS groups, and both groups had significantly higher detection rates compared to traditional culture. Further study showed that 90 patients were positive for mNGS and traditional culture, and the coincidence rates of mNGS and traditional culture in detecting the same pathogen were 81.03% (47/58) for CAS and 84.35% (27/32) for HAS, respectively.

The distribution of infection sites in this study is shown in Figure 2. Obviously, lower respiratory infection was the most common in both the CAS group and HAS group. In each infection site, the positive rate of mNGS was higher than that of culture, as shown in Figure 1B, particularly in lower respiratory infection (adj.$P < 0.001$), bloodstream infection (adj.$P < 0.001$), central nervous system (CNS) infection (adj.$P < 0.001$), and abdominal infection (adj.$P < 0.01$). However, the disparity in cases of urinary infection, skin and soft tissue infection, and pericarditis did not reach statistical significance due to the limited sample size. Similar results were also seen in the CAS group (Figure 1C). Nevertheless, within the HAS cohort, notable variances in detection rates were solely noted in lower respiratory (Figure 1D).

Figure 3 shows the distribution of assorted samples. The most common mNGS specimen was blood (175/387, 45.22%), followed by BALF (72/387, 18.60%), CSF (60/387, 15.50%), sputum (45/387, 11.63%), ascites (14/387, 3.62%), and pus (8/387, 2.07%), pleural effusion (7/387, 1.81%), urine (2/387, 0.52%), skin and soft tissue (2/387, 0.52%), pericardial effusion (1/387, 0.26%), and bone marrow (1/387, 0.26%) in all cases. These results demonstrate a notably higher positive rate of mNGS compared to culture, particularly in blood (adj.$P < 0.001$), sputum (adj.$P < 0.01$), BALF (adj.$P < 0.001$), CSF (adj.$P < 0.001$; Figures 3A–C). The CAS group found similar results in subtypes of blood (adj.$P < 0.001$), sputum (adj.$P < 0.01$), BALF (adj.$P < 0.001$), and CSF (adj.$P < 0.001$; Figures 3D–F). Likewise, similar findings were observed in blood, and BALF in the HAS group, showing statistically significant variances (Figures 3G–I).

Comparison of pathogenic characteristics by mNGS and traditional culture

In this study, mNGS detected 797 pathogens, including 387 bacteria, 252 viruses, 137 fungi, and 21 atypical pathogens. The detection rates of bacteria and fungi using mNGS were markedly superior to those achieved through traditional culture methods (all adj.$P < 0.001$). Regarding the examination of bacteria, the detection rates of mNGS and conventional culture for gram-negative bacteria were notably greater than those for gram-positive bacteria (all adj.$P < 0.001$). Moreover, mNGS identified viruses and unusual pathogens that are undetectable through conventional methods, demonstrating its distinct advantages (Figure 4A, all adj.$P < 0.001$). Similar findings were observed in the CAS and HAS groups, as shown in Figures 4B, C. Positive results from mNGS for different pathogens were significantly higher in both groups than traditional culture methods (all adj.$P < 0.001$). However, it is important to note that atypical pathogens were exclusively identified through mNGS only in the CAS group.

The research also found that mNGS detection of mixed pathogens named after two or more pathogens was more common (174/271, 64.21%), while traditional culture was more frequently used for single pathogen (65/94, 69.15%). In both the total cases and CAS, the positive rates of mNGS significantly exceeded that of traditional culture, whether it was for single pathogen or mixed pathogens (all adj.$P < 0.05$). Similarly, mNGS was superior to traditional culture in detecting mixed pathogens in the HAS group (adj.$P < 0.001$). Nevertheless, when identifying a single pathogen, there was no statistically significant variance found between the two techniques (Figure 4D, adj.$P = 0.2595$). The findings clearly show that the detection rate of mixed pathogens using mNGS was significantly higher than that of single pathogens across different subgroups (adj.$P < 0.001$ for overall cases, adj.$P < 0.001$ for CAS, adj.$P < 0.001$ for HAS).

Analysis of pathogens detected by mNGS and traditional culture

Subsequently, we further studied the common pathogen types in the CAS and HAS groups. In the CAS group, Human herpesvirus ($n = 121$) was found to be the most hackneyed among the top 10 pathogens tested by mNGS, while *Klebsiella pneumoniae* ($n = 39$), and *Aspergillus* ($n = 37$) were the most frequent bacteria and fungi. However, only bacteria and fungi were detected in traditional

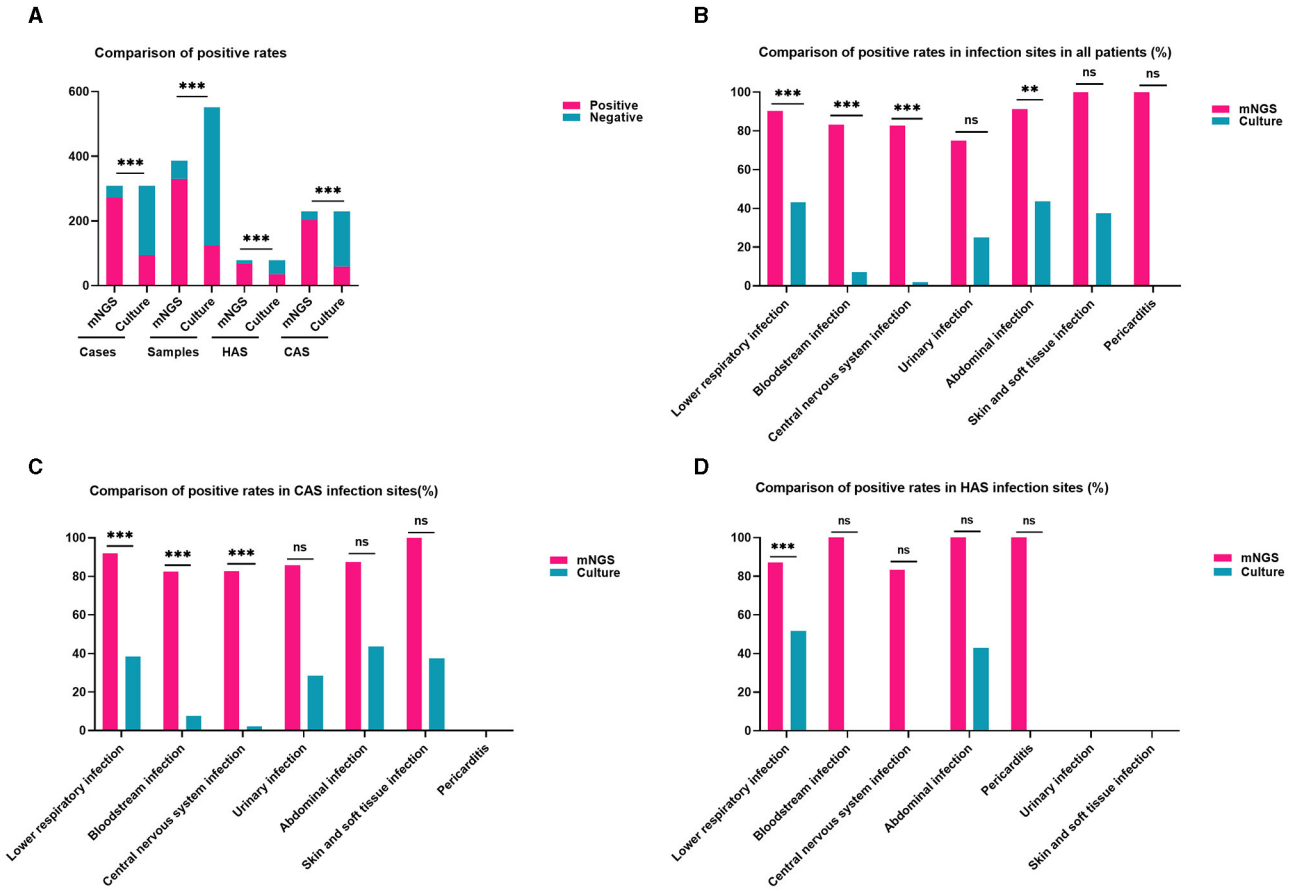


FIGURE 1
Comparison of positivity rates between metagenomic next-generation sequencing (mNGS) and traditional culture methods. **(A)** The positive rates of mNGS outweigh the culture method in all cases, samples, HAS, and CAS (all adj.$P < 0.001$). **(B)** The detection rate of mNGS in different infection locations was higher compared to culture in all patients, particularly in cases of lower respiratory infections, bloodstream infections, central nervous system infections, and abdominal infections (all adj.$P < 0.01$). **(C)** In the CAS group, mNGS showed higher positivity rates than culture in lower respiratory infections, bloodstream infections, and central nervous system infection sites were higher than in culture (all adj.$P < 0.001$). **(D)** The detection rate of mNGS in the lower respiratory tract was higher than that of culture in the HAS group (adj.$P < 0.001$). Significance levels: $^{**}P < 0.01$; $^{***}P < 0.001$; ns, no statistically significant variation.

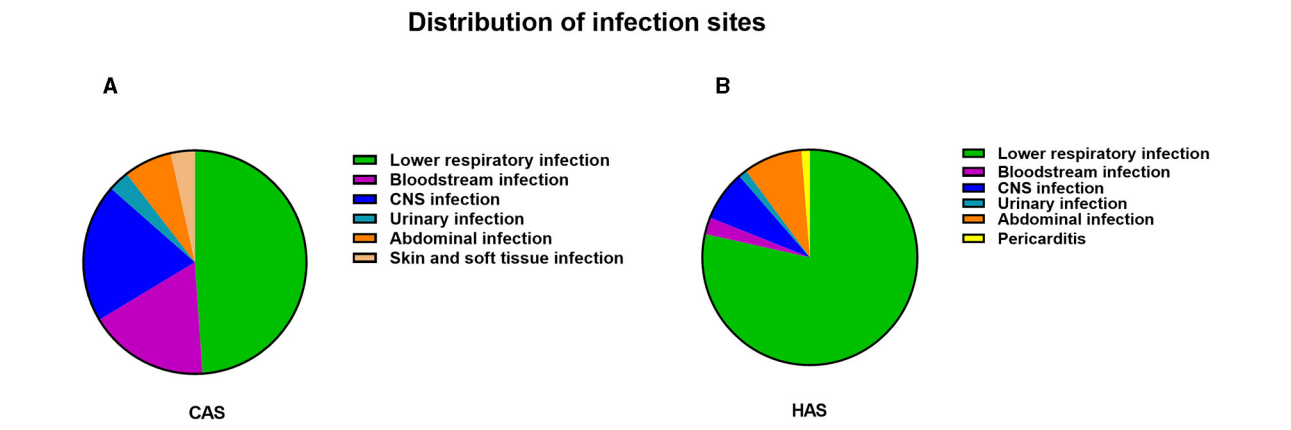


FIGURE 2
The distribution of infection in CAS group and HAS group. **(A)** Distribution of infection sites in CAS group. **(B)** Distribution of infection sites in HAS group. Lower respiratory tract infections were found to be the predominant site of infection, regardless of whether the patients were in the CAS or HAS group.

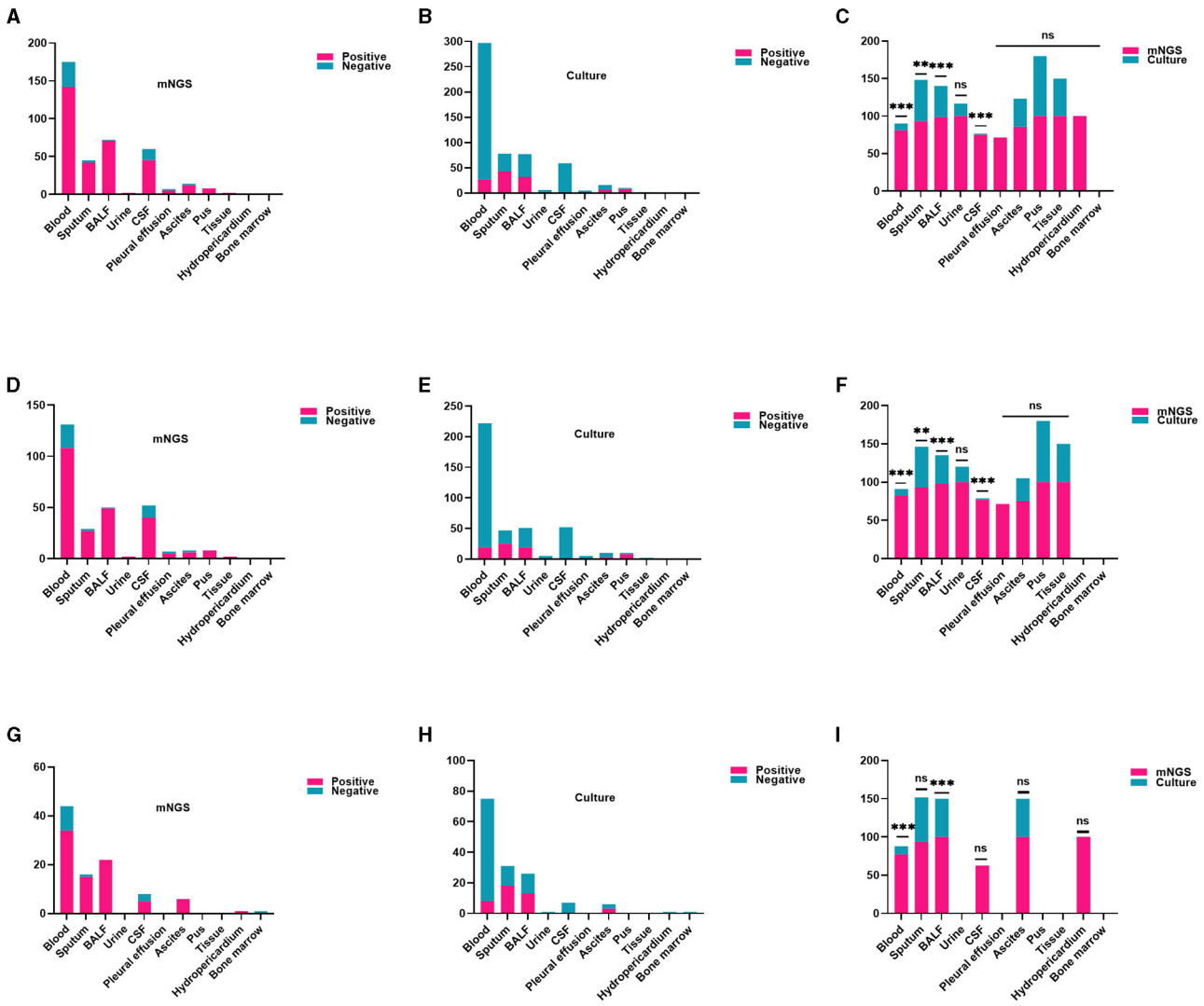


FIGURE 3
Comparison of the positive rate of mNGS and traditional culture in different samples. **(A–C)** In every patient, the detection rate of mNGS in each specimen was notably higher compared to culture, particularly in blood, sputum, BALF, CSF (all adj.$P < 0.01$). **(D–F)** Within the CAS cohort, the detection rate of mNGS was significantly higher than that of traditional culture for various sample types including blood, sputum, BALF, and CSF (all adj.$P < 0.01$). **(G–I)** In the HAS group, mNGS results showed a higher positive rate in blood and BALF (all adj.$P < 0.01$). CSF, cerebrospinal fluid; BALF, bronchoalveolar lavage fluid. Significance levels: $^{**}P < 0.01$; $^{***}P < 0.001$; ns, no statistically significant variation.

cultures, with *K. pneumoniae* ($n = 15$) and *Candida albicans* ($n = 16$) being the commonest. Additionally, mNGS detected 19 cases of *M. tuberculosis*, demonstrating advantages beyond traditional culture. In the HAS group, Human herpesvirus ($n = 42$) remained the most frequently identified pathogen using mNGS, although the bacteria and fungi detected differed compared to those found in the CAS group. The most frequently bacteria were *Acinetobacter baumannii* ($n = 25$), and the fungi detected were *C. albicans* ($n = 23$). Similarly, the most frequently found bacteria and fungi identified through conventional culture methods were comparable to those identified through mNGS. Nevertheless, whether in the CAS group or the HAS group, the rates of bacterial and fungal detection by traditional culture were significantly lower compared to mNGS. The top 10 specific pathogen is illustrated in Figure 5. Following that, we conducted a detailed analysis of prevalent Gram-positive bacteria (top 5), Gram-negative bacteria (top 5), fungi (top 5), various viruses, and atypical pathogens. For further information, please consult Supplementary Figure 1A.

Comparison of pathogen types between single pathogen and mixed pathogens

Next, we conducted an etiological analysis at the level of single pathogen and mixed pathogens. Regarding the single pathogen identified through the aforementioned methodologies, mNGS revealed that bacteria were the predominant pathogen in the CAS group, followed by viruses, fungi, and atypical pathogens. However, traditional culture methods solely detected bacteria and fungi. In the HAS group, mNGS predominantly identified bacteria, followed by fungi and viruses, with no atypical pathogens detected. Conversely, traditional culture methods exhibited limited

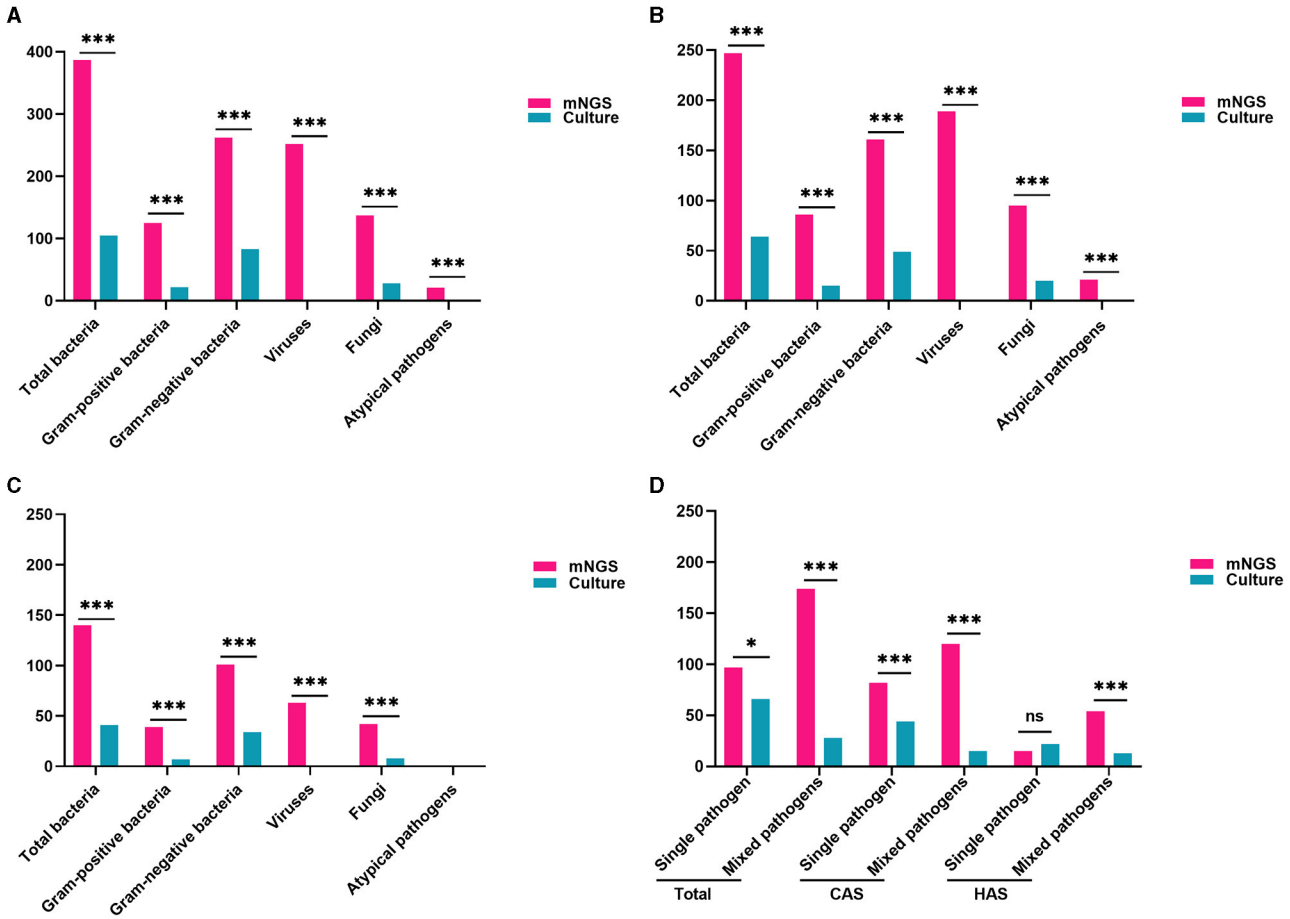


FIGURE 4
Comparison of pathogen traits identified by mNGS and conventional culture methods. **(A)** Comparison of pathogen types and positive rate in all patients (all adj.$P < 0.001$). **(B)** Comparison of pathogen types and positive rate in CAS group (all adj.$P < 0.001$). **(C)** Comparison of pathogen types and positive rates in the HAS group (all adj.$P < 0.001$, except for atypical pathogens). **(D)** mNGS was more common for mixed pathogen infections, while traditional culture was more frequent for single pathogen detection. With the exception of identifying a single pathogen in the HAS group, the positive rate of mNGS significantly surpassed that of traditional culture (all adj.$P < 0.05$). Significance levels: $^{*}P < 0.05$; $^{***}P < 0.001$; ns, no statistically significant variation.

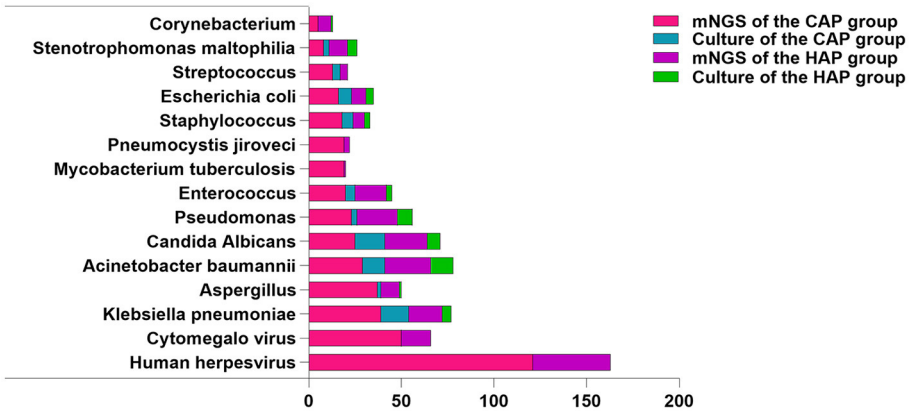


FIGURE 5
Distribution of pathogen species detected by mNGS and culture. Tested pathogens are represented on the *X*-axis by their counts.

detection of pathogenic types, with bacteria being the most frequently identified.

This study demonstrates a higher prevalence of mixed pathogens in the etiology of sepsis. The utilization of mNGS revealed that bacterial-viral co-detection were more frequently observed in both the CAS and HAS groups, whereas traditional culture methods indicated a higher incidence of bacterial-fungal co-infections specifically in the CAS group. Conversely, multiple bacterial infections were more commonly observed in the HAS group. These findings highlight that the single pathogen commonly found in mNGS-based pathogen detection in sepsis is bacteria, and bacterial-viral co-detection is the most prevalent form of mixed pathogens (Table 2). For specific common bacterial strains, please refer to Supplementary Figures 1B–D.

TABLE 2 Comparison of types between single pathogen and mixed pathogens.

Group/type	CAS group		HAS group	
	mNGS	Culture	mNGS	Culture
Single pathogen	82	44	15	22
Bacteria	40	35	9	20
Fungi	11	9	4	2
Viruses	27		2	
Atypical pathogens	4			
Mixed pathogens	120	15	54	13
Multiple bacteria	10	6	11	9
Multiple viruses	5		3	
Multiple fungi	1	1		1
Bacteria combined with fungi	13	8	9	3
Bacteria combined with viruses	41		15	
Bacteria combined with atypical pathogens	2			
Viruses combined with atypical pathogens	7			
Viruses combined with fungi	10		4	
Bacteria, viruses, and fungi	26		12	
Bacteria, viruses, and atypical pathogens	3			
Viruses, fungi, and atypical pathogens	1			
Bacteria, viruses, fungi, and atypical pathogens	1			

Analyze the impact of antibiotic exposure, timing and frequency of mNGS testing on mNGS results

The purpose of this research was to investigate how antibiotics exposure, timing, and frequency of mNGS testing could affect the results of mNGS. The results showed that conducting multiple mNGS tests in individuals with CAS and HAS led to a higher rate of positive results compared to a single test, although this discrepancy was not statistically significant (all adj.$P > 0.05$). Furthermore, the positive rate of mNGS was not affected by antibiotics exposure or the timing of mNGS detection, regardless of patients' group (CAS or HAS). Additionally, it was noted that the mortality rate was elevated in the mNGS positive group when compared to the mNGS negative group within both the CAS and HAS groups. It is important to mention that there was no statistically significant distinction between the two groups, as shown in Supplementary Table 2.

Optimizing antimicrobial therapy based on mNGS and culture results

This study examined the optimization of antimicrobial therapy for sepsis by utilizing both mNGS and culture results. In contrast to the clinical manifestations, inflammatory markers, routine culture findings, and therapeutic outcomes of sepsis, the positive mNGS results in 235 cases demonstrated a favorable influence on clinical management. These findings contribute to the elucidation of pathogenic diagnosis and can be regarded as a well-matched clinical diagnostic group. Conversely, the mNGS results of 73 cases did not exert any discernible impact on clinical treatment. Out of these, 36 cases were positive for mNGS but deemed to be mismatched with the clinical diagnosis group, while 37 cases were negative for mNGS. In the CAS group, there were 126 cases where antimicrobial therapy was optimized based on mNGS and traditional culture results. Among these cases, 41 cases had down-regulated antibiotics, two cases switched to antiviral treatment, and 14 cases switched to anti-tuberculosis therapy. Additionally, within the cohort of patients in the HAS group, a total of 48 mNGS positive patients underwent optimization of their anti-infection regimens, with 17 cases having antibiotic treatment down-regulated. Nevertheless, in the mismatch clinical diagnosis group, there were six cases where the antibiotic regimen was optimized using traditional culture results, of which one case was down-regulated in the CAS and HAS groups, respectively (Table 3).

Analyze the influence of optimizing antimicrobial therapy on mortality rates

When considering the impact of optimizing antimicrobial therapy on mortality rates, it was noted that patients with HAS had a higher mortality rate than patients with CAS, regardless of whether treatment was optimized using mNGS co-culture results or solely mNGS results. However, this difference was not statistically significant (all adj.$P > 0.05$). Due to the small sample size, there was still no significant difference in mortality rates between CAS

10.3389/fmicb.2024.1384166

TABLE 3 Optimizing antimicrobial therapy and mortality rate analysis based on mNGS and culture results.

Project	CAS group		HAS group		*P*-value	adj.*P*
	Number	Mortality rate	Number	Mortality rate		
mNGS positive						
Match clinical diagnosis	169	27.81% (47/169)	66	40.91% (27/66)	0.0612	0.5313
Adjusting treatment	126	30.95% (39/126)	48	45.83% (22/48)	0.0767	0.5843
Adjustment by mNGS and culture	38	50% (19/38)	20	60% (12/20)	0.5826	0.9978
Up-escalated	26		12			
De-escalated	12		8			
Adjustment by mNGS	88	22.73% (20/88)	28	35.71% (10/28)	0.2159	0.9122
Up-escalated	43		19			
De-escalated	29		9			
Adjust to antiviral treatment	2		0			
Adjust to anti-tuberculosis treatment	14		0			
No changes	43	18.60% (8/43)	18	27.78% (5/18)	0.4989	0.996
Mismatch clinical diagnosis	33	24.24% (8/33)	3	33.33% (1/3)	>0.9999	1
Adjustment by culture	4	100% (4/4)	2	50% (1/2)	0.3333	0.974
Up-escalated	3		1			
De-escalated	1		1			
Adjustment by experience	12	8.33% (1/12)	1	0% (0/1)	>0.9999	1
Up-escalated	8		1			
De-escalated	2		0			
Adjust to anti-tuberculosis treatment	2		0			
No changes	17	17.65% (3/17)	0	0% (0/0)	>0.9999	1
mNGS negative	27	22.22% (6/27)	10	10% (1/10)	0.6471	0.9981
Adjustment by experience	17	23.53% (4/17)	7	14.29% (1/7)	>0.9999	1
Up-escalated	10		6			
De-escalated	4		1			
Adjust to antiviral treatment	1		0			
Adjust to anti-tuberculosis treatment	1		0			
Adjust to antibacterial treatment	1		0			
No changes	10	20% (2/10)	3	0% (0/3)	>0.9999	1

and HAS patients in the mNGS positive mismatch clinical diagnosis group when optimizing treatment based on traditional culture results. However, for the patients that were left, there was no significant difference in mortality rates between the two groups after accounting for antimicrobial treatment guided by clinical knowledge (Table 3).

Comparison of medical costs between CAS and HAS

It is widely recognized that sepsis frequently manifests as a critical condition accompanied by substantial treatment expenses, thereby imposing a significant economic burden on both families and society. In order to gain deeper insights into the medical expenses of sepsis patients, this study incorporates various components of medical expenditure, including total hospitalization costs, average daily hospitalization expenses, diagnostic charges, comprehensive medical service fees, treatment expenditures, and consumables expenses. All monetary figures are denominated in US dollars. All medical expenses are expressed in US dollars, utilizing the exchange rate of Renminbi (RMB) to United States dollar (USD) as of September 3, 2023. The findings indicate that the aforementioned expenditures incurred by patients with the HAS group were considerably greater compared to those with the CAS group, and this disparity was statistically significant (all adj.$P < 0.01$). Subsequently, to assess the potential influence

of mNGS results on healthcare costs, the investigation revealed no association between divergent mNGS outcomes and various medical expenditures, irrespective of the CAS or HAS group (Supplementary Tables 3–5, all adj.$P > 0.05$).

Discussion

The guidelines set forth by the Surviving Sepsis Campaign recommend the prompt initiation of antibiotic treatment in adults deemed at risk of sepsis or septic shock, ideally within 1 h of identification (Evans et al., 2021). This practice is widely acknowledged as a crucial intervention to decrease mortality rates in septic patients, as supported by various studies (Kumar et al., 2006; Seymour et al., 2017; Bollinger et al., 2023). Factors that primarily influence the suitable antimicrobial treatment include pathogenic microorganisms, infection source (community or hospital), infection location, immune system status, existing medical conditions, local epidemiological information, and the presence of risk factors for antimicrobial resistance in patients (Gage-Brown et al., 2022). Nevertheless, the low pathogen detection rate of sepsis poses a considerable obstacle to quickly and accurately determining the cause. The impartiality of mNGS renders it highly promising for overcoming diagnostic challenges in scenarios where conventional methods may prove inadequate, such as culture-negative sepsis or polymicrobial infections. Thus, the objective of this study is to assess the efficacy and significance of mNGS in the identification of sepsis, particularly in filling the gap in pathogen detection for both community and hospital-acquired cases, with the aim of informing clinical practice and optimizing patient care.

The findings indicated that the HAS group exhibited elevated scores on the APACHE II and SOFA scales, along with an increased rate of ICU admission, prolonged hospital stays, and a higher likelihood of receiving mechanical ventilation, vasoactive drugs, blood products, and glucocorticoid treatment. These results significantly differed from those of the CAS group, aligning with previous research findings (Westphal et al., 2019; Tonai et al., 2022). The findings suggest a higher incidence of organ dysfunction in patients with HAS. Differently, our analysis identified a greater need for blood products and glucocorticoid therapy in this cohort. Additionally, our investigation uncovered manifestations of anemia, coagulation abnormalities, and renal dysfunction in the HAS group, potentially contributing to the observed distinctions. Meanwhile, our study revealed that there were no statistically significant disparities in the prevalence of past history among sepsis patients originating from CAS and HAS, suggesting that the medical history of sepsis patients remains basic consistent regardless of the infection place. We recommend that clinicians should contemplate employing similar diagnostic and therapeutic approaches for both patient cohorts.

Consistent with prior research (Duan H. et al., 2021; Sun et al., 2022), our study also found a notably elevated detection rate of mNGS in contrast to conventional culture techniques. But our research results also indicate that the high positivity rate of mNGS is not affected by the infection site, different infection sites, or sample types, which is encouraging. In addition, our study emphasizes the broader range of pathogens and higher sensitivity of mNGS in identifying sepsis-causing pathogens compared to conventional culture techniques, especially in detecting viruses and unusual pathogens. This finding underscores the robustness of mNGS in identifying pathogens, regardless of the infection site, sample type, community or nosocomial source, or etiological classification. It should be noted that antibiotic exposure may diminish the sensitivity of blood culture, whereas its impact on mNGS is minimal (Miao et al., 2018; Cheng et al., 2019). Our findings indicate that antibiotic exposure both prior to and following hospitalization, as well as varying detection opportunities, did not significantly impact the positivity rate of mNGS. However, an increase in detection frequency was observed to potentially enhance the positivity rate of mNGS, albeit without statistical significance. Thus, in the context of pathogen detection for sepsis, mNGS demonstrates a superior positive detection rate for pathogens when compared to conventional culture methods. This article posits that variables such as infection site, infection source, sample type, antibiotic exposure, and detection time have minimal influence, thereby suggesting that mNGS is a more efficient approach for pathogen identification.

As a genetic diagnostic tool, mNGS presents the added benefit of not necessitating prior screening for a specific range of etiologies during pathogen identification. This is especially advantageous when traditional culture methods fail to detect microbial agents such as *M. tuberculosis*, mycoplasma, chlamydia, and viruses in a timely manner. Previous retrospective studies have demonstrated that patients with sepsis complicated by viral infection exhibit a more severe clinical presentation and a less favorable prognosis (Duan L. W. et al., 2021). Nevertheless, the identification of viruses presents a significant obstacle for healthcare professionals. Our research revealed that Human herpesvirus was the predominant pathogen in cases of sepsis when utilizing mNGS. In the context of Human herpesviruses, populations generally display susceptibility. A thorough evaluation is necessary in a clinical setting, incorporating clinical manifestations, inflammatory markers, specific viral load quantification, as well as serum levels of immunoglobulin G and immunoglobulin M, to ascertain the existence of active infection. This finding underscores the importance of recognizing viral infection in the diagnosis and management of sepsis, warranting adequate attention. Our research indicated that bacteria, particularly Gram-negative bacilli, were predominant in the identification of sepsis pathogens through mNGS and conventional culture techniques. In cases of CAS and HAS, *K. pneumoniae* and *A. baumannii* were the bacteria most commonly identified using the methods mentioned. These results align with previous studies but highlight the superior sensitivity of mNGS over traditional culture for bacterial detection in sepsis (Geng et al., 2021; Sun et al., 2022; Qin et al., 2023; Zhou et al., 2023).

Sepsis caused by *M. tuberculosis* is commonly seen in individuals with human immunodeficiency virus (HIV), but it can also occur in those without HIV, a fact often overlooked by healthcare professionals. A delay in initiating anti-tuberculosis treatment in cases of sepsis is associated with increased mortality rates (Adegbite et al., 2023). The accurate diagnosis of *M. tuberculosis* septicemia is of significant importance, yet it is frequently misdiagnosed and overlooked in clinical practice. Our study found that mNGS method successfully identified 19 cases of *M. tuberculosis* that were missed by conventional culture methods, leading to improved detection capabilities

and significantly reduced detection time. However, the lack of supporting literature necessitates further investigation through future research collaborations to confirm this finding across various patient populations and medical settings. Overall, mNGS shows distinct benefits in identifying *M. tuberculosis*. In the realm of fungal detection, traditional culture methods are characterized by their time-consuming and labor-consuming, whereas mNGS demonstrates superior sensitivity and specificity. Our research revealed that *Aspergillus* and *C. albicans* were prevalent fungi identified through mNGS and traditional culture, respectively, while mNGS also detected challenging-to-culture strains such as Pneumocystis Jirovecii and Mucoraceae. Overall, discrepancies in the strains detected by mNGS and culture techniques highlight the comprehensive and sensitive nature of mNGS, enabling the identification of elusive strains.

The advancement of detection techniques has led to a gradual rise in the identification rate of atypical pathogens, dispelling the longstanding notion that they are rare pathogen. On the contrary, these pathogens are quite prevalent and often give rise to sporadic or epidemic outbreaks. This study utilized mNGS to identify five atypical pathogens that are challenging to detect using conventional culture methods, demonstrating the efficacy of mNGS for pathogen detection. Notably, Chlamydia psittaci, a rare clinical strain, was detected through mNGS in this study. Chlamydia psittaci, a zoonotic pathogen, frequently presents with atypical symptoms resembling respiratory tract infections, such as high fever, headache, and cough, ultimately progressing to pneumonia and multi-organ failure (Zhang et al., 2020). The shortcomings of traditional etiological and serological detection techniques contribute to low positivity rates and potential misdiagnoses, underscoring the need for alternative methods like mNGS (Zhang et al., 2020; Liang et al., 2022). This research emphasizes the clinical advantages of mNGS in identifying sepsis-causing pathogens, particularly in challenging cases.

Additionally, this study observed that mNGS not only identified the types of pathogens in sepsis patients, but also revealed a higher rate of mixed pathogen infections, which contradicts the findings of traditional culture-based detection methods. Similar findings were also reported in a study investigating pathogen detection in the blood of critically ill patients, suggesting that mNGS outperforms blood culture in detecting mixed infections (Geng et al., 2021). Given these circumstances, it is imperative for clinicians to promptly identify the presence of mixed pathogen infections in sepsis patients and intervene early, as this could potentially benefit the patients. The results of the research showed significant differences in the microorganisms detected using mNGS and conventional culture techniques in cases of sepsis acquired in the community or in hospitals, especially when multiple pathogens were involved. mNGS detection exhibited a higher prevalence of bacterial combined viral among CAS and HAS, whereas traditional culture methods identified a greater number of bacterial combined fungal infections in the CAS group and multiple bacterial infections in the HAS group. These variations may be ascribed to factors such as patient origin, immune status, infection sites, local epidemic strains, and pathogen selectivity, antibiotic usage.

However, in sepsis pathogen detection, distinguishing between infection, colonization, and contamination is challenging for healthcare professionals. To reduce inaccuracies, measures like using positive and negative controls and following standard procedures are taken. Aseptic techniques were used in this study during sample collection and testing to prevent contamination. The identification of pathogenic microorganisms based on comprehensive analysis and judgment of the location of infection, microbial properties, patient symptoms, inflammatory markers, imaging findings, and inspection report (Li et al., 2023a). In the analysis of the mNGS report, pathogenic microorganisms, commensal microorganisms, and contaminants are differentiated based on criteria such as confidence level, specific sequence count, relative abundance, and coverage (Wang, 2021).

Subsequently, we conducted a more comprehensive investigation into the optimization of antimicrobial therapy based on the outcomes obtained from mNGS and/or conventional culture methods. Our findings revealed that among the 180 patients, treatment adjustments were made, with 60 patients reducing antibiotic usage, two patients discontinuing antibiotics in favor of antiviral therapy, and 14 patients transitioning to anti-tuberculosis treatment. These interventions successfully circumvented the misuse and excessive utilization of antibiotics, thereby optimizing their rational application and effectively mitigating the emergence of antibiotic resistance. Nevertheless, when considering the utilization of mNGS co-culture results vs. solely optimizing treatment with mNGS results, it was noted that the mortality rate among HAS patients was higher compared to CAS patients, but the disparity was not deemed statistically significant (all $P > 0.05$). From a clinical standpoint, the elevated mortality rate among HAS patients aligns with previous research trends (Tonai et al., 2022). Furthermore, when considering statistical power calculations, utilizing an effect size of 0.5 and a total sample size of 58 adjusted by mNGS and culture results yielded a calculated statistical power of 0.96. Similarly, with an effect size of 0.5 and a total sample size of 116 adjusted by mNGS, the calculated statistical power was 0.99. Although the observed difference did not reach statistical significance at the present sample size, the calculated statistical power indicates that the sample size may have been insufficient to detect a difference between the two groups. Therefore, the conclusion regarding the higher mortality rate in patients with HAS, while not statistically significant, may still suggest a potential trend. Notably, recent research has demonstrated contrasting findings, suggesting that tailoring antibiotic regimens using mNGS could enhance survival rates in sepsis patients (Zuo et al., 2023). The variations in research findings may be ascribed to factors such as the etiology of sepsis, location of infection, severity of the condition, sample size, and overlooking potential confounding variables such as the results of RNA detection in samples. Consequently, further inquiry is warranted to elucidate this matter.

The cost of sepsis treatment exhibits considerable variation across different countries, generally being quite high (van den Berg et al., 2022). Currently, there is a scarcity of data regarding the medical expenses associated with CAS and HAS. Our study demonstrates a significant disparity in medical costs between patients with HAS and those with CAS, with the former incurring substantially higher expenses. This statistically significant difference underscores the heavier financial burden faced by

patients with HAS, thereby emphasizing the need for clinicians to prioritize the prevention of nosocomial infections. The early identification and prompt treatment of sepsis, as well as the prevention of its progression, are crucial in reducing the overall hospitalization burden associated with sepsis in a clinical setting (Paoli et al., 2018). Consequently, we conducted a comprehensive investigation into the influence of mNGS results on medical expenditures. In the context of expenses associated with sepsis treatment, mNGS does not yield substantial advantages. Given the high sensitivity of mNGS in identifying diverse bacterial species, medical professionals should consider prioritizing its early implementation over alternative approaches.

Our study is subject to certain limitations in terms of research design, interpretation of mNGS results, and evaluation of clinical value. It is crucial to mention that this research is a retrospective study carried out at one center, with a limited sample size. Consequently, while certain trends were observed in the context of CAS and HAS, statistical significance was not achieved. To address this, we plan to expand our sample size by including data from multiple centers in future research. Secondly, our findings indicate that mNGS identified a higher prevalence of mixed etiological infections in the etiological identification of sepsis. This paper solely examines the prevalent pathogen types in mixed infections, neglecting to specifically analyze the composition of mixed etiology, thereby leading to an inadequate comprehension of pathogenic microorganisms. Consequently, it is imperative to undertake further endeavors to scrutinize the precise types of mixed pathogens for enhanced sepsis treatment. Moreover, this study fails to optimize the processes of mNGS and traditional culture in the evaluation of clinical value. It does not provide evidence on the potential impact of early mNGS detection in sepsis on optimizing disease progression, reducing medical intervention and costs, and improving prognosis, further research is warranted.

Conclusion

Overall, mNGS technology demonstrates superiority over traditional culture techniques in identifying the causative agent of sepsis, regardless of factors such as antibiotic exposure, time to detection, sampling frequency, infection site, or sample type. mNGS is particularly effective in detecting polymicrobial infections involving bacteria and viruses, enabling the identification of viral, atypical, and *M. tuberculosis* pathogens that may be overlooked by conventional cultures. Optimizing therapy with mNGS reduces antibiotic overuse without compromising prognosis. Visibly, mNGS presents distinct benefits in the realm of microbial diagnosis and antibiotic selection for sepsis, bearing significant clinical importance. The severe illness and financial burden experienced by patients with HAS underscore the necessity of infection control measures in healthcare settings. Our future research endeavors will focus on an optimization of the clinical implementation of mNGS and conventional culture techniques, with the aim of elucidating the specific effects of mNGS on sepsis outcomes, healthcare delivery, economic implications, and future prospects.

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.

Ethics statement

The studies involving humans were approved by the First Affiliated Hospital of Anhui Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

DZ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. XL: Data curation, Formal analysis, Software, Writing – original draft. YW: Data curation, Writing – original draft. YZ: Data curation, Writing – original draft. HZ: Conceptualization, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The research was funded by the Anhui Provincial Health Research Project (AHWJ2023A10095) and the Anhui Provincial University Natural Science Research Project (2023AH040079).

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.2024.1384166/full#supplementary-material

References

Adegbite, B. R., Elegbede-Adegbite, N. O. M., Edoa, J. R., Honkpehedji, Y. J., Zinsou, J. F., Dejon-Agobé, J. C., et al. (2023). Clinical features, treatment outcomes and mortality risk of tuberculosis sepsis in HIV-negative patients: a systematic review and meta-analysis of case reports. *Infection* 51, 609–621. doi: 10.1007/s15010-022-01950-4

Bollinger, M., Frère, N., Shapeton, A. D., Schary, W., Kohl, M., Kill, C., et al. (2023). Does prehospital suspicion of sepsis shorten time to administration of antibiotics in the emergency department? A retrospective study in one university hospital. *J. Clin. Med.* 12:5639. doi: 10.3390/jcm12175639

Buchman, T. G., Simpson, S. Q., Sciarretta, K. L., Finne, K. P., Sowers, N., Collier, M., et al. (2020). Sepsis among medicare beneficiaries. *Crit. Care Med.* 48, 276–288. doi: 10.1097/CCM.0000000000004224

Cheng, M. P., Stenstrom, R., Paquette, K., Stabler, S. N., Akhter, M., Davidson, A. C., et al. (2019). Blood culture results before and after antimicrobial administration in patients with severe manifestations of sepsis: a diagnostic study. *Ann. Intern. Med.* 171, 547–554. doi: 10.7326/M19-1696

Chiu, C. Y., and Miller, S. A. (2019). Clinical metagenomics. *Nat. Rev. Genet.* 6, 341–355. doi: 10.1038/s41576-019-0113-7

Duan, H., Li, X., Mei, A., Li, P., Liu, Y., Li, X., et al. (2021). The diagnostic value of metagenomic next?generation sequencing in infectious diseases. *BMC Infect. Dis.* 21:62. doi: 10.1186/s12879-020-05746-5

Duan, L. W., Qu, J. L., Wan, J., Xu, Y. H., Shan, Y., Wu, L. X., et al. (2021). Effects of viral infection and microbial diversity on patients with sepsis: a retrospective study based on metagenomic next-generation sequencing. *World J. Emerg. Med.* 12, 29–35. doi: 10.5847/wjem.j.1920-8642.2021.01.005

Evans, L., Rhodes, A., Alhazzani, W., Antonelli, M., Coopersmith, C. M., French, C., et al. (2021). Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med.* 47, 1181–1247. doi: 10.1007/s00134-021-06506-y

Gage-Brown, A., George, C., Maleki, J., Singh, K. P., and Muhi, S. (2022). Is piperacillin-tazobactam an appropriate empirical agent for hospital-acquired sepsis and community-acquired septic shock of unknown origin in australia? *Healthcare* 10:851. doi: 10.3390/healthcare10050851

Geng, S., Mei, Q., Zhu, C., Fang, X., Yang, T., Zhang, L., et al. (2021). Metagenomic next-generation sequencing technology for detection of pathogens in blood of critically ill patients. *Int. J. Infect. Dis.* 103, 81–87. doi: 10.1016/j.ijid.2020.11.166

Gu, W., Deng, X., Lee, M., Sucu, Y. D., Arevalo, S., Stryke, D., et al. (2021). Rapid pathogen detection by metagenomic next-generation sequencing of infected body fluids. *Nat. Med.* 27, 115–124. doi: 10.1038/s41591-020-1105-z

Ishihara, T., Watanabe, N., Inoue, S., Aoki, H., Tsuji, T., Yamamoto, B., et al. (2020). Usefulness of next-generation dna sequencing for the diagnosis of urinary tract infection. *Drug Discov. Ther.* 14, 42–49. doi: 10.5582/ddt.2020.01000

Kim, H., Oh, D. K., Lim, S. Y., Cho, Y., Park, S., Suh, G. Y., et al. (2023). Antibiogram of multidrug-resistant bacteria based on sepsis onset location in korea: a multicenter cohort study. *J. Korean Med. Sci.* 38:e75. doi: 10.3346/jkms.2023.38.e75

Kumar, A., Roberts, D., Wood, K. E., Light, B., Parrillo, J. E., Sharma, S., et al. (2006). Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock*. *Crit. Care Med.* 34, 1589–1596. doi: 10.1097/01.CCM.0000217961.75225.E9

Li, X., Liang, S., Zhang, D., He, M., and Zhang, H. (2023a). The clinical application of metagenomic next-generation sequencing in sepsis of immunocompromised patients. *Front. Cell. Infect. Microbiol.* 13:1170687. doi: 10.3389/fcimb.2023.1170687

Li, X., Yang, L., Li, D., Yang, X., Wang, Z., Chen, M., et al. (2023b). Diagnosis of neurological infections in pediatric patients from cell-free dna specimens by using metagenomic next-generation sequencing. *Microbiol. Spectr.* 11:e0253022. doi: 10.1128/spectrum.02530-22

Liang, Y., Dong, T., Li, M., Zhang, P., Wei, X., Chen, H., et al. (2022). Clinical diagnosis and etiology of patients with chlamydia psittaci pneumonia based on metagenomic next-generation sequencing. *Front. Cell. Infect. Microbiol.* 12:1006117. doi: 10.3389/fcimb.2022.1006117

Lu, X., and Wang, C. (2020). Expert consensus on clinical standardized application of metagenomics next-generation sequencing for detection of pathogenic microorganisms. *Chin. J. Lab Med.* 43, 1181–1195. doi: 10.3760/cma.j.cn114452-20200903-00704

Miao, Q., Ma, Y., Wang, Q., Pan, J., Zhang, Y., Jin, W., et al. (2018). Microbiological diagnostic performance of metagenomic next-generation sequencing when applied to clinical practice. *Clin. Infect. Dis.* 67(suppl_2), S231–S240. doi: 10.1093/cid/ciy693

Miller, S., and Chiu, C. (2022). The role of metagenomics and next-generation sequencing in infectious disease diagnosis. *Clin. Chem.* 68, 115–124. doi: 10.1093/clinchem/hvab173

Niederman, M. S., Baron, R. M., Bouadma, L., Calandra, T., Daneman, N., DeWaele, J., et al. (2021). Initial antimicrobial management of sepsis. *Crit. Care* 25:307. doi: 10.1186/s13054-021-03736-w

Paoli, C. J., Reynolds, M. A., Sinha, M., Gitlin, M., and Crouser, E. (2018). Epidemiology and costs of sepsis in the united states—an analysis based on timing of diagnosis and severity level*. *Crit. Care Med.* 46, 1889–1897. doi: 10.1097/CCM.0000000000003342

Qin, C., Zhang, S., Zhao, Y., Ding, X., Yang, F., Zhao, Y., et al. (2023). Diagnostic value of metagenomic next-generation sequencing in sepsis and bloodstream infection. *Front. Cell. Infect. Microbiol.* 13:1117987. doi: 10.3389/fcimb.2023.1117987

Reinhart, K., Daniels, R., Kissoon, N., Machado, F. R., Schachter, R. D., Finfer, S., et al. (2017). Recognizing sepsis as a global health priority - a who resolution. *N. Engl. J. Med.* 377, 414–417. doi: 10.1056/NEJMp1707170

Rhodes, A., Evans, L. E., Alhazzani, W., Levy, M. M., Antonelli, M., Ferrer, R., et al. (2017). Surviving sepsis campaign. *Crit. Care Med.* 45, 486–552. doi: 10.1097/CCM.0000000000002255

Seymour, C. W., Gesten, F., Prescott, H. C., Friedrich, M. E., Iwashyna, T. J., Phillips, G. S., et al. (2017). Time to treatment and mortality during mandated emergency care for sepsis. *N. Engl. J. Med.* 376, 2235–2244. doi: 10.1056/NEJMoa1703058

Shang, H. W. Y. S. (2015). *National Clinical Laboratory Procedures*. Beijing: People's Medical Publishing House.

Sharon, I., and Banfield, J. F. (2013). Genomes from metagenomics. *Science* 342, 1057–1058. doi: 10.1126/science.1247023

Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., et al. (2016). The third international consensus definitions for sepsis and septic shock (sepsis-3). *Jama-J. Am. Med. Assoc.* 315:801. doi: 10.1001/jama.2016.0287

Sun, L., Zhang, S., Yang, Z., Yang, F., Wang, Z., Li, H., et al. (2022). Clinical application and influencing factor analysis of metagenomic next-generation sequencing (mNGS) in icu patients with sepsis. *Front. Cell. Infect. Microbiol.* 12:905132. doi: 10.3389/fcimb.2022.905132

Tonai, M., Shiraishi, A., Karumai, T., Endo, A., Kobayashi, H., Fushimi, K., et al. (2022). Hospital-onset sepsis and community-onset sepsis in critical care units in japan: a retrospective cohort study based on a japanese administrative claims database. *Crit. Care* 26:136. doi: 10.1186/s13054-022-04013-0

van den Berg, M., van Beuningen, F. E., ter Maaten, J. C., and Bouma, H. R. (2022). Hospital-related costs of sepsis around the world: a systematic review exploring the economic burden of sepsis. *J. Crit. Care* 71:154096. doi: 10.1016/j.jcrc.2022.154096

Wang, H. (2021). Chinese expert consensus on metagenomics next-generation sequencing application on pathogen detection of infectious diseases. *Chin. J. Lab. Med.* 44, 107–120. doi: 10.3760/cma.j.cn114452-20201026-00794

Wang, L. L., Li, S. S., Qin, J. J., Tang, T. T., Hong, J. J., Tung, T. T., et al. (2023). Clinical diagnosis application of metagenomic next-generation sequencing of plasma in suspected sepsis. *Infect. Drug Resist.* 16, 891–901. doi: 10.2147/IDR.S395700

Westphal, G. A., Pereira, A. B., Fachin, S. M., Barreto, A. C. C., Bornschein, A. C. G. J., Caldeira Filho, M., et al. (2019). Characteristics and outcomes of patients with community-acquired and hospital-acquired sepsis. *Rev. Bras. Ter. Intensiva.* 31, 71–78. doi: 10.5935/0103-507X.20190013

WHO (2017). *Improving the prevention, diagnosis and management of sepsis. 140th session EB140.R5. Agenda item 7.2*. Geneva: WHO.

Yang, A., Chen, C., Hu, Y., Zheng, G., Chen, P., Xie, Z., et al. (2022). Application of metagenomic next-generation sequencing (mNGS) using bronchoalveolar lavage fluid (BALF) in diagnosing pneumonia of children. *Microbiol. Spectr.* 10:e0148822. doi: 10.1128/spectrum.01488-22

Zhang, H., Zhan, D., Chen, D., Huang, W., Yu, M., Li, Q., et al. (2020). Next-generation sequencing diagnosis of severe pneumonia from fulminant psittacosis with multiple organ failure: a case report and literature review. *Ann. Transl. Med.* 8:401. doi: 10.21037/atm.2020.03.17

Zhang, Y., Zhou, D., Xia, H., Wang, J., Yang, H., Xu, L., et al. (2023). Metagenomic next-generation sequencing for detection of pathogens in children with hematological diseases complicated with infection. *Mol. Cell. Probes.* 67:101889. doi: 10.1016/j.mcp.2022.101889

Zhou, Y., Shi, W., Wen, Y., Mao, E., and Ni, T. (2023). Comparison of pathogen detection consistency between metagenomic next-generation sequencing and blood culture in patients with suspected bloodstream infection. *Sci Rep.* 13:9460. doi: 10.1038/s41598-023-36681-5

Zuo, Y., Wu, Y., Hu, W., Chen, Y., Li, Y., Song, Z., et al. (2023). The clinical impact of metagenomic next-generation sequencing (mNGS) test in hospitalized patients with suspected sepsis: a multicenter prospective study. *Diagnostics* 13:323. doi: 10.3390/diagnostics13020323

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 17 May 2024
DOI 10.3389/fmicb.2024.1397068

OPEN ACCESS

EDITED BY
Renmao "Tim" Tian,
Illinois Institute of Technology, United States

REVIEWED BY
Etienne Giraud,
Institut National de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), France
Jörg Linde,
Loeffler Institut, Germany
Patricia Alba,
Institute of Experimental Zooprophylactic of the Lazio and Tuscany Regions (IZSLT), Italy
Xiyang Liu,
Illinois Institute of Technology, United States

*CORRESPONDENCE
Bijay K. Khajanchi
✉ bijay.khajanchi@fda.hhs.gov

†These authors have contributed equally to this work

RECEIVED 06 March 2024
ACCEPTED 22 April 2024
PUBLISHED 17 May 2024

CITATION
Felix MA, Sopovski D, Commichaux S, Yoskowitz N, Aljahdali NH, Grim CJ, Abbott CN, Carlton A, Han J, Sanad YM, Zhao S, Wang X, Foley SL and Khajanchi BK (2024) Genetic relatedness and virulence potential of *Salmonella* Schwarzengrund strains with or without an IncFIB-IncFIC(FII) fusion plasmid isolated from food and clinical sources. *Front. Microbiol.* 15:1397068. doi: 10.3389/fmicb.2024.1397068

Genetic relatedness and virulence potential of *Salmonella* Schwarzengrund strains with or without an IncFIB-IncFIC(FII) fusion plasmid isolated from food and clinical sources

Monique A. Felix[1,2†], Danielle Sopovski[1†], Seth Commichaux[3], Noah Yoskowitz[1], Nesreen H. Aljahdali[1,4], Christopher J. Grim[5], Carter N. Abbott[1], Ashlyn Carlton[1,2], Jing Han[1], Yasser M. Sanad[2,6], Shaohua Zhao[7], Xiong Wang[8], Steven L. Foley[1] and Bijay K. Khajanchi[1,3*]

[1]National Center for Toxicological Research, U. S. Food and Drug Administration, Jefferson, AR, United States, [2]University of Arkansas at Pine Bluff, Pine Bluff, AR, United States, [3]Center for Food Safety and Applied Nutrition, Office of Applied Research and Safety Assessment, U. S. Food and Drug Administration, Laurel, MD, United States, [4]Department of Biological Science, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia, [5]Center for Food Safety and Applied Nutrition, U. S. Food and Drug Administration, College Park, MD, United States, [6]Department of Epidemiology, College of Public Health, University of Arkansas for Medical Sciences, Little Rock, AR, United States, [7]Center for Veterinary Medicine, U. S. Food and Drug Administration, Laurel, MD, United States, [8]Minnesota Department of Health, St. Paul, MN, United States

A total of 55 food and clinical *S.* Schwarzengrund isolates were assayed for plasmid content, among which an IncFIB-IncFIC(FII) fusion plasmid, conferring streptomycin resistance, was detected in 17 isolates. Among the 17 isolates, 9 were food isolates primarily collected from poultry meat, and 8 clinical isolates collected from stool, urine, and gallbladder. SNP—based phylogenetic analyses showed that the isolates carrying the fusion plasmid formed a subclade indicating the plasmid was acquired and is now maintained by the lineage. Phylogenetic analysis of the plasmid suggested it is derived from avian pathogenic plasmids and might confer an adaptive advantage to the *S.* Schwarzengrund isolates within birds. IncFIB-IncFIC(FII) fusion plasmids from all food and three clinical isolates were self-conjugative and successfully transferred into *E. coli* J53 by conjugation. Food and clinical isolates had similar virulome profiles and were able to invade human Caco-2 cells. However, the IncFIB-IncFIC(FII) plasmid did not significantly add to their invasion and persistence potential in human Caco-2 cells.

KEYWORDS

***Salmonella* Schwarzengrund, IncFIB-IncFIC(FII) plasmid, fusion plasmid, SNP, conjugation, Caco-2**

Introduction

Salmonella is an enteric pathogen that invades the gut through contaminated food (Hallstrom and McCormick, 2011; Fabrega and Vila, 2013). They are Gram-negative, facultative anaerobes with food animal reservoirs, such as chickens, cows, turkeys, and pigs (Andino and Hanning, 2015). There is an estimated more than 1 million human cases of *Salmonella* infection each year in the U.S. with approximately 20,000 hospitalizations and 400 deaths (Scallan et al., 2011). *Salmonella* infections can be classified as either typhoidal (human specific), or non-typhoidal Salmonellosis (broad host range; Aarestrup et al., 2007). A significant rise of the *Salmonella* Schwarzengrund serovar in countries such as Thailand, Slovakia, New Zealand, Venezuela, the U.S., Japan, and Denmark has been reported (Aarestrup et al., 2007). For example, an increase in *S.* Schwarzengrund infections in humans from 0 to 2.4%, and in chickens from 0.3% to 26.2% from 1993 to 2001 was observed in Thailand (Bangtrakulnonth et al., 2004). Similarly, in Japan, of the serovars collected from broiler chickens, *S.* Schwarzengrund was found to account for 28.1% of isolates in 2005 compared to 0% in 2000. Asai et al. (2009) also reported an increasing trend of isolation of *S.* Schwarzengrund from human patients in Japan. Their findings showed that in just 2 years *S.* Schwarzengrund went from an uncommon serovar to the 10th most commonly reported. A more recent study, on 3,069 cecal samples collected from broiler chickens in a processing plant in Japan from 2013 to 2016, reported 17.8% were positive for *Salmonella* and 21.3% of those were identified as *S.* Schwarzengrund. The prevalence of *S.* Schwarzengrund had increased from the 2.1% reported in their previous study from 2009 to 2012 (Duc et al., 2020).

In the U.S., *S.* Schwarzengrund has risen to become one of the top five *Salmonella* serovars isolated from retail meat (Aarestrup et al., 2007). For example, in 2019, a *S.* Schwarzengrund outbreak linked to ground turkey occurred in three states and resulted in 78,000 pounds of turkey being recalled [Center for Disease Control and Prevention (CDC); https://www.cdc.gov/salmonella/schwarzengrund-03-19/index.html]. In 2007, *S.* Schwarzengrund outbreaks were also linked to dry pet food (Centers for Disease Control and Prevention, 2008).

This increase in the number of *S.* Schwarzengrund infections is not the only important factor to consider; there is also evidence of a higher frequency of antimicrobial resistance (AMR) within the strains of this serovar that is spreading internationally. Using antimicrobial susceptibility testing and pulsed field gel electrophoresis (PFGE) typing, Aarestrup et al. (2007) found that 7 of 14 strains isolated from humans in Denmark shared PFGE patterns with isolates from humans and chicken meat in Thailand, while 22 of 390 human-source isolates from the U.S. also had common profiles to those in Denmark and Thailand. These isolates showed a high frequency of resistance to nalidixic acid, along with a reduced susceptibility to ciprofloxacin (Aarestrup et al., 2007). The first reported instance of fluoroquinolone-resistance in *S.* Schwarzengrund in the U.S. came from a strain isolated from a patient who had traveled to the Philippines (Akiyama and Khan, 2012). In a study conducted on raw chicken samples from marketplaces in Taiwan, most of the *S.* Schwarzengrund strains demonstrated a ACSSuT resistance type [resistant to ampicillin, chloramphenicol, streptomycin, sulfamethoxazole, and tetracycline (Chen et al., 2010)]. Many of these antibiotics are commonly used in a variety of treatments for animal and human infections. Consequently, antimicrobial resistance further complicates the treatment of Salmonellosis caused by *S.* Schwarzengrund, which can lead to higher morbidity and mortality (Nair et al., 2018).

Salmonella enterica possesses a wide range of virulence factors that facilitate the establishment of successful infections in animal and human hosts (Ochman et al., 1996; van der Heijden and Finlay, 2012). The majority of these virulence factors are encoded in the chromosome; however, some are harbored by plasmids (Waters and Crosa, 1991; Ochman et al., 1996; van der Heijden and Finlay, 2012; Khajanchi et al., 2016, 2017; Khajanchi, 2022). A broad range of plasmid incompatibility groups (Inc) have been found in *Salmonella* serovars (Khajanchi et al., 2016, 2017; Khajanchi, 2022). Among them, IncFIB plasmids often possess functions associated with colicin production, the aerobactin siderophore and Sit iron acquisition systems, and persistence in intestinal epithelial cells (Khajanchi et al., 2016, 2017; Khajanchi, 2022).

Johnson et al. (2010) showed that horizontal gene transfer of IncFIB plasmids contributed to the acquisition of antimicrobial resistance. In their study, 902 *Salmonella* isolates that belonged to 59 different serovars were examined for plasmids. The IncFIB plasmids were found in isolates of serovars Kentucky, Typhimurium, and Heidelberg. It was shown that a single PFGE clonal type of *S.* Kentucky harbored these plasmids and the acquisition of the plasmid allowed *S.* Kentucky to be more competitive in colonizing the chicken cecum compared to those lacking the plasmid. Evaluation of sequences from three IncFIB plasmids from *S.* Kentucky isolates that originated in different locations at different times from different sources showed almost identical genetic sequence. These findings point to the IncFIB plasmid being recently attained within the *S.* Kentucky serovar., with a rapid transfer among the population that improved colonization and fitness abilities (Johnson et al., 2010). In other research on horizontal gene transfer, two IncFIB plasmids (pAPEC-O2-ColV and pAPEC-O2-R) were transferred into an avirulent and plasmid-less *E. coli* strain. The *E. coli* strain became virulent toward chick embryos and showed resistance to ampicillin, streptomycin, and several other antibiotics (Johnson et al., 2006).

IncFII and IncFIC plasmids contribute to the horizontal transfer of antimicrobial resistance genes including extended spectrum β-lactamases (ESBL; Yoon and Lee, 2022; de Jesus Bertani et al., 2023). Some IncFIC plasmids are fusion plasmids and carry both the IncFIC and IncFII replicons (Yoon and Lee, 2022). The formation and spread of fusion plasmids in *Enterobacteriaceae* is an emergent problem (Liu et al., 2021). Mobile genetic elements, such as insertion sequences and transposons can contribute to the formation of fusion plasmids (Liu et al., 2021). Therefore, to better understand the spread of fusion plasmids, it is important to characterize *Salmonella* and other enteric bacteria that harbor them. The objectives of the study were: (i) to perform molecular characterization of the IncFIB-IncFIC(FII) fusion plasmid of *S.* Schwarzengrund isolated from food and clinical sources; (ii) to determine the role of IncFIB-IncFIC(FII) fusion plasmid in invasion and persistence of *Salmonella*.

Materials and methods

Bacterial strains

A total of 55 *S.* Schwarzengrund isolates, of which 36 were collected from human patients, were obtained from the Wisconsin

State Lab of Hygiene (WLSH; $n = 15$), Minnesota Department of Health (MDH; $n = 18$) and the Maryland Department of Health (MD; $n = 3$); while 19 food isolates were collected as part of the National Antimicrobial Resistance Monitoring System (NARMS) efforts. Food isolates were primarily collected from chicken, while clinical isolates were collected from stool, urine, and blood (Supplementary Table 1). The sodium azide resistant *Escherichia coli* J53 was used as a recipient for the conjugation studies (Jacoby and Han, 1996).

Whole genome sequencing using short read and long read methods

Of these 55 isolates, 17 fusion plasmid-containing *S.* Schwarzengrund isolates were sequenced using both short read Illumina and long-read Oxford Nanopore Technology (ONT; Sopovski et al., 2024). The remaining 38 isolates were only sequenced using the short read Illumina platform. CheckM was used to determine the completeness and contamination of the short read assemblies. Quality measurements of raw data and assembled genomes were previously published separately (Khajanchi et al., 2019a; Sopovski et al., 2024) and also provided as Supplementary Table 2.

Short-read WGS was performed by a procedure described earlier (Khajanchi et al., 2016). Briefly, genomic DNA from bacterial cells was extracted using a DNeasy Blood and Tissue kit (Qiagen, Valencia, CA, United States). The quality and quantity of the DNA were examined by Nanodrop (ThermoFisher Scientific, Grand Island, NY, United States) and the Qubit BR assay kit (ThermoFisher Scientific). DNA libraries were constructed using 1 ng of DNA from each sample using the Nextera XT DNA library preparation kit (Illumina, San Diego, CA, United States). Samples were multiplexed using combinations of two indexes of Nextera XT Index Kit. DNA samples were diluted, denatured, loaded and sequenced on an Illumina MiSeq instrument with 2×250 pair-end chemistry.

For long-read ONT sequencing, the 1D native barcoding genomic DNA long read selection protocol was used with the SQK-LSK109 kit (Oxford Nanopore, Oxford, UK) as described earlier (Taylor et al., 2019). Briefly, 1 μg of DNA was subjected to end repair and dA-tailing using the NEBNext® Ultra™ II End Repair/dA-Tailing module (New England Biolabs, Ipswich, MA, United States). End-prepped DNA fragments were barcoded using the EXP-NBD104 and EXP-NBD114 kits (Oxford Nanopore). Equimolar amounts of each barcoded sample were pooled together, adapters were ligated, and the resulting library pools were sequenced on a MinION device using a FLO-MIN106 (R9.4.1) flow cell for 48 h.

The nanopore reads were trimmed and filtered using NanoFilt (v2.3.0; De Coster et al., 2018). For NanoFilt, the parameters were set to filter out nanopore reads with a quality score (qscore) of less than 10 or if the read was less than 500 bp long. The short and long-read sequences were assembled by hybrid assembly using UniCycler (v0.4.8; Wick et al., 2017). The FASTA files of the assemblies from each isolate were analyzed using PlasmidFinder (version 2.1) and ResFinder (version 4.1) to identify predicted plasmids and antimicrobial resistance genes, respectively (Supplementary Table 1; Zankari et al., 2012; Carattoli et al., 2014).

Single nucleotide polymorphism analysis

WGS-based SNP analysis of the 55 *Salmonella* genomes was performed using the FDA Center for Food Safety and Applied Nutrition (CFSAN) SNP pipeline (Davis et al., 2015). The CFSAN SNP pipeline was used to find the pairwise SNP distances between the isolates. The hybrid assembly for isolate WLSH7 was used as reference for the SNP analysis. FastTree (v2.1.11), using the general time reversible model, was used to approximate the maximum likelihood phylogeny with 1,000 bootstraps. The phylogeny was visualized in FigTree (v1.4.4).

Plasmid annotation and phylogeny

Platon (v1.6) was used to annotate the assembly contigs for plasmids. NCBI's online BLAST server was used to identify the best hits (≥99% identity and ≥70% query coverage) in the NCBI Nucleotide database to the IncFIB-IncFIC(FII) plasmid, which were downloaded. Genes were annotated in the plasmids using Prokka (v1.14.5). The pangenome was estimated with Roary (v3.12.0). For reference, the pangenome contained 631 genes of which 49 were core. MAFFT (v7.305) was used to align the concatenated 49 core genes. FastTree (v2.1.11), using the general time reversible model, was used to approximate the maximum likelihood phylogeny with 1,000 bootstraps. The phylogeny was visualized in FigTree (v1.4.4). The plasmid annotations were visualized in SnapGene (SnapGene by Dotmatics, Boston, MA, United States). The tanglegram was created using R and the cophyloplot function from the ape package.

Bacterial conjugation

Conjugation experiments were carried out to determine the transferability of IncFIB plasmids in *S.* Schwarzengrund isolates either by plate mating or broth mating approaches (Khajanchi et al., 2019b). In plate mating strategy, IncFIB-IncFIC(FII) positive *S.* Schwarzengrund isolates (donors) and the sodium azide resistant recipient *E. coli* J53 were cross streaked on Luria-Bertani (LB; BD, Franklin Lakes, NJ, United States) agar plates. After 24 h of incubation, the cells from the intersection were collected and re-streaked onto selective plates containing sodium azide (350 μg/mL) and streptomycin (16 μg/mL). Individual colonies were picked and sub-cultured onto MacConkey agar to confirm *E. coli* colonies. Carriage of IncFIB-IncFIC(FII) plasmids by *E. coli* transconjugants were confirmed by PCR. Conjugation experiments that were unsuccessful by plate mating were subjected to a different approach described in our previous study (Khajanchi et al., 2019b). Briefly, a single colony of IncFIB-IncFIC(FII) containing *S.* Schwarzengrund isolates (donor) and *E. coli* J53 (recipient) was grown separately in LB broth overnight at 37°C. The recipient and the donor were subsequently mixed together in a 1:1 proportion and centrifuged at $7,000 \times g$ for 5 min to obtain the pellet. The pellets were dispersed in 250 μL of LB broth and spotted onto a LB agar plate. The plate was incubated for 3–4 h at 37°C in upright position. The growth seen was suspended in 1 mL phosphate buffered saline (PBS) and 100 μL of cell suspension was spread onto a LB agar plate containing sodium azide (350 μg/mL) and streptomycin (16 μg/mL). Pink colonies were

selected after streaking on to selective MacConkey agar plate and presence of IncFIB-IncFIC(FII) was confirmed by PCR.

Virulome and plasmid transfer gene assay

The detection of virulence and plasmid transfer-associated genes encoded by the *Salmonella* isolates were predicted based on their whole genome sequences. Genome sequences from *Salmonella* donors and transconjugants (plasmids in *E. coli* J53) were trimmed, and *de novo* assembly was completed using CLC Genomics Workbench (ver. 9.0, Qiagen, Redwood City, CA, United States). FASTA files of sequence assemblies from each strain were analyzed using the multiple sequence Comparison tool within the FDA Virulence and AMR Plasmid Transfer Factor Database,[1] which targets 594 putative *Salmonella* virulence genes and plasmid transfer genes from key AMR plasmid Inc. groups (Aljahdali et al., 2020; Tate et al., 2022; Algarni et al., 2023). The predicted presence and absence data for the putative virulence and plasmid transfer genes were downloaded from the database, transformed to binary data and imported into BioNumerics for phylogenetic analyses of the virulence and plasmid transfer genes. Based categorical (binary) difference calculations and dendrograms were generated using UPGMA (Applied Maths, Austin, TX, United States). The profiles of the presence/absence of virulence and plasmid transfer genes were further compared in BioNumerics using minimum spanning tree analyses to compare similarities of the wildtype and transconjugant strains.

Bacterial invasion assay

Bacterial invasion assays were performed using human intestinal epithelial cells (Caco-2) as described previously (Khajanchi et al., 2017). Briefly, 10^5 Caco-2 cells per well were seeded in 24-well tissue culture plates and incubated at 37°C overnight in a 5% CO_2 incubator. Cells in one of the wells were counted using a Cellometer Auto T4 (Nexcelom Bioscience, Lawrence, MA, United States) and the Caco-2 cells were infected with different *S.* Schwarzengrund isolates at multiplicity of infection (MOI) of 10. After incubation for 1 h at 37°C, the cells were washed twice with PBS to remove bacteria that had not infected the Caco-2 cells and incubated with 200 μg/mL of gentamicin. After incubation for 1 h at 37°C, the cells were washed twice with PBS and lysed with 0.1% chilled Triton X-100, followed by dilution and plating on trypticase soy agar (TSA) to obtain colony forming unit counts (CFUs) of bacteria following overnight incubation at 37°C. Three replicates per strain were included and the experiments were repeated three times.

Bacterial persistence assay

Caco-2 cells were infected as for the invasion assay. After 1 h incubation, the cells were washed twice with PBS and incubated with 100 μg/mL of gentamicin for 48 h. After incubation, cells were washed, lysed, and CFUs were counted as in the invasion assay procedure, with three replicates per strain and experiments carried out in triplicate.

Results

SNP analyses

SNP analysis (Figure 1) showed that the 17 food and clinical isolates carrying the IncFIB-IncFIC(FII) fusion plasmid clustered within the same subclade (between 0 and 52 SNP differences), separated from the other isolates that lacked the fusion plasmids (Figure 1). This suggests the plasmid was acquired and has been maintained by that subclade lineage. The two exceptions were isolates CVM-6 and WLSH-27 which lacked the fusion plasmid but still clustered with the isolates carrying the fusion plasmid. These two isolates carried an IncFIC plasmid that was near-identical to the IncFIB-IncFIC(FII) fusion plasmid over ~63% of its length, indicating it had not undergone fusion or had lost the IncFIB portion of the fusion plasmid. Within the fusion plasmid subclade, isolates further separated into subclusters by human and chicken isolation sources. However, this difference might be an artifact due to the time of collection because all the chicken isolates were collected in 2013, whereas the clinical isolates were collected between 2013 and 2017.

The concatenated core gene phylogeny of the IncFIB-IncFIC(FII) fusion plasmid

The topology of the IncFIB-IncFIC(FII) fusion plasmid phylogeny differed from the SNP tree in Figure 1, with no clear separation between the human and chicken isolates (Figure 2). Among the best BLAST hits in the NCBI Nucleotide database were plasmids from isolates that were recovered from animals (e.g., chicken, duck, pig, peafowl) with colibacillosis and respiratory disease. Further, several related plasmids were isolated from avian pathogenic *E. coli* (APEC). The gene annotation of the IncFIB-IncFIC(FII) plasmid can be seen in Figure 3. The pangenome of the IncFIB-IncFIC(FII) plasmid contained 631 genes of which 49 were core.

Bacterial conjugation

Among 19 food isolates, nine contained the IncFIB-IncFIC(FII) fusion plasmid (CVM-5, CVM-7, CVM-10, CVM-11, CVM-13, CVM-14, CVM-15, CVM-16, CVM-17; Supplementary Table 1). All nine isolates successfully conjugated with *E. coli* J53. Of the 36 clinical isolates, eight isolates contained the IncFIB-IncFIC(FII) fusion plasmid (MDH29, WLSH-3, WLSH-7, WLSH-13, WLSH-25, WLSH-26, MD-3, MD-4; Supplementary Table 1). Out of the eight isolates, three (WLSH 7, WLSH 13, WLSH 25) successfully conjugated with *E. coli* J53. The conjugation experiment was repeated for the five isolates that did not transfer the plasmid, but conjugation still did not occur, indicating they might not be conjugative under the test conditions. When the sequenced plasmids in the wild type and transconjugant isolates (J53::CVM5, J53::CVM7, J53::CVM10, J53::CVM11, J53::CVM13, J53::CVM14, J53::CVM15, J53::CVM16, J53::CVM17, J53::WLSH-7, J53::WLSH-13, J53::WLSH-25) were

1 https://virulence.fda.gov

10.3389/fmicb.2024.1397068

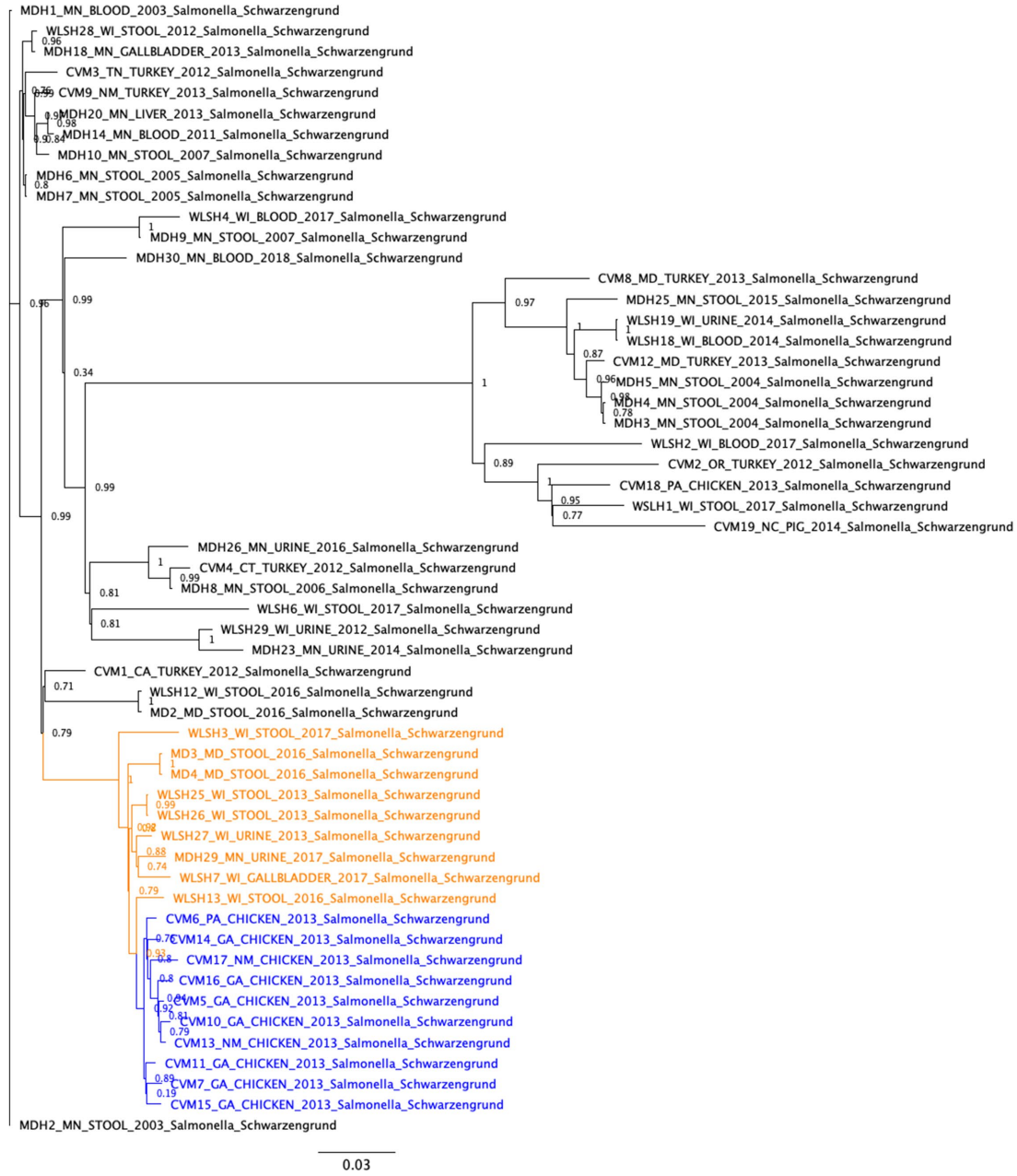


FIGURE 1
Single nucleotide polymorphism (SNP)-based phylogenetic analyses of *Salmonella* Schwarzengrund isolates from food and clinical sources. The SNP differences were identified with the CFSAN SNP Pipeline and the phylogeny was created with FastTree using the general time reversible model and 1,000 bootstraps. The isolates carrying the IncFIB-IncFIC(FII) plasmid formed a closely related cluster (between 0 and 52 SNP differences). These isolates further separated into subclusters by human and chicken isolation sources. The clinical and food isolates carrying the fusion plasmid are highlighted in orange and blue, respectively.

compared, they had the same gene content (Table 1), including antimicrobial resistance and iron acquisition genes, supporting the successful conjugation of the fusion plasmid.

Virulome and plasmid transfer gene analyses

The presence of the *Salmonella*-associated virulence genes are detailed in Supplementary Table 3 and the comparison of the virulence gene profiles are shown in Supplementary Figure 1. Overall, the *S.* Schwarzengrund isolates had very similar virulence factor profiles, with the exception of putative virulence genes associated with the IncFIB-IncFIC(FII) plasmids, including *iucABCD* and *iutA* of the aerobactin operon and *traT*. The transconjugants separated to a distinct clade, as they were *E. coli* and lacking many of the *Salmonella*-associated virulence genes (Supplementary Figure 2). When the plasmid transfer genes were detected, the most common genes detected were the IncFIB-IncFIC(FII)-associated genes that were present in the transconjugants and their corresponding donor strains

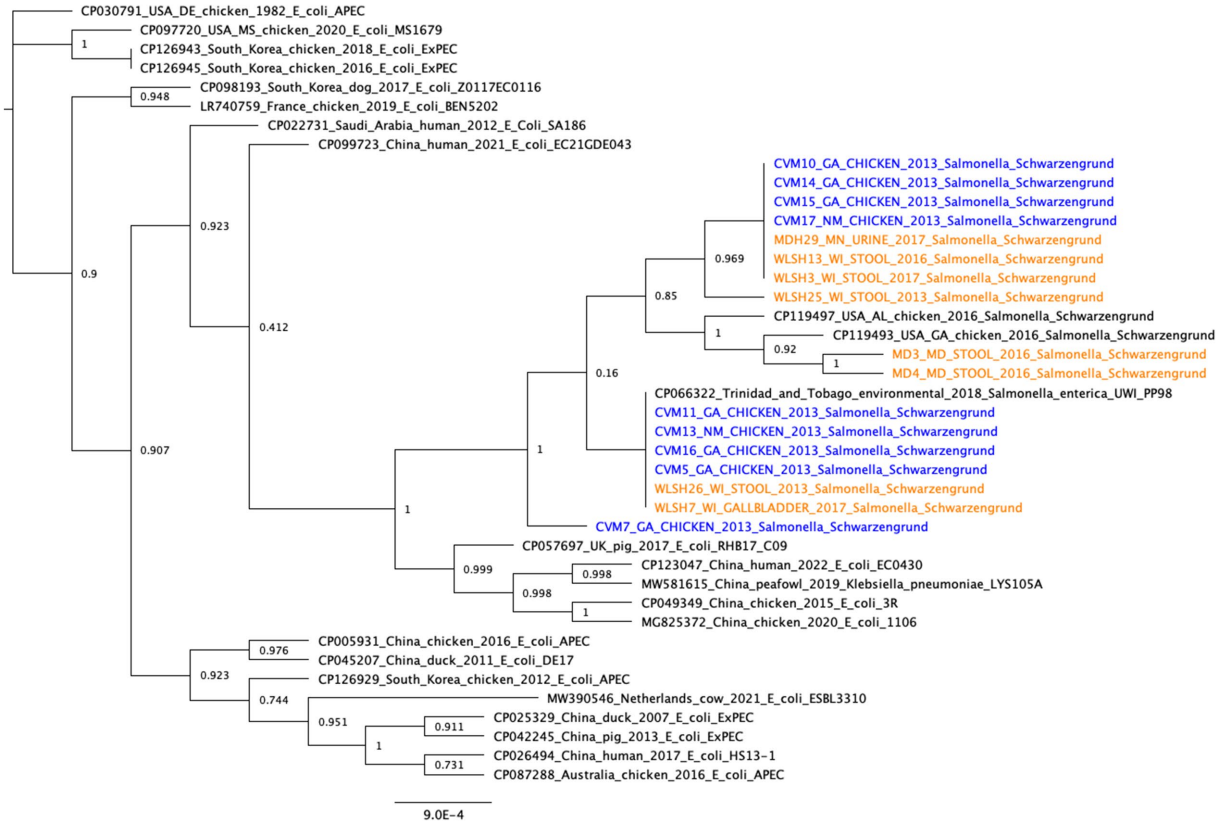


FIGURE 2
Core gene phylogeny of the IncFIB-IncFIC(FII) plasmid. Includes the plasmids from the 17 long-read sequenced isolates and the best BLAST hits from the NCBI Nucleotide database (≥99% identity and ≥70% query coverage). Clinical isolates are in orange and food isolates are in blue. The pangenome of all the plasmids (including the 17 isolates from this study and the best BLAST hits in NCBI) had 631 genes, of which 49 were core. The median number of genes per plasmid sequence was 158.

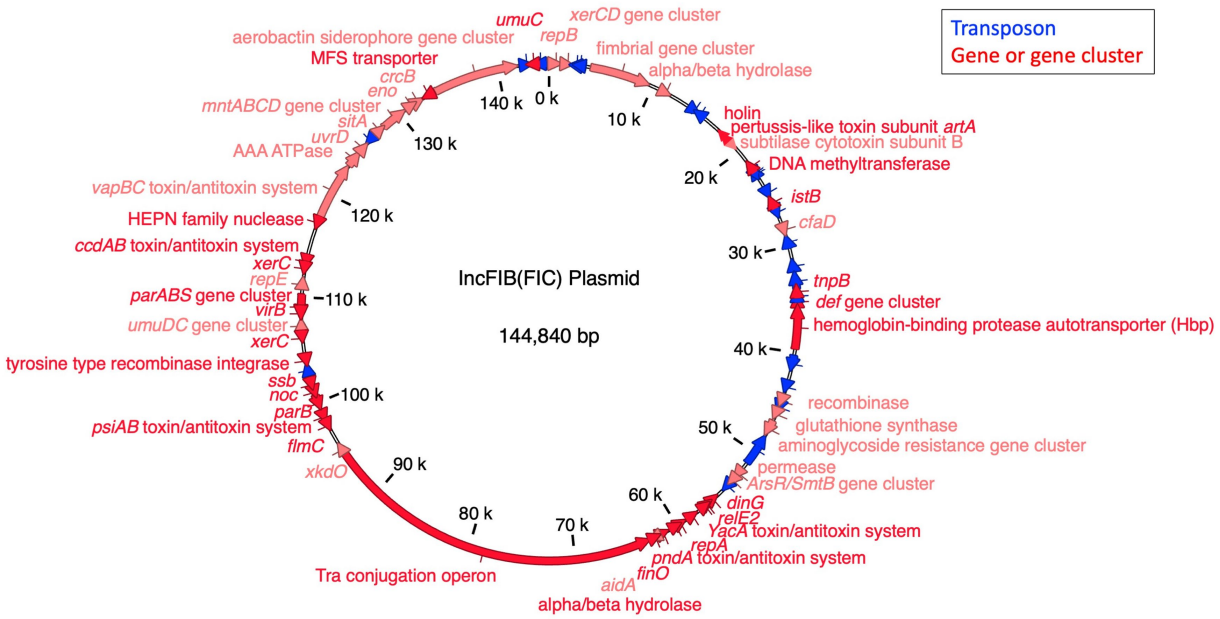


FIGURE 3
The annotated IncFIB-IncFIC(FII) plasmid. The markers for the FIB and FIC replicons co-occurred in the plasmid from 1–344 bp to 59,325–59,519 bp, respectively, indicating it is a fusion plasmid.

10.3389/fmicb.2024.1397068

TABLE 1 Sequence analyses of wild type and transcongugants of 12 IncFIB containing *Salmonella* Schwarzengrund food and clinical isolates.

Wild type isolates				Transconjugants generated using *Escherichia coli* J53 as recipient			
Wildtype Strain ID	Palsmid content	Resistance gene	Iron acquisition genes	Transconjugants ID	Plasmid content	Resistance gene	Iron acquisition genes
CVM −5	IncFIB(AP001918), IncFIC(FII)	*aac(6′)-Iaa, aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ iroNB/sitABCD*	J53::CVM-5	IncFIB(AP001918), IncFIC(FII)	*aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ sitABCD*
CVM-7	IncFIB(AP001918), IncFIC(FII)	*aac(6′)-Iaa, aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ iroNB/sitABCD*	J53::CVM-7	IncFIB(AP001918), IncFIC(FII)	*aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ sitABCD*
CVM-10	IncFIB(AP001918), IncFIC(FII)	*aac(6′)-Iaa, aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ iroNB/sitABCD*	J53::CVM-10	IncFIB(AP001918), IncFIC(FII)	*aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ sitABCD*
CVM-11	IncFIB(AP001918), IncFIC(FII)	*aac(6′)-Iaa, aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ iroNB/sitABCD*	J53::CVM-11	IncFIB(AP001918), IncFIC(FII)	*aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ sitABCD*
CVM-13	IncFIB(AP001918), IncFIC(FII)	*aac(6′)-Iaa, aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ iroNB/sitABCD*	J53::CVM-13	IncFIB(AP001918), IncFIC(FII)	*aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ sitABCD*
CVM-14	IncFIB(AP001918), IncFIC(FII)	*aac(6′)-Iaa, aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ iroNB/sitABCD*	J53::CVM-14	IncFIB(AP001918), IncFIC(FII)	*aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ sitABCD*
CVM-15	IncFIB(AP001918), IncFIC(FII)	*aac(6′)-Iaa, aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ iroNB/sitABCD*	J53::CVM-15	IncFIB(AP001918), IncFIC(FII)	*aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ sitABCD*
CVM-16	IncFIB(AP001918), IncFIC(FII)	*aac(6′)-Iaa, aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ iroNB/sitABCD*	J53::CVM-16	IncFIB(AP001918), IncFIC(FII)	*aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ sitABCD*
CVM-17	IncFIB(AP001918), IncFIC(FII)	*aac(6′)-Iaa, aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ iroNB/sitABCD*	J53::CVM-17	IncFIB(AP001918), IncFIC(FII)	*aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ sitABCD*
WLSH-7	IncFIB(AP001918), IncFIC(FII)	*aac(6′)-Iaa, aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ iroNB/sitABCD*	J53::WLSH-7	IncFIB(AP001918), IncFIC(FII)	*aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ sitABCD*
WLSH-13	IncFIB(AP001918), IncFIC(FII)	*aac(6′)-Iaa, aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ iroNB/sitABCD*	J53::WLSH-13	IncFIB(AP001918), IncFIC(FII)	*aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ sitABCD*
WLSH-25	IncFIB(AP001918), IncFIC(FII)	*aac(6′)-Iaa, aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ iroNB/sitABCD*	J53::WLSH-25	IncFIB(AP001918), IncFIC(FII)	*aph(3″)-Ib, aph(6)-Id*	*iucABCD/iutA/ sitABCD*

(top clade in Supplementary Figure 2). One of the donor strains (WLSH-3) carried the IncFIB-IncFIC(FII) along with IncI1 and IncI2-associated transfer genes. A group of five other strains carried IncI1 plasmid-associated genes, without the IncFIB-IncFIC(FII) plasmids (bottom group of Supplementary Figure 2), and two strains carried an IncHI2 plasmid. These results correlated with the plasmid replicon typing results (Supplementary Table 1). The relatedness of the virulence genes (Figure 4A) and plasmid transfer genes (Figure 4B) can also be displayed with the minimal spanning trees. The three largest groups of *Salmonella* differ by the presence of the aerobactin/*traT* genes (top group) and the absence of *sopE* and *Salmonella* genomic island (SGI)-1-associated insertion genes (*int* and *xis;* group to the right central group). The majority of the SGI-1 genes were absent in all of the strains in the study. The plasmid groups are largely separated by those with the IncFIB-IncFIC(FII) plasmid (top half of Figure 4B) and those lacking the plasmid (bottom groups).

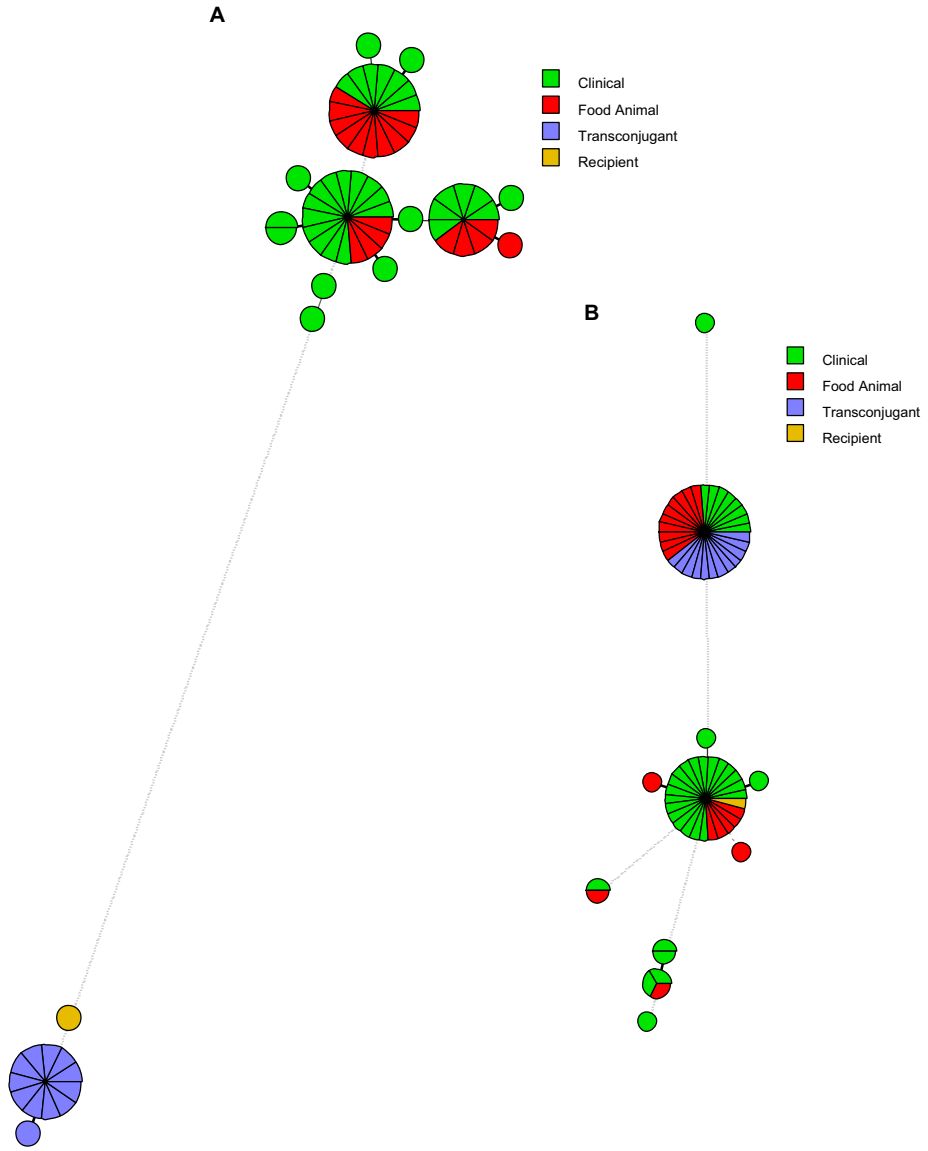


FIGURE 4
Minimum spanning tree analyses of the **(A)** *Salmonella* virulence factors and **(B)** the plasmid transfer genes. The colors correspond to the sample types and the size of the circles are representative of the number of strains sharing common virulence or transfer gene profiles. The distance between the circles is representative of the number of genes differences in the different groups. The closer the circles the more similar the gene profiles. In panel **(A)**, the recipient and transconjugants are *E. coli* strains and as such lack many of the *Salmonella* virulence factors and are distinct from the *Salmonella* isolates. In Panel **(B)**, those grouped closest the letter B are the strains that carry the IncFIB-IncFIC(FII) plasmid.

Invasion and persistence assay

To assess the role of the IncFIB-IncFIC(FII) plasmid in invasion and persistence in host cells, human Caco-2 cells were infected with *S.* Schwarzengrund food and clinical isolates (Figure 5). The general trend, for both food and clinical isolates, was that the amount of surviving CFUs of fusion plasmid containing strains were lower after 48 h (persistence) as compared to 1 h (Invasion). The difference between invasion and persistence rate was statistically significant for food isolates ($p = 0.007$), however; this difference was statistically non-significant for clinical isolates ($p = 0.1192$). We observed that IncFIB-IncFIC(FII) plasmid did not significantly add to the invasion and persistence potential of isolates when compared to those without the plasmids (Figures 5,6).

Discussion

IncFIB plasmids can contain both antimicrobial resistance genes and a wide range of virulence factors, hence dissemination of these plasmids in food pathogens is a public health concern. IncFIB and IncFIB-like plasmids are more commonly harbored by *Salmonella* and APEC, which could serve as reservoir for the spread of these plasmids into other Gram-negative bacteria (Johnson et al.,

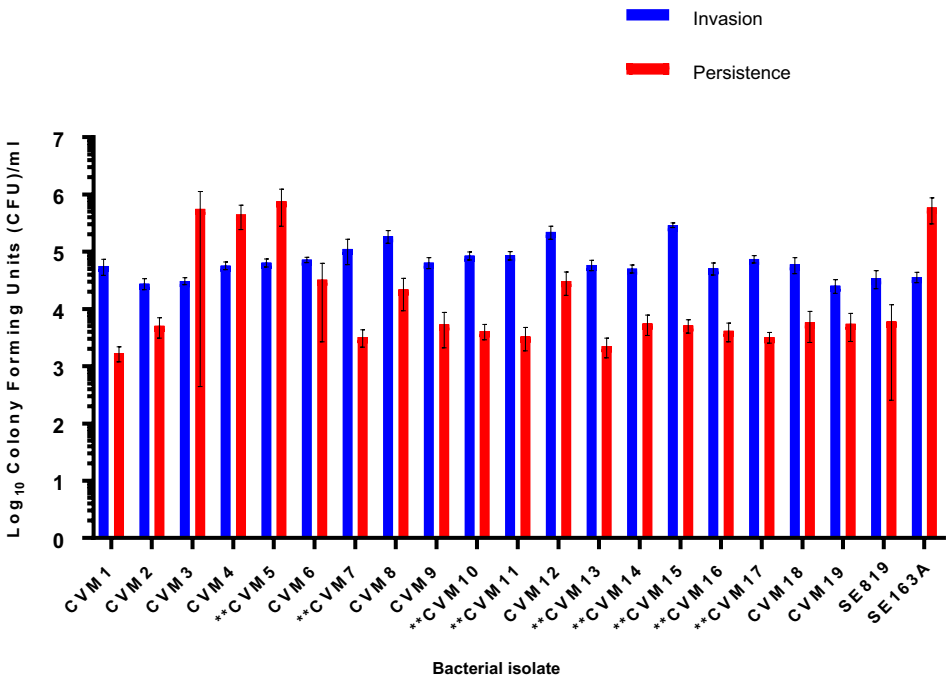


FIGURE 5
Invasion and persistence of 19 wildtype *Salmonella* Schwarzengrund food isolates. Nine food isolates contained IncFIB-IncFIC(FII)are indicated by asterisks. The general trend was that the amount of surviving colony forming units were lower after 48 h (persistence) as compared to 1 h (Invasion). X-axis indicates the number of isolates (CVM 1 to CVM-19), SE819 [less virulent strain that lacked IncFIB-IncFIC(FII)], SE163A (virulent strain that contained IncFIB along with other virulence associated plasmids). Blue bars indicate invasion and red bars indicate persistence. Error bars indicate standard error of mean. The difference between invasion and persistence were analyzed by Student t-test (two-tailed). A $p \leq 0.05$ was considered significantly different between two groups. The difference between invasion and persistence rate for food isolates was statistically significant ($p = 0.007$).

2006, 2010; Khajanchi et al., 2017). In this study, we extensively characterized *S.* Schwarzengrund isolated from food and clinical samples, some of which contained IncFIB-IncFIC(FII) fusion plasmids. In the strains isolated from food samples, all the IncFIB-IncFIC(FII) plasmid containing *S.* Schwarzengrund strains were isolated from chicken sources, indicating that IncFIB-IncFIC(FII) plasmids may be associated with host specific advantages. In clinical isolates, the majority of the IncFIB-IncFIC(FII) plasmid containing *S.* Schwarzengrund strains were isolated from stool and two strains were isolated from urine and the gallbladder. These data suggested that IncFIB-IncFIC(FII) plasmids might play a role in extra-intestinal *Salmonella* infection; however, further study is warranted to provide an experimental basis for this speculation. In addition, acquisition of a similar plasmid such as ColV plasmid by *S.* Kentucky enabled its extraintestinal virulence in chickens (Johnson et al., 2010). It is interesting to note that isolates carrying the IncFIB-IncFIC(FII) fusion plasmid formed a subclade in the SNP phylogeny, separated from the isolates lacking the fusion plasmid. This result implies that the isolates with IncFIB-IncFIC(FII) plasmid is a lineage defined by the acquisition of the IncFIB-IncFIC(FII) plasmid. These data agreed with our previous findings on *S. Typhimurium* in which IncFIB containing *S. Typhimurium* isolates were clustered together and separated from the isolates that did not carry IncFIB plasmids (Khajanchi et al., 2017). Additionally, these findings indicate that there is a possible epidemiological link between IncFIB-IncFIC(FII) containing food and clinical isolates, which warrants further investigation. Additionally, out of eight IncFIB-IncFIC(FII) plasmid containing isolates from humans, two patients reported eating undercooked chicken; while the sources of the remaining six isolates were unknown, with patients either not remembering their exposures to food-related sources or did not agree to an interview.

A tanglegram showed the lack of congruence between the whole genome SNP analysis and the plasmid phylogenies (Supplementary Figure 3). The different topologies of the fusion plasmid carrying isolates in the whole genome phylogeny (Figure 1) and the concatenated core gene plasmid phylogeny (Figure 2) indicates that, although the lineage might have acquired and maintained the plasmid, the plasmid still might have been horizontally transferred within the lineage.

The association of the best NCBI BLAST hit plasmids with human and animal illness supports a possible association with virulence. The virulome analyses showed a similar virulence and plasmid profile for food and clinical isolates. These commonalities are quite evident in Figure 4, where each of the *Salmonella* groupings, for both virulence genes and plasmid genes, have a least three isolates from both food animal and clinical sources. In the virulome analyses, there was some separation of the isolate cluster carrying the IncFIB-IncFIC(FII) fusion plasmids vs. the isolates that lack the fusion plasmid. These differences were due to virulence-associated genes carried on the plasmids (aerobactin operon and *traT*, primarily). Among the isolates not carrying the fusion plasmid, there were two larger subgroups that differed in the carriage of *sopE, int*, and *xis*. The two latter genes are associated with SGI-1. Most of the other SGI-1 associated genes in the

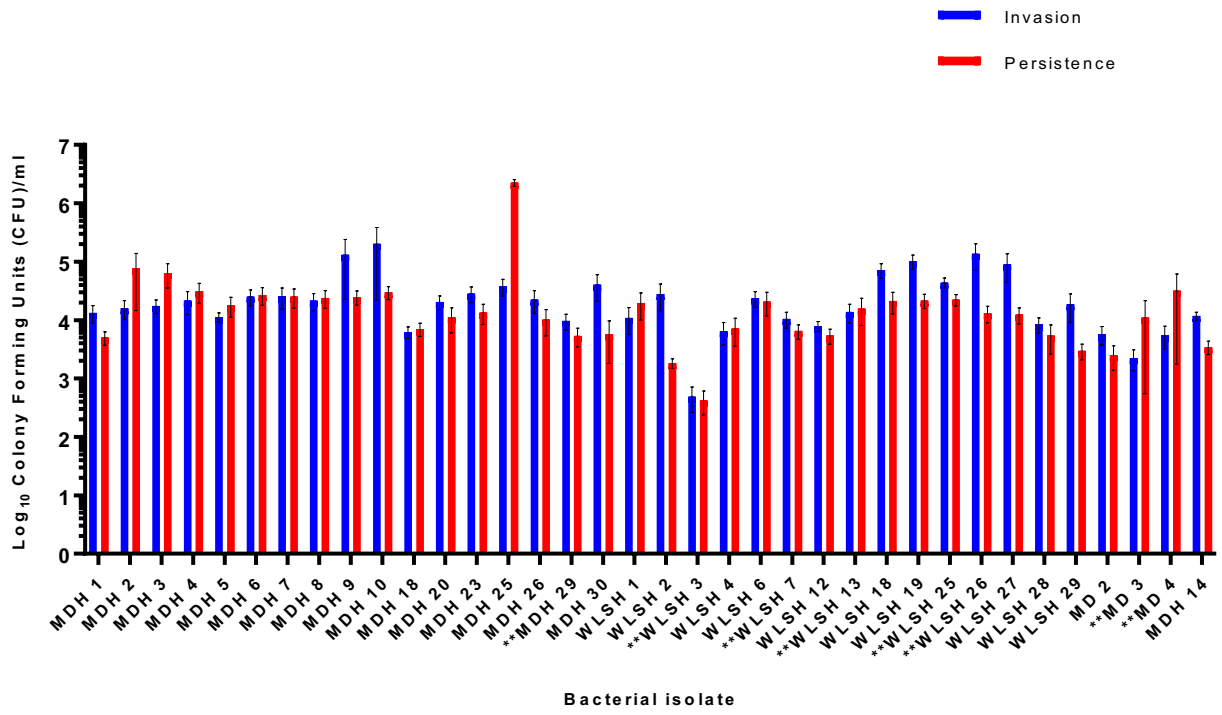


FIGURE 6
Invasion and persistence of 36 wildtype *Salmonella* Schwarzengrund clinical isolates. Eight clinical isolates containing IncFIB-IncFIC(FII) are indicated by asterisks. The general trend was that the number of surviving colony-forming units was lower in persistence as compared to invasion. X-axis indicates the number of isolates with sample ID and Y-axis represents the CFU/ml. Blue bars indicate invasion and red bars indicate persistence. Error bars indicate standard error of mean. The difference between invasion and persistence were analyzed by Student t-test (two-tailed). A $p \leq 0.05$ was considered significantly different between two groups. The difference between invasion and persistence rate for clinical isolates was statistically non-significant ($p = 0.1192$).

database, with the exceptions of *res* and *yidY,* were absent in all of the isolates.

The conjugation of IncFIB and IncFIB-like plasmids has been linked to enhanced virulence in the recipient bacteria. For example, avirulent avian *E. coli* in chick embryo models became virulent after acquiring IncFIB plasmids through transformation (Wooley et al., 1996) and IncFIB transconjugants had increased colonization survival in host compared to recipients (Khajanchi et al., 2017). The *S.* Schwarzengrund IncFIB-IncFIC(FII) fusion plasmids were conjugative as evidenced by their transferability to *E. coli*. The only exception were five clinical isolates whose fusion plasmid did not transfer to *E. coli*, after several attempts. The genetic cause of this could not be determined. Studies suggest that the spread of IncFIB plasmids may lead to improved survival in humans and food animals, increasing the chance for human infection (Johnson et al., 2010; Khajanchi et al., 2017). The spread of antimicrobial resistance genes has also been linked to IncFIB plasmids (Han et al., 2012). Dissemination of IncFIB plasmids along with antimicrobial resistance genes may decrease the effectiveness of antimicrobial therapies for diseases (Khajanchi et al., 2017).

As noted in previous studies, IncFIB plasmids possess several factors that enhance bacterial pathogenicity, such as the production of colicins that kill closely related species, the immunity gene that protects the bacteria from its own bacteriocin, plasmid transfer genes, and the aerobactin system that allows bacteria to combat the host's immunity and to sequester iron for its own survival (Weinberg, 1978; Waters and Crosa, 1991; Khajanchi et al., 2017; Khajanchi, 2022). Various research has been conducted on the contribution of IncFIB plasmids on microbial pathogenesis. However, more work is needed to answer questions about host range, host specificity, environmental sources, and the role of virulence factors encoded by the IncFIB and similar plasmids in bacterial pathogenicity. One aim of this study was to evaluate the virulence capacity of IncFIB-IncFIC(FII) containing *S.* Schwarzengrund by conducting invasion and persistence assessments using tissue culture assay on Caco-2 cells. Previous studies have shown that IncFIB plasmids increased the colonization of chickens by *E. coli* (Ginns et al., 2000), and also enhanced the ability of *S.* Kentucky to colonize chickens (Johnson et al., 2010). Our previous study showed that IncFIB transconjugants had higher invasion and persistence, suggesting that IncFIB plasmids can increase colonization of pathogens in the gut (Khajanchi et al., 2017).

In the present study, Figures 5, 6 show that the IncFIB-IncFIC(FII) plasmids did not significantly contribute to the invasion or persistence of *S.* Schwarzengrund strains isolated from food and clinical sources. The contribution of IncFIB encoded virulence factors in *Salmonella* pathogenesis is still not completely understood. IncFIB or similar plasmids could be diverse in nature and may play distinct roles in different *Salmonella* serovars.

Conclusion

The core gene phylogeny of the IncFIB-IncFIC(FII) fusion plasmid revealed that the *S.* Schwarzengrund isolates might

be descended from the plasmids of avian pathogenic isolates, indicating it might confer adaptation to avian hosts. Further, the fusion plasmid carried several virulence factors that might increase the pathogenicity of its bacterial host. This study shows that the IncFIB-IncFIC(FII) fusion plasmids can be transferred between Enterobacteriaceae species. Though the plasmid has virulence factors such as iron acquisition systems; toxin-antitoxin modules that should increase the pathogenicity of *Salmonella*, our assays showed that there was no difference in invasion and persistence for the isolates with or without of IncFIB-IncFIC(FII). More research is needed to determine the correlation between the virulence factors and the overall pathogenicity of *S.* Schwarzengrund. Future studies will explore the invasion and persistence of *Salmonella* transconjugants in tissue culture. This study highlights that a better understanding of the role of plasmids in *Salmonella* pathogenesis is needed, and that plasmids might be a significant microbiological hazard associated with food.

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 Minnesota Department of Health, Maryland Department of Health, Wisconsin State Lab of Hygiene. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MF: Investigation, Methodology, Writing – original draft. DS: Data curation, Investigation, Writing – original draft. SC: Methodology, Writing – review & editing, Data curation. NY: Methodology, Writing – review & editing, Investigation. NA: Investigation, Methodology, Writing – review & editing. CG: Investigation, Methodology, Writing – review & editing. CA: Investigation, Methodology, Writing – review & editing. AC: Investigation, Methodology, Writing – review & editing. JH: Investigation, Methodology, Writing – review & editing. YS: Investigation, Methodology, Writing – review & editing. SZ: Investigation, Methodology, Writing – review & editing, Resources. XW: Methodology, Resources, Writing – review & editing. SF: Methodology, Resources, Writing – review & editing, Data curation, Funding acquisition, Investigation. BK: Investigation, Methodology, Writing – review & editing, Conceptualization, Supervision, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. CA and NY were supported by Summer Student Research Program (SSRP), Oak Ridge Institute for Science and Education (ORISE). AC and NA were supported through fellowships administered through ORISE. MF was supported by the Master's program at University of Arkansas, Pine Bluff.

Acknowledgments

The authors thank Ahmad Hikal and Saeed Khan for critical review of the manuscript.

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

Author disclaimer

The opinions expressed in this manuscript are solely the authors and do not necessarily represent the official views and policy of the U.S. Food and Drug Administration. Reference to any commercial materials, equipment, or process does not in any way constitute approval, endorsement, or recommendation by the U.S. Food and Drug Administration.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2024.1397068/full#supplementary-material

References

Aarestrup, F. M., Hendriksen, R. S., Lockett, J., Gay, K., Teates, K., McDermott, P. F., et al. (2007). International spread of multidrug-resistant *Salmonella* Schwarzengrund in food products. *Emerg. Infect. Dis.* 13, 726–731. doi: 10.3201/eid1305.061489

Akiyama, T., and Khan, A. A. (2012). Molecular characterization of strains of fluoroquinolone-resistant *Salmonella enterica* serovar Schwarzengrund carrying multidrug resistance isolated from imported foods. *J. Antimicrob. Chemother.* 67, 101–110. doi: 10.1093/jac/dkr414

Algarni, S., Foley, S. L., Tang, H., Zhao, S., Gudeta, D. D., Khajanchi, B. K., et al. (2023). Development of an antimicrobial resistance plasmid transfer gene database for enteric bacteria. *Front Bioinform* 3:1279359. doi: 10.3389/fbinf.2023.1279359

Aljahdali, N. H., Khajanchi, B. K., Weston, K., Deck, J., Cox, J., Singh, R., et al. (2020). Genotypic and phenotypic characterization of incompatibility group FIB positive *Salmonella enterica* Serovar typhimurium isolates from food animal sources. *Genes* 11:1307. doi: 10.3390/genes11111307

Andino, A., and Hanning, I. (2015). *Salmonella enterica*: survival, colonization, and virulence differences among serovars. *ScientificWorldJournal* 2015:520179, 1–16. doi: 10.1155/2015/520179

Asai, T., Murakami, K., Ozawa, M., Koike, R., and Ishikawa, H. (2009). Relationships between multidrug-resistant *Salmonella enterica* Serovar Schwarzengrund and both broiler chickens and retail chicken meats in Japan. *Jpn. J. Infect. Dis.* 62, 198–200. doi: 10.7883/yoken.JJID.2009.198

Bangtrakulnonth, A., Pornreongwong, S., Pulsrikarn, C., Sawanpanyalert, P., and Hendriksen, R. S. (2004). Lo Fo Wong DM, Aarestrup FM: *Salmonella* serovars from humans and other sources in Thailand, 1993-2002. *Emerg. Infect. Dis.* 10, 131–136. doi: 10.3201/eid1001.02-0781

Carattoli, A., Zankari, E., Garcia-Fernandez, 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

Centers for Disease Control and Prevention (2008). Update: recall of dry dog and cat food products associated with human *Salmonella* Schwarzengrund infections--United States. *MMWR Morb. Mortal. Wkly Rep.* 57, 1200–1202,

Chen, M. H., Wang, S. W., Hwang, W. Z., Tsai, S. J., Hsih, Y. C., Chiou, C. S., et al. (2010). Contamination of *Salmonella* Schwarzengrund cells in chicken meat from traditional marketplaces in Taiwan and comparison of their antibiograms with those of the human isolates. *Poult. Sci.* 89, 359–365. doi: 10.3382/ps.2009-00001

Davis, S., Pettengill, J., Lou, Y., Payne, J., Shpuntoff, A., Rand, H., et al. (2015). CFSAN SNP pipeline: an automated method for constructing snp matrices from next-generation sequence data. *PeerJ Comput Sci* 1:e20. doi: 10.7717/peerj-cs.20

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

de Jesus Bertani, A. M., Vieira, T., Reis, A. D., Dos Santos, C. A., de Almeida, E. A., Camargo, C. H., et al. (2023). Whole genome sequence analysis of the first reported isolate of *Salmonella* Agona carrying blaCTX-M-55 gene in Brazil. *Sci. Rep.* 13:2299. doi: 10.1038/s41598-023-29599-5

Duc, V. M., Shin, J., Nagamatsu, Y., Fuhiwara, A., Toyofuku, H., Obi, T., et al. (2020). Increased *Salmonella* Schwarzengrund prevalence and antimicrobial susceptibility of *Salmonella enterica* isolated from broiler chickens in Kagoshima prefecture in Japan between 2013 and 2016. *J. Vet. Med. Sci.* 82, 585–589. doi: 10.1292/jvms.20-0096

Fabrega, A., and Vila, J. (2013). *Salmonella enterica* serovar typhimurium skills to succeed in the host: virulence and regulation. *Clin. Microbiol. Rev.* 26, 308–341. doi: 10.1128/CMR.00066-12

Ginns, C. A., Benham, M. L., Adams, L. M., Whithear, K. G., Bettelheim, K. A., Crabb, B. S., et al. (2000). Colonization of the respiratory tract by a virulent strain of avian *Escherichia coli* requires carriage of a conjugative plasmid. *Infect. Immun.* 68, 1535–1541. doi: 10.1128/IAI.68.3.1535-1541.2000

Hallstrom, K., and McCormick, B. A. (2011). *Salmonella* interaction with and passage through the intestinal mucosa: through the Lens of the organism. *Front. Microbiol.* 2:88. doi: 10.3389/fmicb.2011.00088

Han, J., Lynne, A. M., David, D. E., Tang, H., Xu, J., Nayak, R., et al. (2012). DNA sequence analysis of plasmids from multidrug resistant *Salmonella enterica* serotype Heidelberg isolates. *PLoS One* 7:e51160. doi: 10.1371/journal.pone.0051160

Jacoby, G. A., and Han, P. (1996). Detection of extended-spectrum beta-lactamases in clinical isolates of *Klebsiella pneumoniae* and *Escherichia coli*. *J. Clin. Microbiol.* 34, 908–911. doi: 10.1128/jcm.34.4.908-911.1996

Johnson, T. J., Siek, K. E., Johnson, S. J., and Nolan, L. K. (2006). DNA sequence of a ColV plasmid and prevalence of selected plasmid-encoded virulence genes among avian *Escherichia coli* strains. *J. Bacteriol.* 188, 745–758. doi: 10.1128/JB.188.2.745-758.2006

Johnson, T. J., Thorsness, J. L., Anderson, C. P., Lynne, A. M., Foley, S. L., Han, J., et al. (2010). Horizontal gene transfer of a ColV plasmid has resulted in a dominant avian clonal type of *Salmonella enterica* serovar Kentucky. *PLoS One* 5:e15524. doi: 10.1371/journal.pone.0015524

Khajanchi, B. K. (2022). Foley SL: antimicrobial resistance and increased virulence of *Salmonella*. *Microorganisms* 10:829. doi: 10.3390/microorganisms10091829

Khajanchi, B. K., Han, J., Gokulan, K., Zhao, S., Gies, A., and Foley, S. L. (2016). Draft genome sequences of four *Salmonella enterica* strains isolated from Turkey-associated sources. *Genome Announc.* 4:16. doi: 10.1128/genomeA.01122-16

Khajanchi, B. K., Hasan, N. A., Choi, S. Y., Han, J., Zhao, S., Colwell, R. R., et al. (2017). Comparative genomic analysis and characterization of incompatibility group FIB plasmid encoded virulence factors of *Salmonella enterica* isolated from food sources. *BMC Genomics* 18:570. doi: 10.1186/s12864-017-3954-5

Khajanchi, B. K., Kaldhone, P. R., and Foley, S. L. (2019b). Protocols of conjugative plasmid transfer in *Salmonella*: plate, broth, and filter mating approaches. *Methods Mol. Biol.* 2016, 129–139. doi: 10.1007/978-1-4939-9570-7_12

Khajanchi, B. K., Yoskowitz, N. C., Han, J., Wang, X., and Foley, S. L. (2019a). Draft genome sequences of 27 *Salmonella enterica* Serovar Schwarzengrund isolates from clinical sources. *Microbiol Resour Announc* 8:18. doi: 10.1128/MRA.01687-18

Liu, Y. Y., He, D. D., Zhang, M. K., Pan, Y. S., Wu, H., Yuan, L., et al. (2021). The formation of two hybrid plasmids mediated by IS26 and Tn6952 in *Salmonella enterica* serotype Enteritidis. *Front. Microbiol.* 12:676574. doi: 10.3389/fmicb.2021.676574

Nair, D. V. T., Venkitanarayanan, K., and Kollanoor Johny, A. (2018). Antibiotic-resistant *Salmonella* in the food supply and the potential role of antibiotic alternatives for control. *Food Secur.* 7:167. doi: 10.3390/foods7100167

Ochman, H., Soncini, F. C., Solomon, F., and Groisman, E. A. (1996). Identification of a pathogenicity island required for *Salmonella* survival in host cells. *Proc. Natl. Acad. Sci. U. S. A.* 93, 7800–7804. doi: 10.1073/pnas.93.15.7800

Scallan, E., Hoekstra, R. M., Angulo, F. J., Tauxe, R. V., Widdowson, M. A., Roy, S. L., et al. (2011). Foodborne illness acquired in the United States—major pathogens. *Emerg. Infect. Dis.* 17, 7–15. doi: 10.3201/eid1701.P11101

Sopovski, D., Commichaux, S., Zhao, S., Grim, C. J., Foley, S. L., and Khajanchi, B. K. (2024). Complete genome sequences of 17 *Salmonella enterica* serovar Schwarzengrund isolates carrying an IncFIB-IncFIC(FII) fusion plasmid. *Microbiol Resour Announc* 13:e0106223. doi: 10.1128/mra.01062-23

Tate, H., Hsu, C. H., Chen, J. C., Han, J., Foley, S. L., Folster, J. P., et al. (2022). Genomic diversity, antimicrobial resistance, and virulence gene profiles of Salmonella Serovar Kentucky isolated from humans, food, and animal ceca content sources in the United States. *Foodborne Pathog. Dis.* 19, 509–521. doi: 10.1089/fpd.2022.0005

Taylor, T. L., Volkening, J. D., DeJesus, E., Simmons, M., Dimitrov, K. M., Tillman, G. E., et al. (2019). Rapid, multiplexed, whole genome and plasmid sequencing of foodborne pathogens using long-read nanopore technology. *Sci. Rep.* 9:16350. doi: 10.1038/s41598-019-52424-x

van der Heijden, J., and Finlay, B. B. (2012). Type III effector-mediated processes in *Salmonella* infection. *Future Microbiol.* 7, 685–703. doi: 10.2217/fmb.12.49

Waters, V. L., and Crosa, J. H. (1991). Colicin V virulence plasmids. *Microbiol. Rev.* 55, 437–450. doi: 10.1128/mr.55.3.437-450.1991

Weinberg, E. D. (1978). Iron and infection. *Microbiol. Rev.* 42, 45–66. doi: 10.1128/mr.42.1.45-66.1978

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

Wooley, R. E., Gibbs, P. S., Dickerson, H. W., Brown, J., and Nolan, L. K. (1996). Analysis of plasmids cloned from a virulent avian *Escherichia coli* and transformed into *Escherichia coli* DH5 alpha. *Avian Dis.* 40, 533–539. doi: 10.2307/1592260

Yoon, S., and Lee, Y. J. (2022). Molecular characteristics of ESBL-producing *Escherichia coli* isolated from chickens with colibacillosis. *J. Vet. Sci.* 23:e37. doi: 10.4142/jvs.21105

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644. doi: 10.1093/jac/dks261

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 19 June 2024
DOI 10.3389/fmicb.2024.1424868

OPEN ACCESS

EDITED BY
Xiangyu Fan,
University of Jinan, China

REVIEWED BY
Sankarasubramanian Jagadesan,
University of Nebraska Medical Center, United States
Longxiang Xie,
Henan University, China

*CORRESPONDENCE
Pan Hu
✉ hupan84@163.com

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

RECEIVED 28 April 2024
ACCEPTED 07 June 2024
PUBLISHED 19 June 2024

CITATION
Zhang B, Ren H, Wang X, Han C, Jin Y, Hu X, Shi R, Li C, Wang Y, Li Y, Lu S, Liu Z and Hu P (2024) Comparative genomics analysis to explore the biodiversity and mining novel target genes of *Listeria monocytogenes* strains from different regions. *Front. Microbiol.* 15:1424868. doi: 10.3389/fmicb.2024.1424868

Comparative genomics analysis to explore the biodiversity and mining novel target genes of *Listeria monocytogenes* strains from different regions

Bo Zhang[1†], Honglin Ren[1†], Xiaoxu Wang[2†], Cheng Han[1], Yuanyuan Jin[1], Xueyu Hu[1], Ruoran Shi[1], Chengwei Li[1], Yuzhu Wang[1], Yansong Li[1], Shiying Lu[1], Zengshan Liu[1] and Pan Hu[1*]

[1]State Key Laboratory for Diagnosis and Treatment of Severe Zoonotic Infectious Diseases, Key Laboratory for Zoonosis Research of the Ministry of Education, Institute of Zoonosis, and College of Veterinary Medicine, Jilin University, Changchun, China, [2]Institute of Special Animal and Plant Sciences, Chinese Academy of Agricultural Sciences, Changchun, Jilin, China

As a common foodborne pathogen, infection with *L. monocytogenes* poses a significant threat to human life and health. The objective of this study was to employ comparative genomics to unveil the biodiversity and evolutionary characteristics of *L. monocytogenes* strains from different regions, screening for potential target genes and mining novel target genes, thus providing significant reference value for the specific molecular detection and therapeutic targets of *L. monocytogenes* strains. Pan-genomic analysis revealed that *L. monocytogenes* from different regions have open genomes, providing a solid genetic basis for adaptation to different environments. These strains contain numerous virulence genes that contribute to their high pathogenicity. They also exhibit relatively high resistance to phosphonic acid, glycopeptide, lincosamide, and peptide antibiotics. The results of mobile genetic elements indicate that, despite being located in different geographical locations, there is a certain degree of similarity in bacterial genome evolution and adaptation to specific environmental pressures. The potential target genes identified through pan-genomics are primarily associated with the fundamental life activities and infection invasion of *L. monocytogenes*, including known targets such as *inlB*, which can be utilized for molecular detection and therapeutic purposes. After screening a large number of potential target genes, we further screened them using hub gene selection methods to mining novel target genes. The present study employed eight different hub gene screening methods, ultimately identifying ten highly connected hub genes (*bglF_1, davD, menE_1, tilS, dapX, iolC, gshAB, cysG, trpA, and hisC*), which play crucial roles in the pathogenesis of *L. monocytogenes*. The results of pan-genomic analysis showed that *L. monocytogenes* from different regions exhibit high similarity in bacterial genome evolution. The PCR results demonstrated the excellent specificity of the *bglF_1* and *davD* genes for *L. monocytogenes*. Therefore, the *bglF_1* and *davD* genes hold promise as specific molecular detection and therapeutic targets for *L. monocytogenes* strains from different regions.

KEYWORDS

Listeria monocytogenes, comparative genomics, pan-genomics, biodiversity, target genes

Introduction

Listeria monocytogenes, a foodborne pathogen, is a Gram-positive rod-shaped bacterium belonging to the genus *Listeria* within the phylum Firmicutes et Bacillota. It is a facultative anaerobe (Wang et al., 2021; Lourenco et al., 2022). Currently, there are 30 species of *Listeria* registered in the prokaryotic nomenclature database LPSN (last accessed on November 7, 2023) (Lu et al., 2022). Among the *Listeria* genus, *L. monocytogenes* is commonly considered a pathogenic strain and the most prevalent species (Sheng et al., 2017; Cain et al., 2023). It can cause listeriosis in humans, particularly in immunocompromised individuals such as neonates, elderly, pregnant women, and those with weakened immune systems, and can present with various symptoms including mild diarrhea, meningitis, and sepsis (Locatelli et al., 2017; Radoshevich and Cossart, 2018). The pathogenicity of bacteria and their unique ability to adapt to their habitats have a distinctive genetic basis. Numerous relationships between genes and pathogenic phenotypes have been identified and studied in *L. monocytogenes* (Fox et al., 2016; Disson et al., 2021). However, the genetic basis of *L. monocytogenes* pathogenicity and environmental adaptability is not fully understood and requires further elucidation.

In recent years, with the widespread application of next-generation sequencing and third-generation sequencing technologies, a large number of *L. monocytogenes* genomes have been sequenced and shared (Jordan and McAuliffe, 2018). Comparative genomics analysis of *L. monocytogenes* strains in different regions can deepen our understanding of their genetic mechanisms for adapting to different environments and their pathogenic lifestyles (Lomonaco et al., 2015). The objective of this study was to employ comparative genomics to unveil the biodiversity and evolutionary characteristics of *L. monocytogenes* strains from different regions, screening for potential target genes and mining novel target genes, thus providing significant reference value for the specific molecular detection and therapeutic targets of *L. monocytogenes* strains. Therefore, we conducted comparative genomics research on *L. monocytogenes* strains from different regions including America, Europe, and Asia. The pan-genomes, core genomes, and potential target genes of each *L. monocytogenes* strain were analyzed, and each strain was subjected to multilocus sequence typing (MLST). Functional analysis of the potential target genes in *L. monocytogenes* was conducted using GO and KEGG annotations. Furthermore, a protein-protein interaction (PPI) network was constructed for potential target genes of *L. monocytogenes*, and eight different hub gene analysis methods were utilized to screen novel target genes from the potential target genes. Finally, the virulence genes, antibiotic resistance genes, plasmids, prophages, and CRISPR-Cas systems of each *L. monocytogenes* strain were investigated.

Materials and methods

Data retrieval and management

In this study, a total of 355 genome sequences were retrieved and downloaded from the NCBI genome database (last accessed on November 7, 2023), including 343 *L. monocytogenes* strains from three different regions (223 from America, 91 from Europe, and 29 from Asia), as well as 12 other *Listeria* species and *non-Listeria* bacterial genomes. Detailed information of the studied *L. monocytogenes* genomes, such as GenBank accession numbers, strain names, genome size, GC content, number of contigs and N50, are summarized in Supplementary Tables S1, S2 (Palma et al., 2017; Chiaverini et al., 2021). To ensure greater representativeness, *L. monocytogenes* isolated from cerebrospinal fluid were prioritized, as these strains can induce severe clinical manifestations. In brief, this entails downloading all genomes of *L. monocytogenes* strains isolated from cerebrospinal fluid in NCBI databases pertaining to America, Europe, and Asia for subsequent analysis. To ensure the specificity of the target genes obtained, *non-Listeria* bacterial genomes were selected based on their high coverage and homology with *Listeria* sequences, using Gram-positive reference strains for analysis. Typically, bacteria belonging to the Gram-positive rods exhibit a genomic coverage and homology percentage exceeding 95% (Wang et al., 2022).

Pan-genomic analysis of *Listeria monocytogenes* and non-target bacterial strains from different regions

The analysis of pan-genomic comparison of *L. monocytogenes* and non-target strains can be used to screen potential target genes. The potential target genes refer to those genes that are unique to *L. monocytogenes* strains and are absent in non-target strains (Li et al., 2021a). In brief, all analyzed genome sequences were annotated using Prokka v1.14.6 (Seemann, 2014), and the output results of Prokka were used for pan-genomic analysis with Roary v3.11.2 (Page et al., 2015). A core genome was determined for each isolate using a 99% cutoff, with a BLASTP identity cutoff of 85% (Pang et al., 2019). Genes that matched with all *L. monocytogenes* strains genome sequences were considered highly conserved and used for subsequent comparisons with other *Listeria* species and *non-Listeria* bacterial genomes.

Pan-genome clusters were defined as core-genes: present in all isolates; soft-core genes: present in at least 95% of isolates; shell-genes (accessory genes): present between 15 and 95% of isolates; and cloud-genes (unique genes): present in less than 15% of isolates (Mafuna et al., 2022).

The potential target genes were screened according to the following criteria: 100% presence in *L. monocytogenes* strains and no

presence in non-target bacterial strains. Then, these potential target genes were used screened against the nucleotide collection (nr/nt) databases using the online BLAST program to ensure specificity (Li et al., 2021b).

Multilocus sequence typing analysis

L. monocytogenes was subjected to MLST using 7 housekeeping genes (*abcZ*, *bglA*, *cat*, *dapE*, *dat*, *ldh*, and *lhkA*) as markers (Henri et al., 2016). MLST profiles were obtained from the *Listeria* database hosted by the Pasteur Institute, France. The MLST v.2.18.0 was used to align reads against these profiles to determine the sequence types (STs), Clonal Complex values (CC) and lineage for each genome (Mafuna et al., 2022).

Phylogenetic analysis

To investigate the phylogenetic relationships between the 343 *L. monocytogenes* from different regions, all the core single-copy genes were extracted and aligned using MAFFT v7.490 (Lu et al., 2022). Then, the aligned sequences were concatenated for each strain with a uniform gene order, and GBLOCKS 0.91b was utilized to remove the poorly aligned positions and divergent regions (Lu et al., 2022). MEGA 11 was used to compute the maximum likelihood (ML) phylogenetic tree (Lu et al., 2022). The online tool Interactive Tree of Life (iTOL) v6 was used to visualize the tree with midpoint rooting, and the geographic location, and the ST typing of each strain were annotated on the tree (Lu et al., 2022).

Functional characteristics of potential target genes

In order to investigate the functional characteristics of genes present exclusively in *L. monocytogenes* strains and absent in non-target bacterial strains (potential target genes), annotation analysis was performed using Gene Ontology enrichment analysis (GO analysis) and Kyoto Encyclopedia of Genes and Genomes enrichment analysis (KEGG analysis) (Gao et al., 2021), and the results were integrated.

Protein-protein interaction network analysis and identification of novel target genes

In this study, the STRING database was utilized to construct PPI networks, and these networks were visualized using Cytoscape v3.10.1 (Li et al., 2021c). The CytoHubba function in Cytoscape v3.10.1 was employed to identify hub genes (novel target genes) from the PPI. The CytoHubba function employs eight distinct algorithms to rank genes in the PPI network, which include Degree, Betweenness, BottleNeck, Closeness, Edge Percolated Component (EPC), Maximum Neighborhood Component (MNC), Radiality, and Stress. The top 10 genes with the highest scores are selected as hub genes (Zhang et al., 2019).

Prediction of virulence factors and antibiotic resistance genes of *Listeria monocytogenes*

The prediction of virulence factor-related genes and antibiotic resistance genes in the *L. monocytogenes* genome was conducted to determine their presence. The Virulence Factors of Pathogenic Bacteria (VFDB) database and The Comprehensive Antibiotic Resistance (CARD) database were employed to detect virulence genes and antibiotic resistance genes in the *L. monocytogenes* genome (Tan et al., 2015; Mafuna et al., 2021), and the results are summarized and presented in a heatmap.

Prediction of MGEs of *Listeria monocytogenes*

Mobile Genetic Elements (MGEs) refer to a class of genetic elements capable of spreading or transferring within a genome, such as plasmids and prophages, which can facilitate the evolution of microorganisms (Castro et al., 2021). The plasmid database PLSDB and the PHAge Search Tool-Enhanced Release (PHASTER) prophage database were utilized to detect MGEs in the *L. monocytogenes* genome (Bosi et al., 2017; Matle et al., 2019). The detection results were summarized and presented in a heatmap.

Prediction of CRISPR-Cas systems of *Listeria monocytogenes*

Predict the genome of *L. monocytogenes* to determine the presence of the CRISPR-Cas system. CRISPRCasFinder was used for the detection and typing of the Clustered Regularly Interspaced Short Palindromic Repeats and Cas genes (CRISPR-Cas) system of the *L. monocytogenes* (Parsons et al., 2021). To obtain the functional CRISPR-Cas system, the presence of both the CRISPR sequence and Cas genes was considered as evidence for an actual CRISPR-Cas system and used for the further analysis, and the detection results are summarized and displayed in a heatmap.

Specific primer design and PCR detection conditions for *Listeria monocytogenes*

Primer design for the sequences of *bglF_1* and *davD* genes was performed using Primer Premier 5 software (Table 1) (He et al., 2022). The primers were synthesized by Sangon Biotech Co., Ltd., Shanghai, China. Primer specificity was tested by PCR analysis of strains from the laboratory collection. Total reaction volume was 25 μL, including 12.5 μL of 2 × Es Taq MasterMix (CWBIO, Beijing, China), 1 μL each of forward and reverse primers (10 μM), 8.5 μL of sterile water, and 2 μL of the purified bacterial genomic DNA as a template. An equal volume of sterile distilled water was used instead of the template as a

10.3389/fmicb.2024.1424868

TABLE 1 Specific target genes and primers used for the detection of *L. monocytogenes*.

Gene	Sequence length/bp	Primer	Sequence (5′/3′)	Encoded protein	Product size/bp
bglF_1	1857	bglF_1F	AAGTGGCTGTCATGTTCG	PTS system beta-glucoside-specific EIIBCA component	616
		bglF_1R	ATCGCTACTCCTGCTCCC		
davD	1,467	davDF	AGTTGCGGCCATTACTCC	Glutarate-semialdehyde dehydrogenase	567
		davDR	TTGTCAATCGCATCTTCG		

negative control. PCR thermal cycling involved an initial denaturation step at 95°C for 10 min, followed by 35 cycles of denaturation at 95°C for 30 s, annealing at 56°C for 30 s, and elongation at 72°C for 1 min, with a final elongation at 72°C for 10 min. PCR products were evaluated by 2% agarose electrophoresis.

Results

Genome statistics and general features

By querying the NCBI genomic database, we identified 343 strains of *L. monocytogenes*. We downloaded and curated the whole-genome sequences of these *L. monocytogenes* strains from the NCBI genomic database, along with the corresponding information (Supplementary Tables S1, S2). Among them, there were 223 isolates from America, with an average genome size of 2.98 (2.8–3.2) Mbp, an average GC content of 37.98 (37.5–38.0)%, the number of contigs ≤214 and an average N50 of 334,248. In Europe, there were 91 *L. monocytogenes* isolates, with an average genome size of 2.96 (2.9–3.2) Mbp, an average GC content of 37.98 (37.5–38.0)%, the number of contigs ≤42 and an average N50 of 656,340. In Asia, there were 29 *L. monocytogenes* isolates, among which, one possesses a complete genome, with an average genome size of 2.96 (2.8–3.1) Mbp, an average GC content of 38%, the number of contigs ≤72 and an average N50 of 699,124.

Pan-genomic analysis of *Listeria monocytogenes* strains in different regions

Based on pan-genomic classification, the analysis of *L. monocytogenes* strains from different regions revealed the following gene distribution within the pan-genome: there were 1847 (15%) core genes, 314 (2.6%) soft-core genes, 1,237 (10.1%) shell genes, and 8,860 (72.3%) cloud genes (Figure 1A). The pan-genomic composition of *L. monocytogenes* varies across different regions. In America, *L. monocytogenes* strains possess 1866 core genes, 149 soft-core genes, 1,351 shell genes, and 7,714 cloud genes. In Europe, *L. monocytogenes* strains possess 2,133 core genes, 103 soft-core genes, 1,267 shell genes, and 3,782 cloud genes. In Asia, *L. monocytogenes* strains possess 2,178 core genes, 16 soft-core genes, 1,306 shell genes, and 1,649 cloud genes (Figures 1B,C). By analyzing and summarizing the genomic, pan-genomic, and core genomic features of *L. monocytogenes* strains from different regions, it was revealed that these strains possess open genomes, which provide a genetic basis for their adaptation to diverse environments.

Pan-genomic analysis of *Listeria monocytogenes* strains from different regions and non-target bacterial strains for the screening of potential target genes

To identify potential target genes in *L. monocytogenes* strains from different regions, we conducted pan-genomic analysis of these strains as well as non-target bacterial strains. Among them, due to the presence of non-target strains, the quantities of core genes are 0, the quantity of soft-core genes are 1919, the quantity of shell genes are 1,458, the quantity of cloud genes are 41,390, and the total quantity of genes are 44,767. A total of 357 potential target genes were detected in *L. monocytogenes* strains from different regions (Supplementary Table S3). These potential target genes were present in the *L. monocytogenes* strains included in this study, while being absent in non-target bacterial strains investigated. These potential target genes have the potential to serve as novel target genes for *L. monocytogenes* strains in different regions, but further screening of these potential target genes is still required.

MLST and phylogenetic analysis

To investigate the correlation among *L. monocytogenes* strains from different regions, MLST was employed to genotype the strains at the whole-genome level, aiming to determine the phylogenetic relationships among different sequence types (STs) and their associations with the disease. In America, the most common among *L. monocytogenes* strains was ST1 ($n = 32$, 14.3%), Clonal Complex 1 (CC1) ($n = 34$, 15.2%) and Lineage I ($n = 161$, 72.2%) (Figure 2A; Supplementary Table S4). In Europe, the most common was ST1 ($n = 14$, 15.4%), CC1 ($n = 14$, 15.4%) and Lineage I ($n = 52$, 57.1%) (Figure 2B; Supplementary Table S4). In Asia, the most common was ST8 ($n = 7$, 24.1%), CC8 ($n = 8$, 27.6%) and Lineage II ($n = 18$, 62.1%) (Figure 2C; Supplementary Table S4). In strains of *L. monocytogenes* in America and Europe, a higher proportion is observed for strains of ST1 and CC1 types. In strains of *L. monocytogenes* in Asia, a higher proportion is observed for strains of ST8 and CC8 types.

Medically intriguingly, *Listeria* species inhabit diverse ecological niches, but only *L. monocytogenes* and *L. ivanovii* exhibit pathogenicity (Orsi and Wiedmann, 2016; Lu et al., 2022). To elucidate the evolutionary patterns of *L. monocytogenes* strains from different regions, we conducted a phylogenetic analysis of *L. monocytogenes* using conserved amino acid sequences of all single-copy genes (Figure 2D). Through the phylogenetic tree analysis, we can observe that *L. monocytogenes* strains in different regions share a common ancestor, all falling within this major root of the *Listeria* genus. Although the geographical locations of the *L. monocytogenes* strains

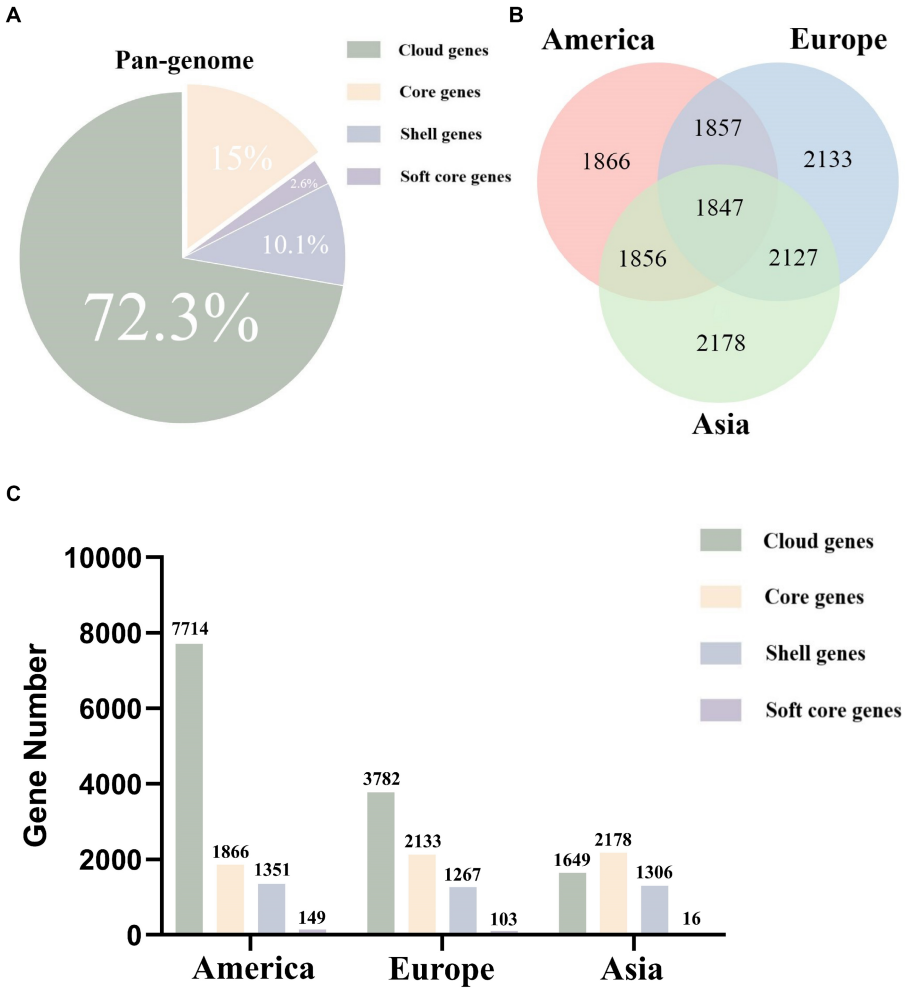


FIGURE 1
The proportion and quantity of various parts of the pan-genome in *L. monocytogenes* from different regions as analyzed through pan-genomics. **(A)** The pan-genome proportion of *L. monocytogenes* in three different regions. **(B)** The core genome Venn diagram of *L. monocytogenes* in three different regions. **(C)** The number of core genes, soft core genes, shell genes, and cloud genes in *L. monocytogenes* from three different regions.

we studied vary significantly, some strains from different locations still cluster within the same branch, indicating they share a relatively similar phylogenetic relationship. This indicates that despite being in different geographic locations and under varying environmental conditions, *L. monocytogenes* exhibits certain similarities in evolutionary mechanisms and genetic variations in response to environmental pressures, displaying a strong adaptability to diverse environments.

Enrichment analysis of the functional characteristics of potential target genes using GO and KEGG

To investigate the functional characteristics of 357 potential target genes in *L. monocytogenes* strains from different regions, we performed functional annotation and classification of these genes using GO and KEGG databases. The detailed information of the potential target genes is presented in Supplementary Table S3. The GO database categorizes gene functions into three main categories, namely Biological Processes (BP), Cellular Components (CC), and Molecular Functions (MF). In the BP category, the most enriched biological processes were cellular process ($n = 124$, 34.7%), nucleobase-containing compound metabolic process ($n = 33$, 9.2%), localization ($n = 31$, 8.7%), and transmembrane transport ($n = 27$, 7.6%). Within the CC category, the most abundant cellular components were cytoplasm ($n = 50$, 14%). In the MF category, the most enriched molecular functions were organic cyclic compound binding ($n = 76$, 21.3%), and purine nucleotide binding ($n = 31$, 8.7%) (Figure 3A). We integrated and ranked all the GO enrichment analysis results of potential target genes, and generated a bubble chart (Figure 3B) to display the top 20 functional features in GO enrichment analysis based on the number of genes and the significance of p values.

The pathway database of KEGG is the most widely used public database for metabolic pathways, which classifies biological metabolic pathways into six categories: Metabolism, Genetic Information Processing, Environmental Information Processing, Cellular Processes, Organismal Systems, and Human Diseases. The potential target genes were annotated using KEGG in six categories. Among these, the most enriched pathways in the Metabolism category were carbohydrate metabolism ($n = 31$, 8.7%), amino acid metabolism ($n = 29$, 8.1%), and metabolism of cofactors and vitamins ($n = 27$, 7.6%). In the Genetic

10.3389/fmicb.2024.1424868

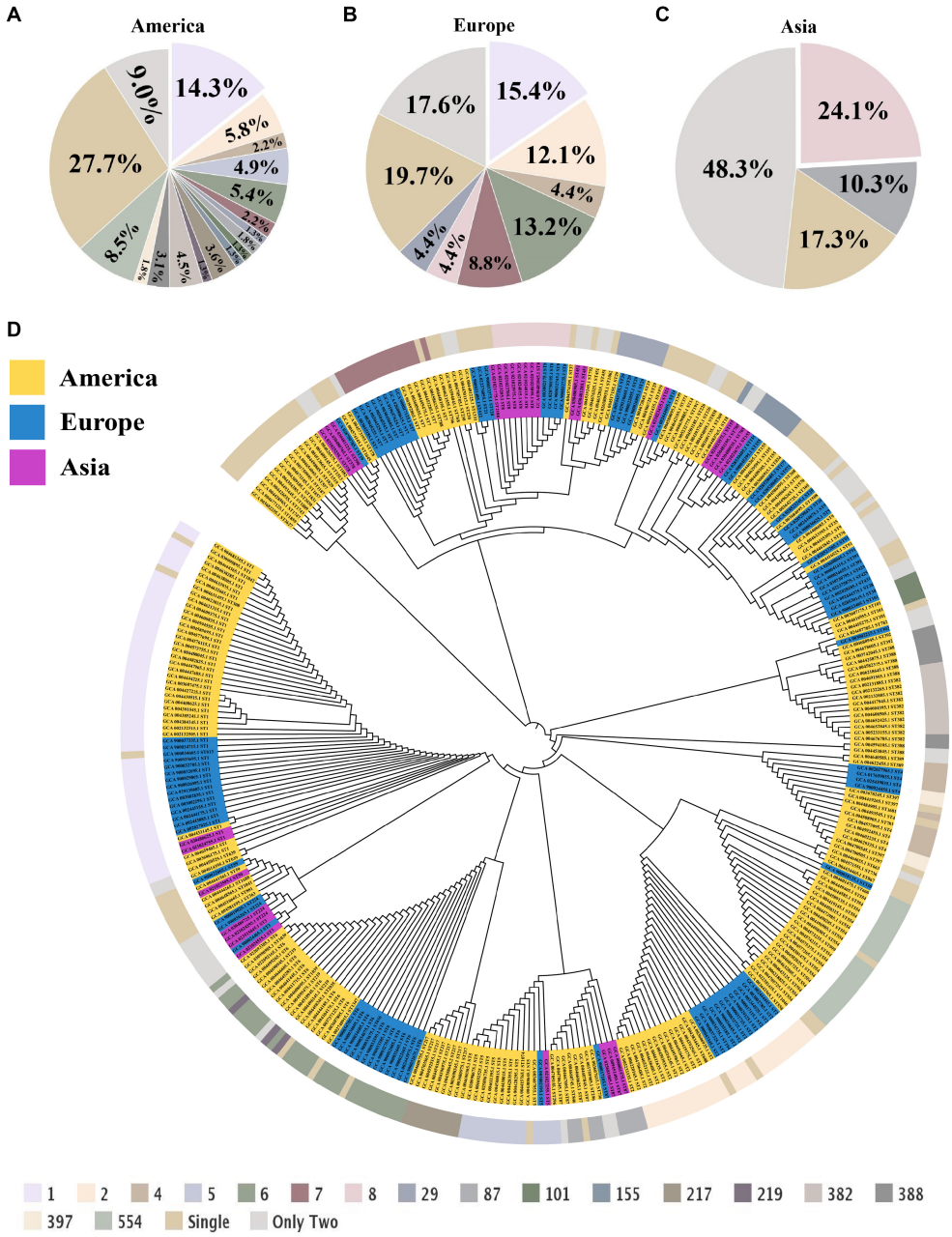


FIGURE 2
Multilocus sequence typing and phylogenetic tree of *L. monocytogenes* from different regions. **(A)** The proportion of STs in *L. monocytogenes* from the Americas region. **(B)** The proportion of STs in *L. monocytogenes* from the European region. **(C)** The proportion of STs in *L. monocytogenes* from the Asian region. **(D)** Phylogenetic trees of *L. monocytogenes* from different regions, with the outer circle indicating the corresponding ST for each strain.

Information Processing category, replication and repair ($n = 10$, 2.8%) were the most enriched pathways. In the Environmental Information Processing category, membrane transport ($n = 19$, 5.3%) and signal transduction ($n = 6$, 1.7%) were the most enriched pathways. In the Cellular Processes category, cellular community – prokaryotes ($n = 4$, 1.1%) and cell growth and death ($n = 2$, 0.6%) were the most enriched pathways. In the Organismal Systems category, endocrine system ($n = 3$, 0.8%) and digestive system ($n = 2$, 0.6%) were the most enriched pathways. In the Human Diseases category, drug resistance: antimicrobial ($n = 4$, 1.1%) and drug resistance: antineoplastic ($n = 2$, 0.6%) were the most enriched pathways (Figure 3C). We integrated and ranked all the KEGG enrichment analysis results of potential target genes, and generated a bubble chart (Figure 3D) to display the top 20 functional features in KEGG enrichment analysis based on the number of genes and the significance of p values.

In summary, the enrichment analysis of functional characteristics of 357 potential target genes using GO and KEGG databases indicates that these genes are primarily associated with metabolic processes, compound binding, protein localization, and transmembrane transport in *L. monocytogenes*. Examples include cellular metabolic

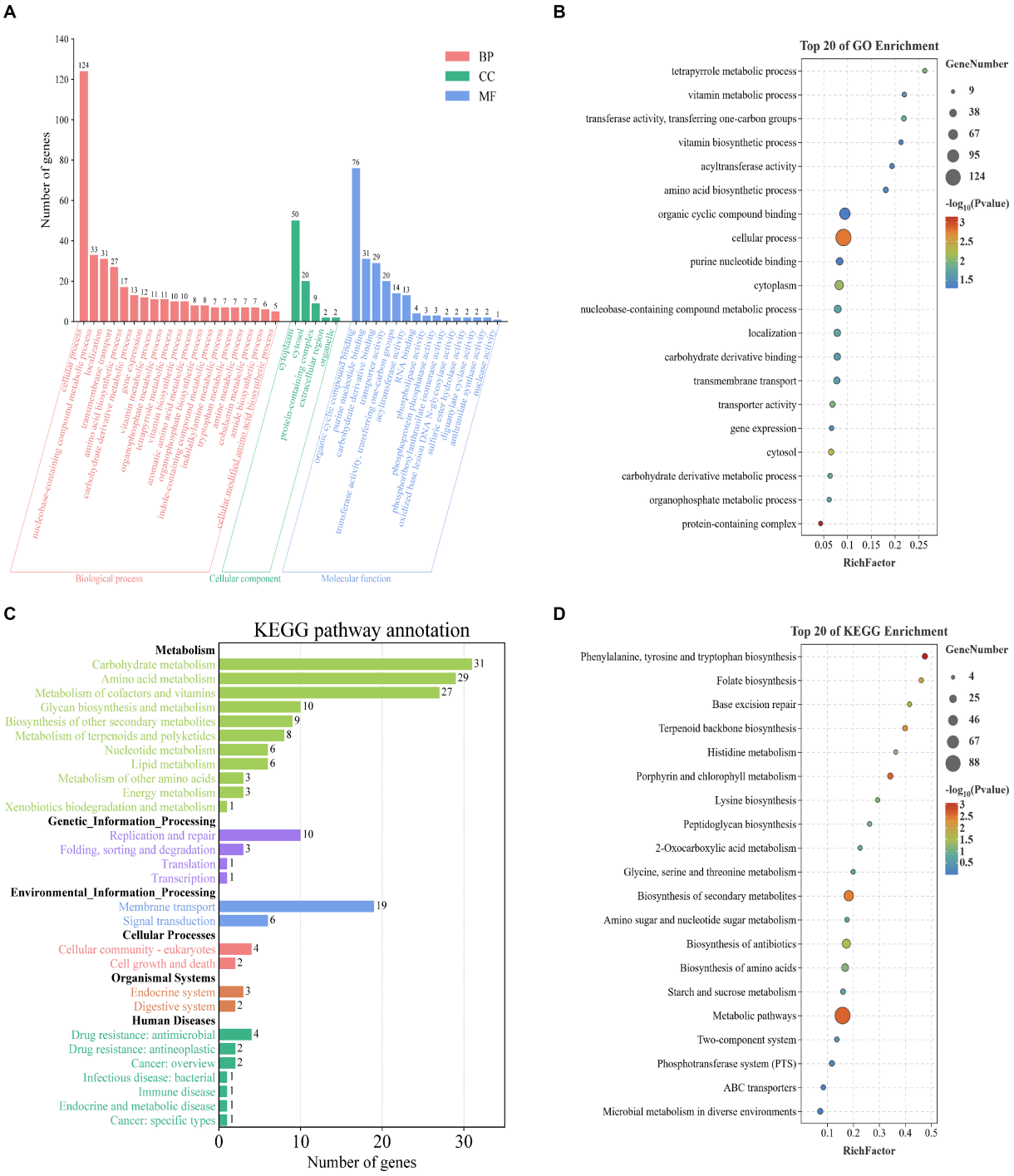


FIGURE 3
Enrichment analysis of potential target genes in *L. monocytogenes* from different regions based on the GO, and KEGG databases. **(A)** Enrichment analysis based on the GO database. **(B)** Enrichment analysis based on GO database with the top 20 enriched terms listed. **(C)** Enrichment analysis based on the KEGG database. **(D)** Enrichment analysis based on KEGG database with the top 20 enriched terms listed.

process, carbohydrate metabolism and organic substance biosynthetic process. However, there were still some genes with unclear functional information, which warrants further investigation in future studies. The potential target genes are closely associated with fundamental biological processes and infection pathogenesis of *L. monocytogenes*, playing crucial roles in sustaining basic life activities, invading the host, and exerting pathogenic effects.

PPI network analysis of potential target genes and identification of novel target genes

The PPI network plays a crucial role in various biological processes within organisms. To further assess the interconnections among potential target genes of *L. monocytogenes* strains in different

regions, PPI analysis was carried out using the STRING database. The PPI network of potential target genes comprised 357 genes, and visualization of the PPI network was performed using Cytoscap_v3.10.1 software. They were clustered together, indicating strong physical interaction or functional association.

To further analyze potential target genes of *L. monocytogenes* for the selection of novel target genes, we employed the CytoHubba function of Cytoscape v3.10.1 software to identify hub genes. Hub genes are key factors in protein-protein interaction networks that exhibit high connectivity in gene expression networks, indicating their ability to regulate multiple genes. The genes in the PPI network were screened using eight different algorithms available in the CytoHubba function. The top 10 genes with the highest scores were selected as hub genes, and their ranking is presented in Table 2. A comprehensive analysis of the results obtained from the eight algorithms was performed, and raincloud plots illustrating the scores of the hub genes for each algorithm was generated (Figure 4B). The top 10 genes with the highest scores identified by the Degree algorithm were ultimately determined as novel target genes, and a PPI network was constructed based on their scores (Figure 4A). Among them, *bglF_1* and *davD* genes had the highest score of 54, followed by *menE_1* gene with a score of 52, *tilS* gene with a score of 50, *dapX* gene with a score of 48, *iolC* gene with a score of 46, *gshAB* gene with a score of 42, *cysG* gene with a score of 42, *trpA* gene with a score of 40, and *hisC* gene with a score of 38. The detailed information regarding these 10 genes, including their functional roles, gene lengths, etc., is provided in Table 3. These 10 hub genes play crucial roles in sustaining basic life activities and infection invasion of *L. monocytogenes* strains, with the potential to become novel target genes for *L. monocytogenes* strains, particularly the top-scoring genes, *bglF_1* and *davD*.

Distribution of virulence genes and antibiotic resistance genes in *Listeria monocytogenes* strains in different regions

To investigate the relationship between *L. monocytogenes* strains in different regions and their pathogenic mechanisms, we predicted virulence factor-encoding genes of the entire genome of *L. monocytogenes*. Based on VFDB prediction and annotation, virulence factors of *L. monocytogenes* were classified into 12 categories including Adherence, Bile resistance, Enzyme, Immune modulator, Intracellular survival, Invasion, Iron uptake, Nucleation-promoting factor, Peptidoglycan modification, Regulation, Surface protein anchoring, and Toxin.

In this study, the virulence genes *dltA*, *fbpA*, *lap*, *plcB*, *stp*, *inlK*, *oppA*, *prsA2*, *inlB*, *lpeA*, *hbp2*, *pdgA*, *agrC*, *cheA*, *lisK*, *lisR*, *prfA*, *virR*, and *virS* were found to be present in 100% of *L. monocytogenes* strains from different regions (Figure 5). Our predictive findings indicate the presence of numerous virulence genes, including *inlB*, and *plcB*, which play pivotal roles in *L. monocytogenes* infection and host invasion, across *L. monocytogenes* strains from various regions. It is well known that *L. monocytogenes* is pathogenic within the *Listeria* genus, owing to its abundance of virulence genes, which contribute to its pathogenicity. Through the prediction of virulence genes, *L. monocytogenes* strains from different regions exhibit high pathogenicity.

With the widespread use of antibiotics, the antimicrobial resistance of foodborne strains has increased in many countries. The distribution of antimicrobial resistance genes in *L. monocytogenes* was investigated using the CARD database. In this study, a total of 10 antimicrobial resistance genes belonging to 9 drug classes and exhibiting 5 resistance mechanisms were identified among the 343 *L. monocytogenes* genomes analyzed. Among the *L. monocytogenes* strains from different regions, 100% were found to harbor four types of antibiotic resistance genes, including phosphonic acid antibiotic gene (*FosX*), glycopeptide antibiotic genes (*vanTG*, *vanYM*), lincosamide antibiotic gene (*lin*), and peptide antibiotic gene (*mprF*) (Figure 6).

Distribution of MGEs in *Listeria monocytogenes* strains in different regions

Currently, many scientists have begun to pay extensive attention to the significant role of horizontal transfer of MGEs in bacterial genome evolution and adaptation to specific environmental pressures. We employed PLSDB and PHASTER databases for the

TABLE 2 Top 10 hub genes ranked by scoring in eight different algorithms.

Catelogy	Rank methods in cytoHubba							
	Degree	Betweenness	BottleNeck	Closeness	EPC	MNC	Radiality	Stress
Gene symbol top 10	bglF_1[a]	bglF_1[a]	davD[b]	davD[b]	dapX	davD[b]	bglF_1[a]	davD[b]
	davD[b]	davD[b]	bglF_1[a]	dapX	hisC	iolC	dapX	bglF_1[a]
	menE_1	iolC	group_1214	bglF_1[a]	bglF_1[a]	menE_1	gshAB	iolC
	tilS	gshAB	gshAB	gshAB	davD[b]	bglF_1[a]	davD[b]	hisC
	dapX	tilS	menE_1	tilS	menE_1	trpA	tilS	menE_1
	iolC	menE_1	hisC	iolC	trpA	hisC	group_1214	gshAB
	gshAB	fni	tilS	menE_1	tilS	tilS	menE_1	cysG
	cysG	cysG	iolC	fni	trpF	ezrA	fni	tilS
	trpA	group_32878	group_8648	hisC	iolC	divIB	hisC	yodC
	hisC	dapX	yceM	group_1214	norG	trpF	hepS	group_3718

[a, b]The two genes highlighted in bold are the most promising candidates to serve as novel target genes.

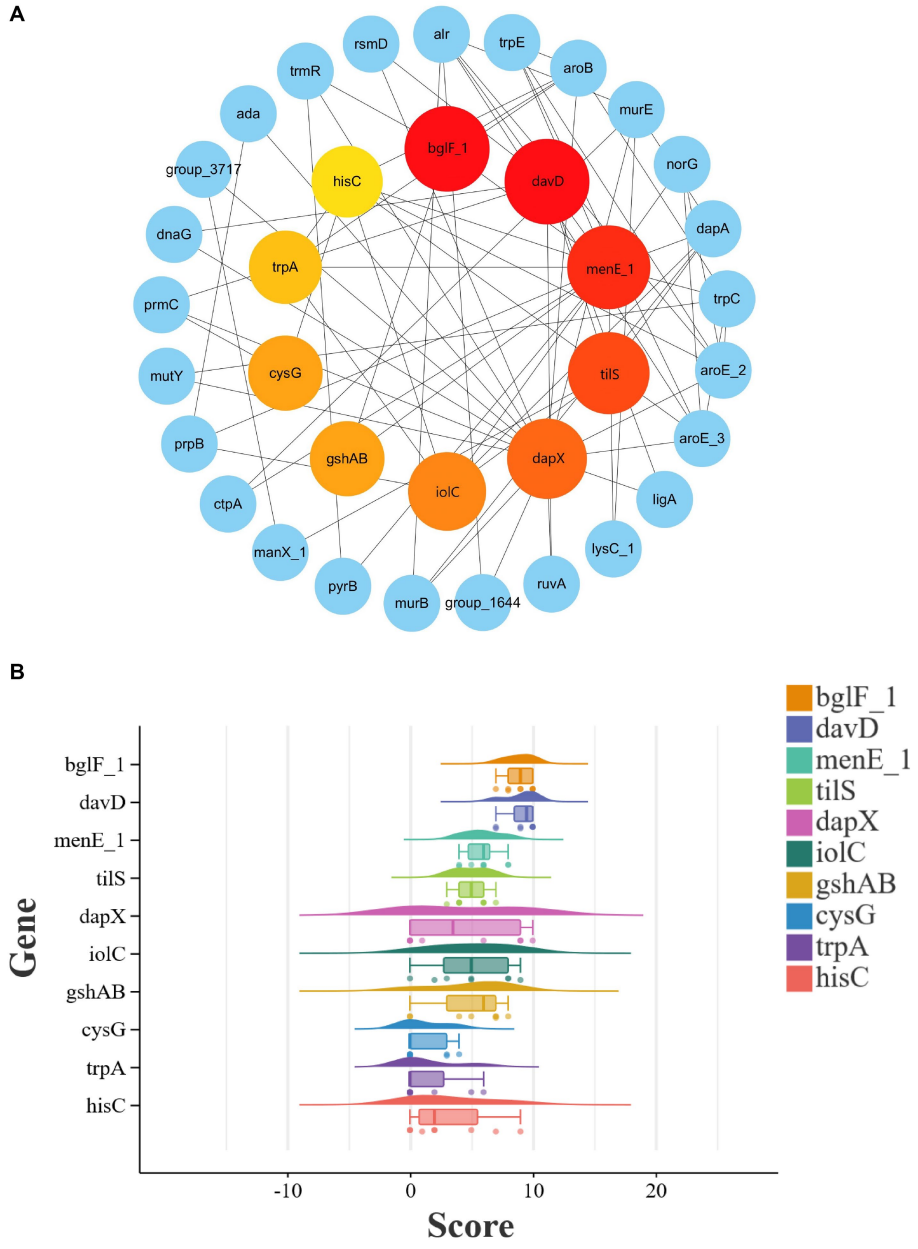


FIGURE 4
PPI network analysis and identification of novel target genes among potential target genes in *L. monocytogenes* from different regions.
(A) Visualization of PPI network among the top 10 hub genes ranked by Degree algorithm. The larger the circle and the deeper the red color, the higher the score of the gene. **(B)** Scoring of hub genes across eight different algorithms.

detection of MGEs, and the results only documented plasmids with an identity score of 1 and intact prophage regions. In this study, the *L. monocytogenes* genome contained a total of 10 plasmids and 7 intact prophage regions. Among them, plasmids *pLmA144*, *pLMR479a*, *pLmcEH-6*, and *pLmcUH29* were exclusively present in ST8, while plasmids *pLIS22* and *pLIS39* were only found in ST3 (Figure 6). The most prevalent prophage was PHAGE_Lister_vB_LmoS_188 [NC_028871] ($n = 64$, 18.7%), followed by PHAGE_Lister_LP_101 [NC_024387] ($n = 48$, 14%) (Figure 6). Intriguingly, despite the geographical disparity of *L. monocytogenes* isolates, identical phage genomes have been detected, suggesting a certain degree of similarity in MGEs across strains from different regions.

Distribution of CRISPR-Cas system types in *Listeria monocytogenes* strains in different regions

The CRISPR-Cas system is a bacterial adaptive immune system that protects bacteria from viral infections, which is also associated with the virulence and pathogenicity of pathogens. In this study, we characterized the CRISPR-Cas systems in 343 *L. monocytogenes* genomes and identified four types of CRISPR-Cas systems. Each CRISPR-Cas system type exhibited distinct cas genes, with a total of 14 cas genes detected (Table 4). The CRISPR-Cas system types detected in *L. monocytogenes* included CAS-TypeIA (4/343), CAS-TypeIB (39/343), CAS-TypeIIA

TABLE 3 Detailed information of top 10 hub genes ranked by scoring in Degree algorithm.

Gene	Name of target genes	Sequence length/bp	Presence profile		Encoded protein	Product size/bp	Source
			In target	In non-target			
bglF_1	lmo1035	1857	343 (100%)	0	PTS system beta-glucoside-specific EIIBCA component	618	This study
davD	lmo0913	1,467	343 (100%)	0	Glutarate-semialdehyde dehydrogenase	488	This study
menE_1	lmo1672	1,409	343 (100%)	0	O-succinylbenzoic acid--CoA ligase	469	This study
tilS	lmo0219	1946	343 (100%)	0	tRNA(Ile)-lysidine synthase	648	This study
dapX	lmo1006	1,146	343 (100%)	0	Putative N-acetyl-LL-diaminopimelate aminotransferase	381	This study
iolC	lmo0385	978	343 (100%)	0	5-dehydro-2-deoxygluconokinase	325	This study
gshAB	lmo2770	2,331	343 (100%)	0	Glutathione biosynthesis bifunctional protein GshAB	776	This study
cysG	lmo1201	1,481	343 (100%)	0	Siroheme synthase	493	This study
trpA	lmo1627	774	343 (100%)	0	Tryptophan synthase alpha chain	257	This study
hisC	lmo1925	1,083	343 (100%)	0	Histidinol-phosphate aminotransferase	360	This study

(21/343), and CAS-TypeIIIA (2/343). Approximately one-fifth of the *L. monocytogenes* genomes (66/343) harbored at least one CRISPR-Cas system, with CAS-TypeIB (11.4%) and CAS-TypeIIA (6%) being the most prevalent (Figure 6). CAS-TypeIA was only detected in ST1, ST5, and ST425, while CAS-TypeIIIA was found exclusively in ST5 and ST392. The CAS-type IA system detected in the *L. monocytogenes* isolates in this study was composed of csa5_TypeIA and casRa_TypeIA. The CAS-TypeIB system was composed of cas5b_TypeIB, cas6_TypeI-III, cas8a1b_TypeIB, cas7b_TypeIB, cas3_TypeI, cas2_TypeI-II-III, cas4_TypeI-II, and cas1_TypeIB. The CAS-TypeIIA system was composed of csn2_TypeIIA, cas2_TypeI-II-III, cas1_TypeII, and cas9_TypeII. The CAS-type IIIA system was composed of csm2_TypeIIIA.

Detection of *Listeria monocytogenes* using specific primers by PCR

To validate the potential of *bglF_1* and *davD* genes as specific molecular detection and therapeutic targets in *L. monocytogenes* strains from different regions, primers were designed for *bglF_1* and *davD* genes, followed by PCR experiments to assess their specificity. The PCR results revealed a distinct band at 616 bp for *L. monocytogenes* in the *bglF_1* gene primer system, while *non-L. monocytogenes* samples showed no band (Figure 7A). Similarly, a clear band at 567 bp was observed for *L. monocytogenes* in the *davD* gene primer system, with no band detected in *non-L. monocytogenes* samples (Figure 7B). The results demonstrated the excellent specificity of the *bglF_1* and *davD* genes for *L. monocytogenes*. Therefore, the *bglF_1* and *davD* genes hold promise as specific molecular detection and therapeutic targets for *L. monocytogenes* strains from different regions.

Discussion

L. monocytogenes, as a significant foodborne pathogen, is widely prevalent worldwide, posing a serious threat to human life and health. Therefore, we conducted comparative genomic analysis of *L. monocytogenes* strains from different regions to explore their biodiversity and evolutionary characteristics, identify potential target genes, and further mining novel target genes, aiming to provide novel specific molecular detection and therapeutic strategies for *L. monocytogenes* strains.

In this study, we conducted a pan-genomic comparative analysis of 343 *L. monocytogenes* strains from different regions to investigate the biodiversity and evolutionary characteristics of strains. To assess the genomic biodiversity of *L. monocytogenes* strains in different regions, we conducted core/pan-genome analysis. Core and accessory genomes were analyzed based on the whole genomes of *L. monocytogenes*. The core genome represents the essential portion necessary for the presence and shared phenotypic features of specific strains, while the accessory genome provides unique characteristics for a species or strain that are not essential for their basic survival, but offer selective advantages for ecological adaptation and antibiotic resistance (Deng et al., 2010; Zhang et al., 2017; Lu et al., 2022). Although the studied strains of *L. monocytogenes* are geographically diverse, there still exist 1847 core genes that constitute the fundamental

10.3389/fmicb.2024.1424868

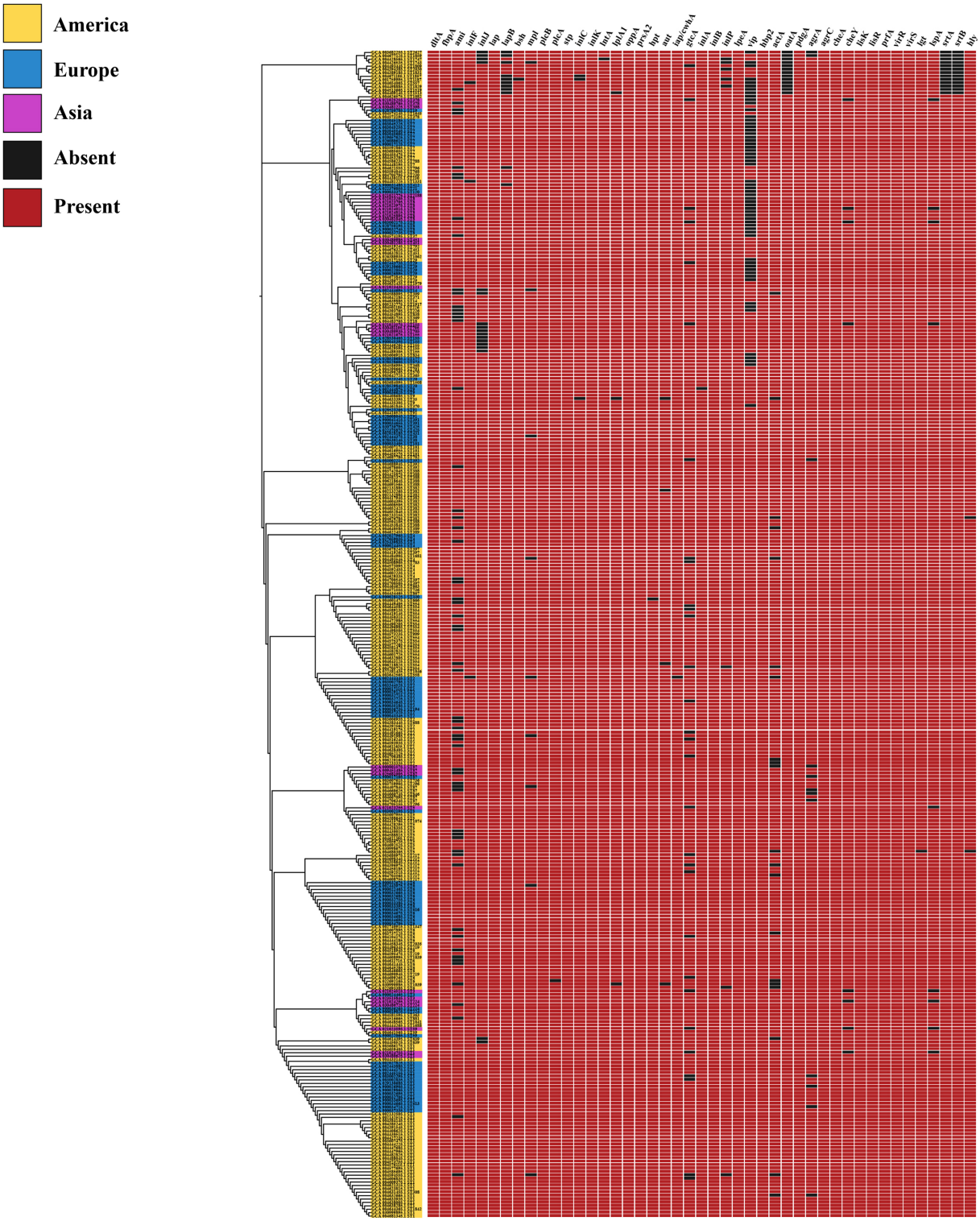


FIGURE 5
The distribution of virulence genes in *L. monocytogenes* from different regions.

components of *L. monocytogenes* survival and development. In America, the number of core genes in *L. monocytogenes* strains is 1866, in Europe it is 2,133, and in Asia it is 2,178. It can be observed that the number of core genes in *L. monocytogenes* strains varies across different regions. Apart from the core genes they share, *L. monocytogenes* strains in different regions also possess unique core genes that are present only in one region and absent in others. These unique core genes may be the primary reason for the distinctiveness of *L. monocytogenes* strains in one region compared to those in other regions. The primary reason for this phenomenon may be attributed

10.3389/fmicb.2024.1424868

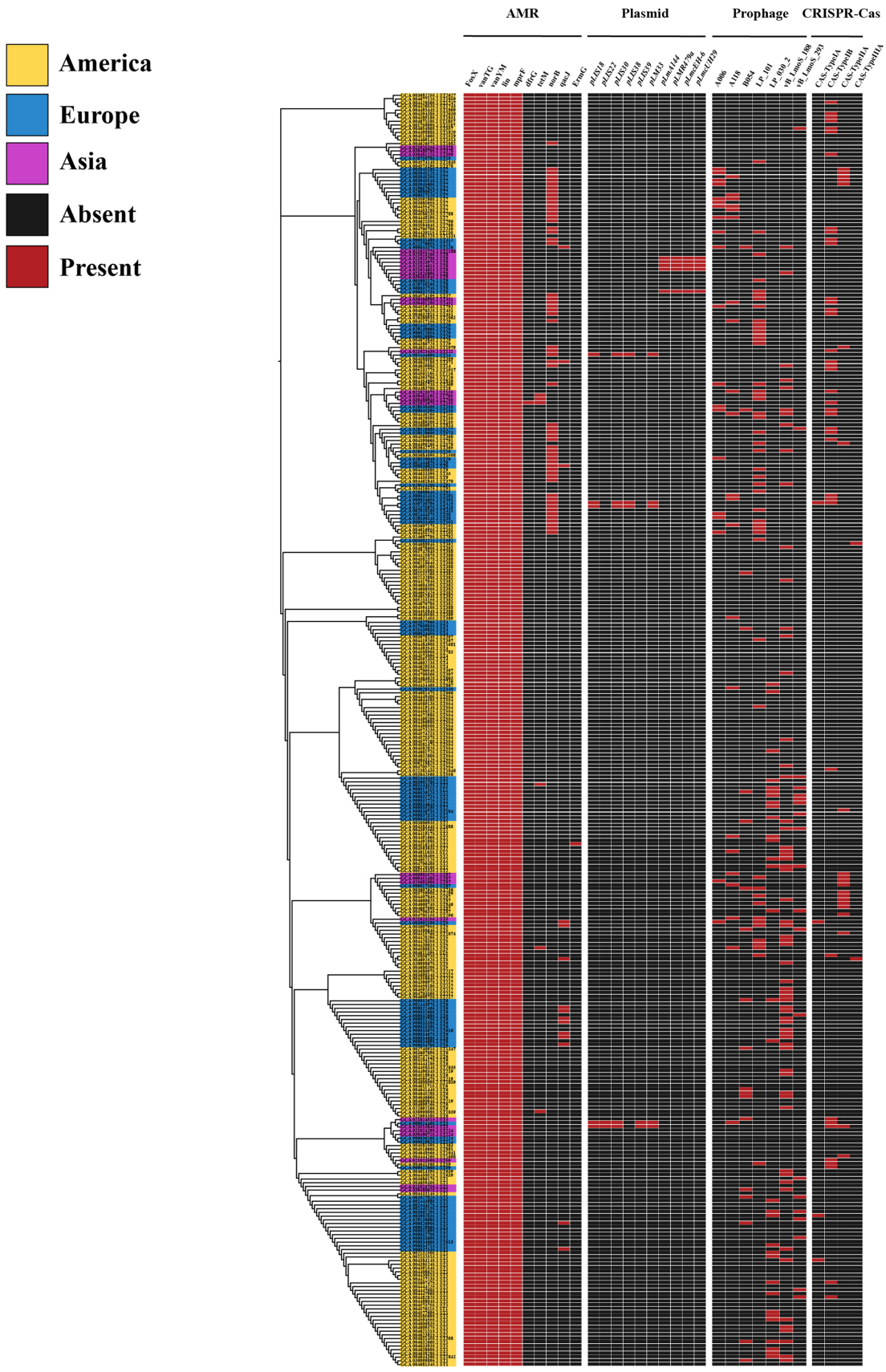


FIGURE 6
The distribution of antibiotic resistance genes, plasmids, prophages, and CRISPR-Cas systems in *L. monocytogenes* from different regions.

TABLE 4 Identification of CRISPR-Cas system types and the corresponding cas genes detected in *L. monocytogenes* strains from different regions.

Types	Cas genes	Area	Number	Total
CAS-type IA	csa5_TypeIA	America	1	2
		Europe	1	
		Asia	0	
CAS-type IA	casRa_TypeIA	America	0	2
		Europe	2	
		Asia	0	
CAS-TypeIB	cas5b_TypeIB	America	3	5
		Europe	1	
		Asia	1	
CAS-TypeIB	cas6_TypeI-III, cas8a1b_TypeIB, cas7b_TypeIB, cas5b_TypeIB, cas3_TypeI, cas2_TypeI-II-III	America	8	13
		Europe	3	
		Asia	2	
CAS-TypeIB	cas6_TypeI-III, cas8a1b_TypeIB, cas7b_TypeIB, cas5b_TypeIB, cas3_TypeI, cas4_TypeI-II, cas1_TypeIB, cas2_TypeI-II-III	America	9	21
		Europe	6	
		Asia	6	
CAS-type IIA	csn2_TypeIIA, cas2_TypeI-II-III, cas1_TypeII, cas9_TypeII	America	11	22
		Europe	7	
		Asia	4	
CAS-type IIIA	csm2_TypeIIIA	America	2	2
		Europe	0	
		Asia	0	
Total				67

to the different environments in which the strains reside. Hence, in order to adapt to these unique environments, the strains have evolved genes that are specific to these environments to counteract environmental pressures (Liao et al., 2023). This is also very intriguing, as it allows for the exploration of the differences among *L. monocytogenes* strains in different regions, analyzing the unique characteristics of strains in different areas, thereby studying the evolutionary patterns of strains in that region, and subsequently devising targeted prevention and control measures for that region. As the number of genomes increases, the pan-genome size continues to rise while the core genome decreases and tends to plateau. This indicates that the studied *L. monocytogenes* possesses an open pan-genome, which provides a genetic basis for the adaptation of *L. monocytogenes* to different environments. The potential target genes are exclusively present in *L. monocytogenes* strains in different regions, while they are absent in non-target strains. This indicates that the potential target genes play a crucial role in the pathogenicity of *L. monocytogenes* strains in different regions. These genes are indispensable for the survival, virulence, and invasion of *L. monocytogenes*, making them essential for maintaining life activities and infection. Therefore, investigating potential target genes can facilitate the analysis of the biodiversity and evolutionary characteristics of *L. monocytogenes*, aiding in the selection of novel specific molecular detection and therapeutic target genes.

To investigate the biodiversity and evolutionary characteristics of *L. monocytogenes* strains in different regions, we conducted MLST typing analysis. The results revealed that *L. monocytogenes* strains from America and Europe were predominantly characterized by ST1 and CC1 types, whereas those from Asia were predominantly characterized by ST8 and CC8 types. Wang et al. (2012) identified the three most common *L. monocytogenes* types in China as ST8, ST9, and ST87. Amarasekara et al. (2024) identified a significant presence of ST1 and CC1 types among *L. monocytogenes* isolates from agricultural markets in the United States. Additionally, Toledo et al. (2018) found that the most prevalent type of *L. monocytogenes* in samples from Chile was ST1. Our results are consistent with the findings reported in the above-mentioned literature. Our findings indicate that the *L. monocytogenes* strains isolated from America, Europe, and Asia exhibit different types. The underlying reasons for this phenomenon could be attributed to variations in the transmission routes and environmental conditions of *L. monocytogenes*, as well as genetic variability among strains. Based on our analysis, although *L. monocytogenes* strains originate from diverse geographical regions, they exhibit relatively similar phylogenetic relationships in the constructed phylogenetic tree. This indicates that while the *L. monocytogenes* strains are present in different environments, they exhibit a certain degree of genetic similarity in terms of bacterial variability. Under various environmental pressures, *L. monocytogenes* gradually evolves into life forms adapted to these specific environments, undergoing extensive genetic variations. Bacterial genetic variations result in distinct predominant types of *L. monocytogenes* strains in America, Europe, and Asia. Interestingly, the predominant sequence type of *L. monocytogenes* in both America and Europe is ST1, which may be attributable to the relatively close

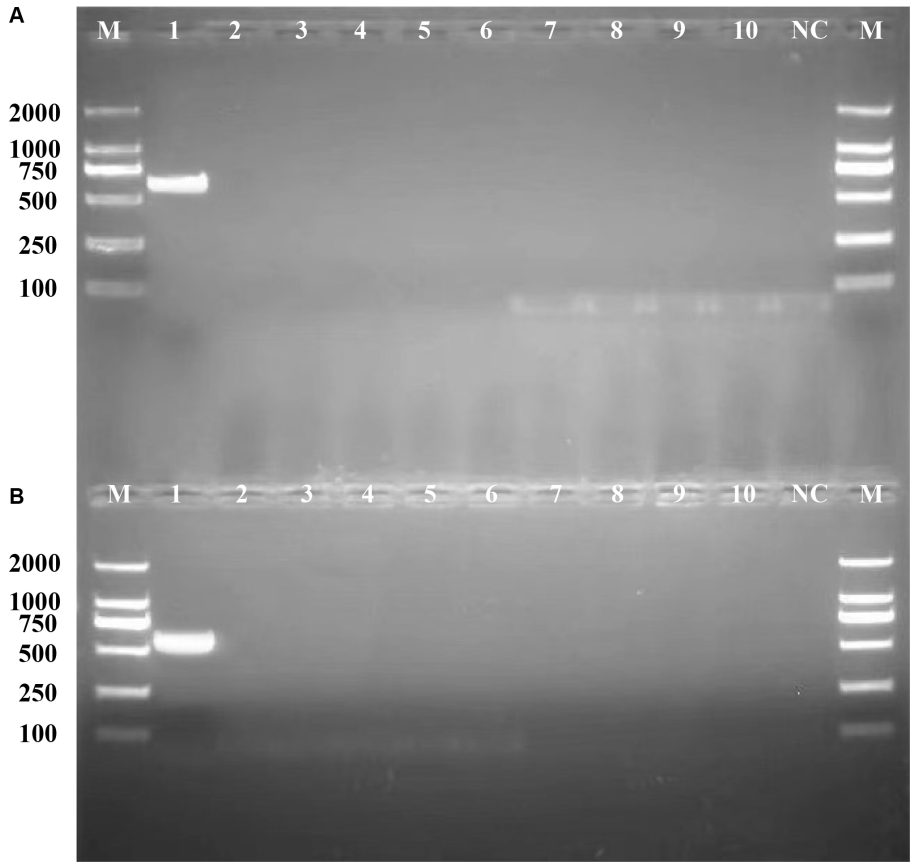


FIGURE 7
Validation of the specificity of the primers for the *bglF_1* and *davD* genes in *L. monocytogenes*. **(A)** The PCR results of the primer system targeting the *bglF_1* gene. **(B)** The PCR results of the primer system targeting the *davD* gene. Lane M: DL DNA 2000 marker, lane NC: negative control, and lanes 1–10: represent 10 different strains, including lane 1: *Listeria monocytogenes*, lane 2: *Listeria innocua*, lane 3: *Listeria ivanovii*, lane 4: *Listeria welshimeri*, lane 5: *Escherichia coli*, lane 6: *Salmonella*, lane 7: *Klebsiella Pneumoniae*, lane 8: *Acinetobacter baumannii*, lane 9: *Pseudomonas aeruginosa*, lane 10: *Pneumocystosis jirovecii*.

geographical proximity of these regions resulting in fewer environmental disparities. Additionally, trade between these regions may contribute to the mutual dissemination of *L. monocytogenes* strains. This discovery provides valuable insights into the reasons for the differences in the predominant ST and CC types of *L. monocytogenes* strains in America, Europe, and Asia.

It is well known that within the genus *Listeria*, only *L. monocytogenes* and *L. ivanovii* are considered pathogenic, with *L. monocytogenes* exhibiting higher pathogenicity. Moreover, the high pathogenicity of *L. monocytogenes* typically relies on a plethora of virulence genes as its foundation. Our analysis findings align with this observation, as strains of *L. monocytogenes* in different regions harbor a significant abundance of virulence genes. During invasion of the host by *L. monocytogenes*, the bacterium first utilizes the *inlA* and *inlB* genes to bind with the E-Cadherin and Met receptors of the host's eukaryotic cell membrane, respectively, thereby inducing bacterial uptake through receptor-mediated endocytosis. After internalization, the bacterium is encapsulated within a vacuole, and releases the *hly*, *plcA*, and *plcB* genes to mediate vacuole escape. Subsequently, the *actA* gene is utilized to induce actin polymerization and generate sufficient force for the bacterium to spread from one cell to another. During the invasion process of *L. monocytogenes*, LIPI-1 (*prfA*, *plcA*, *hly*, *mpl*, *actA*, and *plcB*) and LIPI-2 (*inlA*, *inlB*, *inlC*, *inlE*, *inlF*, *inlG*, *inlH*, *inlJ*, and *inlK*) play pivotal roles (Mejía et al., 2023). Interestingly, in this study, the genes *prfA*, *plcB*, *inlK*, and *inlB* were found to be present in 100% of the selected *L. monocytogenes* strains, whereas *hly*, *actA*, and *inlA* were not always present at 100%, but their presence probability exceeded 99%. This phenomenon could be attributed to prediction errors in the database or possibly due to genetic variations occurring in individual strains under specific environmental conditions (Li et al., 2021d). These virulence genes are essential for infecting and invading hosts, highlighting the high pathogenicity of *L. monocytogenes*. Furthermore, the potential target genes we screened also include these virulence genes. The potential target genes play crucial roles in the fundamental life activities and infective invasion of *L. monocytogenes*. Selecting these virulence genes as molecular detection and therapeutic targets may be a viable option, however, it may lack novelty, as previous studies have validated genes such as *inlA*, *inlB*, and *hly* as targets for the detection and treatment of *L. monocytogenes*. Therefore, although these virulence genes were also selected as potential target genes in this study, they were not directly chosen as targets for detection and treatment. This study employed hub gene screening methods to further select hub genes from numerous potential target genes. As is well known, hub genes, also known as key genes, refer to

genes that play a crucial role in a particular disease or biological process (Li et al., 2022). Therefore, we selected highly scoring hub genes from potential target genes as novel target genes.

In this study, *L. monocytogenes* strains demonstrated relatively high resistance to phosphonic antibiotics, glycopeptide antibiotics, lincosamide antibiotics, and peptide antibiotics. Therefore, it is recommended to avoid selecting these four classes of antibiotics when undergoing treatment. Other types of antibiotics may yield better therapeutic effects, such as ampicillin, gentamicin, and penicillin. Analyzing MGEs can provide insights into the evolution of bacterial genomes. In this study, despite the different geographical locations of the *L. monocytogenes* strains, they exhibited certain similarities at the MGEs level. This suggests that although *L. monocytogenes* is exposed to diverse external environments, there still exists a degree of similarity in terms of bacterial genome evolution.

By performing PPI network analysis and conducting GO and KEGG enrichment analyses on potential target genes, we aimed to understand the role of these genes in *L. monocytogenes* strains in different regions (Adnan et al., 2022). Functional annotation results revealed that the potential target genes encompassed a significant number of transport and metabolism genes, as well as virulence-associated genes, which play crucial roles in the fundamental life activities and pathogenicity of *L. monocytogenes*. However, some genes still lack clear functional information, necessitating further investigation in future studies. Hub genes, which are the most crucial genes in PPI networks, were selected to mining novel target genes (Li et al., 2022). Ten highly connected hub genes (*bglF_1, davD, menE_1, tilS, dapX, iolC, gshAB, cysG, trpA, hisC*) were identified from the pool of potential target genes. These ten hub genes play crucial roles in the fundamental life activities and infective invasion of *L. monocytogenes*. Among them, *bglF_1* and *davD* genes scored the highest and showed closer connections with other proteins, indicating their potential to serve as specific molecular detection and therapeutic targets for *L. monocytogenes* strains. The inhibitors or antagonists of these genes hold promise as novel therapeutic agents. The PCR results demonstrated the excellent specificity of the *bglF_1* and *davD* genes for *L. monocytogenes*. Therefore, the *bglF_1* and *davD* genes hold promise as specific molecular detection and therapeutic targets for *L. monocytogenes* strains from different regions.

Conclusion

In summary, we employed comparative genomic analysis to investigate the biodiversity and evolutionary characteristics of *L. monocytogenes* strains from different regions. Although *L. monocytogenes* strains originate from different regions, they exhibit a high degree of similarity in bacterial genome evolution, harboring numerous potential target genes that sustain the essential life activities and infection invasion of *L. monocytogenes*. Through further exploration of potential target genes and validation of PCR results, the *bglF_1* and *davD* genes emerged as promising candidates for specific molecular detection and therapeutic targets in *L. monocytogenes* strains. This study provides significant reference value for the specific molecular detection and therapeutic targets of *L. monocytogenes* strains.

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

BZ: Data curation, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing. HR: Supervision, Writing – review & editing. XW: Supervision, Writing – review & editing. CH: Validation, Writing – original draft. YJ: Validation, Writing – original draft. XH: Validation, Writing – original draft. RS: Validation, Writing – original draft. CL: Validation, Writing – original draft. YW: Validation, Writing – original draft. YL: Supervision, Writing – review & editing. SL: Supervision, Writing – review & editing. ZL: Supervision, Writing – review & editing. PH: Supervision, Writing – review & editing.

Funding

The work was supported by the National Key Research and Development Program of China (2023YFD1801000).

Acknowledgments

The authors would like to thank all the reviewers who participated in the review, as well as other members in the team of PH.

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.2024.1424868/full#supplementary-material

References

Adnan, M., Siddiqui, A. J., Noumi, E., Hannachi, S., Ashraf, S. A., Awadelkareem, A. M., et al. (2022). Integrating network pharmacology approaches to decipher the multi-target pharmacological mechanism of microbial biosurfactants as novel green antimicrobials against Listeriosis. *Antibiotics (Basel)* 12:5. doi: 10.3390/antibiotics12010005

Amarasekara, N. R., Swamy, A. S., Paudel, S. K., Jiang, W., Li, K., Shen, C., et al. (2024). Hypervirulent clonal complex (CC) of *Listeria monocytogenes* in fresh produce from urban communities. *Front. Microbiol.* 15:1307610. doi: 10.3389/fmicb.2024.1307610

Bosi, E., Fondi, M., Orlandini, V., Perrin, E., Maida, I., Pascale, D., et al. (2017). The pangenome of (Antarctic) Pseudoalteromonas bacteria: evolutionary and functional insights. *BMC Genomics* 18:93. doi: 10.1186/s12864-016-3382-y

Cain, R. J., Scortti, M., Monzó, H. J., and Vázquez, J. A. (2023). *Listeria* InlB expedites vacuole escape and intracellular proliferation by promoting Rab7 recruitment via Vps34. *MBio* 14:e0322122. doi: 10.1128/mbio.03221-22

Castro, H., Douillard, F. P., Korkeala, H., and Lindström, M. (2021). Mobile elements harboring heavy metal and bacitracin resistance genes are common among *Listeria monocytogenes* strains persisting on dairy farms. *mSphere* 6:e0038321. doi: 10.1128/mSphere.00383-21

Chiaverini, A., Guidi, F., Torresi, M., Acciari, V. A., Centorotola, G., Cornacchia, A., et al. (2021). Phylogenetic analysis and genome-wide association study applied to an Italian *Listeria monocytogenes* outbreak. *Front. Microbiol.* 12:750065. doi: 10.3389/fmicb.2021.750065

Deng, X., Phillippy, A. M., Li, Z., Salzberg, S. L., and Zhang, W. (2010). Probing the pan-genome of *Listeria monocytogenes*: new insights into intraspecific niche expansion and genomic diversification. *BMC Genomics* 11:500. doi: 10.1186/1471-2164-11-500

Disson, O., Moura, A., and Lecuit, M. (2021). Making sense of the biodiversity and virulence of *Listeria monocytogenes*. *Trends Microbiol.* 29, 811–822. doi: 10.1016/j.tim.2021.01.008

Fox, E. M., Allnutt, T., Bradbury, M. I., Fanning, S., and Chandry, P. S. (2016). Comparative genomics of the *Listeria monocytogenes* ST204 subgroup. *Front. Microbiol.* 7:2057. doi: 10.3389/fmicb.2016.02057

Gao, Z., Zhong, W., Liu, T., Zhao, T., and Guo, J. (2021). Global proteomic analysis of *Listeria monocytogenes* response to linalool. *Food Secur.* 10:2449. doi: 10.3390/foods10102449

He, P., Wang, H., Yan, Y., Zhu, G., and Chen, Z. (2022). Development and application of a multiplex fluorescent PCR for *Shigella* detection and species identification. *J. Fluoresc.* 32, 707–713. doi: 10.1007/s10895-021-02876-0

Henri, C., Félix, B., Guillier, L., Leekitcharoenphon, P., Michelon, D., Mariet, J. F., et al. (2016). Population genetic structure of *Listeria monocytogenes* strains as determined by pulsed-field gel electrophoresis and multilocus sequence typing. *Appl. Environ. Microbiol.* 82, 5720–5728. doi: 10.1128/AEM.00583-16

Jordan, K., and McAuliffe, O. (2018). *Listeria monocytogenes* in foods. *Adv. Food Nutr. Res.* 86, 181–213. doi: 10.1016/bs.afnr.2018.02.006

Li, J. X., Huang, Y. Y., Huang, Z. M., Cao, X. J., Xie, L. M., and Guo, X. G. (2022). Screening of potential hub genes involved in cutaneous Leishmaniasis infection via bioinformatics analysis. *Acta Trop.* 236:106645. doi: 10.1016/j.actatropica.2022.106645

Li, C., Li, J., and Lu, P. (2021a). Identification of key genes involved in Brg1 mutation-induced cataract using bioinformatics analyses with publicly available microarray data. *Acta Biochim. Pol.* 68, 733–737. doi: 10.18388/abp.2020_5632

Li, F., Ye, Q., Chen, M., Shang, Y., Zhang, J., Ding, Y., et al. (2021b). Real-time PCR identification of *Listeria monocytogenes* serotype 4c using primers for novel target genes obtained by comparative genomic analysis. *LWT* 138:110774. doi: 10.1016/j.lwt.2020.110774

Li, F., Ye, Q., Chen, M., Zhang, J., Xue, L., Wang, J., et al. (2021c). Multiplex PCR for the identification of pathogenic *Listeria* in Flammulina velutipes plant based on novel specific targets revealed by Pan-genome analysis. *Front. Microbiol.* 11:634255. doi: 10.3389/fmicb.2020.634255

Li, F., Ye, Q., Chen, M., Zhou, B., Xiang, X., Wang, C., et al. (2021d). Mining of novel target genes through pan-genome analysis for multiplex PCR differentiation of the major *Listeria monocytogenes* serotypes. *Int. J. Food Microbiol.* 339:109026. doi: 10.1016/j.ijfoodmicro.2020.109026

Liao, J., Guo, X., Li, S., Anupoju, S. M. B., Cheng, R. A., Weller, D. L., et al. (2023). Comparative genomics unveils extensive genomic variation between populations of *Listeria* species in natural and food-associated environments. *ISME Commun.* 3:85. doi: 10.1038/s43705-023-00293-x

Locatelli, A., Lewis, M. A., and Rothrock, M. J. Jr. (2017). The distribution of *Listeria* in pasture-raised broiler farm soils is potentially related to University of Vermont Medium Enrichment Bias toward *Listeria innocua* over *Listeria monocytogenes*. *Front. Vet. Sci.* 4:227. doi: 10.3389/fvets.2017.00227

Lomonaco, S., Nucera, D., and Filipello, V. (2015). The evolution and epidemiology of *Listeria monocytogenes* in Europe and the United States. *Infect. Genet. Evol.* 35, 172–183. doi: 10.1016/j.meegid.2015.08.008

Lourenco, A., Linke, K., Wagner, M., and Stessl, B. (2022). The saprophytic lifestyle of *Listeria monocytogenes* and entry into the food-processing environment. *Front. Microbiol.* 13:789801. doi: 10.3389/fmicb.2022.789801

Lu, Q., Zhu, X., Long, Q., Yi, X., Yang, A., Long, X., et al. (2022). Comparative genomics reveal the utilization ability of variable carbohydrates as key genetic features of *Listeria* pathogens in their pathogenic lifestyles. *Pathogens* 11:1430. doi: 10.3390/pathogens11121430

Mafuna, T., Matle, I., Magwedere, K., Pierneef, R. E., and Reva, O. N. (2021). Whole genome-based characterization of *Listeria monocytogenes* isolates recovered from the food chain in South Africa. *Front. Microbiol.* 12:669287. doi: 10.3389/fmicb.2021.669287

Mafuna, T., Matle, I., Magwedere, K., Pierneef, R. E., and Reva, O. N. (2022). Comparative genomics of *Listeria* species recovered from meat and food processing facilities. *Microbiol. Spectr.* 10:e0118922. doi: 10.1128/spectrum.01189-22

Matle, I., Pierneef, R., Mbatha, K. R., Magwedere, K., and Madoroba, E. (2019). Genomic diversity of common sequence types of *Listeria monocytogenes* isolated from ready-to-eat products of animal origin in South Africa. *Genes (Basel)* 10:1007. doi: 10.3390/genes10121007

Mejía, L., Espinosa-Mata, E., Freire, A. L., Zapata, S., and González-Candelas, F. (2023). *Listeria monocytogenes*, a silent foodborne pathogen in Ecuador. *Front. Microbiol.* 14:1278860. doi: 10.3389/fmicb.2023.1278860

Orsi, R. H., and Wiedmann, M. (2016). Characteristics and distribution of *Listeria* spp., including *Listeria* species newly described since 2009. *Appl. Microbiol. Biotechnol.* 100, 5273–5287. doi: 10.1007/s00253-016-7552-2

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421

Palma, F., Pasquali, F., Lucchi, A., De, C. A., and Manfreda, G. (2017). Whole genome sequencing for typing and characterisation of *Listeria monocytogenes* isolated in a rabbit meat processing plant. *Ital. J. Food Saf.* 6:6879. doi: 10.4081/ijfs.2017.6879

Pang, R., Xie, T., Wu, Q., Li, Y., Lei, T., Zhang, J., et al. (2019). Comparative genomic analysis reveals the potential risk of *Vibrio parahaemolyticus* isolated from ready-to-eat foods in China. *Front. Microbiol.* 10:186. doi: 10.3389/fmicb.2019.00186

Parsons, C., Brown, P., and Kathariou, S. (2021). Use of bacteriophage amended with CRISPR-Cas systems to combat antimicrobial resistance in the bacterial foodborne pathogen *Listeria monocytogenes*. *Antibiotics (Basel)* 10:308. doi: 10.3390/antibiotics10030308

Radoshevich, L., and Cossart, P. (2018). *Listeria monocytogenes*: towards a complete picture of its physiology and pathogenesis. *Nat. Rev. Microbiol.* 16, 32–46. doi: 10.1038/nrmicro.2017.126

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

Sheng, L., Edwards, K., Tsai, H. C., Hanrahan, I., and Zhu, M. J. (2017). Fate of *Listeria monocytogenes* on fresh apples under different storage temperatures. *Front. Microbiol.* 8:1396. doi: 10.3389/fmicb.2017.01396

Tan, M. F., Siow, C. C., Dutta, A., Mutha, N. V., Wee, W. Y., Heydari, H., et al. (2015). Development of *ListeriaBase* and comparative analysis of *Listeria monocytogenes*. *BMC Genomics* 16:755. doi: 10.1186/s12864-015-1959-5

Toledo, V., Bakker, H. C., Hormazábal, J. C., González-Rocha, G., Bello-Toledo, H., Toro, M., et al. (2018). Genomic diversity of *Listeria monocytogenes* isolated from clinical and non-clinical samples in Chile. *Genes (Basel)* 9:396. doi: 10.3390/genes9080396

Wang, Y., Ji, Q., Li, S., and Liu, M. (2021). Prevalence and genetic diversity of *Listeria monocytogenes* isolated from retail pork in Wuhan, China. *Front. Microbiol.* 12:620482. doi: 10.3389/fmicb.2021.620482

Wang, C., Ye, Q., Ding, Y., Zhang, J., Gu, Q., Pang, R., et al. (2022). Detection of *Pseudomonas aeruginosa* serogroup G using real-time PCR for novel target genes identified through comparative genomics. *Front. Microbiol.* 13:928154. doi: 10.3389/fmicb.2022.928154

Wang, Y., Zhao, A., Zhu, R., Lan, R., Jin, D., Cui, Z., et al. (2012). Genetic diversity and molecular typing of *Listeria monocytogenes* in China. *BMC Microbiol.* 12:119. doi: 10.1186/1471-2180-12-119

Zhang, Y., Zheng, Y., Fu, Y., and Wang, C. (2019). Identification of biomarkers, pathways and potential therapeutic agents for white adipocyte insulin resistance using bioinformatics analysis. *Adipocytes* 8, 318–329. doi: 10.1080/21623945.2019.1649578

Zhang, D. F., Zhi, X. Y., Zhang, J., Paoli, G. C., Cui, Y., Shi, C., et al. (2017). Preliminary comparative genomics revealed pathogenic potential and international spread of Staphylococcus argenteus. *BMC Genomics* 18:808. doi: 10.1186/s12864-017-4149-9

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 09 July 2024
DOI 10.3389/fmicb.2024.1428766

OPEN ACCESS

EDITED BY
Xuanyu Tao,
University of Oklahoma, United States

REVIEWED BY
Daniela Hozbor,
Institute of Biotechnology and Molecular Biology (IBBM), Argentina
Michael Payne,
University of New South Wales, Australia
Azadeh Safarchi,
Health and Biosecurity (CSIRO), Australia

*CORRESPONDENCE
Chao Yang
✉ cyang@siii.cas.cn
Chuanqing Wang
✉ chuanqing523@163.com

†These authors have contributed equally to this work

RECEIVED 07 May 2024
ACCEPTED 28 June 2024
PUBLISHED 09 July 2024

CITATION
Fu P, Li Y, Qin J, Xie L, Yang C and Wang C (2024) Molecular epidemiology and genomic features of *Bordetella parapertussis* in Shanghai, China, 2017–2022. *Front. Microbiol.* 15:1428766. doi: 10.3389/fmicb.2024.1428766

Molecular epidemiology and genomic features of *Bordetella parapertussis* in Shanghai, China, 2017–2022

Pan Fu[1,2†], Yijia Li[1†], Jie Qin[1], Li Xie[3], Chao Yang[3*] and Chuanqing Wang[1,2*]

[1]Laboratory of Microbiology, Department of Clinical Laboratory, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai, China, [2]Nosocomial Infection Control Department, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai, China, [3]CAS Key Laboratory of Molecular Virology and Immunology, The Center for Microbes, Development and Health, Shanghai Institute of Immunity and Infection, Chinese Academy of Sciences, Shanghai, China

Background: Pertussis is a highly contagious respiratory illness mainly caused by *Bordetella pertussis* (BP). *Bordetella parapertussis* (BPP) can induce symptoms compatible with pertussis, but has been underdiagnosed and underreported. The current pertussis vaccines offer low protection against BPP. Herein, we aim to reveal the epidemiology and genomic evolution of BPP in Shanghai, China.

Methods: Children diagnosed with BPP infection from January 2017 to December 2022 in Shanghai, China were enrolled. We performed antimicrobial susceptibility testing (AST), multiple locus variable-number tandem repeat analysis (MLVA), and whole genome sequencing (WGS) analysis. A total of 260 international BPP genomes were chosen for comparison to investigate the genomic diversity and phylogenetic characteristics of Chinese strains within a global context.

Results: Sixty patients were diagnosed with BPP infection by culture, with the positive ratio of 3.5‰ (60/17337) for BPP in nasopharyngeal swap samples. The average age of patients was 4.5 ± 0.3 years. BPPs contained four MLVA types including MT6 (65.0%), MT4 (26.7%), untype-1 (6.7%) and MT5 (1.7%), and none of strains showed resistance to macrolides. All strains carried virulence genotype of *ptxP37/ptxA13/ptxB3/ptxC3/ptxD3/ptxE3/fim2-2/fim3-10*. MT4 and MT5 strains carried *prn54*, whereas MT6 and untype-1 BPPs expressed *prn101*. We identified two outbreaks after 2020 caused by MT4 and MT6 strains, each corresponding to distinct WGS-based phylogenetic lineages. The MT4-lineage is estimated to have originated around 1991 and has since spread globally, being introduced to China between 2005 and 2010. In contrast, the MT6-lineage was exclusively identified in China and is inferred to have originated around 2002.

Conclusion: We revealed the genomic diversity of BPPs circulating in Shanghai, China, and reported the outbreaks of MT6 and MT4 BPPs after 2020. This is the first report on the emergence and regional outbreak of MT6 BPPs in the world, indicating that continuous surveillance on BPPs are thus required.

KEYWORDS

***Bordetella parapertussis*, MT6, MT4, children, China**

Introduction

Whooping cough (pertussis) is a highly contagious respiratory disease of humans, which is mainly caused by *Bordetella pertussis* (BP; Feng et al., 2021; Gorgojo et al., 2023). Compared to BP, *B. parapertussis* (BPP) causes a milder whooping cough-like syndrome and is responsible for a smaller proportion (2%~20%) of pertussis (Toubiana et al., 2019). However, BPP infection has been poorly recognized in the world. Pertussis was previously thought to mainly occur in infants (Fu et al., 2019). However, more studies reveal that the prevalence and re-emergence of pertussis has been increasing in older children, adolescents and adults, making a great public threaten in the world (Moore et al., 2019; Zhang et al., 2022).

There are many differences between BBP and BP. For example, BPP lacks the production of the pertussis toxin (Ptx) due to a mutation in the promoter region of the genes encoding this toxin (Arico and Rappuoli, 1987). The World Health Organization (WHO) recommended two types of approach to diagnosis, including direct diagnosis [culture, real-time polymerase chain reaction (RT-PCR)] and indirect diagnosis (serology). These two species can be distinguished based on a number of biochemical characteristics: BPP grow faster and appear grayish; the oxidase test was positive in BP but negative in BPP, etc. Moreover, a series of targets including IS481, IS1001, IS1002, etc. were used to distinguish different *Bordetella* species. For example, IS1001 which presented in all BPPs but was absent in BPs, was widely used to identify BPP (Riffelmann et al., 2005).

BPP might vary from an unrecognized infection to a mild illness or typical pertussis presentation; it is increasingly recognized and reported to public health agencies (Liko et al., 2017). Pertussis vaccines are produced as combination vaccines with diphtheria and tetanus toxoids. In China, a routine immunization schedule of diphtheria, tetanus, whole-cell pertussis vaccine (DTwP) was implemented in the 1960s. Starting in 2005, both DTwP and diphtheria-tetanus-acellular pertussis vaccine (DTaP) were used in China, with DTaP gradually replacing DTwP by 2010 (Wu et al., 2023). Although DTap vaccine significantly reduced the incidence of pertussis, many studies have shown that pertussis vaccination is irrelevant to or just partially protect against BPP infection (Liko et al., 2017). The rodent model showed that aP vaccination, by priming the host response against BP clearance, confers an advantage to BPP by interfering with optimal immune clearance and resulting in increased lung colony-forming units (Long et al., 2010).

Until now, a series of studies on BP strains were reported in China, including Zhejiang province (Lin et al., 2022), Shanghai (Fu et al., 2023b), Shenzhen (Wu et al., 2021), and Beijing (Zhou et al., 2024). However, systematic studies or reports on BPP strains are very scarce in the world. National surveillance on BPP strains are largely lacking in China. In this study, we performed a continuous surveillance on *Bordertella* spp. based on culture, and collected a total of 60 BPP strains from January 2017 to December 2022 in Shanghai, China. We systematically analyzed the clinical and epidemiology features, the antimicrobial resistance (AMR) profiles, and the genomic evolution of those strains.

Materials and methods

Enrollment of pertussis cases

From January 2017 to December 2022, there were 740 children diagnosed as pertussis by bacterial culture in Shanghai, China. Their nasopharyngeal swab (NP) samples were collected for *Bordetella* spp. culture and antimicrobial resistance testing. Their basic information, clinical diagnosis, and X-ray imaging were collected based on the electronic medical records. The laboratory testing results were collected and analyzed in this study, including white blood cell counts (WBC, $\times 10^9$/L) and C-reactive protein (CRP, mg/L). All data collection and analysis were anonymous. This study was approved by the Ethics Committee of the Children's Hospital of Fudan University (no. 2022-66).

Culture and antimicrobial susceptibility testing of BPP strains

NP samples were delivered to clinical microbiology laboratory and immediately spread onto charcoal agar (OXOID, United Kingdom) plates supplemented with 10% defibrinated sheep blood and cephalexin (40 mg/L). The plates were incubated in a humidified incubator at 35°C for 3 to 5 days. Different *Bordetella* species were verified by Gram staining, biochemical tests, and Matrix assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS, Bruker, Germany).

The BPP isolates were suspended equivalent to a 0.5 McFarland standard and inoculated onto charcoal agar containing 10% sheep blood without cephalexin. The minimum inhibitory concentrations of four antimicrobial agents, including erythromycin, azithromycin, clarithromycin, and sulfamethoxazole/trimethoprim, were determined by the E-test after 72 h of incubation at 35°C. The standardized interpretation criteria are based on our previous report (Fu et al., 2019).

Whole genome sequencing and analysis

Genomic DNA of BPP strains were extracted using QIAamp DNA mini kit (QIAGEN) and whole-genome sequencing were performed on Illumina NovaSeq platform. Sequencing data were analyzed as previously described (Yang et al., 2022; Fu et al., 2023b). Briefly, species identification was performed using Kraken 2 based on sequencing data (Lu et al., 2022).[1] Genome assembly was performed using shovill pipeline. The genome characteristics of newly sequenced data were calculated using Quast v5.0.2 (Gurevich et al., 2013). The prevalence of insertion sequence (IS) IS1001 were detected by searching against assembled genome sequences using BLASTN. It was considered present if the BLASTN hit coverage and identity were at least 90%. Core-genome single-nucleotide-polymorphisms (SNPs) were identified using the Snippy pipeline,[2] with strain 12822 [accession

1 https://github.com/tseemann/shovill

2 https://github.com/tseemann/snippy

number: (NC_002928.3)] as the reference genome. Maximum-likelihood phylogenetic trees were constructed using RAxML-NG based on core-genome SNPs (Kozlov et al., 2019). The maximum-likelihood tree was rooted using the midpoint method. The dated phylogenetic trees were automatically rooted based on temporal signal using a root-to-tip linear regression with BactDating (Didelot et al., 2018). New sequencing data have been deposited in NCBI Sequence Read Archive (SRA) under accession number PRJNA1060880.

A total of 260 publicly available international genomes were downloaded from NCBI GenBank or SRA database, with accession numbers listed in the appendix (Supplementary Table S1). The international BPP strains included France (118), USA (91), Spain (29), Germany (3), Austria (2), United Kingdom (1), Australia (1), Japan (1), Iran (1), and unknow (13).

Multiple locus variable-number tandem repeat analysis, multilocus sequence typing, and *Bordetella* spp. virulence genotyping analysis

Genomic DNA of BPP isolates was prepared by a QIAamp DNA mini kit (QIAGEN). Multiple locus variable-number tandem repeat analysis (MLVA) was performed following the procedures according to the report of Kamachi et al. (2019). Four loci (VNTR4, VNTR13, VNTR14, and VNTR15) were amplified by PCR. The number of repeats at each VNTR locus was calculated from the DNA fragment length. The assignment of an MLVA type (MT) was based on the combination of repeat counts for VNTR4, VNTR13, VNTR14, and VNTR15 according to previous reports (Kamachi et al., 2019).

Multilocus sequence typing (MLST) were analyzed by seven housekeeping genes (*adk*, *fumC*, *glyA*, *tyrB*, *icd*, *pepA*, and *pgm*). BPP genomes data were matched on the website.[3] The alleles at each of the seven loci defined the allelic profile or sequence type (ST).

Assembled BPP genome sequences were used for virulence genotyping by searching against BIGSdb-Pasteur genomic platform for *Bordetella*.20. The virulence-related genes included pertussis toxin (PTX) promoter (*ptxp*), five *ptx* genes (*ptxA, ptxB, ptxC, ptxD*, *ptxE*), pertactin (*prn*), filamentous hemagglutinin B (*fhaB*), and fimbrial proteins (*fim2, fim3*).

Statistical analysis

All statistical analyses were performed using the GraphPad Prism software version 8.0. The *t* test and Bonferroni correction were performed to compare the differences of clinical characteristics and laboratory testing results between two groups. A *p*-value of less than 0.05 was considered statistically significant.

3 https://bigsdb.pasteur.fr/bordetella/

Results

Distributions and detection of BPP cases from 2017 to 2022

As shown in Figure 1A, there were 740 children diagnosed as pertussis by bacterial culture, and 91.4% (676 patients), 8.1% (60 patients) and 0.5% (4 patients) of the pertussis were caused by BP, BPP, and *Bordetella bronchitis* (BB), respectively.

BP strains were continuously isolated from 2017 to 2019 (ranging from 4 strains to 97 strains per quarter), but only one BP strain was collected in 2020. Notably, BP infection was re-emerged from 2021 to 2022, and there were two peaks of BP infection at 4th quarter of 2021 (40 strains) and 1st quarter of 2022 (96 strains). Compared to BP cases, BPP infection was quite scarce before 2021. Only four BPP strains were identified at 3rd quarter of 2017. Concurrent with the BP re-emergence time, BPP infection had outbreaks at 4th quarter of 2021 (25 strains) and 1st quarter of 2022 (29 strains; Figure 1B).

Among 17,337 patients who received nasopharyngeal swap samples culture, there were 60 children diagnosed with BPP infection, with the positive ratio of 3.5‰ (60/17337). The average ages were (4.5 ± 0.3) years old. The infants (≤1 year), toddler (>1~3 years), preschool (>3~5 years), school age (>5~8 years) and adolescents (>8~11 years) accounted for 10% (6), 16.7% (10), 41.7% (25), 30% (18), and 1.7% (1), respectively (Figure 1C). Most of patients came from the internal medicine clinic (80%, 48; Figure 1D).

Clinical and laboratory characteristics of BPP cases

As shown in Table 1, most of the patients (76.7%, 46 patients) presented paroxysmal cough and phlegm, and the average cough period was (27.0 ± 4.6) days. Moreover, there were 56.7% (34 patients), 30% (18 patients), 21.7% (13 patients), 13.3% (8 patients) and 6.7% (4 patients) of BPP cases presented rhinorrhea, fever, vomiting, wheezing and spasmodic cough, respectively. We further compared the difference of younger children (1.7 ± 0.3 years old, 2 months to 3 years) and older children (5.6 ± 0.2 years old, >3 years to 11 years). It is noted that younger children presented more severe clinical symptoms than older children, including longer cough periods [(35.3 ± 9.6) days vs. (20.0 ± 3.2) days, $p = 0.03$), vomit (56.3% vs.9.1%, $p < 0.01$), and wheezing (31.3% vs.6.8%, $p = 0.04$].

There were 50% (30 patients) and 40% (24 patients) of BPP cases treated by macrolides and cephalosporins, respectively. However, no BPP strains isolated from the patients showed resistance to macrolides, and all strains presented sensitive to sulfamethoxazole/trimethoprim. The chest X-ray of BPP cases presented either bronchitis (45.0%, 27), bronchopneumonia (13.3%, 8) or pneumonia (5.0%, 3), and there was no difference between younger children and older children. WBC was slightly increased to $(11.4 \pm 0.4) \times 10^9$/L, and 23.3% (14) of patients presented abnormal CRP (>8 mg/L; Table 1).

Prevalence of different BPP types from 2017 to 2022

All 60 BPP strains belonged to one MLST type 19 (ST19, 100%), whereas MLVA analysis further revealed four different BPP subtypes

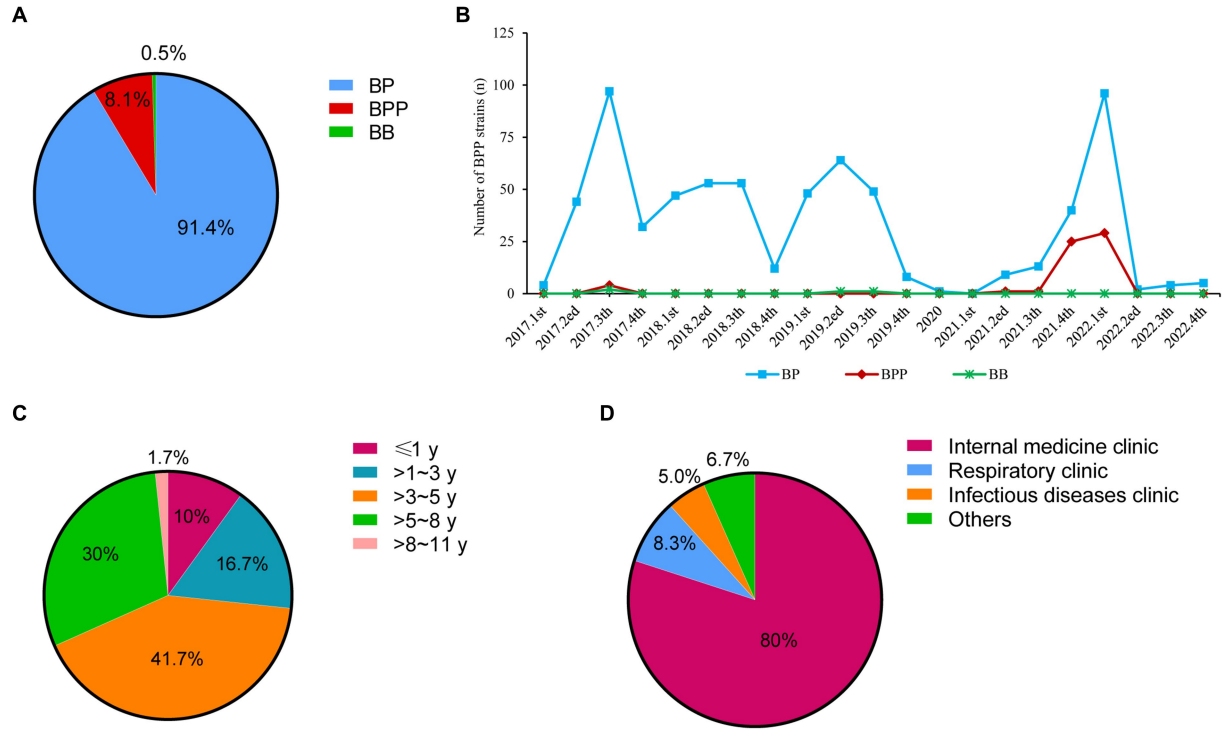


FIGURE 1
Distributions of *Bordetella* species and basic information of BPP infection cases from 2017 to 2022. **(A)** Distributions of different *Bordetella* species; **(B)** Detection ratios of three *Bordetella* species from 2017 to 2022; **(C)** Age distributions of BPP infection cases; **(D)** Department distributions of BPP infection cases. BP, *Bordetella pertussis*; BPP, *Bordetella parapertussis*; BB, *Bordetella bronchitis*.

in this study. MLVA type 6 (MT6) with the VNTR profiles of 4-5-13-5 was the major subtype (65.0%, 39 strains) of BPPs in this study, followed by MT4 (VNTRs: 3-7-18-4, 26.7%, 16 strains). Other subtype including MT5 (VNTRs: 3-7-21-4) and untype-1 (VNTRs: 4-5-10-4) were less frequently detected, with the ratios of 1.7% (1 strain) and 6.7% (4 strains), respectively (Figure 2A).

MT6 and MT4 were mostly isolated at 4th quarter, 2021 (56.4% and 12.5%) and 1st quarter, 2022 (41.0% and 81.3%), respectively (Figure 2B). There were MT6-BPP outbreaks at 4th quarter, 2021 (22 strains) and 1st quarter, 2022 (16 strains), and MT4-BPP outbreak at 1st quarter, 2022 (13), respectively (Figure 2C).

Genomic characteristics and evolution of Shanghai BPP strains

For the newly sequenced strains, the average GC content, number of contigs and size of assemblies were 68.17% (68.16–68.17%), 80 (70–92) and 4.72 Mb (4.72–4.73), with an average of 92-fold (87–141) depth for each genome. The marker sequence of BPP, IS1001, was found in all Chinese BPP strains.

We compared 60 Shanghai BPPs with 260 public genomes of global BPPs to reveal the phylogenetic relationship of those strains. After integrated the MLVA subtypes of 60 BPPs with WGS analysis, we defined two lineages: MT4-lineage and MT6-lineage, corresponding to MT4 and MT6 strains that caused disease outbreaks. As shown in Figure 3A, MT4-lineage isolates in Shanghai were closely related to those isolated from the France, United States, Austria, and Spain. MT6-lineage strains isolated after 2020 in Shanghai were quite different to other strains. Figure 3B showed that MT4-lineage was estimated to have originated at 1991 [95% confidence interval (CI): 1985~1996] probably from USA, and has spread to multiple regions, including Europe (France and Spain), Australia and China. This lineage was inferred to have been introduced to China around 2005 to 2010, much earlier than our first MT4 strain isolated in 2021. MT6-lineage was different to either the MT4-lineage strains or other international strains. Notably, this lineage was estimated to have originated in China at 2002 (95% CI: 1970~2013) and was only identified in Shanghai, China (Figure 3C).

We further identified 74 SNP sites that can be used to distinguish MT4-lineage and MT6-lineage, with alleles that are completely different in the two lineage strains. Among these SNPs, 64 located within gene regions (64 genes, each with one SNP), and an additional 10 SNPs located in intergenic regions (Supplementary Table S3). SNP sites can be used to distinguish MT4- and MT6-lineages, exhibiting alleles that are completely different in the two lineage strains. For example, the SNP at position 130,756 has two alleles: A and C. All MT4-lineage strains carried allele A, while all MT6-lineage strains carried allele C.

In this study, BPPs expressed a series of *Bordetella* virulence factors, with the profiles of *ptxP37/ptxA13/ptxB3/ptxC3/ptxD3/ptxE3/fim2-2/fim3-10*. All of MT4 and MT5 strains carried *prn54*, whereas MT6 and untype-1 BPs expressed *prn101* (Supplementary Table S2). The *fhaB* genes included four different alleles: *fhaB-45* (48.3%), *fhaB-3* (18.3%), *fhaB-22* (18.3%), and *fhaB-44* (15.1%), all of which were not related to any MT subtypes.

TABLE 1 Clinical and laboratory characteristics of BPP cases during 2021 to 2022.

Characteristics		Total	Younger children	Older children	*p*-value
		(*n* = 60)	(≤3 years old, *n* = 16)	(>3 years old, *n* = 44)	
Gender	Male	32 (53.3%)	11 (68.8%)	21 (47.7%)	
	Female	28 (46.7%)	5 (31.3%)	23 (52.3%)	0.15
Ages (Years)		4.5 ± 0.3	1.7 ± 0.3	5.6 ± 0.2	**<0.01**
Clinical symptoms	Paroxysmal cough	46 (76.7%)	12 (75.0%)	34 (77.3%)	0.87
	Cough days	27.0 ± 4.6	35.3 ± 9.6	20.0 ± 3.2	**0.03**
	Spasmodic cough	4 (6.7%)	2 (12.5%)	2 (4.5%)	0.58
	Vomit	13 (21.7%)	9 (56.3%)	4 (9.1%)	**<0.01**
	phlegm	46 (76.7%)	13 (81.3%)	33 (75.0%)	0.87
	Rhinorrhea	34 (56.7%)	10 (62.5%)	24 (54.5%)	0.58
	Wheezing	8 (13.3%)	5 (31.3%)	3 (6.8%)	**0.04**
	Fever	18 (30.0%)	5 (31.3%)	13 (29.5%)	0.9
Antimicrobial usage	Macrolides	30 (50%)	9 (56.3%)	19 (43.2%)	0.37
	Cephalosporins	24 (40%)	9 (56.3%)	15 (34.1%)	0.12
Antimicrobial resistant ratios	Macrolides	0 (0%)	0 (0%)	0 (0%)	>0.99
	Sulfamethoxazole/trimethoprim	0 (0%)	0 (0%)	0 (0%)	>0.99
X-ray	Bronchitis	27 (45.0%)	11 (68.8%)	16 (36.4%)	0.54
	Bronchopneumonia	8 (13.3%)	3 (18.8%)	5 (11.4%)	0.75
	Pneumonia	3 (5.0%)	0 (0%)	3 (6.8%)	0.56
Laboratory testing	WBC	11.4 ± 0.4	12.2 ± 0.9	11.2 ± 0.5	0.28
	Abnormal CRP	14 (23.3%)	4 (25.0%)	10 (22.7%)	0.97

White blood cell counts (WBC, 10^9/L), C-reactive protein (CRP, mg/L), Abnormal CRP was defined as CRP ≥ 8 mg/L. Statistical analyses were performed with GraphPad Prism 8.00. The t test and Bonferroni correction were performed to compare the differences of clinical characteristics and laboratory testing between two groups. A *p*-value of less than 0.05 was considered statistically significant, and was bolden in the table.

We further compared the clinical and laboratory characteristics of MT4-lineage and MT6-lineage BPPs (Supplementary Table S4). There were no big differences between those two BPP lineages in Shanghai ($p > 0.05$), including the clinical symptoms, the inflammatory factors, X-rays, antimicrobial treatment history and AMR profiles. However, MT6-lineage BPP cases presented longer cough periods of (31.2 ± 5.6) days, which was two or more times as long as MT4-lineage (14.2 ± 2.7 days, $p = 0.04$).

Discussion

BPP can cause whooping cough in human, but the epidemiology of respiratory illness caused by BPP strains has been poorly recognized in the world (Mastrantonio et al., 1998). Herein, we collected a total of 60 BPPs from 2017 to 2022 in Shanghai, China, and systematically analyzed the clinical and epidemiologic features and the genomic evolution of those strains. BPP accounted for 8.1% of whooping cough in this study, and caused more severe clinical symptoms in younger children (0~3 year). BPPs circulating in Shanghai contained four different MT types. It is noted that there was the outbreak of BPP infection after 2020, and the major MLVA types were MT4 and MT6. These two subtypes have evolved independently: MT4-lineage was highly homogeneous to international BPPs which were estimated to originate at USA and introduced to China around 2005 to 2010, whereas MT6-lineage was estimated to originate in China and was only identified in Shanghai, China.

After aP vaccine replaced wP vaccine by 2010 in China, there are two types of DTaP formulations licensed in China: one is the two-component DTaP containing Ptx and Fha, and another is the three-component DTaP containing Ptx, Fha and Prn. The current vaccine in Shanghai contains Ptx and Fha. However, BP vaccines fail to or partially induce protection against BPPs and the incidence of this species has been rising over the years (Long et al., 2010; Liko et al., 2017). BPP has been circulating worldwide and causes outbreaks despite high pertussis vaccine coverage of young children. Unlike BP infections which were primarily identified in infants before 2020 in Shanghai, China (Fu et al., 2019, 2023b), the average ages of BPP cases were (4.5 ± 0.3) years, and only 10% patients aged less than 1-year-old. The average cough period of BPP infection was much longer than our previous report of BP infection (27.0 ± 4.6 days vs. 15.5 ± 0.8 days; Fu et al., 2023a). Moreover, younger children (0–3 years) presented more severe clinical symptoms than those aged more than 3-years-old. Therefore, we must pay attention to BPP infections among children, especially the younger children.

Previously, erythromycin has been the mainstay of antibiotic therapy for pertussis as it decreases the transmission of infection and ameliorates symptoms particularly in younger and more severely affected infants (Mortensen and Rodgers, 2000). However, after the erythromycin-resistant BPs in China was firstly

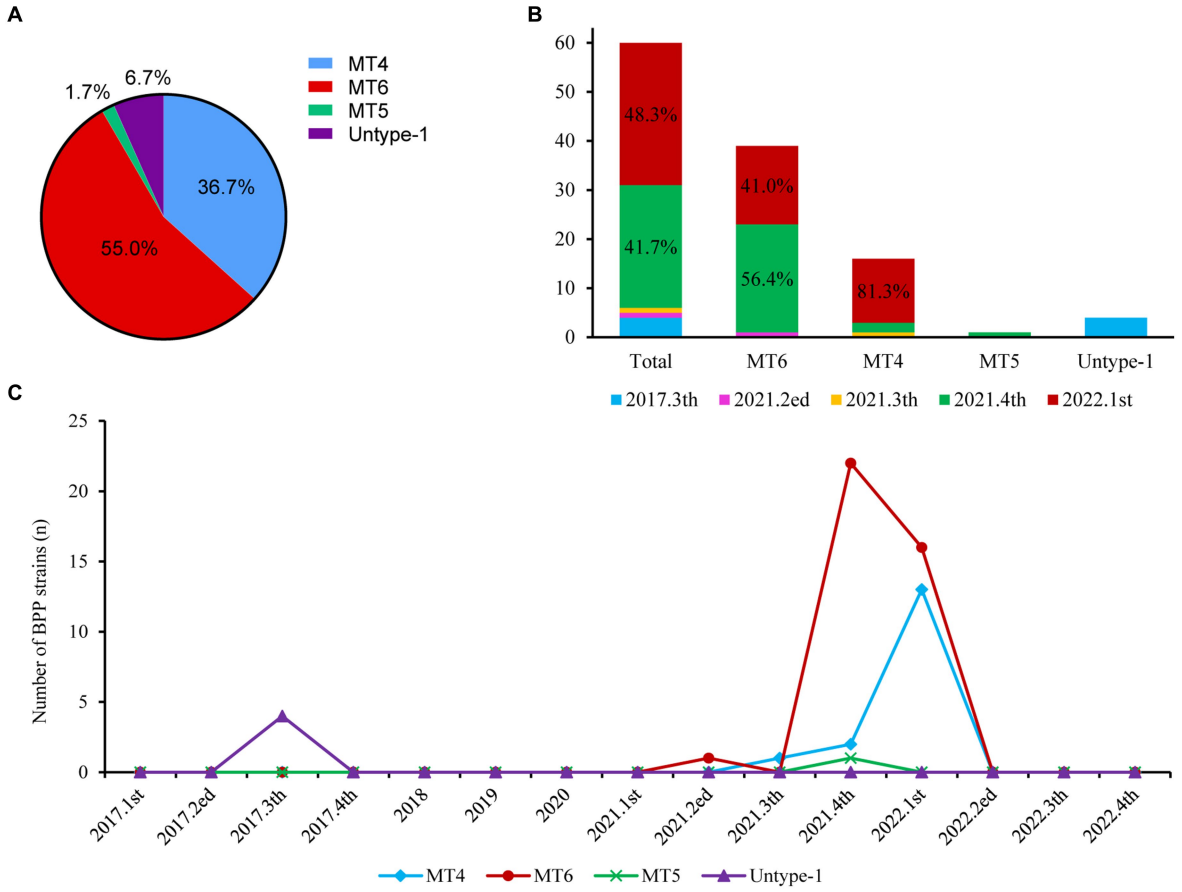


FIGURE 2
Prevalence of different BPP subtypes from 2017 to 2022. **(A)**. MT subtypes of all BPPs; **(B)**. Distributions of different MLVA types over time; **(C)**. Detection of different BPP subtypes per quarter from 2017 to 2022.

isolated in Shandong Province, China in 2011, more macrolides-resistant BP (MRBP) strains were reported in China, making macrolides less effective against BP infection (Zhang et al., 2013). Antimicrobials such as macrolides and sulfamethoxazole/trimethoprim recommended for BP infection have also been used for treating and preventing BPP infection (Mortensen and Rodgers, 2000). This study revealed a high proportion of macrolides treatment (50%) against BPP infection, but none of the BPPs were resistant to macrolides. This is quite different with high macrolides resistance of BPs in China (Feng et al., 2021; Wu et al., 2022; Fu et al., 2023b), indicating that macrolides are still effective against BPP infection. The resistance mechanism in BPs and BPPs are different: The 23S rRNA A2047G mutation is considered the major mechanism of resistance to macrolides in BP strains. However, the macrolides resistance mechanism in BPP is still unclear. Lately, Fong et al. (2022) reported that the macrolides resistance in BPP was probably related to the upregulation of an efflux pump mechanism, but it still needs further investigation. Therefore, we hypothesize that the macrolides resistance due to A2047G mutation was stable and has the potential spread capability than any other resistance mechanisms.

BPP can caused regional or national outbreaks. For example, Koepke et al. (2015) reported the concurrent outbreak of BPPs and BPs during 2011 to 2012 in Wisconsin, USA, and the BPPs accounted for nearly 6.0% of pertussis cases. In this study, we identified BPP outbreak after 2020 in Shanghai, China, which is greatly consistent with BP re-emergence time after 2020. We hypothesized that the potential re-emerging of BPPs and BPs in this study was related to the suppressed spread or circulation of respiratory pathogens during COVID-19. There were two major MLVA types including MT6 and MT4, the VNTR profiles of MT6 and MT4 were quite heterogenous, revealing the independent spread and genomic evolution of these two subtypes. Recently, Kamachi et al. (2019) constructed the MLVA analysis method of BPP strains, and identified one MT6 strain isolated at 2010 in Taiwan, China, and two MT4 strains isolated at 2010 in Taiwan, China and at 1988 in France, respectively.

BPP and BP share the same virulence factors including Prn, dermonecrotic toxin, Fha and adenylate cyclase (Mastrantonio et al., 1998). However, the Ptx as one of the major virulence factors is only expressed in BP since the *ptx* operon in BPP is dysfunctional (Arico and Rappuoli, 1987). In this study, many of the virulence factors characterized in BP strains are commonly expressed in BPPs. BPPs in Shanghai carried a series of virulence factors, including *ptxP37/ptxA13/ptxB3/ptxC3/ptxD3/ptxE3/fim2-2/fim3-10*. It is noted that although BPPs carry *ptx* genes in the genomes, they cannot express and secrete the pertussis toxin due to mutation in Ptx promoter region. *Prn* allele was diverse among different BPP subtypes: MT4 and MT5 strains all carried *prn54*, whereas MT6 and untype-1 BPs

10.3389/fmicb.2024.1428766

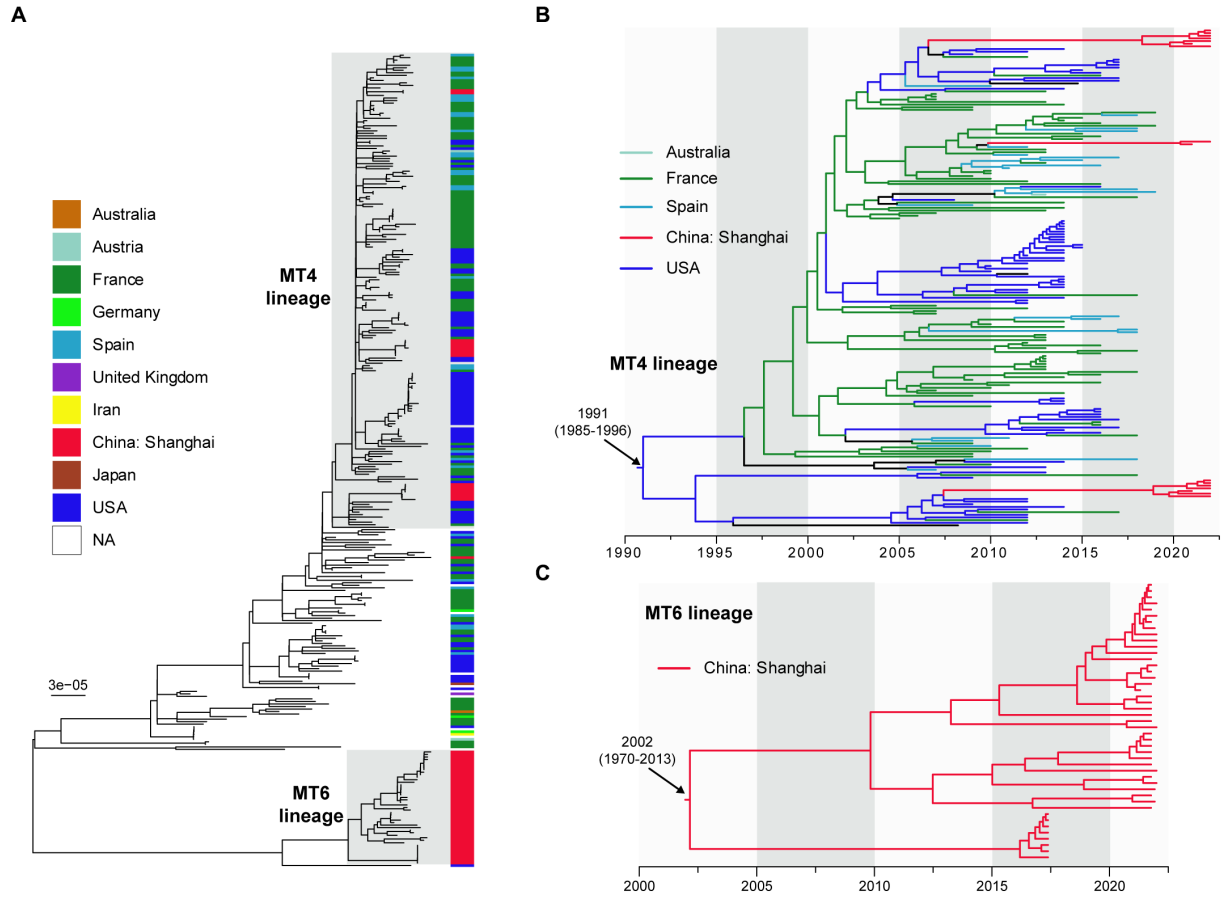


FIGURE 3
The dated maximum-likelihood phylogenetic tree of 60 Shanghai, China and 260 international BPP strains **(A)**. Different colors presented different geographic areas and red color indicates Shanghai strains. MT4-lineage **(B)** and MT6 lineage **(C)** are enlarged for visualization. NA, not applicable. The maximum-likelihood tree **(A)** was rooted using the midpoint method. The dated phylogenetic trees **(B,C)** were automatically rooted based on temporal signal.

expressed *prn101*, revealing the heterogeneity of virulence genes among different BPP subtypes. Prn-deficient BPs have widely been reported in countries using aP vaccines, such as the United States (85%), Australia (>80%), Sweden (69%), and Italy (55%; Byrne and Slack, 2006; Martin et al., 2015; Zomer et al., 2018; Weigand et al., 2019; Ma et al., 2021). Herein, all BPPs expressed *prn* gene without any mutation or disruption, and none of *prn*-negative BPP was identified. The aP vaccines containing Prn as an immunogen was thought to be the selection pressure for *prn*-negative BP strains (Ma et al., 2021). The current ACVs in Shanghai only contain Ptx and Fha, so we hypothesized that *prn* expression was not influenced by vaccine pressure because the current ACVs used in Shanghai contain no Prn antigen.

We collected 260 international BPPs for comparison, revealing the different genomic characteristics and molecular evolution of MT4-lineage and MT6-lineage. Firstly, 64 SNPs in gene regions and 10 SNPs in intergenic regions were identified among these two lineages. Secondly, the dated phylogenetic trees further revealed evolutionary differences between MT4- and MT6-lineage: MT4-lineage which originated at 1991 from USA was introduced to China around 2005~2010; MT6-lineage presented genomic heterogeneity to any other BPP strains, and was exclusively identified in Shanghai, China. We further compared the clinical and laboratory features of MT4-lineage and MT6-lineage BPPs. Although most of the clinical manifestations and inflammatory factors between these two lineages showed no significant differences, MT6-lineage cases presented the longer cough period (31.2 ± 5.6 days), which was about two or more times as long as MT4-linage cases. Therefore, it is thus important to keep continuous surveillance of MT-6 lineage BPP strains.

In summary, we revealed the genomic diversity and molecular evolution of different BPP subtypes circulating in China, and reported the emergence and outbreak of MT6 and MT4 BPPs after 2020 in Shanghai, China. To the best of our knowledge, it is the first report on the emergence and regional outbreak of MT6-lineage BPP in the world, highlighting that continuous surveillance and effective detection on BPP strains are thus required.

Data availability statement

The original contributions presented in the study are publicly available. 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 Children's Hospital of Fudan University (no. 2022-66). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

PF: Funding acquisition, Supervision, Writing – original draft, Writing – review & editing. YL: Data curation, Investigation, Methodology, Writing – review & editing. JQ: Investigation, Methodology, Writing – review & editing. LX: Methodology, Writing – review & editing. CY: Supervision, Writing – original draft, Writing – review & editing. CW: Supervision, Writing – original draft, Writing – review & editing.

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 National Natural Science Foundation of China (grant no. 82202567) and the Shanghai municipal three-year action plan for strengthening the construction of the public health system (2023–2025) GWVI-2.1.2.

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.2024.1428766/full#supplementary-material

References

Arico, B., and Rappuoli, R. (1987). Bordetella parapertussis and *bordetella bronchiseptica* contain transcriptionally silent pertussis toxin genes. *J. Bacteriol.* 169, 2847–2853. doi: 10.1128/jb.169.6.2847-2853.1987

Byrne, S., and Slack, A. T. (2006). Analysis of *bordetella pertussis* pertactin and pertussis toxin types from Queensland, Australia, 1999-2003. *BMC Infect. Dis.* 6:53. doi: 10.1186/1471-2334-6-53

Didelot, X., Croucher, N. J., Bentley, S. D., Harris, S. R., and Wilson, D. J. (2018). Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Res.* 46:e134. doi: 10.1093/nar/gky783

Feng, Y., Chiu, C. H., Heininger, U., Hozbor, D. F., Tan, T. Q., and von Konig, C. W. (2021). Emerging macrolide resistance in *bordetella pertussis* in mainland China: findings and warning from the global pertussis initiative. *Lancet Reg Health West Pac.* 8:100098. doi: 10.1016/j.lanwpc.2021.100098

Fong, W., Timms, V., Sim, E., Pey, K., Nguyen, T., and Sintchenko, V. (2022). Genomic and transcriptomic variation in bordetella spp. following induction of erythromycin resistance. *J. Antimicrob. Chemother.* 77, 3016–3025. doi: 10.1093/jac/dkac272

Fu, P., Wang, C., Tian, H., Kang, Z., and Zeng, M. (2019). *Bordetella pertussis* infection in infants and young children in shanghai, China, 2016-2017: clinical features, genotype variations of antigenic genes and macrolides resistance. *Pediatr. Infect. Dis. J.* 38, 370–376. doi: 10.1097/INF.0000000000002160

Fu, P., Zhou, J., Meng, J., Liu, Z., Nijiati, Y., He, L., et al. (2023a). Emergence and spread of mt28 ptxp3 allele macrolide-resistant *bordetella pertussis* from 2021 to 2022 in China. *Int. J. Infect. Dis.* 128, 205–211. doi: 10.1016/j.ijid.2023.01.005

Fu, P., Zhou, J., Yang, C., Nijiati, Y., Zhou, L., Yan, G., et al. (2023b). Molecular evolution and increasing macrolide resistance of *bordetella pertussis*, shanghai, China, 2016-2022. *Emerg. Infect. Dis.* 30, 29–38. doi: 10.3201/eid3001.221588

Gorgojo, J. P., Carrica, M., Baroli, C. M., Valdez, H. A., Alvarez, H. J., and Rodriguez, M. E. (2023). Adenylate cyclase toxin of *bordetella parapertussis* disrupts the epithelial barrier granting the bacterial access to the intracellular space of epithelial cells. *PLoS One* 18:e291331. doi: 10.1371/journal.pone.0291331

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

Kamachi, K., Otsuka, N., Fumimoto, R., Ozawa, K., Yao, S. M., Chiang, C. S., et al. (2019). A novel multilocus variable-number tandem repeat analysis for *bordetella parapertussis*. *J. Med. Microbiol.* 68, 1671–1676. doi: 10.1099/jmm.0.001095

Koepke, R., Bartholomew, M. L., Eickhoff, J. C., Ayele, R. A., Rodd, D., Kuennen, J., et al. (2015). Widespread *bordetella parapertussis* infections-Wisconsin, 2011-2012: clinical and epidemiologic features and antibiotic use for treatment and prevention. *Clin. Infect. Dis.* 61, 1421–1431. doi: 10.1093/cid/civ514

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). Raxml-ng: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35, 4453–4455. doi: 10.1093/bioinformatics/btz305

Liko, J., Robison, S. G., and Cieslak, P. R. (2017). Do pertussis vaccines protect against *bordetella parapertussis*? *Clin. Infect. Dis.* 64, 1795–1797. doi: 10.1093/cid/cix221

Lin, L. N., Zhou, J. S., Hua, C. Z., Bai, G. N., Mi, Y. M., and Zhou, M. M. (2022). Epidemiological and clinical characteristics of pertussis in children and their close contacts in households: a cross-sectional survey in Zhejiang province, China. *Front. Pediatr.* 10:976796. doi: 10.3389/fped.2022.976796

Long, G. H., Karanikas, A. T., Harvill, E. T., Read, A. F., and Hudson, P. J. (2010). Acellular pertussis vaccination facilitates *bordetella parapertussis* infection in a rodent model of bordetellosis. *Proc. Biol. Sci.* 277, 2017–2025. doi: 10.1098/rspb.2010.0010

Lu, J., Rincon, N., Wood, D. E., Breitwieser, F. P., Pockrandt, C., Langmead, B., et al. (2022). Metagenome analysis using the kraken software suite. *Nat. Protoc.* 17, 2815–2839. doi: 10.1038/s41596-022-00738-y

Ma, L., Caulfield, A., Dewan, K. K., and Harvill, E. T. (2021). Pertactin-deficient *bordetella pertussis*, vaccine-driven evolution, and reemergence of pertussis. *Emerg. Infect. Dis.* 27, 1561–1566. doi: 10.3201/eid2706.203850

Martin, S. W., Pawloski, L., Williams, M., Weening, K., DeBolt, C., Qin, X., et al. (2015). Pertactin-negative *bordetella pertussis* strains: evidence for a possible selective advantage. *Clin. Infect. Dis.* 60, 223–227. doi: 10.1093/cid/ciu788

Mastrantonio, P., Stefanelli, P., Giuliano, M., Herrera, R. Y., Ciofi, D. A. M., Anemona, A., et al. (1998). *Bordetella parapertussis* infection in children: epidemiology, clinical symptoms, and molecular characteristics of isolates. *J. Clin. Microbiol.* 36, 999–1002. doi: 10.1128/JCM.36.4.999-1002.1998

Moore, A., Harnden, A., Grant, C. C., Patel, S., and Irwin, R. S. (2019). Clinically diagnosing pertussis-associated cough in adults and children: chest guideline and expert panel report. *Chest* 155, 147–154. doi: 10.1016/j.chest.2018.09.027

Mortensen, J. E., and Rodgers, G. L. (2000). In vitro activity of gemifloxacin and other antimicrobial agents against isolates of bordetella pertussis and *bordetella parapertussis*. *J. Antimicrob. Chemother.* 45, 47–49. doi: 10.1093/jac/45.suppl_3.47

Riffelmann, M., Wirsing, V. K. C., Caro, V., and Guiso, N. (2005). Nucleic acid amplification tests for diagnosis of bordetella infections. *J. Clin. Microbiol.* 43, 4925–4929. doi: 10.1128/JCM.43.10.4925-4929.2005

Toubiana, J., Azarnoush, S., Bouchez, V., Landier, A., Guillot, S., Matczak, S., et al. (2019). *Bordetella parapertussis* bacteremia: clinical expression and bacterial genomics. *Open Forum Infect. Dis.* 6:z122. doi: 10.1093/ofid/ofz122

Weigand, M. R., Williams, M. M., Peng, Y., Kania, D., Pawloski, L. C., and Tondella, M. L. (2019). Genomic survey of *bordetella pertussis* diversity, United States, 2000-2013. *Emerg. Infect. Dis.* 25, 780–783. doi: 10.3201/eid2504.180812

Wu, X., Du, Q., Li, D., Yuan, L., Meng, Q., Fu, Z., et al. (2022). A cross-sectional study revealing the emergence of erythromycin-resistant *bordetella pertussis* carrying ptxp3 alleles in China. *Front. Microbiol.* 13:901617. doi: 10.3389/fmicb.2022.901617

Wu, S., Hu, Q., Yang, C., Zhou, H., Chen, H., Zhang, Y., et al. (2021). Molecular epidemiology of bordetella pertussis and analysis of vaccine antigen genes from clinical isolates from Shenzhen, China. *Ann. Clin. Microbiol. Antimicrob.* 20:53. doi: 10.1186/s12941-021-00458-3

Wu, D., Jing, R., Zheng, H., He, K., Li, Y., Yu, W., et al. (2023). Health and economic evaluation of vaccination against pertussis in China: a 40-year analysis. *Value Health* 26, 666–675. doi: 10.1016/j.jval.2022.10.011

Yang, C., Li, Y., Jiang, M., Wang, L., Jiang, Y., Hu, L., et al. (2022). Outbreak dynamics of foodborne pathogen *vibrio parahaemolyticus* over a seventeen year period implies hidden reservoirs. *Nat. Microbiol.* 7, 1221–1229. doi: 10.1038/s41564-022-01182-0

Zhang, J., Deng, J., and Yang, Y. (2022). Pertussis vaccination in chinese children with increasing reported pertussis cases. *Lancet Infect. Dis.* 22, 21–22. doi: 10.1016/S1473-3099(21)00752-0

Zhang, Q., Li, M., Wang, L., Xin, T., and He, Q. (2013). High-resolution melting analysis for the detection of two erythromycin-resistant *bordetella pertussis* strains carried by healthy schoolchildren in China. *Clin. Microbiol. Infect.* 19, E260–E262. doi: 10.1111/1469-0691.12161

Zhou, G., Li, Y., Wang, H., Wang, Y., Gao, Y., Xu, J., et al. (2024). Emergence of erythromycin-resistant and pertactin- and filamentous hemagglutinin-deficient *bordetella pertussis* strains—Beijing, China, 2022-2023. *China CDC Wkly.* 6, 437–441. doi: 10.46234/ccdcw2024.085

Zomer, A., Otsuka, N., Hiramatsu, Y., Kamachi, K., Nishimura, N., Ozaki, T., et al. (2018). *Bordetella pertussis* population dynamics and phylogeny in Japan after adoption of acellular pertussis vaccines. *Microb Genom.* 4:180. doi: 10.1099/mgen.0.000180

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 10 July 2024
DOI 10.3389/fmicb.2024.1410651

OPEN ACCESS

EDITED BY
Renmao "Tim" Tian,
Illinois Institute of Technology, United States

REVIEWED BY
Adeline Pastuszka,
Université de Tours, France
Leandro Simões,
Federal University of Rio de Janeiro, Brazil
Chienchung Lee,
Linkou Chang Gung Memorial Hospital, Taiwan

*CORRESPONDENCE
Uzma Basit Khan
✉ uk1@sanger.ac.uk

RECEIVED 01 April 2024
ACCEPTED 19 June 2024
PUBLISHED 10 July 2024

CITATION
Khan UB, Dyster V, Chaguza C, van Sorge NM, van de Beek D, Man WK, Bentley SD, Bijlsma MW and Jamrozy D (2024) Genetic markers associated with host status and clonal expansion of Group B Streptococcus in the Netherlands. *Front. Microbiol.* 15:1410651. doi: 10.3389/fmicb.2024.1410651

Genetic markers associated with host status and clonal expansion of Group B Streptococcus in the Netherlands

Uzma Basit Khan[1]*, Victoria Dyster[1], Chrispin Chaguza[2], Nina M. van Sorge[3,4], Diederik van de Beek[5], Wing Kit Man[5], Stephen D. Bentley[1], Merijn W. Bijlsma[5,6], Dorota Jamrozy[1] and on behalf of the NOGBS study group

[1]Parasites and Microbes Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom, [2]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, Yale University, New Haven, CT, United States, [3]Department of Medical Microbiology and Infection Prevention, Amsterdam Infection and Immunity, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands, [4]Netherlands Reference Laboratory for Bacterial Meningitis, Amsterdam UMC Location AMC, Amsterdam, Netherlands, [5]Department of Neurology, Amsterdam Neuroscience, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands, [6]Department of Paediatrics, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands

Objectives: Certain Group B Streptococcus (GBS) genotypes are associated with invasive disease in neonates. We conducted a comparative genomic analysis of GBS isolates from neonatal disease and maternal carriage in the Netherlands to determine distribution of genetic markers between the two host groups.

Methods: Whole genome sequencing was used to characterise 685 neonatal invasive isolates (2006–2021) and 733 maternal carriage isolates (2017–2021) collected in the Netherlands.

Results: Clonal complex (CC) 17 and serotype III were significantly more common in disease while carriage isolates were associated with serotypes II, IV, V as well as CC1. Previously reported CC17-A1 sub-lineage was dominant among disease isolates and significantly less common in carriage. The phiStag1 phage, previously associated with expansion of invasive CC17 isolates in the Netherlands, was more common among disease isolates compared to carriage isolates overall, however it was equally distributed between CC17 isolates from carriage and disease. Prevalence of antimicrobial resistance genes was overall lower in disease compared to carriage isolates, but increased significantly over time, mediated by rise in prevalence of a multidrug resistance element ICESag37 among disease isolates.

Conclusion: There is a stable association between certain GBS genotypes and invasive disease, which suggests opportunities for developing more precise disease prevention strategies based on GBS targeted screening.

In contrast, GBS mobile genetic elements appear less likely to be correlated with carriage or disease, and instead are associated with clonal expansion events across the GBS population.

KEYWORDS

Group B Streptococcus, neonatal invasive disease, maternal carriage, serotypes, clonal complexes, mobile genetic elements

Introduction

Streptococcus agalactiae (Group B Streptococcus, GBS) is a common coloniser of the vaginal and gastrointestinal tracts of healthy adults. Carriage of GBS during pregnancy represents a risk factor for the development of invasive disease in the newborn and GBS is a leading cause of invasive infection in neonates worldwide (Gonçalves et al., 2022). Beta-lactams represent the first choice for intrapartum antibiotic prophylaxis (IAP) during labour and treatment of GBS disease. While most GBS isolates remain susceptible to beta-lactams (Kobayashi et al., 2021), prevalence of resistance to second-line antibiotics such as erythromycin and clindamycin has been increasing (Slotved and Hoffmann, 2020; Kekic et al., 2021; Sabroske et al., 2023).

Group B Streptococcus isolates are often grouped based on their capsular polysaccharide (CPS), with 10 different serotypes described to date: Ia, Ib, and II–IX (Berti et al., 2014). GBS CPS is a major virulence factor of GBS and a number of GBS multivalent vaccines targeting CPS are currently under development (Absalon et al., 2022). GBS isolates are also characterised using multi-locus sequence typing (MLST), which has revealed that five GBS clonal complexes (CCs) are associated with colonisation and disease in humans: CC1, CC10, CC17, CC19, and CC23 (Björnsdóttir et al., 2016; Khan et al., 2022). Some GBS lineages are associated with specific CPS serotypes, for instance CC17 isolates express predominantly serotype III (Teatero et al., 2016). Associations between GBS molecular markers and different host groups have been observed, with CC17-serotype III dominant among neonatal GBS invasive disease (Teatero et al., 2016; Bianchi-Jassir et al., 2020; Jamrozy et al., 2020), while CC1 often associated with disease in the adult population (Flores et al., 2015).

We have previously reported that CC17 prevalence has increased among GBS isolates from neonatal disease in the Netherlands, which was associated with expansion of particular CC17 clonal groups and with acquisition of a novel phage phiStag1 (Jamrozy et al., 2020). It has been unclear whether the increasing prevalence of these CC17 clones occurred only among the disease-associated GBS isolates, or was reflective of a more broad expansion across the GBS population. To address this, we have used whole genome sequencing (WGS) to analyse and contrast population structures of GBS isolates from maternal carriage and neonatal disease, collected in the Netherlands. Furthermore, to better understand the genetic variability between isolates from the two at-risk populations, we compared the distribution of key GBS molecular markers such as serotype, CC, antimicrobial resistance (AMR) genes and the intra-lineage population structure within the major CCs.

Materials and methods

GBS isolates

The collection consisted of 685 neonatal (<90 days old) invasive GBS isolates collected between 2006 and 2021, and 733 maternal carriage GBS isolates collected between 2017 and 2021 in the Netherlands. Isolates from neonatal disease were derived from a nationwide surveillance of bacterial meningitis and infant bacteraemia conducted by the Netherlands Reference Laboratory for Bacterial Meningitis (NRLMB). Disease isolates collected between 2006 and 2016 were described previously (Jamrozy et al., 2020). The infections were classified as early onset disease (EOD) at age 0–6 days, and as late onset disease (LOD) at age 7–89 days. Maternal carriage isolates were collected from pregnant women in hospitals in Amsterdam, The Hague, Utrecht, Hengelo, and Arnhem, for the Netherlands observational study on GBS disease, bacterial virulence and protective serology (NOGBS). Isolates were cultured from the vagina (n = 528) or urine (n = 205) according to local hospital protocols.

Whole-genome sequencing and post processing

Genomic DNA was extracted using either the Wizard® Genomic DNA Purification Kit or the Maxwell® RSC Cultured Cells DNA Kit (AS1620) from Promega. Tagged DNA libraries were created using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina. Whole-genome sequencing was performed on the Illumina NovaSeq 6000 platform with 150 bp paired-end reads. Sequence reads were used to create assemblies using SPAdes v3.10.0 (Bankevich et al., 2012). Annotated assemblies were produced as described previously (Page et al., 2016).

Whole-genome sequence data analysis

The sequence data was assessed using GBS QC pipeline v1.0.3[1]. Sequences that have passed QC were analysed using the GBS typer pipeline v1.0.10[2] to determine sequence type (ST), serotype, and AMR gene carriage. Novel MLST alleles and ST profiles were

1 https://github.com/sanger-bentley-group/GBS_QC_nf

2 https://github.com/sanger-bentley-group/GBS-Typer-sanger-nf

deposited in the MLST database[3]. Isolates were assigned to a CC using the geoBURST algorithm in PHYLOVIZ v2.0 (Nascimento et al., 2017) and a single locus variant for group definition. To determine the presence of a phiStag1 (Jamrozy et al., 2020) and ICESag37 elements, sequence reads were mapped to reference sequences (phiStag1: GenBank accession PP091924; ICESag37: accession no. CP019978, 629058-702486) with SRST2 v0.2 using default parameters (Inouye et al., 2014).

Phylogenetic analyses were performed as detailed in Supplementary Methods. CC17 isolates from the Netherlands were supplemented with publicly available CC17 genomes to reconstruct a global, time-calibrated phylogeny as detailed in Supplementary Methods.

Statistical analysis

Fisher's exact test was used to determine significant association between host status and GBS genotypes, P-value < 0.001 was considered statistically significant.

Results

Serotype, ST, and CC distribution among GBS from carriage and disease

The dataset consisted of 733 maternal carriage and 685 neonatal disease isolates. The majority of neonatal isolates were from EOD (62%) with the remainder derived from LOD (38%; Supplementary Table 1).

Based on the *in silico* analysis, nine capsular serotype genotypes were identified (Ia, Ib, II–VII, and IX), while six isolates were non-typeable (Figure 1A). The most common serotypes among carriage isolates were III (25%), V (19%), II (17%), and Ia (17%), while disease isolates were predominantly serotype III (59%), followed by Ia (22%) (Figure 1B).

We identified 149 unique STs (Supplementary Table 1). The most common STs among carriage isolates were: ST1 (11%), ST17 (11%), ST23 (9%), ST19 (9%), ST28 (6%), and ST24 (5%) (Supplementary Figure 1). In contrast, disease isolates were dominated by ST17 (39%), followed by less common ST23 (15%) and ST19 (10%). The STs were grouped into 12 CCs. The main CCs among all GBS isolates were CC17 (29%), CC19 (19%), CC23 (15%), CC1 (13%), and CC8 (8%) (Figure 1C). In line with ST assignment, the majority of disease isolates belonged to CC17 (45%), followed by CC23 (18%) and CC19 (16%). The carriage isolates showed a more diverse CC distribution, spread across the five main CCs: CC19 (21%), CC1 (18%), CC17 (15%), CC23 (13%), and CC8 (11%) (Figure 1D).

We analysed associations between CCs and serotypes (Figure 2 and Supplementary Figure 2), which showed that CC17 and CC23 carried a single dominant serotype, III and Ia, respectively, while the other main CCs had a higher serotype diversity (Figure 2). Most serotypes were associated with multiple CCs, except for VI and VII

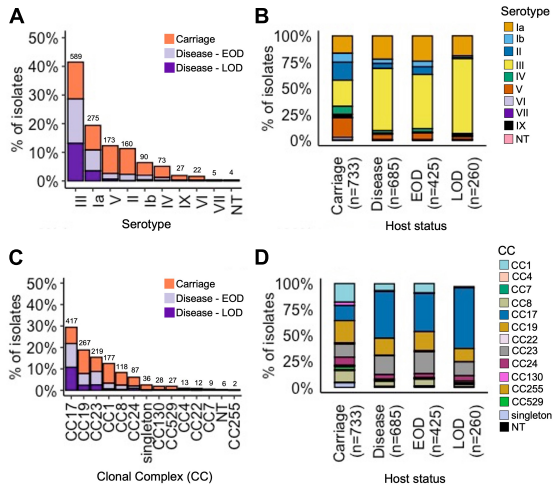


FIGURE 1
Serotype and CC distribution among the GBS isolates by host status. **(A)** Proportion of all isolates representing each serotype, stratified by host status. **(B)** Relative serotype distribution in each host status group. **(C)** Proportion of all isolates representing each CC, stratified by host status. **(D)** Relative CC distribution in each host status group. **(A,C)** Total number of isolates for each serotype and CC, respectively, is displayed above the bars.

which were only identified in CC1, while serotype IX was found only in CC130 isolates (Supplementary Figure 2).

We wished to compare the distribution of genotypes between isolates from carriage and disease. However, since our dataset was not fully temporally matched, we needed to account for the possibility of sampling bias due to the previously reported temporal changes in the prevalence of certain GBS lineages among isolates from neonatal invasive disease in the Netherlands (Jamrozy et al., 2020). To account for the likelihood of a continuing temporal trend in frequency of GBS genotypes, we have evaluated the differences between carriage and disease isolates by comparing a full dataset as well as a subset consisting only of isolates that were collected during overlapping collection years (2018–2021). As such, the latter included only the most recently collected disease isolates.

Across the full dataset we observed that serotype III was significantly more common in disease while serotypes Ib, II, IV, and V were more prevalent in carriage isolates ($P < 0.001$; Table 1). Among temporarily matched datasets, serotypes II, IV, and V remained more common in carriage although this was not statistically significant, while serotype III was still significantly associated with disease isolates.

Among all isolates, ST17 was significantly more common in disease, while ST1, ST28, ST291, and ST569 were significantly associated with carriage isolates ($P < 0.001$; Table 1). In time-matched datasets, these carriage-associated STs were still more prevalent among carriage isolates but this was not statistically significant. In contrast, ST17 was still significantly more common among disease isolates. In line with these associations, CC17 was significantly associated with disease while CC1 with the carriage isolates ($P < 0.001$), which was observed across the full and time-matched datasets. Additionally, CC8 isolates were more common in carriage although this was statistically significant only for the full dataset. We also observed that CC19 was significantly ($P < 0.001$)

3 https://pubmlst.org/sagalactiae/

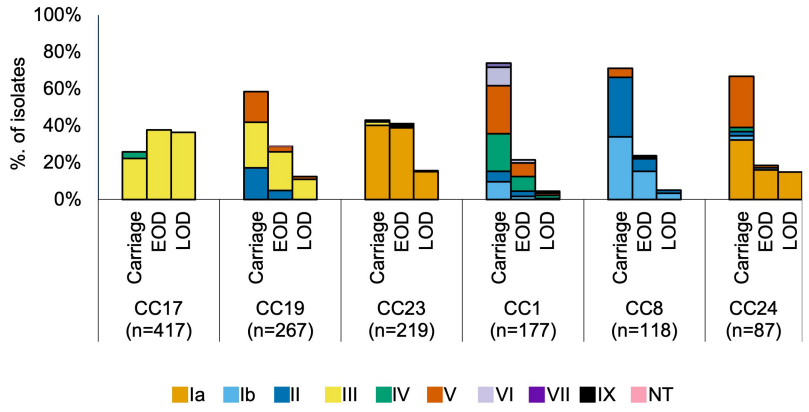


FIGURE 2
Serotype distribution by host status and CC. EOD, early onset disease; LOD, late onset disease.

more common in carriage but only within the time-matched dataset, due to a substantial drop in its prevalence in the most recent disease isolates. Regarding CC-serotype associations, isolates from CC24-serotype V and CC17-serotype IV were found exclusively in carriage isolates except for a single CC24-serotype V identified in disease isolate (Table 1).

We also compared the distribution of genotypes between maternal carriage isolates collected from vagina and urine and observed no variation in prevalence of serotypes and CCs between the two isolation sources (Supplementary Figure 3).

Phylogenetic structure and host status associations within GBS CC

Intra-lineage population structure was analysed by clustering each of the five major GBS CCs into phylogenetic clades (Figure 3). We have previously reported a clonal expansion of specific CC17 clades (CC17-A1 and CC17-A2) among GBS isolates from neonatal disease in the Netherlands (Jamrozy et al., 2020) and wished to compare their distribution among carriage and disease isolates, together with a broader comparison of GBS population between the two host groups. The phylogenetic trees of CC17 and CC23 revealed a single dominant clade (CC17-A and CC23-A, respectively), while the phylogenies of other CCs were more diverse, revealing between 4 and 6 distinct clades each. To identify the CC17 clades associated with the previously reported expansion, the dominant CC17 clade, CC17-A, was partitioned further into three sub-clades: CC17-A1, CC17-A2, and CC17-A0. For each clade identified, we calculated its prevalence across all carriage and disease isolates to identify dominant clusters within each host group and to compare their distribution (Figure 4).

The most common clades among the carriage isolates were CC19-B (13%), CC23-A (12%), and CC1-A (9%). In disease isolates, the most prevalent clades were CC17-A1 (24%) and CC23-A (18%). Additionally, clades CC1-A, CC8-C, CC19-B, and CC19-D were significantly more common among carriage while CC17-A0, CC17-A1, CC17-B, and CC19-A were associated with the disease isolates ($P < 0.001$; Table 1). Those associations remained significant in time-matched datasets only for CC17-A0 and CC17-A1.

To better understand the variable CC clade distribution between carriage and disease isolates, we also compared the prevalence of these clades within corresponding CC (Supplementary Figure 4). This has revealed that for CC1, CC8, and CC23 the distribution of clades was similar between carriage and disease isolates. For instance, CC1-A, CC8-C, and CC23-A represented dominant CC1, CC8, and CC23, respectively, clades in both carriage and disease. In contrast, for CC17 and CC19, we observed that variable clade distribution was associated with differences in CC17 and CC19 population structure between carriage and disease. As such, CC17-A1 was the dominant CC17 clade in disease isolates, while CC17 isolates from carriage showed an equal distribution of CC17-A1 and CC17-A2. The dominant CC19 clade in carriage isolates was CC19-B, while in disease the majority of isolates belonged to CC19-A.

Previous analysis of CC17 isolates from neonatal invasive disease in the Netherlands also revealed acquisition of a novel phage, phiStag1 (GenBank accession PP091924), which correlated with the clonal expansion of clade CC17-A1 (Jamrozy et al., 2020). In the current dataset, phiStag1 phage was found in 26% of all isolates, and it was significantly more common in disease (32%) in comparison to carriage (21%) isolates (Table 1). The phage was found predominantly in CC17 isolates where it was mostly associated with CC17-A1 and CC17-A2 (Supplementary Figure 5). Despite being more common in disease isolates overall, the phage was equally distributed among CC17 isolates from carriage and disease (Supplementary Figure 5). The phiStag1 phage was also detected in other dominant CCs: CC19 (10%), CC23 (28%), CC1 (5%), and CC8 (19%), where it was mostly equally distributed between carriage and disease (Supplementary Figure 5).

GBS resistome

Tetracycline resistance genes (*tet*M, *tet*O, and *tet*L) were the most prevalent AMR determinants, observed in 86% of all GBS isolates. They were equally represented in disease and carriage isolates (Supplementary Table 2).

10.3389/fmicb.2024.1410651

TABLE 1 Prevalence of genotypes found to be differentially distributed between GBS from carriage and disease.

	Full			Time-matched subset		
	Carriage	Disease	*P*-value	Carriage	Disease	*P*-value
Serotype						
Ib	8% (62)	4% (28)	<0.001	8% (59)	9% (6)	1
II	17% (127)	5% (33)	<0.001	17% (120)	6% (4)	0.01
III	25% (183)	59% (406)	<0.001	25% (177)	57% (40)	<0.001
IV	8% (56)	2% (17)	<0.001	8% (53)	0	0.01
V	19% (137)	5% (36)	<0.001	19% (130)	6% (4)	0.004
MLST						
ST1	11% (77)	3% (23)	<0.001	10% (71)	3% (2)	0.05
ST17	11% (82)	39% (266)	<0.001	11% (80)	43% (30)	<0.001
ST28	6% (41)	1% (7)	<0.001	5% (38)	1% (1)	0.25
ST291	2% (15)	0	<0.001	2% (12)	0	0.61
ST569	2% (15)	0	<0.001	2% (15)	0	0.38
CC						
CC1	18% (131)	7% (46)	<0.001	18% (125)	3% (2)	<0.001
CC8	11% (84)	5% (34)	<0.001	11% (80)	6% (4)	0.16
CC17	15% (108)	45% (309)	<0.001	15% (103)	53% (37)	<0.001
CC19	21% (156)	16% (111)	0.02	21% (149)	6% (4)	<0.001
CC-serotype						
CC17-IV	2% (15)	0	<0.001	2% (12)	0	0.04
CC24-V	3% (24)	0.1% (1)	<0.001	3% (24)	1% (1)	0.63
Clades						
CC1-A	9% (68)	3% (22)	<0.001	9% (64)	3% (2)	0.07
CC8-C	5% (37)	2% (11)	<0.001	5% (34)	3% (2)	0.76
CC17-A0	0.4% (3)	7% (50)	<0.001	0.4% (3)	9% (6)	<0.001
CC17-A1	5% (37)	24% (167)	<0.001	5% (37)	21% (15)	<0.001
CC17-B	2% (13)	5% (36)	<0.001	2% (12)	9% (6)	0.004
CC19-A	1% (8)	7% (46)	<0.001	1% (8)	3% (2)	0.23
CC19-B	13% (94)	6% (42)	<0.001	13% (90)	1% (1)	0.002
CC19-D	6% (45)	1% (7)	<0.001	6% (42)	1% (1)	0.17
MGE						
phiStag1	21% (156)	32% (218)	<0.001	21% (148)	39% (27)	0.002
AMR						
MLS_B	24% (179)	15% (102)	<0.001	25% (172)	26% (18)	0.88
ICESag37	4% (31)	5% (37)	0.27	4% (30)	17% (12)	<0.001

The total number of isolates from carriage/disease with corresponding genotype is shown in brackets. The prevalence of genotypes is shown for the full and time-matched (2018–2021) datasets.

The second most common were genes conferring resistance to macrolides, lincosamides, and streptogramin B (MLS_B) antibiotics (*erm*B, *erm*A, *erm*T, *mef* A/*msr*D, *lnu*B, *lsa*C, and *lsa*E), which were present in 20% of all GBS isolates (Supplementary Table 2). The most common MLS_B resistance determinants were *ermB* (12%), *mef* A/*msr*D (4%), and *erm*A (4%; Supplementary Table 2). Across the collection, the highest prevalence of MLS_B resistance genes was observed in isolates belonging to CC19 (32%), CC1 (30%), and CC17 (19%) (Supplementary Table 2), and the majority were from clades CC19-B, CC1-A and CC17-A2, respectively (Figures 3A, C, E). Across the full dataset, MLS_B resistance genes were more common in carriage (24%) in comparison to disease (15%) isolates. This was no longer observed in a time-matched dataset, which showed a comparable frequency of MLS_B resistance genes in carriage (25%) and disease isolates (26%).

Overall, 6% of all GBS isolates carried aminoglycoside resistance genes, with similar prevalence in isolates from carriage (7%) and disease (6%) in a full dataset (Supplementary Table 2).

10.3389/fmicb.2024.1410651

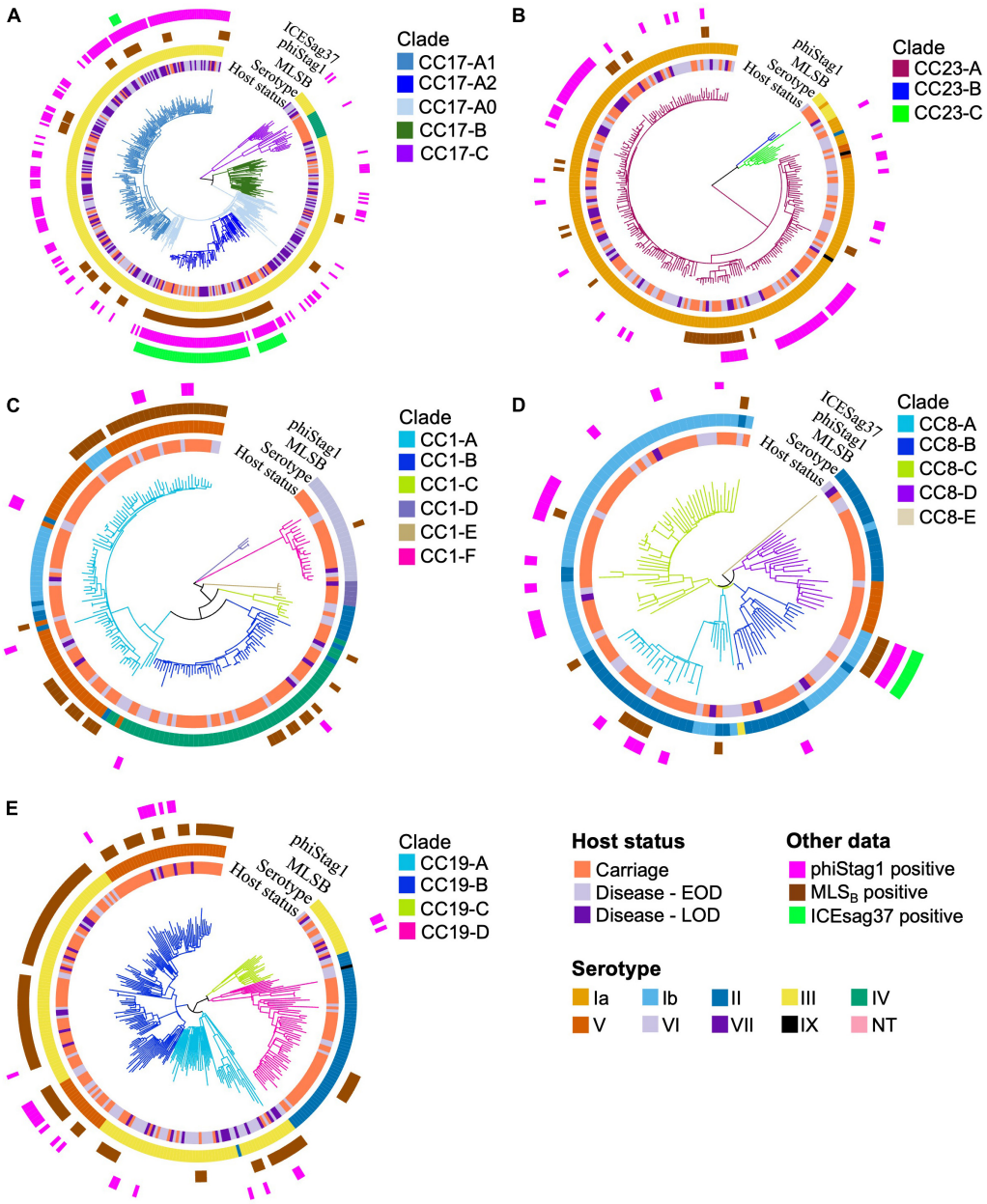


FIGURE 3
Phylogenetic trees of the five major CCs. The branches of each tree are coloured in accordance with CC-specific clusters ID. Each tip is annotated with (from the innermost circle): host status, serotype, carriage of MLS_B resistance genes, phiStag1 and ICESag37 (where applicable). Phylogenetic trees of **(A)** CC17, **(B)** CC23, **(C)** CC1, **(D)** CC8, and **(E)** CC19.

However, in a time-matched dataset they became more common in disease isolates (17%). Low frequency of chloramphenicol resistance genes (1%) was observed, mostly in carriage isolates (2%; Supplementary Table 2).

In CC17, a number of AMR determinants [*ant*(6-Ia), *aph*(3′-III), *aad*E, *erm*B, *tet*O] were carried by clonally related isolates (Supplementary Figure 6). Further analysis revealed that these resistance genes were located on a single, previously defined mobile genetic element (MGE), ICESag37 (Zhou et al., 2017). The majority of CC17 isolates carrying ICESag37 belonged to CC17-A2 (94%; Figure 3A). The ICESag37 element was also detected in CC8 isolates, exclusively in clade CC8-B (Figure 3D). The prevalence of ICESag37 was similar in carriage (4%) and disease (5%) isolates in a full dataset. However, its prevalence increased substantially in more recent disease isolates and in a time-matched dataset it was significantly more prevalent in disease ($P < 0.001$; Table 1).

Global CC17 phylogeny and prevalence of ICESag37

To further investigate the apparent association between ICESag37 element and CC17-A2 isolates, we combined our CC17 sequence data ($n = 229$) with publicly available CC17 genomes

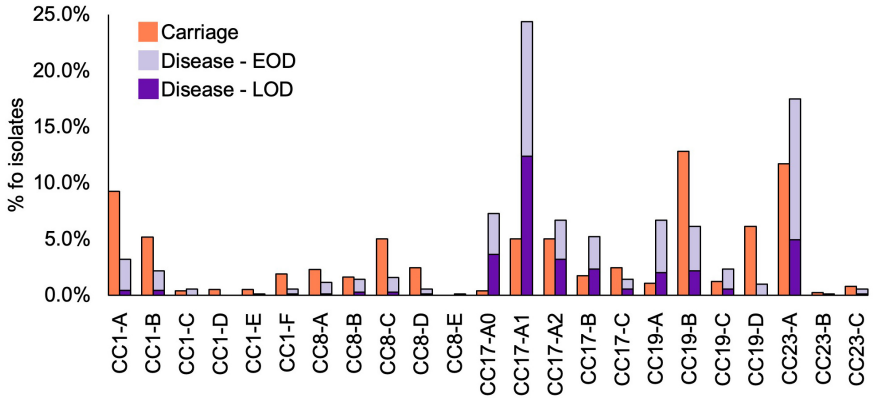


FIGURE 4
Prevalence of CC1, CC8, CC17, CC19, and CC23 clades among GBS isolates by host status. Disease isolates are stratified by disease onset (EOD, early onset disease; LOD, late onset disease).

(n = 650) (Supplementary Table 3) and reconstructed a time-calibrated, global CC17 phylogeny (Figure 5). The non-Dutch CC17 isolates represented 19 countries and most were derived from disease (83%; Supplementary Table 3).

The global CC17 isolates clustered into the three previously observed clades: CC17-A, CC17-B, and CC17-C (Figure 5). The majority of CC17-A isolates were represented by clade CC17-A1 (45%). The ICESag37 element was identified in 10% of non-Dutch CC17 genomes and only in isolates belonging to CC17-A, predominantly in CC17-A2 (63%) but also in CC17-A1 (12%) (Figure 5 and Supplementary Figure 7). The ICESag37-positive CC17 isolates were globally distributed and clustered into three distinct sub-clades, indicating multiple independent acquisition events followed by clonal expansion (Figure 5 and Supplementary Figure 7). It was estimated that all ICESag37-positive sub-clades emerged in the 1990s. Based on this dataset, the first ICESag37 positive CC17 isolates were collected in 2010 in Canada and China, with the first isolation in the Netherlands in 2011 (Supplementary Figure 8). Regardless of the country of origin, the majority of globally derived CC17-A2 isolates collected between 2010 and 2021 were ICESag37 positive (Supplementary Figure 8). ICESag37 sequence from all globally distributed CC17 isolates shared significant nucleotide identity (93%–100%, median 99.8%) (Supplementary Figure 9).

Discussion

Intrapartum antibiotic prophylaxis currently represents the main strategy for the prevention of early onset GBS disease. This prevention strategy assumes an equal risk of neonatal invasive disease from any identified colonising GBS isolate. However, our and previous research clearly showed that some GBS genotypes carry a higher risk of neonatal disease. More studies are needed to investigate the pathophysiological mechanisms that drive these differences in invasive potential and evaluate the added value of GBS genotype determination to more precisely target GBS prevention. Our work has shown that, in line with previous reports, CC17-serotype III strains were significantly more common in disease (Kekic et al., 2021). while serotypes II, IV, V, and CC1 were associated with maternal carriage. We have also identified variable prevalence of some lineage-serotype combinations between the two host groups. This included isolates representing CC24-serotype V and CC17-serotype IV, which were associated with carriage. This suggests that the association between CC17 and neonatal disease is serotype III dependent. Although other serotypes have emerged within this GBS lineage, they appear less likely to cause neonatal infection as none of the CC17-serotype IV were observed among disease isolates in our collection. In contrast, serotype III remained associated with neonatal disease even after exclusion of all CC17 isolates ($P < 0.001$).

Our previous work has shown expansion of specific CC17 sub-clades, CC17-A1 and CC17-A2, among isolates from neonatal invasive disease in the Netherlands, which correlated with a rise in disease incidence in the country. A matched collection of isolates from maternal carriage from the Netherlands was not previously available, which hindered further investigation of the epidemiology of these clones in a wider GBS population. In this work, we addressed this data gap and compared the prevalence of different clades from major CCs, including CC17, between carriage and disease isolates. Overall, CC17-A1 clade was the most prevalent sub-lineage among all disease isolates, suggesting an increased capacity to cause disease. However, although it was considerably less common among all carriage isolates, the CC17 population from carriage was dominated by CC17-A1 and CC17-A2 isolates. This suggests that the previously reported rise in the frequency of these clusters in GBS from neonatal disease likely reflected their expansion in the carriage GBS CC17 population, which resulted in a spillover to invasive GBS population.

We also reported previously and in this work that the expanding CC17 sub-clades, CC17-A1 and CC17-A2, are associated with certain MGEs that might contribute to their prevalence. One is a novel phage, previously termed phiStag1, which emerged suddenly in the CC17 population around the mid-1990s (Jamrozy et al., 2020). A recent study has shown that the phage belongs to a novel group of phages designated streptococcal mobilisable prophages (SMphages) (Huang et al., 2023). The phage carries a putative virulence gene, which was termed Alp-P1 and was shown to promote the adhesion and invasion of bovine and human cells. These findings further indicate that phiStag1 might

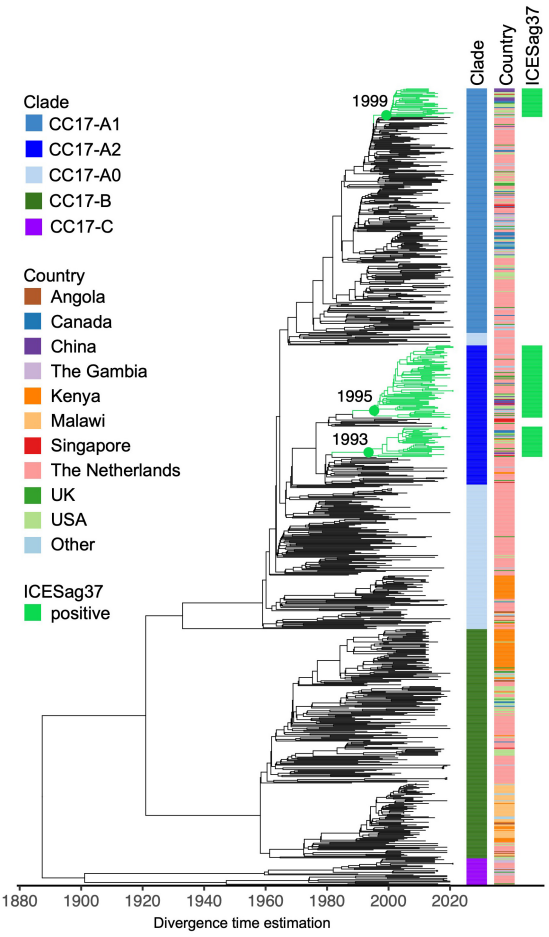


FIGURE 5
Core genome time-calibrated maximum likelihood phylogeny of global CC17 GBS isolates. The tree consists of external (n = 650) and Dutch (n = 229) CC17 GBS genomes. Each tip is annotated with CC17 clade ID and country of isolation ("Other": countries represented by less than 10 isolates). Branches of clusters carrying ICESag37 are coloured in green.

provide some selective advantage to its host and thus promote clonal expansion of CC17-A1 and CC17-A2. In our dataset, we found phiStag1 to be overall more common among disease isolates. However, among CC17 isolates, the phage was equally distributed among carriage and disease. Further work is needed to better understand phiStag1's role in GBS disease. While it was found more common in isolates from disease, this was likely driven by its association with CC17 and the dominance of this lineage within disease. It remains unclear if presence of this phage contributes to maternal colonisation, transmission to the infant or neonatal invasive disease.

We have also observed a high prevalence of the ICESag37 element among CC17 isolates. This MGE confers resistance to erythromycin, tetracycline and aminoglycosides (Zhou et al., 2017). It was first identified in the Sag37 strain, which represents ST12. In our dataset, ICESag37 was most common in CC17 (15%), followed by CC8 (4%), which includes ST12. Carriage of a MDR ICESag element, corresponding to ICESag37, has been reported previously in CC17 (Campisi et al., 2016). Our analysis of a global CC17 phylogeny has confirmed that ICESag37-positive CC17 isolates are widely distributed and have been found in Asia, Europe, and North America. We also observed that carriage of this MGE within CC17 is associated mostly with sub-clade CC17-A2. Within the Dutch GBS collection, CC17-A2 accounted for 87% of all isolates carrying ICESag37. As such, ICESag37-positive CC17-A2 isolates resistant to both macrolides and aminoglycosides might pose a clinical threat due to reduced options for first- and second-line antimicrobial treatment of GBS infections.

Limitations of our study include a temporal sampling bias, with disease and carriage isolates collected over different time periods, with only a 4-year overlap between the two collections (2018–2021). To account for this, we conducted a parallel analysis of full and time-matched datasets. While some genotypes showed statistically significant associations across both datasets, for many the differences between carriage and disease isolates were no longer statistically significant in time-matched dataset, which is likely partly due to much lower disease sample size in the latter. However, the analysis also showed that the prevalence of AMR genes was higher in most recently collected disease isolates, which was associated with increase in frequency of isolates carrying the ICESag37 element. Finally, the maternal carriage isolates were recovered from vagina and urine, with the latter potentially associated with asymptomatic bacteriuria. However, we observed no variation in genotype distribution between isolates from these sources suggesting that GBS isolates from urine are acquired from the rectovaginal site and represent the same GBS population.

Here we report that the previously observed clonal expansion of CC17-A1 and CC17-A2 clades as well as the emergence of phiStag1 phage among GBS isolates from neonatal invasive disease in the Netherlands likely reflect changes in the maternal carriage population. Overall, our findings reinforce the importance of comparing GBS isolates from healthy individuals and patients to identify pathogen genotypes that might be associated with increased capacity to cause disease. Altogether this will provide pathogenicity markers that can be targeted in disease prevention strategies as well as molecular markers for surveillance of high-risk clones that demonstrate enhanced dissemination across GBS population irrespective of the host status.

Data availability statement

All GBS isolates analyzed in this study are publicly available. The ENA accession numbers for each GBS isolate are provided in Supplementary Table 1.

Ethics statement

Ethical approval was not required for the studies involving humans because it involved the analysis of anonymized whole genome sequence data from Group B Streptococcus (GBS) isolates collected in the Netherlands. The data used in this research were obtained from a collection database maintained by the Netherlands Reference Laboratory for Bacterial Meningitis, which serves as the national reference laboratory for community-acquired invasive bacterial infections, including neonatal GBS infections. As the data were anonymized and did not contain

any identifiable patient information, the study did not involve direct interaction with human subjects. Furthermore, the research adhered to all relevant ethical guidelines and regulations governing the use of anonymized genomic data for scientific research purposes. Therefore, ethical approval was not deemed necessary for this type of retrospective, anonymized data analysis study. The studies were conducted in accordance with the local legislation and institutional requirements. 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 because in accordance with guidance from the Netherlands Reference Laboratory for Bacterial Meningitis, patient data utilized in this study were obtained from an anonymized collection database. As a result, individual patients are not identifiable or traceable, aligning with Dutch Law requirements.

Author contributions

UK: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft. VD: Software, Writing – review & editing. CC: Methodology, Writing – review & editing. NS: Conceptualization, Writing – review & editing. DB: Conceptualization, Writing – review & editing. WM: Methodology, Writing – review & editing. SB: Conceptualization, Writing – review & editing. MB: Conceptualization, Writing – review & editing. DJ: Conceptualization, Methodology, Project administration, Supervision, Validation, Writing – review & editing.

Members of the NOGBS study group

Eveline van Asbeck (Department of Obstetrics and Gynaecology, Tergooi Hospital, Hilversum, Netherlands), Rolanda Baars (Department of Paediatrics, St. Jansdal Hospital, Harderwijk, Netherlands), Desireé V. Ballegoy (Department of Medical Microbiology and Immunology, St. Antonius Hospital, Nieuwegein, Netherlands), Ron van Beek (Department of Paediatrics, Amphia Hospital, Breda, Netherlands), Vincent Bekker (Department of Paediatrics, Leiden University Medical Centre, Leiden, Netherlands), Maartje van den Berg (Department of Paediatrics, Haaglanden Medical Centre, The Hague, Netherlands), Geert Jan Blok (Department of Neonatology, Northwest Clinics, Alkmaar, Netherlands), Mijke Breukels (Department of Paediatrics, Elkerliek Hospital, Helmond, Netherlands), Alwin F. J. Brouwer (Department of Paediatrics, Hospital of Nij Smellinghe, Drachten, Netherlands), Matthijs Brouwer (Department of Neurology, Amsterdam UMC, Amsterdam, Netherlands), Renske Cornelisse-van Vugt (Department of Paediatrics, Canisius Wilhelmina Hospital, Nijmegen, Netherlands), Dick van Dam (Laboratory for Medical Microbiology and Public Health, Hengelo, Netherlands), Karin van Dijk (Department of Medical Microbiology and Infection Control, Amsterdam University Medical Centre, Amsterdam Infection and Immunity Institute, Amsterdam, Netherlands), Luçan C. Delemarre (Department of Paediatrics, Amstelland Hospital, Amstelveen, Netherlands), Anouk Dings (Department of Paediatrics, Gelre Hospitals, Apeldoorn, Netherlands), Rienus A. Doedens (Department of Paediatrics, Martini Hospital, Groningen, Netherlands), Stefan M. van Dorth (Department of Paediatrics, Tjongerschans Hospital, Heerenveen, Netherlands), Gertjan Driessen (Department of Paediatrics, Haga Hospital, The Hague, Netherlands), Erika van Elzakker (Department of Medical Microbiology and Infection Prevention, Amsterdam University Medical Centre, University of Amsterdam, Amsterdam, Netherlands), Arie van der Ende (Department of Medical Microbiology and Infection Prevention, Amsterdam University Medical Centre, University of Amsterdam, Amsterdam, Netherlands), Katja de Graaff (Department of Obstetrics and Gynaecology, Reinier de Graaf Hospital, Delft, Netherlands), Hester M. Havers (Department of Paediatrics, Alrijne Hospital, Leiderdorp, Netherlands), Jojanneke Heidema (Department of Paediatrics, St. Antonius Hospital, Utrecht, Netherlands), Marieke A. C. Hemels (Department of Neonatology, Isala Clinics, Zwolle, Netherlands), Maartje E. N. van den Heuvel (Department of Paediatrics, Onze Lieve Vrouwe Gasthuis, West Location, Amsterdam, Netherlands), Marion E. van Hoorn (Department of Obstetrics and Gynaecology, HagaZiekenhuis, Den Haag, Netherlands), Marlies van Houten (Department of Paediatrics, Spaarne Gasthuis, Haarlem, Netherlands), Flip van der Hulst (Department of Paediatrics, Zaans Medical Centre, Zaandam, Netherlands), Monique A. M. Jacobs (Department of Paediatrics, Slingeland Hospital, Doetinchem, Netherlands), Arieke Janse (Department of Paediatrics, Gelderse Vallei Hospital, Ede, Netherlands), Miranda de Jong (Department of Paediatrics, Albert Schweitzer Hospital, Dordrecht, Netherlands), Anton H. van Kaam (Department of Neonatology, Amsterdam University Medical Centre, University of Amsterdam, Amsterdam, Netherlands), Ageeth Kaspers (Department of Paediatrics, Medisch Spectrum Twente, Twente, Netherlands), Merel N. van Kassel (Department of Neurology, Amsterdam University Medical Centre, University of Amsterdam, Amsterdam, Netherlands), Anne A. M. W. van Kempen (Department of Paediatrics, Onze Lieve Vrouwe Gasthuis, East Location, Amsterdam, Netherlands), Kristine Klúčovská (Department of Paediatrics, Treant Hospital Group, Hoogeveen, Netherlands), Karen Korbeek (Department of Paediatrics, St. Jansdal Hospital, Harderwijk, Netherlands), René F. Kornelisse (Department of Paediatrics, Erasmus Medical Centre, Rotterdam, Netherlands), Taco W. Kuijpers (Department of Paediatric Immunology, Rheumatology and Infectious Diseases, Amsterdam University Medical Centre, University of Amsterdam, Amsterdam, Netherlands), Elisabeth van Leeuwen (Department of Obstetrics and Gynaecology, Amsterdam University Medical Centre, University of Amsterdam, Amsterdam, Netherlands), Ineke Linde (Department of Infectious Diseases, GGD Amsterdam, Amsterdam, Netherlands), Jeannette von Lindern (Department of Paediatrics, Groene Hart Hospital, Gouda, Netherlands), Karen van Mechelen (Department of Neonatology, Maastricht University Medical Centre, Maastricht), Netherlands), Clemens B. Meijssen (Department of Paediatrics, Meander Medical Centre, Amersfoort, Netherlands), Jeroen Noordzij (Department of Paediatrics, Reinier de Graaf Hospital, Delft, Netherlands), Annemarie Oudshoorn (Department of Paediatrics, Gelre Hospitals, Apeldoorn, Netherlands), Frans B. Plötz (Tergooi Hospital, Hilversum, Netherlands), Maarten Rijpert (Department of Paediatrics, Zaans Medical Centre,

Zaandam, Netherlands), Maaike van Rossem (Department of Paediatrics, Rijnstate Hospital, Arnhem, Netherlands), Machteld van Scherpenzeel (Department of Paediatrics, Medical Centre Leeuwarden, Leeuwarden, Netherlands), Maarten Schijffelen (Laboratory for Medical Microbiology and Public Health, Hengelo, Netherlands), George Shabo (Department of Paediatrics, Hospital Group Twente, Twente, Netherlands), Jacqueline van der Sluijs (Department of Neonatology, Maxima Medical Centre, Veldhoven, Netherlands), Linde Snoek (Department of Neurology, Amsterdam UMC, Amsterdam, Netherlands), Jacqueline U. M. Termote (Department of Neonatology, University Medical Centre Utrecht, Utrecht, Netherlands), Gerdien A. Tramper-Stranders (Department of Paediatrics, Franciscus Gasthuis, Rotterdam, Netherlands), Gavin W. ten Tusscher (Department of Paediatrics, Dijklander Hospital, Hoorn, Netherlands), Saraa J. Vainio (Department of Medical Microbiology and Immunology, St. Antonius Hospital, Nieuwegein, Netherlands), Joost van de Ven (Department of Obstetrics and Gynaecology, Elkerliek Hospital, Helmond, Netherlands), Mirjam van Veen (Department of Paediatrics, Haga Hospital, The Hague, Netherlands), Marlies Vermaas (Department of Paediatrics, Admiraal de Ruyter Hospital, Goes, Netherlands), Marjoke Verweij (Department of Paediatrics, Viecuri Medical Centre, Venlo, Netherlands), Douwe H. Visser (Department of Neonatology, Amsterdam University Medical Centre, University of Amsterdam, Amsterdam, Netherlands), Karlijn C. Vollebregt (Department of Obstetrics and Gynaecology, Spaarne Gasthuis, Haarlem, Netherlands), Wouter J. de Waal (Department of Paediatrics, Diakonesse Hospital, Utrecht, Netherlands), Anne-Marie van Wermeskerken (Department of Paediatrics, Flevohospital, Almere, Netherlands), Janneke Wilms (Department of Paediatrics, BovenIJ Hospital, Amsterdam, Netherlands), Tom F. W. Wolfs (Department of Paediatrics, University Medical Centre Utrecht, Utrecht, Netherlands), Maurice G. A. J. Wouters (Department of Obstetrics and Gynaecology, Amsterdam University Medical Centre, University of Amsterdam, Amsterdam, Netherlands).

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by a Clinical Research Grant of the Amsterdam Institute for Infection and Immunity and by Stichting Steun Emma Kinderziekenhuis and ItsME Foundation to MB and an NWO-Vici-Grant (918.19.627) to DB. UK and DJ were supported by the Bill and Melinda Gates Foundation (grant INV-010426). SB was supported by the Wellcome Trust (grant 220540/Z/20/A).

Acknowledgments

We thank the technicians of Netherlands Reference Laboratory for Bacterial Meningitis for the collection of samples and characterisation of the group B streptococcal isolates. We would like to thank the Sequencing Operations and Pathogen Informatics teams at the Wellcome Sanger Institute for sequencing and informatics support.

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.2024.1410651/full#supplementary-material

References

Absalon, J., Simon, R., Radley, D., Giardina, P. C., Koury, K., Jansen, K. U., et al. (2022). Advances towards licensure of a maternal vaccine for the prevention of invasive group B streptococcus disease in infants: A discussion of different approaches. *Hum. Vacc. Immunother.* 18:2037350. doi: 10.1080/21645515.2022.2037350

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. Comp. Biol.* 19, 455–477.

Berti, F., Campisi, E., Toniolo, C., Morelli, L., Crotti, S., Rosini, R., et al. (2014). Structure of the type IX group B *Streptococcus* capsular polysaccharide and its evolutionary relationship with types V and VII. *J. Biol. Chem.* 289, 23437–23448. doi: 10.1074/jbc.M114.567974

Bianchi-Jassir, F., Paul, P., To, K.-N., Carreras-Abad, C., Seale, A. C., Jauneikaite, E., et al. (2020). Systematic review of Group B Streptococcal capsular types, sequence types and surface proteins as potential vaccine candidates. *Vaccine* 38, 6682–6694. doi: 10.1016/j.vaccine.2020.08.052

Björnsdóttir, E., Martins, E., Erlendsdóttir, H., Haraldsson, G., Melo-Cristino, J., Kristinsson, K., et al. (2016). Changing epidemiology of group B streptococcal infections among adults in Iceland: 1975–2014. *Clin. Microbiol. Infect.* 22, 379.e9–e16. doi: 10.1016/j.cmi.2015.11.020

Campisi, E., Rosini, R., Ji, W., Guidotti, S., Rojas-López, M., Geng, G., et al. (2016). Genomic analysis reveals multi-drug resistance clusters in group b streptococcus cc17 hypervirulent isolates causing neonatal invasive disease in Southern Mainland China. *Front. Microbiol.* 7:1265. doi: 10.3389/fmicb.2016.01265

Flores, A. R., Galloway-Peña, J., Sahasrabhojane, P., Saldaña, M., Yao, H., Su, X., et al. (2015). Sequence type 1 group B Streptococcus, an emerging cause of invasive

disease in adults, evolves by small genetic changes. *Proc. Natl. Acad. Sci. U.S.A.* 112, 6431–6436. doi: 10.1073/pnas.1504725112

Gonçalves, B. P., Procter, S. R., Paul, P., Chandna, J., Lewin, A., Seedat, F., et al. (2022). Group B streptococcus infection during pregnancy and infancy: Estimates of regional and global burden. *Lancet Glob. Health* 10, e807–e819. doi: 10.1016/S2214-109X(22)00093-6

Huang, J., Dai, X., Wu, Z., Hu, X., Sun, J., Tang, Y., et al. (2023). Conjugative transfer of streptococcal prophages harboring antibiotic resistance and virulence genes. *ISME J.* 17, 1467–1481. doi: 10.1038/s41396-023-01463-4

Inouye, M., Dashnow, H., Raven, L. A., Schultz, M. B., Pope, B. J., Tomita, T., et al. (2014). SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* 6:90. doi: 10.1186/s13073-014-0090-6

Jamrozy, D., Bijlsma, M. W., de Goffau, M. C., van de Beek, D., Kuijpers, T. W., Parkhill, J., et al. (2020). Increasing incidence of group B streptococcus neonatal infections in the Netherlands is associated with clonal expansion of CC17 and CC23. *Sci. Rep.* 10:9539. doi: 10.1038/s41598-020-66214-3

Kekic, D., Gajic, I., Opavski, N., Kojic, M., Vukotic, G., Smitran, A., et al. (2021). Trends in molecular characteristics and antimicrobial resistance of group B streptococci: A multicenter study in Serbia, 2015-2020. *Sci. Rep.* 11:540. doi: 10.1038/s41598-020-79354-3

Khan, U. B., Jauneikaite, E., Andrews, R., Chalker, V. J., and Spiller, O. B. (2022). Identifying large-scale recombination and capsular switching events in *Streptococcus agalactiae* strains causing disease in adults in the UK between 2014 and 2015. *Microb. Genom.* 8:000783. doi: 10.1099/mgen.0.000783

Kobayashi, M., McGee, L., Chochua, S., Apostol, M., Alden, N. B., Farley, M. M., et al. (2021). Low but increasing prevalence of reduced beta-lactam susceptibility among invasive group B streptococcal isolates, US population-based surveillance, 1998-2018. *Open Forum Infect. Dis.* 8:ofaa634. doi: 10.1093/ofid/ofaa634

Nascimento, M., Sousa, A., Ramirez, M., Francisco, A. P., Carriço, J. A., and Vaz, C. (2017). PHYLOViZ 2.0: Providing scalable data integration and visualization for multiple phylogenetic inference methods. *Bioinformatics* 33, 128–129. doi: 10.1093/bioinformatics/btw582

Page, A. J., De Silva, N., Hunt, M., Quail, M. A., Parkhill, J., Harris, S. R., et al. (2016). Robust high-throughput prokaryote de novo assembly and improvement pipeline for illumina data. *Microb. Genom* 2:e000083. doi: 10.1099/mgen.0.000083

Sabroske, E. M., Iglesias, M. A. S., Rench, M., Moore, T., Harvey, H., Edwards, M., et al. (2023). Evolving antibiotic resistance in group B streptococci causing invasive infant disease: 1970-2021. *Pediatr. Res.* 93, 2067–2071. doi: 10.1038/s41390-022-02375-3

Slotved, H. C., and Hoffmann, S. (2020). The epidemiology of invasive group B *Streptococcus* in Denmark From 2005 to 2018. *Front. Public Health* 8:40. doi: 10.3389/fpubh.2020.00040

Teatero, S., Ramoutar, E., McGeer, A., Li, A., Melano, R. G., Wasserscheid, J., et al. (2016). Clonal Complex 17 group B *Streptococcus* strains causing invasive disease in neonates and adults originate from the same genetic pool. *Sci. Rep.* 6:20047.

Zhou, K., Xie, L., Han, L., Guo, X., Wang, Y., and Sun, J. (2017). ICESag37, a novel integrative and conjugative element carrying antimicrobial resistance genes and potential virulence factors in *Streptococcus agalactiae*. *Front. Microbiol.* 8:1921. doi: 10.3389/fmicb.2017.01921

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 29 August 2024
DOI 10.3389/fmicb.2024.1452007

OPEN ACCESS

EDITED BY
Renmao "Tim" Tian,
Illinois Institute of Technology, United States

REVIEWED BY
Yasmine Hasanine Tartor,
Zagazig University, Egypt
Wanderson Marques Da Silva,
National Scientific and Technical Research Council (CONICET), Argentina

*CORRESPONDENCE
David L. Erickson
david_erickson@byu.edu

†PRESENT ADDRESS
Michael A. Olson,
Department of Biological Sciences, Snow College, Ephraim, UT, United States

RECEIVED 20 June 2024
ACCEPTED 19 August 2024
PUBLISHED 29 August 2024

CITATION
Olson MA, Cullimore C, Hutchison WD, Grimsrud A, Nobrega D, De Buck J, Barkema HW, Wilson E, Pickett BE and Erickson DL (2024) Genes associated with fitness and disease severity in the pan-genome of mastitis-associated *Escherichia coli*.
Front. Microbiol. 15:1452007.
doi: 10.3389/fmicb.2024.1452007

Genes associated with fitness and disease severity in the pan-genome of mastitis-associated *Escherichia coli*

Michael A. Olson[1†], Caz Cullimore[1], Weston D. Hutchison[1], Aleksander Grimsrud[1], Diego Nobrega[2], Jeroen De Buck[2], Herman W. Barkema[2], Eric Wilson[1], Brett E. Pickett[1] and David L. Erickson[1*]

[1]Department of Microbiology and Molecular Biology, Brigham Young University, Provo, UT, United States, [2]Faculty of Veterinary Medicine, University of Calgary, Calgary, AB, Canada

Introduction: Bovine mastitis caused by *Escherichia coli* compromises animal health and inflicts substantial product losses in dairy farming. It may manifest as subclinical through severe acute disease and can be transient or persistent in nature. Little is known about bacterial factors that impact clinical outcomes or allow some strains to outcompete others in the mammary gland (MG) environment. Mastitis-associated *E. coli* (MAEC) may have distinctive characteristics which may contribute to the varied nature of the disease. Given their high levels of intraspecies genetic variability, virulence factors of commonly used MAEC model strains may not be relevant to all members of this group.

Methods: In this study, we sequenced the genomes of 96 MAEC strains isolated from cattle with clinical mastitis (CM). We utilized clinical severity data to perform genome-wide association studies to identify accessory genes associated with strains isolated from mild or severe CM, or with high or low competitive fitness during *in vivo* competition assays. Genes associated with mastitis pathogens or commensal strains isolated from bovine sources were also identified.

Results: A type-2 secretion system (T2SS) and a chitinase (ChiA) exported by this system were strongly associated with pathogenic isolates compared with commensal strains. Deletion of *chiA* from MAEC isolates decreased their adherence to cultured bovine mammary epithelial cells.

Discussion: The increased fitness associated with strains possessing this gene may be due to better attachment in the MG. Overall, these results provide a much richer understanding of MAEC and suggest bacterial processes that may underlie the clinical diversity associated with mastitis and their adaptation to this unique environment.

KEYWORDS

mastitis, *Escherichia coli*, GWAS, chitinase, ExPEC

Introduction

Bovine mastitis often results from bacterial infection. Mastitis-associated *E. coli* (MAEC), abundant in the dairy environment, are the most important cause of this disease. These bacteria cycle between the bovine digestive tract and soils and beddings of stalls, from which they may gain

access to the mammary gland (MG) via the teat canal. Once established within the MG, MAEC can induce a range of clinical presentations. Subclinical mastitis is usually defined as increased somatic cell counts in milk and transient inflammation caused by cytokine release. MAEC infections are typically cleared rapidly without complication or need for antibiotic intervention (Roberson et al., 2004; Fuenzalida and Ruegg, 2019). Conversely, severe clinical mastitis (CM) can damage the MG through sustained inflammation and high bacterial loads. These cows often suffer permanent udder damage, and the bacteria occasionally disseminate beyond the MG leading to sepsis (Wenz et al., 2001; Suojala et al., 2013). Some mastitis cases are characterized by mild acute disease followed by extended periods of chronic or recurrent infections. Occasionally MAEC strains gain access through the teat canal to the MG during the non-lactating period. Symptoms of CM develop shortly after the next lactation begins, which also tends toward chronic infections (Bradley and Green, 2000; Lippolis et al., 2014).

Features that distinguish MAEC from other *E. coli* strains have been difficult to identify and remain incompletely understood. Although they may come from any of the diverse *E. coli* lineages, MAEC strains often fall within *E. coli* phylogroups A and B1, which are also the most common phylogroups of commensal strains. Nevertheless, commensal strains are unable to cause acute clinical or chronic mastitis (Blum et al., 2015), suggesting that there are fundamental differences between commensals and MAEC strains that have yet to be discovered. MAEC frequently belong to sequence types (MLST) 10, 58, 95, and 1125 (Kempf et al., 2016; Leimbach et al., 2017; Blum and Leitner, 2013; Nuesch-Inderbinen et al., 2019; Freitag et al., 2017; Keane, 2016) but this also does not distinguish them from other strains. Genes predicted to be associated with MAEC have been examined in numerous PCR-based surveys as well as more extensive genomic studies (Blum et al., 2015; Nuesch-Inderbinen et al., 2019; Jamali et al., 2018; Ismail and Abutarbush, 2020; Lippolis et al., 2018).

Phenotypes distinctive of MAEC strains include relatively robust resistance to the complement system and greater motility than other *E. coli* (Lippolis et al., 2018; Guerra et al., 2020; Lippolis et al., 2016). The ferric dicitrate transport system encoded by the *fecABCDE* genes is also highly expressed and much more consistently found in MAEC genomes compared with other *E. coli* (Jung et al., 2021; Lin et al., 1999). Previously, we conducted a functional genetic screen using transposon insertion sequencing to identify the fitness factors of a single MAEC strain (Olson et al., 2018). This work demonstrated that the *fec* genes are needed to colonize lactating mouse MGs and implicated the high-affinity zinc transport system and several genes involved in other metabolic pathways in fitness in MGs. However, different genes may be required for fitness in other MAEC strain backgrounds. A more thorough understanding of MAEC genomics may help identify strains capable of causing mastitis from the varied strains found in agricultural settings, as well as those more likely to cause severe CM.

Mastitis severity depends on several host factors that contribute to disease outcome. For instance, the stage of lactation when the MG becomes infected has a strong influence on whether the infecting bacteria are efficiently cleared. At the time of parturition and earlier stages of lactation, cows are more prone to severe CM, which has been attributed to immunological dysfunction of neutrophils and lymphocytes during this time (Burvenich et al., 2007; Sordillo, 2005; Vangroenweghe et al., 2005; Hill, 1981; Cai et al., 1994; Burvenich et al., 1994).

In addition to host factors, bacterial factors may also influence CM severity (Guerra et al., 2020; Lehtolainen et al., 2003; Wenz et al., 2006). Some MAEC strains carry virulence factors often found in extraintestinal pathogenic *E. coli* (ExPEC) such as toxins, siderophores, capsules and adhesins (Olson et al., 2021). For the most part, the influence of these virulence factors during intramammary infections has not been determined. Thus far, the only trait with a demonstrated association with CM severity is swarming motility, which is higher in MAEC strains isolated from severe CM cases than those strains isolated from mild and moderate CM (Guerra et al., 2020). Gene expression comparisons of MAEC isolates from transient infections and persistent CM also demonstrated that flagella gene expression and motility are generally higher in the persistent CM strains (Lippolis et al., 2018). Persistent strains are also more resistant to serum complement and express the *fec* operon genes at higher levels than transient strains.

Previous genomic analyses have focused on identifying genes that are unique to MAEC isolates compared to non-pathogenic strains inhabiting the same niches or commensal strains belonging to the same phylogroups. These analyses have uncovered putative marker genes that could distinguish MAEC from other strains, including those that may function in niche-specific metabolic pathways, gene regulation, and virulence (Jung et al., 2021; Goldstone et al., 2016). In this study, we sequenced 96 MAEC genomes and implemented a comparative genomics approach to uncover genes more likely to be present in strains isolated from cattle with mild or severe CM. We then extended this analysis to identify genes associated with either mastitis isolates or commensal *E. coli* strains from cattle. We employed mouse infection and milk growth assays to separate MAEC strains with higher or lower fitness to identify genes that are associated with these phenotypes.

Materials and methods

Bacterial strains and growth conditions

E. coli strains M22 through M117 (Supplementary Dataset 1) were isolated from individual quarter milk samples from cattle with clinical mastitis as part of the Canadian National Cohort of Dairy Farms as previously described (Reyher et al., 2011). In brief, 89 herds across Canada (Alberta, Ontario, Quebec, and the Maritime provinces Prince Edward Island, New Brunswick, and Nova Scotia) were selected to be representative of their respective province in terms of housing type, bulk tank somatic cell count, cattle breed, and milking schedule, and were followed from February 2007 to December 2008. At the time of collection, farmers evaluated the clinical signs presented in each affected animal and assigned a clinical score, as follows: mastitis score 1 (mild = abnormal milk only), mastitis score 2 (moderate = abnormal milk and local inflammation signs), or mastitis score 3 (severe = abnormal milk, local inflammation and systemic clinical signs) (Sears and McCarthy, 2003). From this collection (Dufour et al., 2019), strains isolated from cows in the middle to late stages of their lactation cycle were selected in order to focus on bacterial differences and minimize the effect that early lactation has on mastitis severity (Hyvonen et al., 2010). Selection was also designed to include isolates from different herds and provinces. Six strains isolated from CM cases in the United States were also included in this study

(Supplementary Dataset 1). Strains M3, M6, M9, M11, and M12 were previously isolated from quarter milk samples of CM cases (Olson et al., 2018). Clinical severity data were not available for these isolates. Strain G1 was supplied by Jennifer Wilson (Jersey Girls, Jerome ID) and was isolated from a cow with severe, gangrenous mastitis that necessitated culling of the animal.

The *E. coli* strains obtained from the mastitis pathogen culture collection (Dufour et al., 2019) were verified phenotypically by colony morphology on MacConkey agar (Difco, United States) plates and confirmed by whole-genome sequencing. Bacteria were routinely grown in Luria-Bertani (LB) medium at 37°C. For milk cultures, whole, unpasteurized cow's milk was obtained from a local supplier and used immediately or stored at −80°C until use. To determine growth yields of individual MAEC isolates, bacteria from overnight LB cultures were added to 100 µL milk to a concentration of 10^3 CFU/mL in a 96-well format. Plates were incubated without shaking at 37°C. A sample was immediately removed ($T = 0$), serially diluted and plated on MacConkey agar to determine the starting concentration. Bacterial concentrations were also measured at 4 h and 8 h post-inoculation. The change in CFU/ml at 4 h and 8 h relative to $T = 0$ was calculated for three biological replicates for each strain.

Genome sequencing, assembly, and annotation

Total DNA was isolated from MAEC strains using a ZR Fungal/Bacterial DNA MiniPrep kit (Zymoresearch). DNA sequencing libraries were prepared using the Illumina Nextera DNA Library Prep kit as previously described (Baym et al., 2015). DNA libraries were sequenced by Genewiz, Inc. (South 330Plainfield, NJ), and Illumina paired-end reads of 150 bp were generated on a MiSeq with version 2 chemistry. Quality control, contig assembly, *in silico* determination of phylogroup, multi-locus sequence typing, and GrapeTree analysis were performed within EnteroBase (Zhou et al., 2020; Zhou et al., 2018) and genomes were annotated with Prokka (version 1.14.6) (Seemann, 2014). GrapeTree analysis was performed using the Achtman 7 Gene MLST scheme (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, *recA*) and the MSTree V2 algorithm. Accession numbers for genome assemblies are found in Supplementary Dataset 1 and are publicly available at Enterobase.[1]

Core and accessory genome determination, alignment, phylogenetic trees, and pan-genome analysis

PIRATE (version 1.0.4 with default parameters) was used to perform a pan-genome analysis on all the GFF annotation files for all MAEC genomes and the results were then converted to the ROARY gene presence/absence format. The SNPs of all genes that comprised the core genome were concatenated in the same linear order prior to reconstructing a maximum-likelihood phylogenetic tree using IQ-Tree, and the subsequent tree was visualized using Interactive Tree of Life (Letunic and Bork, 2019; Minh et al., 2013; Nguyen et al., 2015). For the IQ-Tree phylogenetic reconstruction, ModelFinder (Kalyaanamoorthy et al., 2017) was employed to select the generalized time reversible model (GTR + R10), and an ultrafast bootstrap approximation (UFboot) with 1,000 bootstrap replicates was used. The PIRATE output was then input into the SCOARY (version 1.6.16 with default parameters) tool (Brynildsrud et al., 2016), which predicts clusters of orthologous genes across the core and accessory genome. Specifically, SCOARY identified members of the accessory genome that had a statistically significant association between gene presence and the tested phenotype. For the severity analysis, we selected 90 isolates belonging to mastitis scores 1 or 3 (44 mild, 46 severe). SCOARY was also used to identify genes in the pan-genome associated with competitive fitness in milk and in mouse MGs (using barcoded strains, see below). For this analysis, the top and bottom 30% of strains for each condition were separated based on their competition index (CI) values (regardless of whether they came from mild or severe CM cases). To identify genes associated with pathogenic or commensal strains, 220 genomes for bovine commensal strains were downloaded from NCBI using "bos taurus commensal" with the *E. coli* species tag. Similarly, 188 MAEC genomes were downloaded from NCBI using "bovine mastitis" with the *E. coli* species tag or from (Alawneh et al., 2020) using the python downloading programs at: https://github.com/SomeoneNamedCaz/E.-Coli-genome-analysis. SCOARY input files consisted of a gene absence/presence Rtab file generated by PIRATE and a custom trait file which assigned a discrete phenotype for each strain.

Hierarchical clustering analysis

The MD Anderson Cancer Center Next-Generation Clustered Heat Map (NG-CHM) builder was used for hierarchical clustering[2] based on the ExPEC virulence gene carriage. The Euclidian distance and single-linkage methods were used to generate the clusters.

Detection of plasmids, antimicrobial resistance (AMR), and virulence genes

The PlasmidFinder 2.1 database at the Center for Genomic Epidemiology[3] was used to detect and type plasmids found in the MAEC strains in this study (Supplementary Dataset 1). Assembled reads for each strain were searched using the most recent Enterobacteriaceae plasmid database using 90% minimum identity and 60% minimum length coverage cutoffs. Incompatibility (Inc) groups that were found in each strain were recorded. Each Inc. group was counted individually even when multiple Inc. groups were detected in a single strain. ABRicate software[4] was used to find genes implicated in AMR from the Resfinder database (Zankari et al., 2012). Virulence genes were identified using the VirulenceFinder 2.0

1 https://enterobase.warwick.ac.uk

2 https://build.ngchm.net/NGCHM-web-builder/

3 http://genepi.food.dtu.dk/resfinder

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

program[5] using 90% minimum identify and 60% minimum length cutoffs.

Development of barcoded plasmids and barcode sequencing

The low copy pACYC184 plasmid was modified by designing PCR primers incorporating a partial Illumina adapter flanking 12 random nucleotides and eliminating the tetracycline resistance gene, yielding a 2,102 bp product (Supplementary Figure 1). The left primer incorporated the random sequences and the first Illumina Truseq adapter, and both primers included SalI overhang sequences. The PCR product was digested with SalI, ligated and transformed into chemically competent *E. coli* DH5α. Ninety-six unique plasmids were isolated, and their barcode sequence determined by Sanger sequencing. These plasmids were then transformed into individual MAEC isolates. Four strains were unable to be transformed because of pre-existing chloramphenicol resistance, leaving 92 strains that were successfully transformed. An inoculum was prepared by growing each strain individually in LB to an $A_{600} = 1.0$, mixing 20 µL of each strain, together, and diluting the mixture to a final concentration of 5×10^6 CFU/ml in PBS. This inoculum was frozen for use in competition tests and was sequenced as the input library.

The input library was grown in duplicate in LB broth and in whole unpasteurized cow milk (pooled from healthy cows from a local commercial supplier) for 8 h. The input library was also injected into lactating mouse MGs (see below). After growth in LB, milk and MG infections, plasmids were isolated from the bacterial population in each sample. The barcodes were amplified using primers that also added the remainder of the Illumina adapter as well as sample-specific identifying sequence. Illumina paired-end reads of 150 bp were generated on MiSeq version 2 sequencer by Genewiz, Inc. (South Plainfield, NJ) and CD-Genomics, Inc. (Shirly, NY). A custom grep function was used to identify and count barcodes for each strain from the sequence reads. Fitness scores were calculated as the number of reads for each barcode in an output sample as a proportion of the total reads (all barcodes) in that sample, divided by the ratio of that same barcode to the total reads in the inoculum (input) library.

Mouse infections

Lactating CD-1 IGS mice between 9 and 12 weeks of age and 10 to 11 days postpartum were infected as previously described (Olson et al., 2018). Protocol 16–0302 was reviewed and approved by the Institutional Animal Care and Use Committee of Brigham Young University. Briefly, a 50 µL volume of bacteria containing 500 CFU of each strain for total of ~50,000 CFU was suspended in phosphate-buffered saline (PBS). The inoculum was injected directly through the teat canal into the ductal network of the 4th left and 4th right MGs of five individual mice using a 33-gage needle with a beveled end. Pups were removed for 1–2 h after injections and then reunited with the mother and allowed to nurse normally. Mice were euthanized 24 h post infection and each individual gland was separately homogenized in 1 mL PBS. The tissue homogenate was added to LB broth containing chloramphenicol (10 µg/mL) to recover the bacteria and isolate plasmid DNA. Sequencing of each MG sample (as described above) was successful for nine of the 10 individual glands. The barcode frequencies in each gland ($n = 9$) were determined and used to determine competitive fitness of each strain.

Deletion and complementation of *chiA*

An allelic exchange plasmid was created using the pAX1 plasmid (Wiles et al., 2018). Upstream and downstream regions (500 bp) of *chiA* in strain M45 were amplified, stitched together using overlap extension, and inserted into the SalI and AvrII sites of the pAX1 plasmid. The resulting plasmid was transformed via electroporation into the donor *E. coli* strain MFDλpir. Integration of the suicide plasmid to create merodiploid recipients and excision to produce unmarked deletions of *chiA* in each of the strains was performed as described (Wiles et al., 2018). Complementation of *chiA* gene was done by amplifying *chiA* from strain M45 including 300 bp upstream to include the putative promoter and ligating the resulting product into pJET1.2. The resulting plasmid pWH01 was verified by sequencing and transformed via electroporation into each deletion mutant.

MAC-T cell culture and media

Bovine mammary alveolar epithelial cells (MAC-T cells) were generously provided by Dr. Janos Zempleni (University of Nebraska-Lincoln). They were grown in T-75 flasks with 40% (v/v) Dulbecco's Modified Eagle Medium (DMEM), 40% (v/v) Ham's F12 Medium, and 10% (v/v) bovine serum (FBS). FBS was heat-inactivated prior to use in media by incubating in a 56°C water bath for 30 min with periodical mixing. This was supplemented with 5 µg/mL bovine insulin, 1 µg/mL hydrocortisone, 23 mM HEPES buffer, 2.2 g/L sodium bicarbonate, and 40 mM L-glutamine. Penicillin, streptomycin (100 U/mL and 100 µg/mL, respectively) and amphotericin B (2.5 µg/mL) were added into the media for routine growth. MAC-T cells were grown at 37°C and 5% CO_2 (v/v).

Adhesion assays

Adhesion was measured as previously described (Almeida et al., 1996) with modifications. MAC-T cells were seeded into a 12-well plate and grown to ≥95% confluency. The density of epithelial cells was determined by trypan blue exclusion and counting using Cell Counter model R1 automated cell counter (Olympus). Approximately 6×10^5 cells were present in each well. The number of cells and viability did not vary throughout the assays as determined by trypan blue exclusion. On the day of each assay, spent media was removed from each well and the cells were washed 3 times with 1 mL of sterile PBS to remove residual antibiotics from media. Media without antibiotics was added into each well and cells were allowed to incubate for ≥2 h prior to inoculation with bacteria.

Overnight cultures of bacteria were diluted in sterile PBS to an OD_{600} of 0.5. Bacteria were added to a multiplicity of infection (MOI)

5 https://cge.food.dtu.dk/services/VirulenceFinder

of 10:1. The 12-well plates were then centrifuged at 300 x g for 5 min at room temperature to synchronize contact of bacteria with the MAC-T cells. Plates were then incubated at 37°C and 5% CO_2 (v/v) for 1 h. Following incubation, media was aspirated and wells washed 3x with 1 mL of sterile PBS to dislodge non-and weakly adhered bacteria. Triton X-100 (500 µL, 0.1%) in PBS was then added into each well and incubated at room temperature for 5 min to lyse cells. The resulting suspension of bacteria was then homogenized, serially diluted and plated on LB agar overnight and CFUs were counted.

Statistical analyses

Statistical analyses were carried out using Prism9 (GraphPad) or SCOARY. A $p < 0.05$ was considered statistically significant. For genome-wide association (GWAS) studies, Fisher's exact test was used to determine significance for each gene. Associations of the Inc. group with a strain type (mild or severe CM) were assessed using Fisher's exact test. Distribution of AMR genes between mild and severe CM isolates was compared using the Mann–Whitney *T*-test. Correlations between fitness scores were determined by Spearman rank correlation analysis or Mann–Whitney *T*-test. Differences in adherence to MAC-T cells by wild-type, *chiA* mutant, or complemented strains were analyzed using a one-way ANOVA.

Results

Bacterial growth in milk is not associated with clinical mastitis severity

The ability to efficiently utilize the nutritional components present in milk and thus grow rapidly may be a determining factor for some strains ability to colonize MGs successfully. To determine if there is a link between *in vitro* growth in milk and CM severity, we compared the growth of a subset of mild and severe MAEC strains in whole unpasteurized cow's milk (Supplementary Figure 2). Replication was measured after 4 or 8 h of growth. MAEC strains varied widely in their growth yields, with some strains replicating 100-fold within 4 h and up to 100,000-fold by 8 h, whereas others did not increase. However, replication was not different between strains isolated from mild and severe CM at either time point.

Genome analysis of mild and severe clinical mastitis isolates

Complete genome sequences for 90 MAEC strains (46 severe and 44 mild CM isolates) were assembled and annotated (Supplementary Dataset 1). From these annotated sequences, we identified a total of 15,395 unique genes representing the pan-genome. Of these, 3,177 (20.6%) were considered core genes (>98% prevalence), 263 (1.7%) were considered soft core genes (95–98% prevalence), 1,523 were considered shell genes (15–95% prevalence), and 10,432 (67.8%) were considered cloud genes (<15% prevalence) (Supplementary Table 1).

In silico phylogroup analysis demonstrated that most strains belong to phylogroups A and B1. Multi-locus sequence typing based on seven house-keeping genes was also used to more precisely assess the phylogenetic backgrounds of the MAEC strains (Supplementary Dataset 1). A total of 30 different STs were identified, with many STs being represented by only one strain each. In our study, ST10, ST58, ST1121 and ST1125 were the most abundant, representing 45% of the total strains (Figure 1A). MLST groupings did not correlate with mild or severe CM.

To explore whether CM severity can be predominantly attributed to differences in the core genome, a phylogenetic tree was constructed based on the core genome sequence alignments. In this analysis, several ExPEC strain, one enterohemorrhagic strain, and additional MAEC strains without clinical severity data were included as references (Figure 1B). The phylogenetic tree reveals the high diversity of MAEC strains. Compared to the relatively tight clustering of other ExPEC, the MAEC strains belong to a broad range of backgrounds, including some closely related to other ExPEC strains. Of 10 strains that clustered closely with other ExPEC strains, nine were isolated from severe CM and one from mild CM. However, there was no consistent clustering into clades based on CM severity. There was also no association between the core genome and the geographical region where the strains were isolated (Supplementary Figure 3).

ExPEC are difficult to define based on any single or group of virulence genes. However, they often carry genes related to iron acquisition (yersiniabactin *ybtP*, aerobactin *iutA*, salmochelin *iroN* siderophores, ferric citrate *fecA*, and *sit* ferrous iron transport), group 2 or 3 capsules (*kps*), adhesive and invasive factors [P fimbriae; *papC,* S fimbriae; *sfaA*, *focC*, afimbiral adhesins (*afaD*), and toxins (*sat*, *hlyA*, *cdtA*)] (Johnson and Stell, 2000; Johnson et al., 2003; Royer et al., 2023; Welch, 2016; Sarowska et al., 2019). These genes were detected in several of the MAEC genomes (Figure 2). Mean number of virulence genes carried by each strain was 3.32 ± 2.15 or 3.17 ± 2.23 for mild and severe isolates, respectively, demonstrating that simple abundance of ExPEC virulence genes is not predictive of CM severity. For example, strain M93 (mild isolate) carries 10 of these ExPEC virulence genes. As CM severity could be associated with specific combinations of virulence genes, hierarchical clustering was then performed based on the presence or absence of each gene (Figure 2). This analysis demonstrated no apparent clustering of strains based on CM severity and ExPEC virulence gene content.

Plasmids contribute to bacterial diversity such as that demonstrated by our study population, in part because they can contribute to homologous or non-homologous recombination. Plasmids also frequently carry specific fitness or AMR genes, which could influence CM severity. Incompatibility (Inc) typing was performed for each of the MAEC strains, and 22 different Inc. types were detected, suggesting a wide diversity of plasmid content (Figure 3 and Supplementary Dataset 1). Multiple Inc. types were frequently detected in the same strain, suggesting the possibility of hybrid plasmids. When compared to the severe CM strains, more Inc. types were detected among the mild CM strains and were more frequent overall, whereas 15.2% of severe CM isolates and 4.5% of mild CM isolates did not carry any plasmids (Figure 3). IncF1B (APO1918) was the most abundant Inc. type and was detected in the majority of mild and severe CM strains (65.9 and 56.5% respectively).

Carriage of AMR genes among mild and severe CM strains was also assessed *in silico* by searching a repository of AMR genes. This analysis

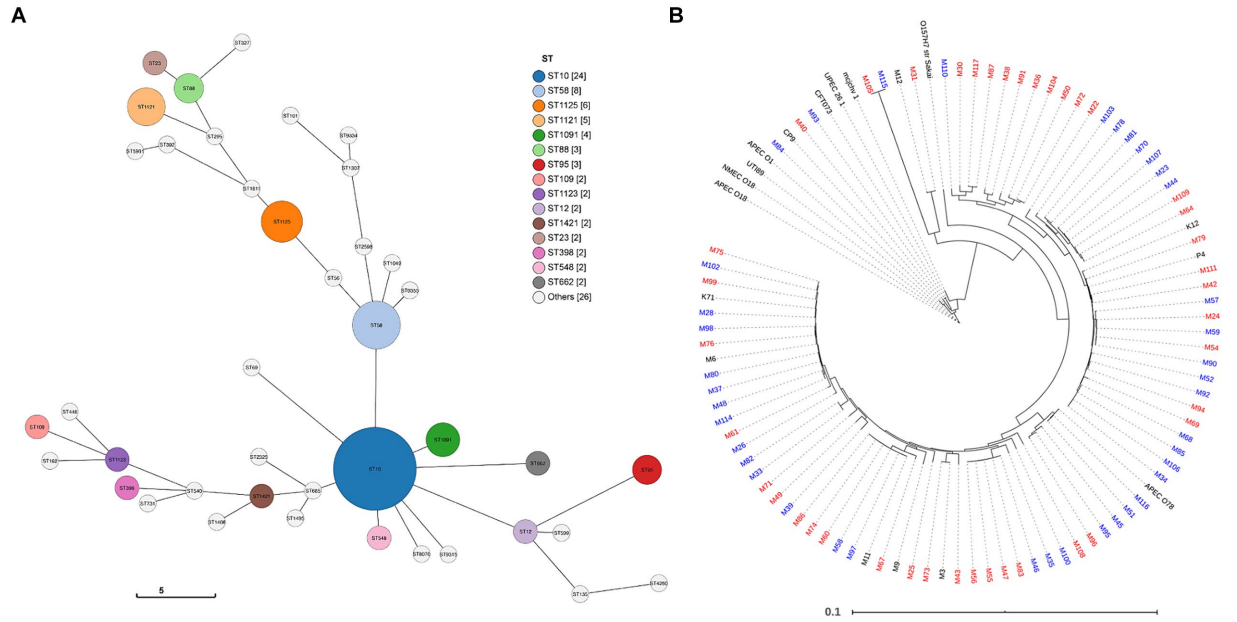


FIGURE 1
Genome comparison of MAEC strains. **(A)** Sequence type (ST) prediction using the GrapeTree function with the MSTree V2 algorithm to visualize strain relatedness (Zhou et al., 2020; Zhou et al., 2018). Overall, 30 unique STs were detected. The legend displays STs represented more than once in the population (node diameter scaled to frequency of detection). **(B)** Core genome relatedness among strains isolated from severe mastitis or mild mastitis. Strains isolated from mild CM cases are denoted in blue and severe CM in red. Strains labeled black were included as references or did not have clinical severity data and were not part of the analysis.

demonstrated that 12 strains carried one or more genes conferring resistance to aminoglycosides, beta-lactams, anti-folates, macrolides, phenicols, and tetracyclines (Supplementary Table 2). Aminoglycoside and anti-folate resistance genes were the most abundant. Notably, two severe CM isolates (M79 and M96) had nine AMR genes each. However, the distribution of AMR genes was not significantly different between mild and severe CM isolates ($p = 0.6029$).

Genes associated with clinical mastitis severity

Next, we analyzed each gene in the pan-genome and scored it according to the association with mild or severe CM phenotypes. Each gene with an apparent association with the phenotype was reanalyzed, incorporating information about the phylogenetic structure to implicate genes associated with CM severity. Using a *p*-value cutoff of 0.05, we detected 25 genes that were associated with severe CM and 79 genes associated with mild CM (Supplementary Dataset 1 and Table 1). Among those associated with severe CM were two genes likely involved in O-antigen or capsule biosynthesis (*wbpI* and *capD*) and several genes involved in producing Yad fimbriae. These genes were detected in eight strains isolated from severe CM cases and were not present in any strains from mild CM cases. Genes encoding fimbrial protein subunits YadM, YadL, YadK and YadN as well as the fimbrial usher HtrE and chaperone YadV were all positively associated with severe CM.

Most genes associated with mild CM were hypothetical genes. However, genes for producing the biofilm PGA exopolysaccharide and several tyrosine recombinases were associated with mild CM. The complete *pgaABCD* operon was detected in nearly all (39 of 40) strains with a mild CM score whereas just *pgaD* was missing in eight strains from severe CM. Three tyrosine recombinase alleles (Blum and Leitner, 2013; Goldstone et al., 2016; Dufour et al., 2019) were associated with mild CM strains, while distinct alleles (Fuenzalida and Ruegg, 2019; Lippolis et al., 2016; Cai et al., 1994) were associated with severe CM strains (Table 1).

Genes associated with mastitis vs. commensal strains

We conducted a complementary analysis of a larger publicly available set of *E. coli* genomes to identify genes that may distinguish commensal from mastitis-causing strains. These included strains isolated from the gastrointestinal tracts of cattle, which were designated as commensals. Likewise, publicly available genome sequences for all strains with an identifiable mastitis designation as well as strains from this study or published by Alawneh et al. (2020) were designated as pathogens. As expected, genes within the ferric dicitrate receptor (*fec*) operon were among the most strongly associated with the mastitis isolates. The *gadB* gene encoding the glutamate decarboxylase B enzyme was more often absent from the mastitis isolates (detected in 30.3% genomes) whereas the commensal strains typically contained *gadB* (84.5%). Genes encoding the accessory type II secretion system (*gsp*) and the linked *chiA* gene were also strongly associated with disease-causing strains (Table 2).

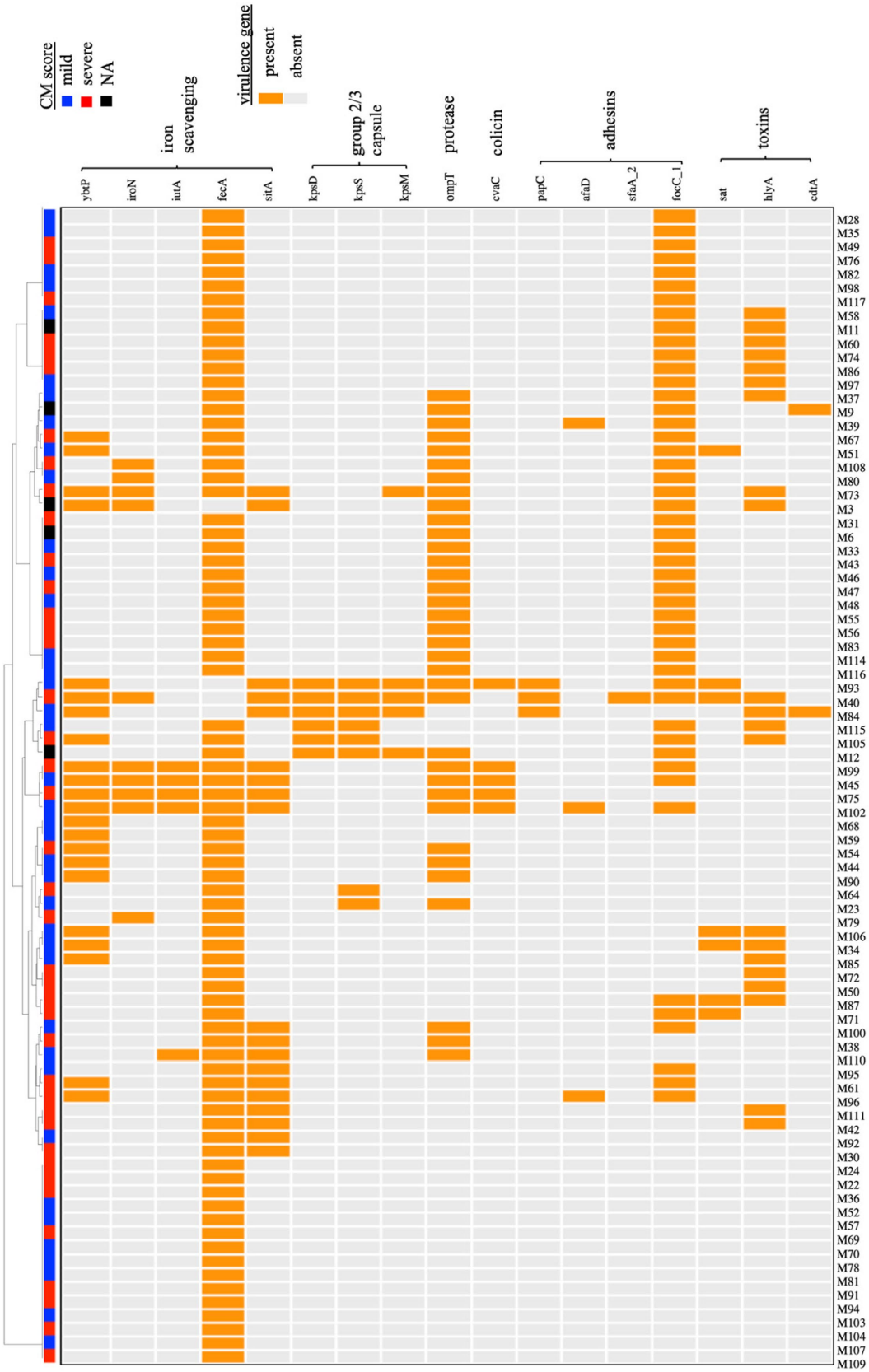


FIGURE 2
Hierarchical clustering of MAEC strains based on the carriage of virulence genes associated with the ExPEC phenotype. A dendogram was built based on the presence or absence of each gene. Both mild and severe CM isolates were present in each clade, and no relationship between CM severity and particular combinations of virulence genes was detected. NA-CM severity score not available.

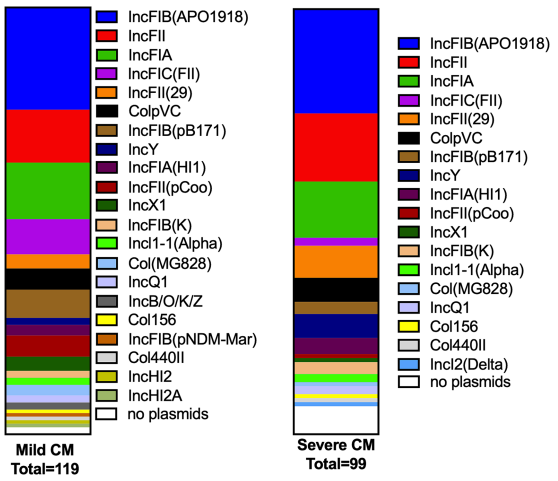


FIGURE 3
Plasmid replicons (Inc types) found in mild and severe CM strains. The number of plasmid replicon genes detected in mild and severe strains is presented as a proportion of the total for each group. More unique replication genes were detected in the mild CM strains compared with severe CM strains, whereas more of the severe CM strains contained no detectable Inc. types than the mild CM strains.

Genes associated with fitness in milk and mammary glands

Next, we tested whether strains isolated from severe or mild CM have different fitness levels within the MG environment. We competed 92 MAEC strains against each other and used GWAS to identify genes associated with fitness. In these assays, pools of barcoded MAEC strains were grown in LB media, whole unpasteurized milk, and mouse MGs. Seventy-three bacterial strains were successfully recovered (sequencing failed for 19 barcodes and these strains were excluded from the analysis). From these 73 remaining strains, the abundance of each barcode was used to calculate the competition index (CI) as a measurement of fitness for each strain (Supplementary Dataset 1).

The CI scores that were calculated for each strain in duplicate LB and milk samples were highly consistent with each other ($R^2 = 0.99$ and 0.94 respectively). The replicate samples for the nine MG infections were much more variable than in LB or milk. However, several strains consistently outcompeted others in mouse MGs (Supplementary Figure 4). For instance, strain M65 exhibited a mean CI of 225. The mean CI of each strain during growth in LB and milk, (Figure 4A), milk and MGs (Figure 4B), and LB and MGs (Figure 4C) were positively correlated. Correlation between CI milk and MGs was slightly stronger ($r = 0.53$) than between LB and MGs ($r = 0.38$). No association between the diagnosed CM severity (mild vs. severe) and fitness in mouse MGs was evident (Figure 4D).

Genes in the pan-genome that are associated with competitive fitness in milk and in mouse MGs were then identified. For this analysis, the top and bottom 30% of strains for each condition were separated based on their CI scores. Only five genes were positively associated with higher growth in milk, including a predicted inner membrane protein *yjeO,* three genes with unknown function, and an IS3 family transposase, *insK*. In mouse MGs, 38 genes were positively associated with increased competitive fitness (Supplementary Dataset 1). These included the bacterioferritin (*bfr*) and bacterioferritin-associated ferredoxin (*bfd*) genes that are involved in intracellular iron storage and mobilization. Notably, 14 of these genes are involved in the accessory type 2 secretion system found in two operons (*gspC-O* and *gspAB*) as well as the chitinase encoded by *chiA* (Table 3). The type 2 secretion system and *chiA* were also more frequently associated with mastitis-causing strains than with commensals (Table 2).

These associations suggested that ChiA might promote fitness of pathogenic strains. As ChiA proteins are known to enhance adherence of other bacteria to host cells or tissues, the role of ChiA in adhesion to cultured mammary alveolar epithelial (MAC-T) cells was investigated. We chose to investigate three strains that were in our original study population that possess *chiA* (M45, M93, M111) as well as one additional strain that was not part of the original analysis (G1). M45 was among the top 30% most fit strains in MGs. M45 and M93 were isolated from mild cases of CM while M111 was isolated from a severe case. Strain G1 was isolated from a case of severe, gangrenous mastitis.

Wild type, *chiA* deletion mutants, or their complemented strains were added to MAC-T cells (MOI = 10) for 1 h and then the proportion of adherent cells measured. M45Δ*chiA* demonstrated an approximately 2-fold reduction compared to the wild-type parent (3.1×10^5 vs. 7.2×10^5 CFUs, respectively) in attachment (Figure 5). Attachment was restored upon reintroduction of *chiA* by plasmid complementation. A greater reduction was observed in M93 (>6-fold) between the wild type and mutant (Figure 5), which was also able to be complemented. The slight reduction in attachment of M111Δ*chiA* compared to the wild-type strain was not statistically significant ($p = 0.06$) and complementation did not change adherence of this strain (Figure 5). Deletion of *chiA* in G1 caused the largest decrease in attachment of all the MAEC strains we investigated. G1Δ*chiA* presented almost a 10-fold decrease when compared to the wild type as shown in Figure 5. Wild-type levels of adherence were restored upon plasmid complementation in G1.

Discussion

The features of pathogenic *E. coli* that differentiate them from non-pathogens remain incompletely understood, as is the relationship between bacterial fitness and the clinical disease that occurs during infection. Several recent studies have employed genome-wide association tools paired with clinical or experimental data to identify accessory genes associated with bacterial virulence, niche adaptation, AMR, and environmental persistence (Bazinet, 2017; Maury et al., 2019; Gouliouris et al., 2018; Zong et al., 2018; He et al., 2018; Fritsch et al., 2019; Gori et al., 2020). In this study we sequenced 96 MAEC genomes, with the goal of identifying bacterial genes associated with differences in manifestation of bovine CM. We also identified genes associated with commensal strains in comparison with clinical bovine mastitis isolates. We developed a barcoding system where multiple strains can be tracked as they compete in different conditions. This allowed us to quantify differences in fitness in specific controlled conditions and use these differences to identify genes associated with high and low-fitness strains. This approach enabled the identification of novel genes that may influence bacterial growth in multiple host environments. Our study differs from most bacterial GWAS in that it

10.3389/fmicb.2024.1452007

TABLE 1 Top genes* associated with MAEC isolated from mild or severe clinical mastitis (CM).

		Proportion of strains		
Gene name	Annotation	Severe CM	Mild CM	*P*-value
Mild				
xerC_5	Tyrosine recombinase XerC	0.429	0.756	2.96E-03
xapA	Purine nucleoside phosphorylase 2	0.667	0.902	1.05E-02
xapB	Xanthosine permease	0.667	0.902	1.05E-02
pgaD	Biofilm PGA synthesis protein PgaD	0.810	0.976	1.18E-02
xerC_1	Tyrosine recombinase XerC	0.048	0.244	1.31E-02
hcaR_2	Hca operon transcriptional activator HcaR	0.690	0.902	2.01E-02
traA	Pilin	0.476	0.732	2.20E-02
Severe				
wbpI	UDP-2,3-diacetamido-2,3-dideoxy-D-glucuronate 2-epimerase	0.190	0.000	5.36E-03
capD	UDP-glucose 4-epimerase	0.190	0.000	5.36E-03
yadV	Putative fimbrial chaperone YadV	0.190	0.000	5.36E-03
yadM	Putative fimbrial-like protein YadM	0.190	0.000	5.36E-03
yadK	Putative fimbrial-like protein YadK	0.190	0.000	5.36E-03
htrE	Outer membrane usher protein HtrE	0.190	0.000	5.36E-03
cfaE	CFA/I fimbrial subunit E	0.214	0.024	1.45E-02
ygcB	CRISPR-associated endonuclease/helicase Cas3	0.286	0.073	2.01E-02
casB	CRISPR system Cascade subunit CasB	0.286	0.073	2.01E-02
hpcD	5-carboxymethyl-2-hydroxymuconate Delta-isomerase	0.500	0.244	2.20E-02
hscC_2	Chaperone protein HscC	0.143	0.000	2.57E-02
fliD	Flagellar hook-associated protein 2	0.571	0.317	2.59E-02
ygbF	CRISPR-associated endoribonuclease Cas2	0.310	0.098	2.74E-02
casE	CRISPR system Cascade subunit CasE	0.310	0.098	2.74E-02
casD	CRISPR system Cascade subunit CasD	0.310	0.098	2.74E-02
ygbT	CRISPR-associated endonuclease Cas1	0.310	0.098	2.74E-02
casC	CRISPR system Cascade subunit CasC	0.310	0.098	2.74E-02
yadL	Putative fimbrial-like protein YadL	0.190	0.024	2.91E-02
yadN	Putative fimbrial-like protein YadN	0.190	0.024	2.91E-02

*Excluding hypothetical proteins.

combines clinical and mouse infection assay data to identify a gene that was validated by functional studies.

Our data confirm the diversity of MAEC strains (Figure 1) and reveals the large pan-genome associated with these bacteria. In previous genomic studies of MAEC strains, most were classified as ST10, ST23, ST58, ST88, or ST1125 in phylogroups A or B1, which is consistent with our findings (Leimbach et al., 2017; Blum and Leitner, 2013; Nuesch-Inderbinen et al., 2019). Within the accessory genome, we identified genes encoding adhesive and extracellular matrix structures associated with strains isolated from mastitis cases diagnosed as either mild or severe (Table 1). This includes the Yad fimbriae, which were positively associated with mastitis severity, and conversely, the PGA biofilm exopolysaccharide was negatively associated with mastitis severity. Yad fimbriae are also enriched among ExPEC strains of avian and human origin and are thought to promote adherence (Ren et al., 2016; Wurpel et al., 2013; Dziva et al., 2013; Verma et al., 2016; Larsonneur et al., 2016; Elpers and Hensel, 2020). Sustained inflammation in the MG is a main trigger for severe CM and bacterial factors that promote adherence would likely increase immune detection and signaling, leading to phagocytic cell recruitment. Alternatively, PGA may lessen CM severity by promoting biofilm formation, shielding inflammatory molecules on the surface of the bacteria and/or reducing invasion into epithelial cells.

MAEC are typically excluded from discussion of ExPEC strains more broadly. ExPEC exact an outsized disease burden in both humans and animals and are difficult to distinguish reliably from other *E. coli*. We did find several instances of MAEC strains possessing genes with demonstrated roles in ExPEC virulence (Figure 2). Although these virulence factors were not associated with CM severity generally, it does not rule out the possibility that they contribute to the virulence of individual strains in MGs, as we have previously demonstrated for the capsule and zinc uptake gene clusters of strain M12 (Olson et al., 2018; Olson et al., 2021). Furthermore, additional factors continue to be recognized that contribute to ExPEC infections,

10.3389/fmicb.2024.1452007

TABLE 2 Top genes* associated with mastitis or commensal bovine *E. coli* isolates.

		Proportion of strains		
Gene name	Annotation	Commensal	Mastitis	*P*-value
Commensal				
gadB	Glutamate decarboxylase B subunit	0.846	0.303	7.84E-35
tufB	Elongation factor Tu	0.865	0.420	3.11E-25
cytR	Cytidine repressor	0.765	0.415	6.53E-15
Mastitis				
fecB	Ferric dicitrate ABC transporter - periplasmic binding protein	0.360	0.878	1.07E-31
fecD	Ferric dicitrate ABC transporter - membrane subunit	0.354	0.872	1.33E-31
fecE	Ferric dicitrate ABC transporter - ATP binding subunit	0.354	0.872	1.33E-31
fecI	RNA polymerase sigma 19 factor	0.379	0.888	2.48E-31
fecR	Regulator for fec operon periplasmic	0.379	0.888	2.48E-31
fecA	Ferric citrate outer membrane porin FecA	0.379	0.888	2.48E-31
fecC	Ferric dicitrate ABC transporter - membrane subunit	0.360	0.872	4.61E-31
yjhV	KpLE2 phage-like element predicted protein	0.344	0.835	8.14E-28
mqsA	Antitoxin of the MqsRA toxin-antitoxin system and DNA-binding transcriptional repressor	0.273	0.644	5.53E-16
yafN	Antitoxin of the YafO-YafN toxin-antitoxin system	0.280	0.649	6.72E-16
yafO	Ribosome-dependent mRNA interferase toxin	0.273	0.638	1.23E-15
mqsR	mRNA interferase toxin of the MqsR-YgiT toxin-antitoxin system	0.273	0.638	1.23E-15
ygiS	Putative transporter subunit	0.328	0.691	3.16E-15
ybfP	Putative pectinase	0.251	0.590	
gspO	Type 4 prepilin-like proteins leader peptide-processing enzyme	0.289	0.633	6.79E-14
gspK	Putative type II secretion system protein K	0.292	0.633	8.86E-14
gspA	Putative general secretion pathway protein	0.292	0.633	8.86E-14
gspF	Putative type II secretion system protein F	0.292	0.633	8.86E-14
gspH	Putative type II secretion system protein H	0.292	0.633	8.86E-14
chiA	Putative bifunctional chitinase/lysozyme	0.292	0.633	8.86E-14
gspM	Putative type II secretion system protein M	0.292	0.633	8.86E-14
gspJ	Putative type II secretion system protein J	0.292	0.633	8.86E-14

*Excluding hypothetical proteins.

and these may also be selected by the dairy environment and enriched in MAEC strains. The ferric dicitrate iron acquisition system is one such example (Frick-Cheng et al., 2022). While their role in intramammary infection has been established, they were recently demonstrated to enhance ExPEC urovirulence, suggesting that the *fec* system may be a factor in zoonotic spread of *E. coli* since it promotes fitness in multiple hosts and tissue types (Frick-Cheng et al., 2022). The presence of these classic and newly appreciated ExPEC virulence genes in MAEC strains suggests that they occasionally infect humans and cause bloodstream or urinary tract infections.

We identified many different Inc. plasmid groups in our MAEC strains and found that mild CM strains had significantly more Inc. groups than severe CM strains (Figure 3). However, we did not investigate whether any virulence or AMR genes were carried on these plasmids as is frequently the case. The majority of the plasmids we detected belong to the IncF family which are usually conjugative (Douarre et al., 2020), illustrating their potential to spread resistance in these populations. AMR is a significant health and environmental concern. However, only 12 strains carried one or more AMR genes. Two strains carried genes that confer resistance to six different classes of antimicrobials. Carriage of AMR genes did not appear to be associated with disease severity.

Unsurprisingly, our results indicate that those strains that are highly competitive *in vitro* tend to outcompete other strains *in vivo*, due to more rapid growth, direct antibacterial antagonism, or both (Figure 4). Interestingly, the correlation between competitive fitness in milk with MG infections than was slightly stronger than the correlation between fitness in LB and MG, suggesting that the ability to utilize nutrients or resist antimicrobial substances found in milk contributes to growth in lactating MGs. However, the lack of correlation we observed between bacterial fitness and CM severity (Figure 4D) illustrates that successful pathogens may replicate to high numbers without triggering deleterious responses in the host. Similarly, we observed no relationship between *in vitro* growth rates of individual strains in milk with CM severity (Supplementary Figure 2), which has also been reported by other researchers (Kornalijnslijper et al., 2004).

The *chiA* gene encoding a putative chitinase/chitin-binding protein was enriched in MAEC with higher fitness during mouse

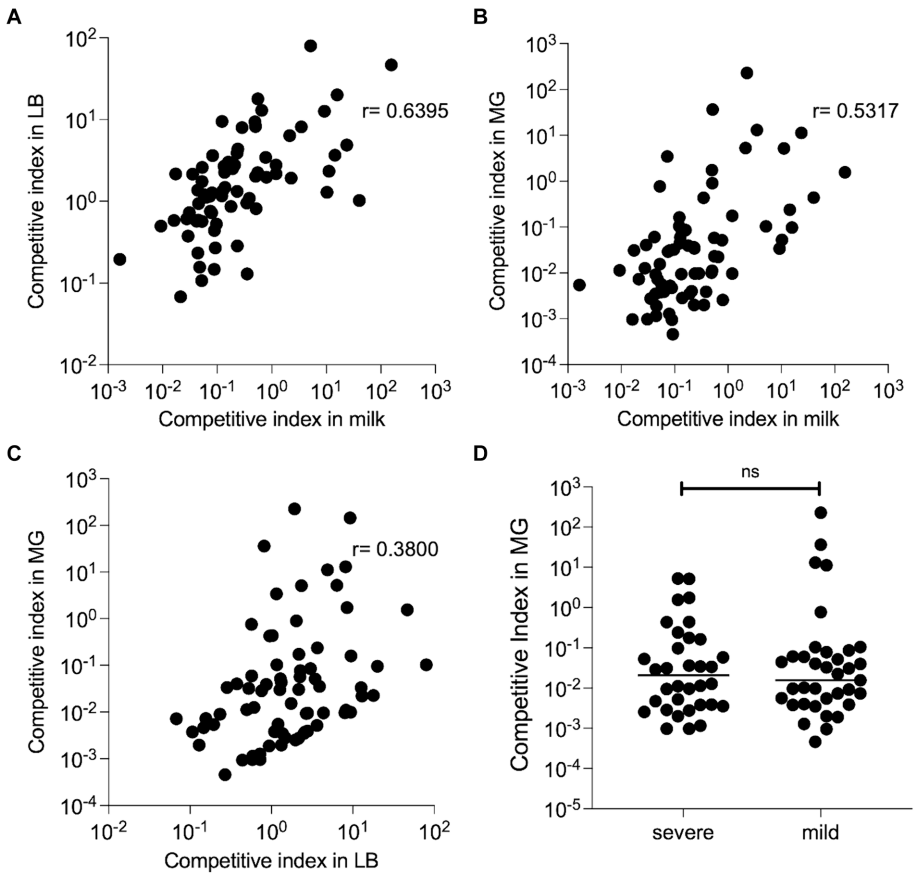


FIGURE 4
Competitive fitness of barcoded MAEC strains. All strains were inoculated together and grown in LB or unpasteurized cow's milk (in duplicate) for 8 h, or in nine lactating MGs for 24 h. The bacteria were recovered, and their barcode plasmids were sequenced to determine their competitive indexes. **(A)** CI during *in vitro* growth in LB is strongly correlated with CI during growth in milk ($p < 0.0001$ by Spearman rank correlation test). Each dot represents a single MAEC strain plotted at the mean competition index in both environments. **(B)** A slightly weaker positive correlation was detected between CI in milk and MGs ($p < 0.0001$) and in **(C)** LB and MGs ($p = 0.0008$) by Spearman rank correlation test. **(D)** CI in mouse MG infections for MAEC strains isolated from mild or severe CM cases. No significant difference was detected in these groups Student's *t*-test with Mann–Whitney correction ($p = 0.904$).

MG infections. This gene, along with the type II secretion system linked with it, was also associated with pathogenicity in the larger cohort of bovine strains. However, they were not associated with CM severity. In the non-pathogenic K12 strain, transcription of this type II secretion system is normally repressed by the Hns protein, and the full *gsp* locus is needed for proper *chiA* secretion (Francetic et al., 2000). ChiA has been implicated in the virulence of some adherent/invasive *E. coli* strains that cause colitis. This is not due to their chitinolytic activity, but rather because of chitin-binding domains that are found in the N-terminus. These domains mediate binding to chitinase-3-like-1 (CHI3L1), which is expressed on the surface of intestinal epithelial cells, leading to subsequent invasion of the bacteria (Low et al., 2013; Mizoguchi, 2006). Interestingly, expression of host chitin-like proteins is also induced by some bacterial pathogens (Lee et al., 2011). CHI3L1 regulates innate immune defenses against *Streptococcus pneumoniae* and *Pseudomonas aeruginosa* lung infections through inhibition of caspase-1-dependent macrophage pyroptosis (Dela Cruz et al., 2012; Marion et al., 2016). Conversely, CHI3L1 expressed by intestinal epithelial cells during inflammatory bowel disease helps facilitate enteric bacterial infection.

Recently, CHI3L1 was also found in the milk secretions of quarters with bovine coliform mastitis (Breyne et al., 2018). CHI3L1 gene expression is also increased in mouse MGs following *E. coli* infection. In knockout mice lacking CHI3L1, bacterial growth is not affected, but the influx of neutrophils into the lumen of the infected gland is reduced (Breyne et al., 2018). It also promotes increased proliferation of mammary epithelial cells and reduces apoptosis (Anand et al., 2016). In the absence of CHI3L1, migration, maturation, and activation of macrophages is significantly impaired (Hughes et al., 2012; He et al., 2013). It seems likely that some MAEC strains may bind to CHI3L1 via ChiA, which could promote bacterial attachment and invasion and suppression of its inflammatory functions. In this way, bacterial fitness may be increased while limiting disease severity.

Alternatively, it is possible that ChiA contributes to infection through glycosidase activity, independently of its binding to host proteins. *Salmonella* chitinases modify glycans present in the extracellular matrix, uncovering mannose residues present on the epithelial surface and making them available for attachment through type I fimbriae (Chandra et al., 2022; Devlin et al., 2022). They also increase survival inside phagocytes by dampening the expression of host antimicrobial responses in dendritic cells and macrophages

10.3389/fmicb.2024.1452007

TABLE 3 Top genes* associated with MAEC fitness in mouse MG infections.

		Proportion of strains		
		Most fit	Least fit	*P*-value
rpiB	Ribose-5-phosphate isomerase B	0.812	0.313	3.85E-03
bfd	Bacterioferritin-associated ferredoxin	0.812	0.375	2.90E-02
bfr	Bacterioferritin	0.812	0.375	2.90E-02
gspB	Putative general secretion pathway protein B	0.812	0.375	2.90E-02
gspC	Putative type II secretion system protein C	0.812	0.375	2.90E-02
gspA	Putative general secretion pathway protein A	0.812	0.375	2.90E-02
gspF	Putative type II secretion system protein F	0.812	0.375	2.90E-02
gspG	Putative type II secretion system protein G	0.812	0.375	2.90E-02
gspD	Putative secretin GspD	0.812	0.375	2.90E-02
gspE	Putative type II secretion system protein E	0.812	0.375	2.90E-02
gspJ	Putative type II secretion system protein J	0.812	0.375	2.90E-02
gspH	Putative type II secretion system protein H	0.812	0.375	2.90E-02
gspI	Putative type II secretion system protein I	0.812	0.375	2.90E-02
gspO	Type 4 prepilin-like proteins leader peptide-processing enzyme	0.812	0.375	2.90E-02
gspL	Putative type II secretion system protein L	0.812	0.375	2.90E-02
gspM	Putative type II secretion system protein M	0.812	0.375	2.90E-02
gspK	Putative type II secretion system protein K	0.812	0.375	2.90E-02
chiA	putative bifunctional chitinase/lysozyme	0.812	0.375	2.90E-02
alsB	D-allose-binding periplasmic protein	0.812	0.375	2.90E-02

*Excluding hypothetical proteins.

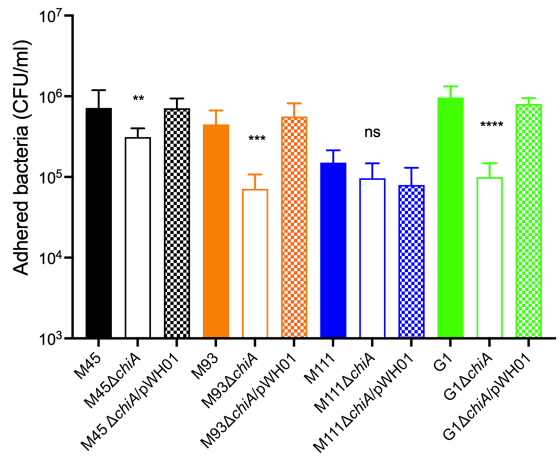


FIGURE 5
Role of ChiA in adherence of MAEC to mammary epithelial cells. Wild-type MAEC strains, their isogenic Δ*chiA* mutants, or the complemented mutant strains were tested for their ability to adhere to MAC-T cells at an MOI = 10. One-way ANOVA with Tukey's correction was used to determine significant differences between the wild-type and each Δ*chiA* mutant (***p* = 0.0082, ****p* = 0.0003, *****p* < 0.0001). These results are a representative experiment that was performed twice with six replicates per strain.

(Chandra et al., 2022). ChiA may play similar roles in MAEC colonization of MGs.

Our study has several limitations. First, the barcoding plasmid that we used to track mixed populations of bacteria may not behave identically in each strain's unique genetic background. Although the plasmid has only one coding sequence for chloramphenicol resistance and is unlikely to affect gene expression broadly, the presence of the plasmid may interfere with stability of other native plasmids, which we have not examined. This may influence the fitness of these bacteria in unexpected ways. Secondly, competitive fitness of MAEC strains may be most relevant in natural environments in the presence of many other bacterial species other than *E. coli*. The ability to outcompete other *E. coli* strains may be less critical for MAEC than their ability to defend themselves against other diverse bacteria in the cattle GI tract, in soil, or ascending the MG teat canal. Future work should test what genes contribute to fitness in these environments and whether they can explain why some MAEC strains are more common, particularly those STs that we and others have identified. Finally, the results of our study may have been influenced by random factors such as when CM was diagnosed and scored. For example, a case that was detected early may have been diagnosed as mild whereas if diagnosed a few hours later may have been scored as moderate or severe. Diagnosis of CM is also inherently subjective, and the criteria may have been interpreted differently by individual farmers.

The wide range of severity and clinical presentation of MG infections by MAEC is impressive, and more studies directly characterizing the putative virulence factors of these strains are required. In this study, we have identified new accessory genes that could play a role in host specificity of these bacteria and influence disease outcomes. The role of ChiA in colonizing MGs as well as its functions in other ExPEC strains deserves further study, including its potential role in immune suppression and interaction with host structures. The ability of some MAEC strains to cross host species

barriers and colonize different tissues underscores the importance of better understanding the diversity among this group of bacteria.

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 animal study was approved by the Institutional Animal Care and Use Committee, Brigham Young University. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

MO: Conceptualization, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. CC: Data curation, Formal analysis, Investigation, Software, Writing – review & editing. WH: Investigation, Methodology, Writing – original draft. AG: Investigation, Writing – review & editing. DN: Data curation, Investigation, Writing – review & editing. JB: Conceptualization, Investigation, Methodology, Resources, Writing – review & editing. HB: Data curation, Resources, Supervision, Writing – review & editing. EW: Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Writing – review & editing. BP: Conceptualization, Investigation, Software, Writing – review & editing. DE: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing.

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 the NIH R15AI159847-01A1 (DE).

Acknowledgments

We thank Paolo Moroni and Jennifer Wilson for providing MAEC isolates and Janos Zempleni for providing MAC-T cells.

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.2024.1452007/full#supplementary-material

References

Alawneh, J. I., Vezina, B., Ramay, H. R., Al-Harbi, H., James, A. S., Soust, M., et al. (2020). Survey and sequence characterization of bovine mastitis-associated *Escherichia coli* in dairy herds. *Front. Vet. Sci.* 7:582297. doi: 10.3389/fvets.2020.582297

Almeida, R. A., Luther, D. A., Kumar, S. J., Calvinho, L. F., Bronze, M. S., and Oliver, S. P. (1996). Adherence of *Streptococcus uberis* to bovine mammary epithelial cells and to extracellular matrix proteins. *Zentralbl. Veterinarmed. B* 43, 385–392. doi: 10.1111/j.1439-0450.1996.tb00330.x

Anand, V., Jaswal, S., Singh, S., Kumar, S., Jena, M. K., Verma, A. K., et al. (2016). Functional characterization of mammary gland Protein-40, a chitinase-like glycoprotein expressed during mammary gland apoptosis. *Apoptosis* 21, 209–224. doi: 10.1007/s10495-015-1196-z

Baym, M., Kryazhimskiy, S., Lieberman, T. D., Chung, H., Desai, M. M., and Kishony, R. (2015). Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One* 10:e0128036. doi: 10.1371/journal.pone.0128036

Bazinet, A. L. (2017). Pan-genome and phylogeny of *Bacillus cereus* sensu lato. *BMC Evol. Biol.* 17:176. doi: 10.1186/s12862-017-1020-1

Blum, S. E., Heller, E. D., Sela, S., Elad, D., Edery, N., and Leitner, G. (2015). Genomic and phenomic study of mammary pathogenic *Escherichia coli*. *PLoS One* 10:e0136387. doi: 10.1371/journal.pone.0136387

Blum, S. E., and Leitner, G. (2013). Genotyping and virulence factors assessment of bovine mastitis *Escherichia coli*. *Vet. Microbiol.* 163, 305–312. doi: 10.1016/j.vetmic.2012.12.037

Bradley, A. J., and Green, M. J. (2000). A study of the incidence and significance of intramammary enterobacterial infections acquired during the dry period. *J. Dairy Sci.* 83, 1957–1965. doi: 10.3168/jds.S0022-0302(00)75072-7

Breyne, K., Steenbrugge, J., Demeyere, K., Lee, C. G., Elias, J. A., Petzl, W., et al. (2018). Immunomodulation of host Chitinase 3-like 1 during a mammary pathogenic *Escherichia coli* infection. *Front. Immunol.* 9:1143. doi: 10.3389/fimmu.2018.01143

Brynildsrud, O., Bohlin, J., Scheffer, L., and Eldholm, V. (2016). Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* 17:238. doi: 10.1186/s13059-016-1108-8

Burvenich, C., Bannerman, D. D., Lippolis, J. D., Peelman, L., Nonnecke, B. J., Kehrli, M. E. Jr., et al. (2007). Cumulative physiological events influence the inflammatory response of the bovine udder to *Escherichia coli* infections during the transition period. *J. Dairy Sci.* 90, E39–E54. doi: 10.3168/jds.2006-696

Burvenich, C., Paape, M. J., Hill, A. W., Guidry, A. J., Miller, R. H., Heyneman, R., et al. (1994). Role of the neutrophil leucocyte in the local and systemic reactions during experimentally induced *E. coli* mastitis in cows immediately after calving. *Vet. Q.* 16, 45–50. doi: 10.1080/01652176.1994.9694416

Cai, T. Q., Weston, P. G., Lund, L. A., Brodie, B., Mckenna, D. J., and Wagner, W. C. (1994). Association between neutrophil functions and periparturient disorders in cows. *Am. J. Vet. Res.* 55, 934–943. doi: 10.2460/ajvr.1994.55.07.934

Chandra, K., Roy Chowdhury, A., Chatterjee, R., and Chakravortty, D. (2022). GH18 family glycoside hydrolase Chitinase a of *Salmonella* enhances virulence by facilitating invasion and modulating host immune responses. *PLoS Pathog.* 18:e1010407. doi: 10.1371/journal.ppat.1010407

Dela Cruz, C. S., Liu, W., He, C. H., Jacoby, A., Gornitzky, A., Ma, B., et al. (2012). Chitinase 3-like-1 promotes *Streptococcus pneumoniae* killing and augments host tolerance to lung antibacterial responses. *Cell Host Microbe* 12, 34–46. doi: 10.1016/j.chom.2012.05.017

Devlin, J. R., Santus, W., Mendez, J., Peng, W., Yu, A., Wang, J., et al. (2022). *Salmonella enterica* serovar typhimurium chitinases modulate the intestinal glycome and promote small intestinal invasion. *PLoS Pathog.* 18:e1010167. doi: 10.1371/journal.ppat.1010167

Douarre, P. E., Mallet, L., Radomski, N., Felten, A., and Mistou, M. Y. (2020). Analysis of COMPASS, a new comprehensive plasmid database revealed prevalence of multireplicon and extensive diversity of IncF plasmids. *Front. Microbiol.* 11:483. doi: 10.3389/fmicb.2020.00483

Dufour, S., Labrie, J., and Jacques, M. (2019). The mastitis pathogens culture collection. *Microbiol. Resour. Announc.* 8, 10–128. doi: 10.1128/MRA.00133-19

Dziva, F., Hauser, H., Connor, T. R., van Diemen, P. M., Prescott, G., Langridge, G. C., et al. (2013). Sequencing and functional annotation of avian pathogenic *Escherichia coli* serogroup O78 strains reveal the evolution of *E. coli* lineages pathogenic for poultry via distinct mechanisms. *Infect. Immun.* 81, 838–849. doi: 10.1128/IAI.00585-12

Elpers, L., and Hensel, M. (2020). Expression and functional characterization of various chaperon-usher fimbriae, Curli fimbriae, and type 4 pili of enterohemorrhagic *Escherichia coli* O157:H7 Sakai. *Front. Microbiol.* 11:378. doi: 10.3389/fmicb.2020.00378

Francetic, O., Belin, D., Badaut, C., and Pugsley, A. P. (2000). Expression of the endogenous type II secretion pathway in *Escherichia coli* leads to chitinase secretion. *EMBO J.* 19, 6697–6703. doi: 10.1093/emboj/19.24.6697

Freitag, C., Michael, G. B., Kadlec, K., Hassel, M., and Schwarz, S. (2017). Detection of plasmid-borne extended-spectrum beta-lactamase (ESBL) genes in *Escherichia coli* isolates from bovine mastitis. *Vet. Microbiol.* 200, 151–156. doi: 10.1016/j.vetmic.2016.08.010

Frick-Cheng, A. E., Sintsova, A., Smith, S. N., Pirani, A., Snitkin, E. S., and Mobley, H. L. T. (2022). Ferric citrate uptake is a virulence factor in Uropathogenic *Escherichia coli*. *MBio* 13:e0103522. doi: 10.1128/mbio.01035-22

Fritsch, L., Felten, A., Palma, F., Mariet, J. F., Radomski, N., Mistou, M. Y., et al. (2019). Insights from genome-wide approaches to identify variants associated to phenotypes at pan-genome scale: application to *L. monocytogenes*' ability to grow in cold conditions. *Int. J. Food Microbiol.* 291, 181–188. doi: 10.1016/j.ijfoodmicro.2018.11.028

Fuenzalida, M. J., and Ruegg, P. L. (2019). Negatively controlled, randomized clinical trial to evaluate intramammary treatment of nonsevere, gram-negative clinical mastitis. *J. Dairy Sci.* 102, 5438–5457. doi: 10.3168/jds.2018-16156

Goldstone, R. J., Harris, S., and Smith, D. G. (2016). Genomic content typifying a prevalent clade of bovine mastitis-associated *Escherichia coli*. *Sci. Rep.* 6:30115. doi: 10.1038/srep30115

Gori, A., Harrison, O. B., Mlia, E., Nishihara, Y., Chan, J. M., Msefula, J., et al. (2020). Pan-GWAS of *Streptococcus agalactiae* highlights lineage-specific genes associated with virulence and niche. *ADA Forecast* 11, 10–1128. doi: 10.1128/mBio.00728-20

Gouliouris, T., Raven, K. E., Ludden, C., Blane, B., Corander, J., Horner, C. S., et al. (2018). Genomic surveillance of *Enterococcus faecium* reveals limited sharing of strains and resistance genes between livestock and humans in the United Kingdom. *mBio* 9, 10–1128. doi: 10.1128/mBio.01780-18

Guerra, S. T., Orsi, H., Joaquim, S. F., Guimaraes, F. F., Lopes, B. C., Dalanezi, F. M., et al. (2020). Short communication: investigation of extra-intestinal pathogenic *Escherichia coli* virulence genes, bacterial motility, and multidrug resistance pattern of strains isolated from dairy cows with different severity scores of clinical mastitis. *J. Dairy Sci.* 103, 3606–3614. doi: 10.3168/jds.2019-17477

He, Q., Hou, Q., Wang, Y., Li, J., Li, W., Kwok, L. Y., et al. (2018). Comparative genomic analysis of *Enterococcus faecalis*: insights into their environmental adaptations. *BMC Genomics* 19:527. doi: 10.1186/s12864-018-4887-3

He, C. H., Lee, C. G., Dela Cruz, C. S., Lee, C. M., Zhou, Y., Ahangari, F., et al. (2013). Chitinase 3-like 1 regulates cellular and tissue responses via IL-13 receptor alpha2. *Cell Rep.* 4, 830–841. doi: 10.1016/j.celrep.2013.07.032

Hill, A. W. (1981). Factors influencing the outcome of *Escherichia coli* mastitis in the dairy cow. *Res. Vet. Sci.* 31, 107–112. doi: 10.1016/S0034-5288(18)32532-3

Hughes, K., Wickenden, J. A., Allen, J. E., and Watson, C. J. (2012). Conditional deletion of Stat3 in mammary epithelium impairs the acute phase response and modulates immune cell numbers during post-lactational regression. *J. Pathol.* 227, 106–117. doi: 10.1002/path.3961

Hyvonen, P., Haarahiltunen, T., Lehtolainen, T., Heikkinen, J., Isomaki, R., and Pyorala, S. (2010). Concentrations of bovine lactoferrin and citrate in milk during experimental endotoxin mastitis in early-versus late-lactating dairy cows. *J. Dairy Res.* 77, 474–480. doi: 10.1017/S0022029910000579

Ismail, Z. B., and Abutarbush, S. M. (2020). Molecular characterization of antimicrobial resistance and virulence genes of *Escherichia coli* isolates from bovine mastitis. *Vet. World* 13, 1588–1593. doi: 10.14202/vetworld.2020.1588-1593

Jamali, H., Krylova, K., and Aider, M. (2018). Identification and frequency of the associated genes with virulence and antibiotic resistance of *Escherichia coli* isolated from cow's milk presenting mastitis pathology. *Anim. Sci. J.* 89, 1701–1706. doi: 10.1111/asj.13093

Johnson, J. R., Kuskowski, M. A., Owens, K., Gajewski, A., and Winokur, P. L. (2003). Phylogenetic origin and virulence genotype in relation to resistance to fluoroquinolones and/or extended-spectrum cephalosporins and cephamycins among *Escherichia coli* isolates from animals and humans. *J. Infect. Dis.* 188, 759–768. doi: 10.1086/377455

Johnson, J. R., and Stell, A. L. (2000). Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. *J. Infect. Dis.* 181, 261–272. doi: 10.1086/315217

Jung, D., Park, S., Ruffini, J., Dussault, F., Dufour, S., and Ronholm, J. (2021). Comparative genomic analysis of *Escherichia coli* isolates from cases of bovine clinical mastitis identifies nine specific pathotype marker genes. *Microb. Genom.* 7:000597. doi: 10.1099/mgen.0.000597

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Keane, O. M. (2016). Genetic diversity, the virulence gene profile and antimicrobial resistance of clinical mastitis-associated *Escherichia coli*. *Res. Microbiol.* 167, 678–684. doi: 10.1016/j.resmic.2016.06.011

Kempf, F., Slugocki, C., Blum, S. E., Leitner, G., and Germon, P. (2016). Genomic comparative Study of bovine mastitis *Escherichia coli*. *PLoS One* 11:e0147954. doi: 10.1371/journal.pone.0147954

Kornalijnslijper, J. E., Daemen, A., van Werven, T., Niewold, T. A., Rutten, V., and Noordhuizen-Stassen, E. N. (2004). Bacterial growth during the early phase of infection determines the severity of experimental *Escherichia coli* mastitis in dairy cows. *Vet. Microbiol.* 101, 177–186. doi: 10.1016/j.vetmic.2004.04.005

Larsonneur, F., Martin, F. A., Mallet, A., Martinez-Gil, M., Semetey, V., Ghigo, J. M., et al. (2016). Functional analysis of *Escherichia coli* Yad fimbriae reveals their potential role in environmental persistence. *Environ. Microbiol.* 18, 5228–5248. doi: 10.1111/1462-2920.13559

Lee, C. G., Da Silva, C. A., Dela Cruz, C. S., Ahangari, F., Ma, B., Kang, M. J., et al. (2011). Role of chitin and chitinase/chitinase-like proteins in inflammation, tissue remodeling, and injury. *Annu. Rev. Physiol.* 73, 479–501. doi: 10.1146/annurev-physiol-012110-142250

Lehtolainen, T., Pohjanvirta, T., Pyorala, S., and Pelkonen, S. (2003). Association between virulence factors and clinical course of *Escherichia coli* mastitis. *Acta Vet. Scand.* 44, 203–205. doi: 10.1186/1751-0147-44-203

Leimbach, A., Poehlein, A., Vollmers, J., Gorlich, D., Daniel, R., and Dobrindt, U. (2017). No evidence for a bovine mastitis *Escherichia coli* pathotype. *BMC Genomics* 18:359. doi: 10.1186/s12864-017-3739-x

Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Lin, J., Hogan, J. S., and Smith, K. L. (1999). Antigenic homology of the inducible ferric citrate receptor (FecA) of coliform bacteria isolated from herds with naturally occurring bovine intramammary infections. *Clin. Diagn. Lab. Immunol.* 6, 966–969. doi: 10.1128/CDLI.6.6.966-969.1999

Lippolis, J. D., Brunelle, B. W., Reinhardt, T. A., Sacco, R. E., Nonnecke, B. J., Dogan, B., et al. (2014). Proteomic analysis reveals protein expression differences in *Escherichia coli* strains associated with persistent versus transient mastitis. *J. Proteome* 108, 373–381. doi: 10.1016/j.jprot.2014.06.008

Lippolis, J. D., Brunelle, B. W., Reinhardt, T. A., Sacco, R. E., Thacker, T. C., Looft, T. P., et al. (2016). Differential gene expression of three mastitis-causing *Escherichia coli* strains grown under planktonic, swimming, and swarming culture conditions. *mSystems* 1:e00064-16. doi: 10.1128/mSystems.00064-16

Lippolis, J. D., Holman, D. B., Brunelle, B. W., Thacker, T. C., Bearson, B. L., Reinhardt, T. A., et al. (2018). Genomic and transcriptomic analysis of *Escherichia coli* strains associated with persistent and transient bovine mastitis and the role of Colanic acid. *Infect. Immun.* 86:e00566-17. doi: 10.1128/IAI.00566-17

Low, D., Tran, H. T., Lee, I. A., Dreux, N., Kamba, A., Reinecker, H. C., et al. (2013). Chitin-binding domains of *Escherichia coli* ChiA mediate interactions with intestinal epithelial cells in mice with colitis. *Gastroenterology* 145:e9, 602–612. doi: 10.1053/j.gastro.2013.05.017

Marion, C. R., Wang, J., Sharma, L., Losier, A., Lui, W., Andrews, N., et al. (2016). Chitinase 3-like 1 (Chil1) regulates survival and macrophage-mediated interleukin-1beta and Tumor necrosis factor alpha during *Pseudomonas aeruginosa* pneumonia. *Infect. Immun.* 84, 2094–2104. doi: 10.1128/IAI.00055-16

Maury, M. M., Bracq-Dieye, H., Huang, L., Vales, G., Lavina, M., Thouvenot, P., et al. (2019). Hypervirulent *Listeria monocytogenes* clones' adaption to mammalian gut accounts for their association with dairy products. *Nat. Commun.* 10:2488. doi: 10.1038/s41467-019-10380-0

Minh, B. Q., Nguyen, M. A., and von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195. doi: 10.1093/molbev/mst024

Mizoguchi, E. (2006). Chitinase 3-like-1 exacerbates intestinal inflammation by enhancing bacterial adhesion and invasion in colonic epithelial cells. *Gastroenterology* 130, 398–411. doi: 10.1053/j.gastro.2005.12.007

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Nuesch-Inderbinen, M., Kappeli, N., Morach, M., Eicher, C., Corti, S., and Stephan, R. (2019). Molecular types, virulence profiles and antimicrobial resistance of *Escherichia coli* causing bovine mastitis. *Vet. Rec. Open* 6:e000369. doi: 10.1136/vetreco-2019-000369

Olson, M. A., Grimsrud, A., Richards, A. C., Mulvey, M. A., Wilson, E., and Erickson, D. L. (2021). Bile salts regulate zinc uptake and capsule synthesis in a mastitis-associated extraintestinal pathogenic *Escherichia coli* strain. *Infect. Immun.* 89:e0035721. doi: 10.1128/IAI.00357-21

Olson, M. A., Siebach, T. W., Griffitts, J. S., Wilson, E., and Erickson, D. L. (2018). Genome-wide identification of fitness factors in mastitis-associated *Escherichia coli*. *Appl. Environ. Microbiol.* 84:e02190-17. doi: 10.1128/AEM.02190-17

Ren, Y., Palusiak, A., Wang, W., Wang, Y., Li, X., Wei, H. T., et al. (2016). A high-resolution typing assay for uropathogenic *Escherichia coil* based on fimbrial diversity. *Front. Microbiol.* 7:623. doi: 10.3389/fmicb.2016.00623

Reyher, K. K., Dufour, S., Barkema, H. W., Des Coteaux, L., Devries, T. J., Dohoo, I. R., et al. (2011). The national cohort of dairy farms—a data collection platform for mastitis research in Canada. *J. Dairy Sci.* 94, 1616–1626. doi: 10.3168/jds.2010-3180

Roberson, J. R., Warnick, L. D., and Moore, G. (2004). Mild to moderate clinical mastitis: efficacy of intramammary amoxicillin, frequent milk-out, a combined intramammary amoxicillin, and frequent milk-out treatment versus no treatment. *J. Dairy Sci.* 87, 583–592. doi: 10.3168/jds.S0022-0302(04)73200-2

Royer, G., Clermont, O., Marin, J., Condamine, B., Dion, S., Blanquart, F., et al. (2023). Epistatic interactions between the high pathogenicity island and other iron uptake systems shape *Escherichia coli* extra-intestinal virulence. *Nat. Commun.* 14:3667. doi: 10.1038/s41467-023-39428-y

Sarowska, J., Futoma-Koloch, B., Jama-Kmiecik, A., Frej-Madrzak, M., Ksiazczyk, M., Bugla-Ploskonska, G., et al. (2019). Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: recent reports. *Gut Pathog.* 11:10. doi: 10.1186/s13099-019-0290-0

Sears, P. M., and McCarthy, K. K. (2003). Diagnosis of mastitis for therapy decisions. *Vet. Clin. North Am. Food Anim. Pract.* 19, 93–108. doi: 10.1016/S0749-0720(02)00074-9

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

Sordillo, L. M. (2005). Factors affecting mammary gland immunity and mastitis susceptibility. *Livest. Prod. Sci.* 98, 89–99. doi: 10.1016/j.livprodsci.2005.10.017

Suojala, L., Kaartinen, L., and Pyorala, S. (2013). Treatment for bovine *Escherichia coli* mastitis - an evidence-based approach. *J. Vet. Pharmacol. Ther.* 36, 521–531. doi: 10.1111/jvp.12057

Vangroenweghe, F., Lamote, I., and Burvenich, C. (2005). Physiology of the periparturient period and its relation to severity of clinical mastitis. *Domest. Anim. Endocrinol.* 29, 283–293. doi: 10.1016/j.domaniend.2005.02.016

Verma, R., Rojas, T. C. G., Maluta, R. P., Leite, J. L., da Silva, L. P. M., Nakazato, G., et al. (2016). Fimbria-encoding gene *yadC* has a pleiotropic effect on several biological characteristics and plays a role in avian pathogenic *Escherichia coli* pathogenicity. *Infect. Immun.* 84, 187–193. doi: 10.1128/IAI.01138-15

Welch, R. A. (2016). Uropathogenic *Escherichia coli*-associated exotoxins. *Microbiol. Spectr.* 4:1128. doi: 10.1128/microbiolspec.UTI-0011-2012

Wenz, J. R., Barrington, G. M., Garry, F. B., Ellis, R. P., and Magnuson, R. J. (2006). *Escherichia coli* isolates' serotypes, genotypes, and virulence genes and clinical coliform mastitis severity. *J. Dairy Sci.* 89, 3408–3412. doi: 10.3168/jds.S0022-0302(06)72377-3

Wenz, J. R., Barrington, G. M., Garry, F. B., McSweeney, K. D., Dinsmore, R. P., Goodell, G., et al. (2001). Bacteremia associated with naturally occuring acute coliform mastitis in dairy cows. *J. Am. Vet. Med. Assoc.* 219, 976–981. doi: 10.2460/javma.2001.219.976

Wiles, T. J., Wall, E. S., Schlomann, B. H., Hay, E. A., Parthasarathy, R., and Guillemin, K. (2018). Modernized tools for streamlined genetic manipulation and comparative study of wild and diverse proteobacterial lineages. *MBio* 9, 10–1128. doi: 10.1128/mBio.01877-18

Wurpel, D. J., Beatson, S. A., Totsika, M., Petty, N. K., and Schembri, M. A. (2013). Chaperone-usher fimbriae of *Escherichia coli*. *PLoS One* 8:e52835. doi: 10.1371/journal.pone.0052835

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644. doi: 10.1093/jac/dks261

Zhou, Z., Alikhan, N. F., Mohamed, K., Fan, Y., Agama Study, G., and Achtman, M. (2020). The EnteroBase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity. *Genome Res.* 30, 138–152. doi: 10.1101/gr.251678.119

Zhou, Z., Alikhan, N. F., Sergeant, M. J., Luhmann, N., Vaz, C., Francisco, A. P., et al. (2018). GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res.* 28, 1395–1404. doi: 10.1101/gr.232397.117

Zong, Z., Fenn, S., Connor, C., Feng, Y., and McNally, A. (2018). Complete genomic characterization of two *Escherichia coli* lineages responsible for a cluster of carbapenem-resistant infections in a Chinese hospital. *J. Antimicrob. Chemother.* 73, 2340–2346. doi: 10.1093/jac/dky210

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 09 September 2024
DOI 10.3389/fmicb.2024.1456637

OPEN ACCESS

EDITED BY
Renmao "Tim" Tian,
Illinois Institute of Technology, United States

REVIEWED BY
Steven L. Foley,
National Center for Toxicological Research (FDA), United States
Salvatore Walter Papasergi,
National Research Council (CNR), Italy

*CORRESPONDENCE
Michelle Qiu Carter
✉ michelle.carter@usda.gov

RECEIVED 28 June 2024
ACCEPTED 14 August 2024
PUBLISHED 09 September 2024

CITATION
Carter MQ, Carychao D and Lindsey RL (2024) Conditional expression of flagellar motility, curli fimbriae, and biofilms in Shiga toxin- producing *Escherichia albertii*. *Front. Microbiol.* 15:1456637. doi: 10.3389/fmicb.2024.1456637

Conditional expression of flagellar motility, curli fimbriae, and biofilms in Shiga toxin-producing *Escherichia albertii*

Michelle Qiu Carter[1]*, Diana Carychao[1] and Rebecca L. Lindsey[2]

[1]Produce Safety and Microbiology Research Unit, U.S. Department of Agriculture, Agricultural Research Service, Western Regional Research Center, Albany, CA, United States, [2]Enteric Diseases Laboratory Branch, Centers for Disease Control and Prevention, Atlanta, GA, United States

Escherichia albertii is an emerging foodborne pathogen. We previously reported that some avian Shiga toxin-producing *E. albertii* strains exhibited higher or comparable cytotoxicity in Vero-d2EGFP cells with several enterohemorrhagic *E. coli* (EHEC) outbreak strains. To better understand the environmental persistence of this pathogen, comparative genomics and phenotypic assays were applied to assess adhesion capability, motility, and biofilm formation in *E. albertii*. Among the 108 adherence-related genes, those involved in biogenesis of curli fimbriae, hemorrhagic *E. coli* pilus, type 1 fimbriae, and Sfm fimbriae were conserved in *E. albertii*. All 20 *E. albertii* strains carried a complete set of primary flagellar genes that were organized into four gene clusters, while five strains possessed genes related to the secondary flagella, also known as lateral flagella. Compared to EHEC strain EDL933, the eight chemotaxis genes located within the primary flagellar gene clusters were deleted in *E. albertii*. Additional deletion of motility genes *flhABCD* and *motBC* was identified in several *E. albertii* strains. Swimming motility was detected in three strains when grown in LB medium, however, when grown in 5% TSB or in the pond water-supplemented with 10% pigeon droppings, an additional four strains became motile. Although all *E. albertii* strains carried curli genes, curli fimbriae were detected only in four, eight, and nine strains following 24, 48, and 120 h incubation, respectively. Type 1 fimbriae were undetectable in any of the strains grown at 37°C or 28°C. Strong biofilms were detected in strains that produced curli fimbriae and in a chicken isolate that was curli fimbriae negative but carried genes encoding adhesive fimbriae K88, a signature of enterotoxigenic *E. coli* strains causing neonatal diarrhea in piglets. In all phenotypic traits examined, no correlation was revealed between the strains isolated from different sources, or between the strains with and without Shiga toxin genes. The phenotypic variations could not be explained solely by the genetic diversity or the difference in adherence genes repertoire, implying complex regulation in expression of various adhesins. Strains that exhibited a high level of cytotoxicity and were also proficient in biofilm production, may have potential to emerge into high-risk pathogens.

KEYWORDS

***Escherichia albertii*, foodborne pathogen, biofilm, adhesins, fimbriae, flagella, motility**

1 Introduction

Escherichia albertii, an emerging foodborne pathogen, is the most divergent lineage among the other *Escherichia* species and clades (Walk et al., 2009; Ooka et al., 2015). Due to similar biochemical properties and possession of the intimin gene located on the locus of enterocyte effacement (LEE) pathogenicity island, many *E. albertii* isolates have been misidentified as enteropathogenic *E. coli* (EPEC), or enterohemorrhagic *E. coli* (EHEC) (Ooka et al., 2012). *E. albertii* causes diarrhea, abdominal pain, and high fever in humans, although bacteremia and extraintestinal infections were also reported (Gomes et al., 2020). The well-known virulence factors in *E. albertii* include LEE encoded intimin and its Tir receptor, responsible for the initial adherence of pathogen cells to the host epithelial cell surfaces, as well as the LEE-encoded type three secretion system (T3SS) and the effector proteins. Other common virulence factors include Shiga toxin (Stx), cytolethal distending toxin (CDT), type six secretion systems (T6SS), and the vacuolating autotransporter toxin Vat (Carter et al., 2023a). Sporadic infections and outbreaks of foodborne gastroenteritis caused by *E. albertii* have been reported worldwide (Konno et al., 2012; Ooka et al., 2013; Ori et al., 2018; Masuda et al., 2020; Bengtsson et al., 2023; Iguchi et al., 2023). Transmission of *E. albertii* is thought to occur via contaminated food or water although in most outbreaks the transmission vehicles were not identified (Masuda et al., 2020; Muchaamba et al., 2022).

Growing evidence supports that *E. albertii* has a wide habitat range. *E. albertii* strains have been isolated from domestic and wild animals, various foods, and aquatic environments (Muchaamba et al., 2022). Among the reported animal hosts, birds appear to be one of the main reservoirs/carriers (Oaks et al., 2010; Hinenoya et al., 2021; Hinenoya et al., 2022; Wang et al., 2022; Barmettler et al., 2023; Xu et al., 2024). Presence of *E. albertii* in various water bodies and food products including chicken, pork, duck meat, mutton, and oysters has been reported (Felfoldi et al., 2010; Maheux et al., 2014; Lindsey et al., 2015; Maeda et al., 2015; Wang et al., 2016; Arai et al., 2022), however, little is known about the contamination routes and the environmental prevalence and persistence of this emerging foodborne pathogen. A recent study investigating the survival of *E. albertii* in foods and water revealed that *E. albertii* grew faster in chicken than in pork or in oysters but had low viability in warm environmental water (Hirose et al., 2024). Induction of flagellar biosynthesis and swimming motility was observed in some strains when cells were exposed to hypoosmotic pressure or at ambient temperature, suggesting a role of flagellar motility in the survival of *E. albertii* in aquatic environments (Ikeda et al., 2020).

Biofilm is a common microbial lifestyle in natural environments (Watnick and Kolter, 2000). Compared with planktonic cells, biofilm-associated cells are better at coping with environmental stresses and have increased resistance to toxic substances including antibiotics and chemical sanitizers. Therefore, biofilm formation by enteropathogenic bacteria would increase their survival and persistence in natural environments and may serve as a source of contamination. Biofilm formation involves multiple steps, including initial surface contact, transient association, attachment, maturation, and dispersion (O'Toole et al., 2000). Numerous bacterial adherence factors including surface adhesive appendages and autotransporter proteins play a role in biofilm formation. In *E. coli* K-12 strains, flagellar motility, curli fimbriae, as well as FimH adhesin were found to be important for initial surface contact and attachment. Additionally, flagellar motility was found playing a role in biofilm dispersion (Pratt and Kolter, 1998; Reisner et al., 2003; Karatan and Watnick, 2009).

Shiga toxin-producing *E. coli* (STEC) produces diverse fimbrial and nonfimbrial adhesins that facilitate the attachment to and/or colonization by STEC cells in diverse ecological niches (McWilliams and Torres, 2014; Vogeleer et al., 2014). In STEC O157:H7 strains, curli fimbriae were found to mediate binding to, and invasion of epithelial cells, and promote the attachment of pathogens to plant and abiotic surfaces (Gophna et al., 2001; Fink et al., 2012; Carter et al., 2016). The hemorrhagic *E. coli* pilus (HCP), originally identified in STEC O157:H7 as a colonization factor (Xicohtencatl-Cortes et al., 2007), contributed to the biofilm formation of STEC O157:H7 strains on abiotic surfaces (Xicohtencatl-Cortes et al., 2009). In *E. coli* and other enteric pathogens, expression of type 1 fimbriae is controlled by a phase variation mechanism, which reversibly switches between the "ON" and "OFF" state of *fim* genes transcription (Abraham et al., 1985). This switch is mediated by an invertible DNA element, *fimS*, and two site-specific recombinases. Inversion of *fimS* abolishes the transcription of *fimA*, which encodes the major subunit of type 1 fimbriae. Expression of type 1 fimbriae were detected in STEC non-O157 strains, but not in O157:H7 strains (Roe et al., 2001). In STEC O157:H7 strains, transcription of *fimA* is locked at the "OFF" state due to a 16-bp deletion within the *fimS* (Iida et al., 2001). Type 1 fimbriae contributed to the attachment of STEC cells to abiotic surfaces in a O128:H2 strain and contributed to the biofilm formation when the *fim* genes of the STEC O157:H7 strain Sakai were expressed in a nonpathogenic *E. coli* strain (Cookson et al., 2002; Elpers and Hensel, 2020).

Knowledge about environmental persistence of *E. alberti* is scarce. Biofilm formation by *E. albertii* was reported in only a few clinical strains at 37°C although the efficiency of biofilm formation was much lower than that of *E. coli* strain 042 (Lima et al., 2019). Understanding prevalence and persistence of *E. albertii* in nonhost environments will provide valuable information for risk assessment and to bridge gaps in understanding the epidemiology of this emerging human pathogen. We previously reported genomic features and virulence genes repertoire of Shiga toxin-producing *E. albertii* strains isolated from wild birds in an agricultural region in California and revealed that some bird strains exhibited higher or comparable cytotoxicity with several EHEC outbreak strains (Carter et al., 2023a). To gain insight into the persistence of *E. albertii* in nonhost environments, we systematically evaluated the adhesion capability and several phenotypic traits known to contribute to bacterial biofilm formation in a set of *E. albertii* avian and clinical strains. Our study revealed great genetic diversity in genes encoding fimbrial and nonfimbrial adhesins in *E. albertii* as well as vast strain variations in expression of curli fimbriae, swimming motility, and in biofilm formation. Our study provides a foundation into further understanding how *E. albertii* senses and responds to environmental stimuli for improved survival in the changing environments.

2 Materials and methods

2.1 Bacterial strains and growth media

Bacterial strains and their sources are listed in Table 1. The complete genome sequences of *E. albertii* strains were reported previously (Carter et al., 2023a). The strains were grown routinely in Luria-Bertani (LB) broth (10 g tryptone, 5 g yeast extract, and 5 g NaCl per liter) unless noted.

2.2 Sequence analysis

The flagellar genes in *E. albertii* strains were identified by using BLASTn searches with the flagellar genes of the *E. coli* K-12 sub-strain MG1655 and the EHEC strain EDL933 (Supplementary Table 1). Additional flagellar genes were identified from the *E. albertii* genome annotations as described previously (Carter et al., 2023a) (Supplementary Table 2). *E. coli* genes related to fimbriae and pili biogenesis and genes encoding protein adhesins (Supplementary Table 3) were used as queries of BLASTn to identify homologs of adherence-related genomic loci in *E. albertii* strains. The BLASTn was performed in Geneious Prime® with a threshold of 65% for gene coverage and 70% or 25% for sequence identity at nucleotides or amino acids level, respectively. Homologs of each gene or the entire operons were extracted from the corresponding bacterial genomes. DNA sequences were aligned using Clustal Omega in Geneious Prime® (2024.0.3) and neighbor-joining consensus trees were constructed with the following parameters: Genetic Distance Model, Jukes-Cantor; Resampling Method, bootstrap; and number of replicates, 10,000.

2.3 Motility tests

Swimming motility was examined for each strain grown on soft agar (0.25%) in rich medium (LB), diluted TSB (5%), and sterile pond water containing 10% pigeon droppings as described previously (Murakami et al., 2020) with modification. The pond water was collected from a public accessible creek in Albany, California (37°53′43.86″N, 122°18′16.68″W). The pigeon droppings were collected near a train station in El Cerrito, California (37°54′9.63″N, 122°17′56.17″W). To prepare the 10% pigeon-droppings suspension, pigeon droppings were first suspended in nine volumes of pond water and then filtered through a 0.22-μm filter followed by adding agar to 0.25% prior to autoclaving. Single colonies of each *E. albertii* strain were point-inoculated on soft agar plates using sterile toothpicks. The plates were incubated at 30°C for three days prior to observing the motility.

2.4 Detection of curli fimbriae

Curli fimbriae were examined by growing each strain at 26°C for 1, 2, and 5 days on Congo Red indicator (CRI) plates, consisting of LB agar plates without sodium chloride (LBNS) and supplemented with 40 μg/ml of Congo Red dye and 10 μg/ml of Coomassie Brilliant Blue, as described previously (Carter et al., 2011). Curli-producing strains were indicated by red colonies whereas curli-deficient strains were indicated by white colonies on CRI plates.

2.5 Detection of type 1 fimbriae

Production of type 1 fimbriae was examined by hemagglutination for each strain grown in LBHS broth statically at 37°C or in LBNS broth statically at 28°C for two days. Cells were collected by centrifugation at 8,000 g for 3 min and resuspended in 1x PBS buffer at a final concentration about 3 x 10^8 cells/ml. Fifty μl of bacterial suspension was then mixed with 50 μl of guinea pig red blood cells (Innovative Research Inc) at room temperature in the presence or absence of 1% D mannose as previously described (Biscola et al., 2011). *E. coli* strain DH5a was used as a positive control and EHEC strain EDL933 was used as a negative control.

2.6 Biofilm formation and quantification

Biofilm assays were carried out as described previously (Carter et al., 2023b). Briefly, overnight cultures of *E. albertii* grown in LB at 37°C were inoculated in LBNS broth at a final concentration of $1x10^6$ cells/ml. One ml of inoculated LBNS broth was aliquoted into a borosilicate glass tube and then incubated statically at 28°C for 1, 2, and 5 days. At the end of each incubation, the planktonic cells were removed carefully, and the tubes were rinsed twice with one ml sterile distilled water and then stained with one ml 0.1% crystal violet at room temperature for 30 min. The dye was then removed gently, and the tubes were washed twice with sterile distilled water. The crystal violet that bound to the glass tube was solubilized in 0.5 ml of 33% acetic acid and the absorbance was determined at 570 nm using a microplate reader (SpectraMax 340; Molecular Devices, Sunnyvale, CA). Tubes with uninoculated media served as negative controls. Each data set was the average of results from at least three biological replicates. All data were first evaluated for normal distribution by the Shapiro-Wilk test using Graph Pad Prism 10 Version 10.2.3 (Dotmatics). The differences in biofilm formation, represented by the absorbance at 570 nm, among the strains were assessed by the adjusted *P*-value of the Tukey's multiple comparisons test after a One-way ANOVA test ($P \leq 0.05$). Similarly, the differences in biofilm formation of each strain at various incubation times were assessed by the adjusted *P*-value of the Tukey's multiple comparisons test after a One-way ANOVA test.

3 Results

3.1 *E. albertii* flagellar genes

Of the 48 genes related to flagella biosynthesis and motility in strain EDL933, homologs of 40 and 35 genes were identified in 13 and seven *E. albertii* strains, respectively (Supplementary Table 1). In strain EDL933, these 48 flagellar genes are distributed

at four genomic locations, with a size of 11.5 Kb, 15.6 Kb, 6.7 Kb, and 11.0 Kb for Regions 1–4, respectively. Examining the genomic locations of the flagellar genes in *E. albertii* revealed a similar genes organization as in strain EDL933 (Figure 1A). Among the four genomic locations, the greatest sequence variation was detected in Region 2. In strain EDL933, Region 2 contained seven flagellar genes and eight chemotaxis genes (*cheZYBR*, *tap*, *tar*, and *cheWA*). Unlike strain EDL933, the eight chemotaxis genes were deleted in all *E. albertii* strains examined. Furthermore, an additional deletion of genes *flhBCD* and *motBA* was detected in a subset of *E. albertii* strains including five avian and two clinical strains. This deletion appeared to be mediated by a recombination between the sites within genes *flhA* and *otsA* since a 183-bp *otsA* gene fragment was located immediately upstream of a truncated *flhA* gene. In contrast, the flagellar genes located in the other three regions in strain EDL933 were all conserved in *E. albertii* strains.

Interestingly, a Flag-2 locus, which encodes a secondary flagellar system that resembles the lateral flagella in *Aeromonas hydrophila* and *Vibrio parahaemolyticus* (Ren et al., 2005), was identified in one avian (RM9973) and four clinical *E. albertii* strains. The Flag-2 loci in *E. albertii* varied in size from 33 Kb in the clinical strain 2010C-3449 to 44 Kb in the clinical strain 05-3106. Like the Flag-2 locus in the enteroaggregative *E. coli* (EAEC) strain 042, Flag-2 genes were organized into three gene clusters, separated by the two variable regions, VR1 and VR2 (Figure 1B). In strain 042, the first gene cluster contains 14 genes that are involved in regulation and expression of flagellar basal body components. Homologs of these 14 genes were detected in the avian strain RM9973 and in the clinical strain 05-3106. Genes *lfhB* and *lfiR* were deleted in the strain 07-3866, while genes *lafK* and *lfiEFGHIJ* were deleted in both strains 54-2045 and 2010C-3449 (Supplementary Table 2). The second gene cluster in strain 042 also contains 14 genes encoding flagellar structural proteins and the third gene cluster carries nine genes that are mainly involved in flagellar filament synthesis. Homologs of all genes within the second and the third gene clusters were detected in the Flag-2 positive *E. albertii* strains. In the Flag-2 negative *E. albertii* strains, this region was about 2.5 Kb, containing the truncated two border genes, *flhA* and *lafU* (Figure 1B). A highly similar truncated Flag-2 locus was detected in *E. coli* strains EDL933 and K-12 strain MG1655 (% Identity > 90).

3.2 Motility in *E. albertii*

When grown in LB at 30°C for three days, motility was observed in two avian strains, RM10507 and RM10705, and one clinical strain 07-3866 (Table 1). Both strains RM10507 and RM10705 were isolated from brown-headed cowbird and were Flag-2 negative. These three strains remained motile when grown in 5% TSB or in pond water supplemented with 10% pigeon droppings (Table 1). Interestingly, four nonmotile strains when grown in LB, became motile when grown in 5% TSB or in the pond water supplemented with 10% pigeon droppings (Table 1). These four strains included two avian strains, RM9973 and RM9976 that were both isolated from American crow, the chicken isolate 2014C-4356, and the clinical strain 05-3106. The majority of nonmotile phenotypes could be explained by the mutations identified in the flagellar genes, including the deletion of *motAB* and *flhBCD* in avian strains RM15112-RM15116 and in clinical strains 2014C-4015 and 2014EL-1348 (Figure 1B), point deletions in *fliF* of the strains RM9974 and 2011C-4180 and in *motA* of the strain 2013C-4143, and an amber mutation in *flgG* and *flhA* of the strains 54-2045 and 2010C-3449, respectively (Supplementary Table 1).

3.3 *E. albertii* fimbrial genes

Homologs of genes encoding 12 fimbriae and pili implicated in adherence, biofilm formation, and pathogenesis in diverse *E. coli* pathotypes were examined in *E. albertii*. All genes are listed in Supplementary Table 3. Homologs of genes encoding curli fimbriae, type 1 fimbriae, and hemorrhagic *E. coli* pilus (HCP) were detected in all strains while homologs of genes encoding adhesive fimbriae, Sfm fimbriae, and P fimbriae were detected in a subset of *E. albertii* strains examined.

3.3.1 Curli genes and expression of curli fimbriae

Like *E. coli*, genes related to biogenesis of curli fimbriae in *E. albertii* are organized in two divergent operons, *csgDEFG* and *csgBAC*, and located upstream of tRNA gene *serX* (Figure 2A). Sequence analysis revealed that all *E. albertii* curli genes were placed in the same clade that was separated from the *E. coli* curli genes (Figure 2B). The curli genes of the *E. albertii* strains shared a high sequence similarity (> 95%) with each other, except for the clinical strain 2010C-3449, in which, both *csgE* and *csgD* were truncated due to an IS insertion, while *csgA* carried an amber mutation.

Unexpectedly, production of curli fimbriae varied greatly among the *E. albertii* strains (Figure 2C). Among the 10 avian strains, production of curli fimbriae was observed in strains RM9973, RM9974, RM9976 and RM10705, although all avian strains carried intact coding sequences for all curli genes. As expected, no curli fimbriae were observed for clinical strain 2010C-3449. Among the other clinical strains, production of curli fimbriae was detected in three out of four stx_{2f} positive strains (2012EL-1823B, 2014C-4015, 2014EL-1348) and in the stx_{2f} negative strain 07-3866. Colonies of strain 2013C-4143 exhibited pink and light red color following 48 h and 120 h incubation, respectively, suggesting that this strain could produce curli fimbriae under the condition examined but with less amount compared with the other curli-positive strains (Figure 2C).

3.3.2 Type 1 fimbriae genes and expression of type 1 fimbriae

In *E. coli*, the type 1 fimbriae genes (*fimB*, *fimE*, and *fimAICDFGH*) are located on an 8.8-Kb DNA fragment. Expression of the type 1 fimbriae is controlled by a phase variation mechanism, in which, transcription of *fimA* is switched to "ON" or "OFF" by an invertible DNA element, *fimS*, and two site-specific recombinases encoded by genes *fimB* and *fimE*, respectively (Figure 3A). Homologs of the nine *fim* genes were identified in all 20 *E. albertii* strains examined. Additionally, the invertible element, *fimS*, flanked by two 9-bp inverted repeats (IRs), was also detected in all *E. albertii* strains. The IRs (5′-TTGGGGCCA-3′) in *E. albertii* strains were identical to the IRs in *E. coli* strains EDL933 and K-12 strain MG1655, except for the avian strain RM9973, in which a single base substitution of G to A occurred at position 6.

FIGURE 1
Sequence analyses of *E. albertii* flagellar genes. **(A)** Gene organization and genomic locations of the primary flagellar genes. Numbers indicate the corresponding chromosomal positions of the four flagellar gene clusters in *E. coli* O157:H7 strain EDL933 and *E. albertii* avian strains RM9973 and RM15112. Green arrows represent the flagellar genes detected in *E. albertii* and grey arrows represent the bordering genes or the hypothetical genes. Yellow arrows represent *E. coli* genes that are missing in *E. albertii* strains. **(B)** Gene organization and genomic locations of the secondary flagellar genes (Flag-2). Numbers indicate the corresponding chromosomal positions in EAEC strain 042 and avian strain RM9973. The three flagellar gene clusters are indicated by the red arrows. In the strains lacking a Flag-2, such as EDL933 and RM9974, the corresponding chromosomal sites were uniformly occupied by the two truncated genes, *flhA* and *lafU*. Genes labeled with an "*" indicate those carrying mutations within the coding sequences.

Expression of type 1 fimbriae in *E. coli* O157:H7 strains including EDL933 is silenced due to a 16 bp deletion in *fimS* (Roe et al., 2001). This deletion locks the transcription of *fimA* at the "OFF" orientation. Comparative analyses of *E. albertii fimS* genes with the EDL933 *fimS* revealed an intact *fimS* in *E. albertii* strains, like the *fimS* in *E. coli* K-12 strain MG1655 (Figure 3B). Sequence analyses of other *fim* genes placed all *E. albertii* strains in the same clade, separated from the *fim* genes in the *E. coli* strains (Figure 3C). Various mutations including point deletions and IS insertions were revealed in the *fimA, fimC, and fimD genes* of clinical strain 2010C-3449, the *fimD* of the clinical strain 54-2045, and the *fimI* of the clinical strain 2013C-4143 (Supplementary Table 3).

However, a mannose-sensitive hemagglutination assay failed to detect type 1 fimbriae in any of the 20 *E. albertii* strains examined when they were grown in LBHS at 37°C or in LBNS at 28°C, like strain EDL933. Production of type 1 fimbriae was detected in *E. coli* DH5a cells under both testing conditions.

3.3.3 Other fimbriae genes

Homologs of genes encoding hemorrhagic *E. coli* pilus (*hcpABC*) were identified in all 20 *E. albertii* strains, and mutations in *hcpB* (annotated as *gspE* in *E. albertii*), encoding the type II secretion system protein GspE, were present in seven out of the 20 strains examined (Supplementary Table 3). Similarly, homologs of genes encoding the Sfm fimbriae (*sfmACDHF* and *sfmZ*) were identified in the 18 out of the 20 strains examined. Mutations were most common in *sfmD*, encoding a fimbrial biogenesis usher protein (Supplementary Table 3). In the clinical strains 54-2045 and 2010C-3449, only a homolog of *sfmA* was present. Among the adhesive fimbriae genes examined, homologs of *cfaABCD* genes, which are often present on the chromosomes of EHEC strains, were identified in 10 out of the 14 stx_{2f}-positive *E. albertii* strains, while homologs of *faeCDEFGHIJ* genes, which are often present on the plasmids of enterotoxigenic *E. coli* (ETEC) strains, were identified in the clinical strain 07-3866 and in the chicken isolate 2014C-4356 (Supplementary Table 3). In *E. coli*, there are 12 genes (*papXGFEKJDCHABI*) related to biogenesis of P fimbriae. Homologs of seven genes, *papEKJDCHA*, were identified in seven out of the 20 *E. albertii* strains, homologs of five genes, *papJDCHA*, were identified in the chicken isolate 2014C-4356, and homologs of four genes, *papDCHA*, were identified in the clinical strain 2012EL-1823B.

3.4 Nonfimbrial adhesin genes and their genetic diversity

The most common autotransporter adhesin genes detected in *E. albertii* were *paa*, *ehaC*, *eaeH*, *ehaB*, and *sinB*

TABLE 1 *E. albertii* strains used in this study and the motility test.

Strains	Sources/Year of isolation	GenBank BioSample Number	[a]Swimming Motility		
stx_{2f}-positive *E. albertii*			LB	5% TSB	Pond water with 10% pigeon droppings
RM9973	American crow (*Corvus brachyrhynchos*)/2009	SAMN12620691	–	+	+
RM9974	American crow (*Corvus brachyrhynchos*)/2009	SAMN12620692	–	–	–
RM9976	American crow (*Corvus brachyrhynchos*)/2009	SAMN12620693	–	+	+
RM10507	Brown-headed cowbird (*Molothrus ater*)/2009	SAMN12620694	+	+	+
RM10705	Brown-headed cowbird (*Molothrus ater*)/2009	SAMN12620697	+	+	+
RM15112	Oregon Junco (*Junco hyemalis*)/2011	SAMN12620700	–	–	–
RM15113	Oregon Junco (*Junco hyemalis*)/2011	SAMN12620701	–	–	–
RM15114	Oregon Junco (*Junco hyemalis*)/2011	SAMN12620702	–	–	–
RM15115	White-Breasted Nuthatch (*Sitta carolinensis*)/2011	SAMN12620703	–	–	–
RM15116	Oregon Junco (*Junco hyemalis*)/2011	SAMN12620704	–	–	–
2011C-4180	Human/2011	SAMN03019926	–	–	–
2012EL-1823B	Human/2012	SAMN04498560	–	–	–
2014C-4015	Human/2014	SAMN04505646	–	–	–
2014EL-1348	Human/2014	SAMN04505647	–	–	–
stx-negative *E. albertii*					
2014C-4356	Chicken Carcass/2009	SAMN07159041	–	+	+
05-3106	Human/2005	SAMN08199278	–	+	+
07-3866	Human/2007	SAMN07159045	+	+	+
54-2045 (NCTC 9362)	Human/1954	SAMN09534374	–	–	–
2010C-3449	Human/2010	SAMN07159044	–	–	–
2013C-4143	Human/2013	SAMN08172567	–	–	–

[a]Swimming motility was observed after incubation at 30°C for three days.

(Supplementary Table 3). *paa* encodes an AcfC family adhesin. A homolog of *paa* was present in nearly all *E. albertii* strains examined and exhibited > 80% sequence identity with the *paa* gene in EHEC strain EDL933. The gene *ehaC* encodes an AIDA-I family autotransporter adhesin. A homolog of *ehaC* was detected in all *E. albertii* strains examined, although a point deletion and a point insertion were present in the avian strain RM10705 and the clinical strain 2010C-3449, respectively (Supplementary Table 3). The *E. albertii ehaC* genes exhibited ~ 75% sequence identity with the *ehaC* gene in strain EDL933. *eaeH* encodes an intimin-like adhesin FdeC. A distant homolog (~ 80% length in CDS and 27.6% identity in amino acids) was identified in all *E. albertii* strains examined, although mutations were detected in the clinical strain 2011C-4180 (IS insertion), and avian strains RM10507 (Insertion of 5'-GTCTG-3') and RM10705 (a point deletion). A homolog of *ehaB*, ranging in size from 2430 bp to 2979 bp was detected in *E. albertii* strains. Interestingly, the *ehaB* genes in avian strains RM9973 and RM9976, and in clinical strains 05-3106, 54-2045, and 2014C-4356, displayed higher sequence similarity with the EDL933 *ehaB* gene compared with the *ehaB* genes in other *E. albertii* strains. The gene *sinH*, encoding an intimin-like inverse autotransporter, was present in all *E. albertii* strains examined. In fact, the gene *sinH* appeared to be widespread in *E. albertii* but only present in a subset of *E. coli* strains. Homologs of *sinH* were not identified in EHEC strain EDL933.

3.5 Biofilm formation

Following the initial 24 h incubation, a visible ring was observed for avian strain RM9974 and four clinical strains, 2012EL-1823B, 2014C-4015, 2014EL-1348, and 07-3866 (Figure 4A). Consistently,

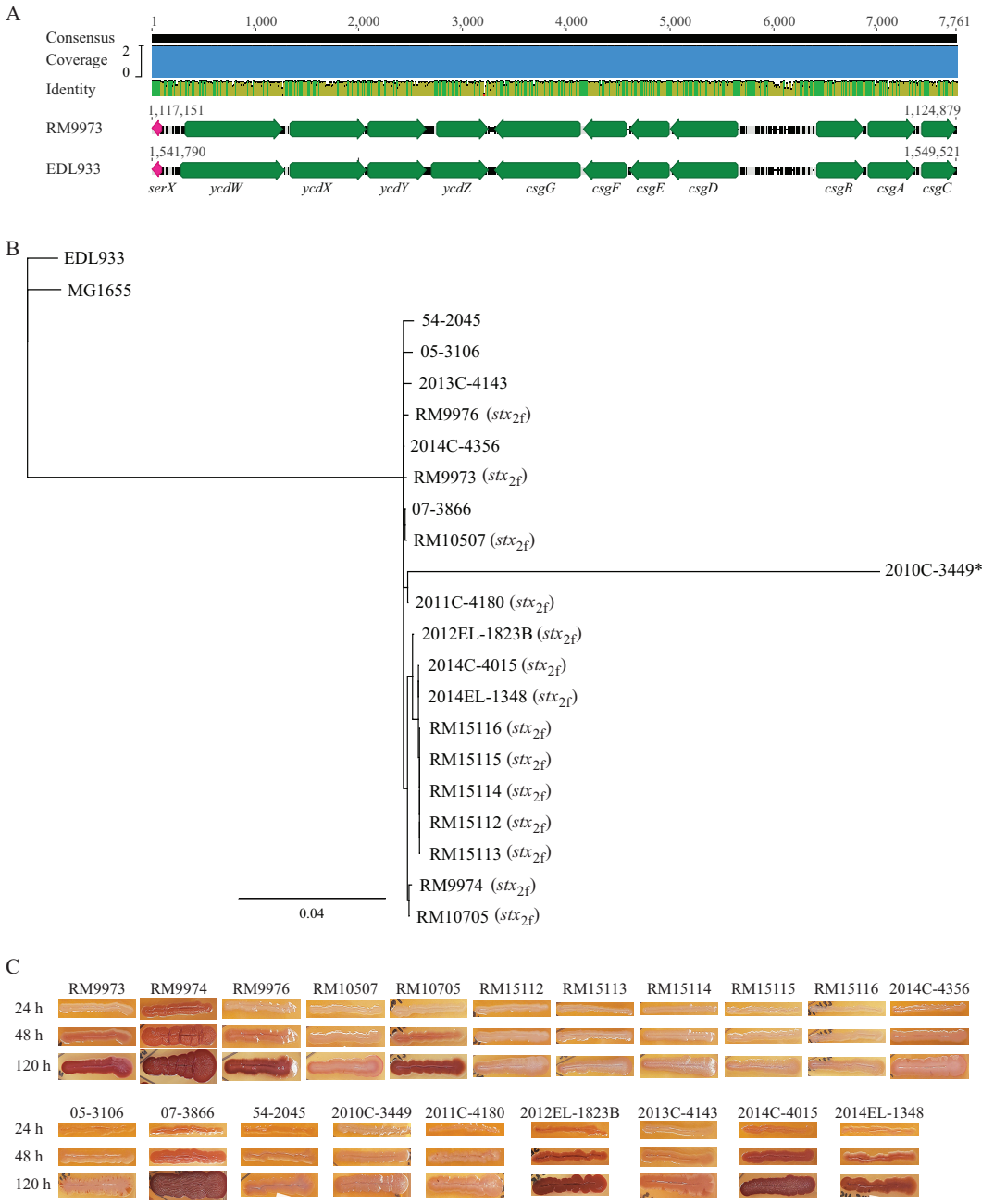


FIGURE 2
Sequence analyses of curli genes and detection of curli fimbriae in *E. albertii*. **(A)** Chromosomal locations of curli operons and pairwise comparison of the curli genes between the *E. albertii* avian strain RM9973 and *E. coli* O157:H7 strain EDL933. Numbers indicate the corresponding chromosomal positions in each strain. Green arrows represent the annotated genes, and the pink arrows represents tRNA gene *serX*. **(B)** Sequence analysis of curli genes. The curli operons were identified by BLASTn search of a database containing all genomes examined using a 4.4-Kb DNA fragment containing the seven curli genes of the *E. coli* strain MG1655 as a query in Geneious Prime®. The sequences of the curli genes were extracted from corresponding genomes and aligned using Clustal Omega alignment in Geneious Prime®. A consensus tree was constructed with the following parameters: Genetic Distance Model, Jukes-Cantor; Resampling tree method: Bootstrap; Number of Replicates: 10,000; Support Threshold: 50%. The stx_{2f} positive strains are indicated in parentheses. The strain marked with an "*" indicates presence of mutations within the coding sequences of curli genes. **(C)** Detection of curli fimbriae on CRI plates. Curli fimbriae were examined by growing each strain on the CRI plates at 26°C for 24 h, 48 h, and 120 h. Production of curli fimbriae is indicated by red colonies which resulted from the binding of CR dye supplemented in growth medium.

quantitative analysis revealed that the attached biomass for the above five strains were all significantly greater than the rest of the strains except the comparison between strains 2014EL-1348 and RM15112 (One-way ANOVA, adjust $P < 0.05$) (Supplementary Table 4). Among the five biofilm producing strains, strains RM9974, 2014C-4015, and 07-3866 produced significantly greater amounts of biofilm than the other two strains (Figure 4B). Following 48 h incubation, the attached biomass for the five biofilm producing strains were all significantly greater than the corresponding biofilms at 24 h (Figure 4C) (One way ANOVA

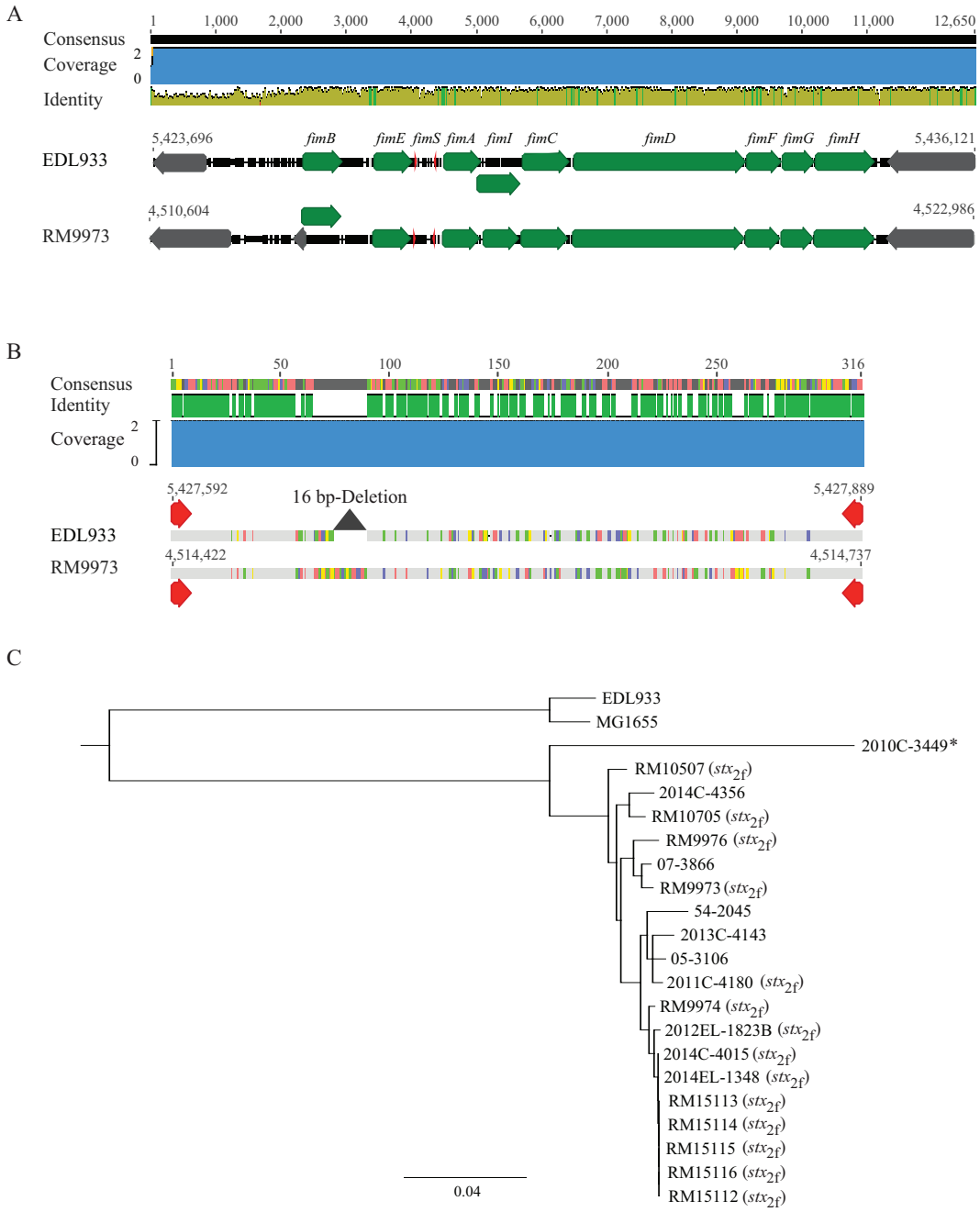


FIGURE 3
Sequence analyses of *E. albertii* type 1 fimbriae genes. **(A)** Chromosomal locations and pairwise comparison of the type 1 fimbriae genes between the *E. albertii* avian strain RM9973 and *E. coli* O157:H7 strain EDL933. Numbers indicate the chromosomal positions in each strain. Green arrows represent the *fim* genes; Gray arrows represent the neighbor genes; and the red arrows represents right and left inverted repeats (IRs) within *fimS*. The 9-bp IR in strain EDL933 is 5′-ttggggcca-3′ while in strain RM9973, the 9-bp IR is 5′-ttgggacca-3′. **(B)** Pairwise alignment of the *cis* element *fimS* of the *E. albertii* avian strain RM9973 and *E. coli* O157:H7 strain EDL933. Red arrows represent the IRs. The grey triangle represents the 16 bp-deletion in EDL933 *fimS*. **(C)** Sequence analysis of *E. albertii* type 1 fimbriae genes with the *fim* genes in *E. coli* strains EDL933 and MG1655. The *fim* genes were identified by BLASTn search of a database containing all genomes examined in this study using an 8.7-Kb DNA fragment containing the nine *fim* genes of the *E. coli* strain EDL933 as a query in Geneious Prime®. The sequences of the *fim* genes were extracted from corresponding genomes and aligned using Clustal Omega alignment in Geneious Prime®. A consensus tree was constructed using Geneious Tree Builder with the following parameters: Genetic Distance Model, Jukes-Cantor; Resampling tree method: Bootstrap; Number of Replicates: 10,000; Support Threshold: 50%. The stx_{2f} positive strains are indicated in parentheses. The strain marked with an "*" carries mutations within the coding sequences of type 1 fimbriae genes.

test, adjust $P < 0.05$). Among the strains that did not produce any detectable biofilm at 24 h, a visible ring was observed for strains RM9973 and RM10705 (Figure 4A). Quantitative analysis revealed that strains 07-3866, 2014C-4015, 2014EL-1348, and RM9974 produced significantly greater amounts of biofilm than that of the strain 2012EL-1823B (One way ANOVA test, adjust

$P < 0.05$) (Figure 4C). Among the strains that did not produce any visible biofilms at 48 h, strain 2014C-4356 produced a considerable amount of biomass on the glass surface following 120 h incubation (Figure 4A). Among the strains that produced biofilms following 48 h incubation, a significant increase in attached biomass was observed for all strains following 120 h incubation (One-way ANOVA, adjust $P < 0.05$) (Figure 4D). At 120 h post inoculation, quantitative analysis revealed that avian strain RM9974, chicken isolate 2014C-4356 along with the clinical strain 2014C-4015 were the strongest biofilm producers, followed by the clinical strains 2014EL-1348 and 2012EL-1823B, and the avian strain RM9973, which all produced significantly greater amounts of biofilm than the clinical strain 07-3866 (One-Way ANOVA, adjust $P < 0.05$). For avian strains RM9976 and RM10705, although a visible ring was detected on glass surfaces, they were not significantly different from those of non-biofilm producing strains (Figure 4D).

4 Discussion

Flagellar motility allows bacteria to move rapidly towards nutrients and away from toxic substances, thus it plays an essential role in bacteria to explore new niches and to establish colonization. Moreover, flagella also serve as a virulence factor in many enteric pathogens, contributing to adhesion, invasion, and host colonization (Moens and Vanderleyden, 1996; Colin et al., 2021). In *E. coli*, nearly 50 genes are involved in flagella assembly and function. Expression of the flagellar genes is tightly regulated by a three-tiered transcriptional hierarchy to ensure production of flagellum at the right time and under the applicable conditions (Khan et al., 2020). The group I genes, *flhDC*, encode the master transcriptional regulator FlhDC that activates the expression of the group II genes. In *E. coli*, there are nearly 30 genes belonging to the group II and many of these genes encode components of the flagellum basal body and hook and the sigma factor 28 FliA. FliA regulates the expression of group III genes, which are involved in synthesis of complete flagellum and chemotaxis systems. The regulation of flagellar gene expression in *E. albertii* is unknown, although a similar hierarchical regulation fashion is expected considering the close phylogenetic relationship between the two species. Originally, *E. albertii* was thought nonmotile and lacked flagella although 74% *E. albertii* strains were reported to carry a complete set of flagellar biosynthesis genes (Abbott et al., 2003; Ooka et al., 2015). Induction of flagellar motility by low osmotic pressure was observed in 27 out of the 59 *E. albertii* strains tested (Ikeda et al., 2020); similarly, induction of swimming motility by nutrients derived from pigeon droppings was observed in six out of the 12 strains examined (Murakami et al., 2020), implying strain variation in expression of flagellar motility in *E. albertii*. Consistently, our study revealed great diversity in flagellar genes repertoire and conditionally expressing swimming motility in *E. albertii*. Among the 20 *E. albertii* strains examined, three were motile regardless of the growth conditions, while four were motile only when grown in pond water supplemented with pigeon droppings or in the diluted TSB medium. Deletion of a large DNA fragment containing genes *flhAB*, *motBA*, and *flhCD* is likely the molecular basis of nonmotile phenotypes observed in five avian and two clinical strains. Furthermore, a loss-of-function mutation in genes *fliC*, *fliF*, *fliJ*, *flgD*, *flgG*, *flhA*, and *motA* may explain some but not all non-motile phenotypes observed in our study. Considering the highly complex regulation in flagellar gene expression, comparative transcriptomic studies may provide insight into the molecular basis of the strain variation in expression of flagellar motility.

Our study revealed the presence of the Flag-2 locus in *E. albertii*. The Flag-2 locus appears to be widespread among the *Enterobacterales* (De Maayer et al., 2020) and serve as a hot spot for gene insertions and deletions. Consistently, great sequence variation was observed among the five Flag-2 loci identified, including large deletions, point mutations and transposon insertions. However, several VR1 genes that are predicated on having a role in posttranslational regulation of flagellar biosynthesis are conserved in *E. albertii*, including the glycosyltransferase gene and the lysine-N-methylase gene. Unlike the primary flagellar system, the function of the Flag-2 locus is not fully understood. Expression of Flag-2 genes was observed in *Yersinia enterocolitica* with a maximal level at 20°C and, in *Plesiomonas shigelloides,* the Flag-2 locus encoded lateral flagella appeared to be essential for swarming motility (Bresolin et al., 2008; Merino et al., 2015). Systematic analyses of cargo genes located in the VR2 in Flag-2 loci suggested a role in secretion of virulence factor and in inter-bacterial competition (De Maayer et al., 2020). Searching other *E. albertii* genomes deposited in public databases as of April 2024 revealed that about 25% of genomes carry a Flag-2 locus. Additional studies are needed to elucidate any physiological roles or ecological benefits conferred by this secondary flagellar system in *E. albertii*.

Among the 12 fimbriae/pili that are commonly present in *E. coli*, genes encoding curli fimbriae, hemorrhagic *E. coli* pilus, type 1 fimbriae, and Sfm fimbriae were identified in most of the *E. albertii* strains examined, while genes related to biogenesis of adhesive fimbriae, or P fimbriae were only present in a subset of strains. Curli, also known as bacterial amyloid, is an important colonization factor involved in initial surface attachment, biofilm formation, and induction of the host inflammatory response (Barnhart and Chapman, 2006). Although 19 out of the 20 *E. albertii* strains examined in our study carried intact curli genes, production of curli fimbriae was detected in only nine strains. This strain variation could not be explained solely by the differences in the coding sequences of the curli genes or the differences in the intergenic regions between the two curli operons, including the promoters of *csgD* and *csgB*, since some curli-deficient strains shared the identical intergenic sequences with the curli expressing strains (Data not shown). Strain variation in curli production were reported in *E. coli* and *Salmonella enterica* (Romling et al., 1998; Dyer et al., 2007), which both have served as the model organisms for studying curli biogenesis and regulation. In both *E. coli* and *S. enterica*, expression of curli is regulated by a complex regulation network involving multiple transcriptional regulators, two-component regulatory systems, and in some isolates cyclic dinucleotide 3',5'-cyclic di-GMP (Barnhart and Chapman, 2006; Blomfield and van der Woude, 2007). Therefore, mutations in any of these regulators or the target sequences that interact with the regulators directly or indirectly could have an impact on the expression of curli fimbriae. For example, mutations in genes encoding the transcriptional regulators *rpoS* or *rcsB* were reported to be the molecular bases of strain variation in curli production in EHEC O157:H7 (Carter et al., 2012; Carter et al., 2014) and

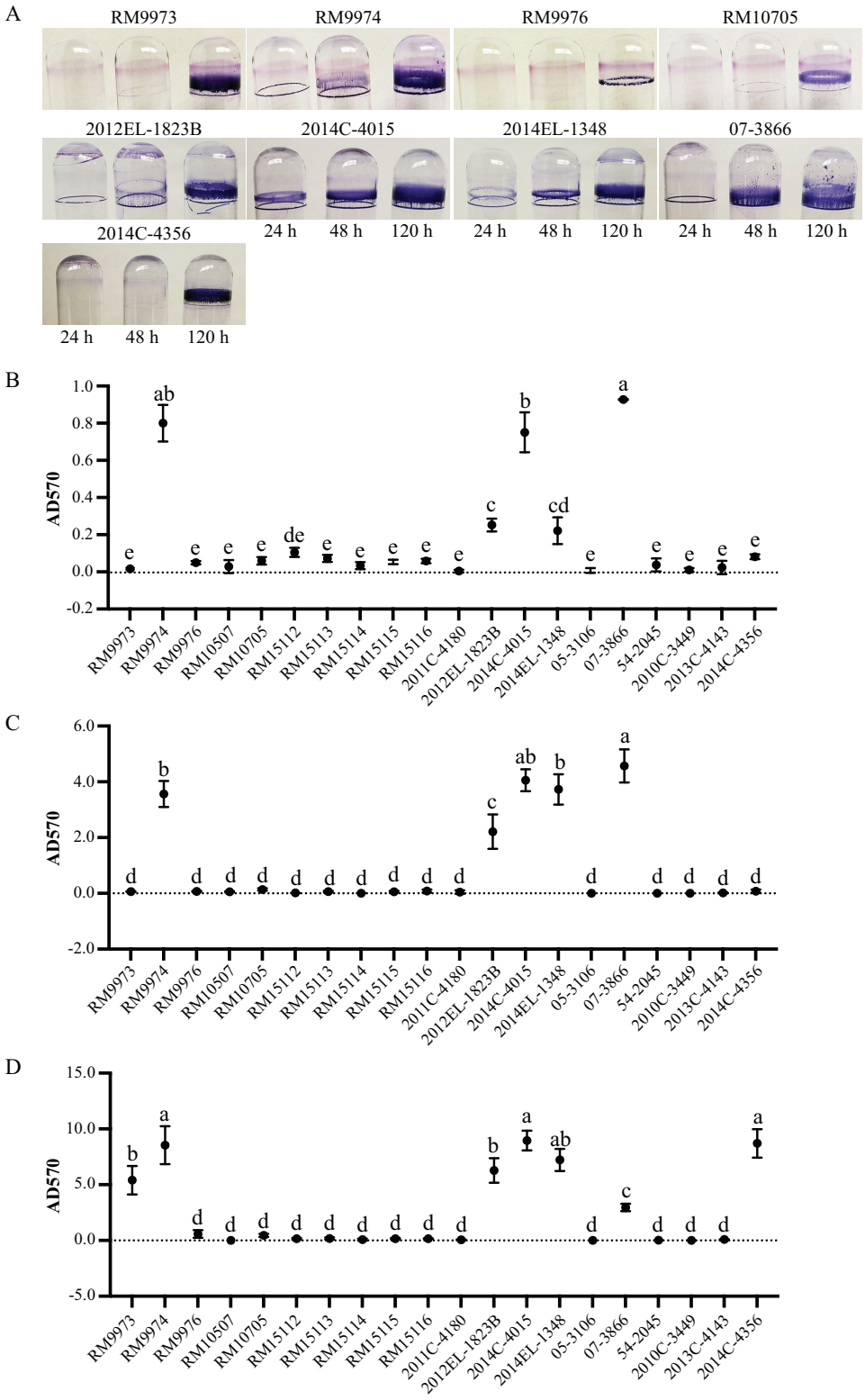


FIGURE 4
Biofilm formation by *E. albertii* strains on glass surfaces. **(A)** Crystal violet staining the attached biomass on the glass surfaces under a static growth condition for 24 h, 48 h, and 120 h. Only strains that can produce visible rings are shown here. **(B–D)** quantitative analyses of biofilms under a static growth condition for 24 h **(B)**, 48 h **(C)**, and 120 h **(D)**. Each data set represents the mean and SD of three biological replicates. Differences that are statistically significant (One-way ANOVA followed by a Tukey's multiple comparisons test, adjust $P < 0.05$) are indicated by different letters. The detailed results of the statistical analyses are presented in Supplementary Table 4.

mutations in the *csgD* promoter could lead to overproduction of curli fimbriae (Uhlich et al., 2002). Knowledge about the regulation of curli fimbriae in *E. albertii* is limited. Examining genes encoding putative transcriptional regulators of the curli genes in *E. albertii* revealed a loss-of-function mutation in the *rpoS gene* in strains 05-3106 and 54-2045, while no mutations were identified in genes encoding Crl, MlrA, CpxRA, OmpR-EnvZ, or RcsBC. Additional studies are required to gain a comprehensive understanding of curli regulation network in *E. albertii* as well as the environmental and physiological signals that may induce or repress the expression of curli fimbriae.

Although the majority of *E. albertii* strains examined in our study carried functional *fim* genes and an intact *fimS*, none of them displayed a mannose-sensitive hemagglutination (MSHA) phenotype under the growth conditions examined. Expression of type 1 fimbriae was reported to be dependent on the growth conditions. For example, optimal production of a predominantly type 1 fimbriae positive population in *Shigella* required serial passage every 48 to 72 h in unshaken brain heart infusion broth at 37°C (Snellings et al., 1997). Since the goal of our study was to reveal if the type 1 fimbriae contributed to the biofilm formation in *E. albertii*, the conditions tested for production of type 1 fimbriae were the conditions used for examining biofilm formation, which may not be optimal for expression of type 1 fimbriae. Additionally, the phase variation of type 1 fimbriation is regulated at multiple levels. In *E. coli*, switch of *fimS* is required but not sufficient for biosynthesis of type 1 fimbriae. Besides FimB and FimE, other transcriptional regulators including IHF, Lrp, and H-NS were reported to be involved in *fimS* switch (Blomfield and van der Woude, 2007). Variations in the activities of FimB and FimE, cross talks between fimbrial operons, as well as the presence of other recombinases can all contribute to variation in expression of type 1 fimbriae. Additional studies are required to understand growth conditions, physiological cues, and environmental signals for induction of type 1 fimbriae in *E. albertii*.

Among the nonfimbrial adhesins examined, genes encoding the autotransporter (AT) adhesins were predominant in *E. albertii*. For example, homologs of *ehaA*, *ehaB*, *ehaC*, and *upaH* that all encode an AIDA-I type autotransporter (AT) adhesin were identified in all or most of the strains examined and a homolog of *ehaG*, encoding a trimeric AT adhesin was identified in all strains. Other AIDA-1 type adhesins genes in *E. albertii* included *aatA*, *aidA*, *agn43*, and *cah*. In *E. coli*, Ag43 is the most prevalent AIDA-1 type AT adhesin, however, in *E. albertii*, the *agn43* was identified only in a clinical strain. Other commonly detected non-fimbrial adhesin genes were *eaeH*, *paa,* and *sinH*. The gene *eaeH* encodes an intimin-like adhesin that facilitates adhesion of bacterial cells, delivery of heat-labile toxin, and colonization of the small intestine in ETEC (Sheikh et al., 2014). The gene *paa*, encoding an AcfC family adhesin, is widespread in both EHEC and ETEC strains. Paa contributes to the formation of A/E lesions in animal hosts and thus is an important virulence factor in various *E. coli* pathotypes (An et al., 1999; Batisson et al., 2003). The *paa* gene appears to be widespread in *E. albertii* and in some strains, there are two *paa* loci, including the bird strains isolated in Poland (BioSample numbers: SAMN33094111, SAMN33094114, SAMN33094099, SAMN33094102, and SAMN33094112), and a poultry strain isolated in China (SAMN17525956). The gene *sinH* encodes an intimin-like inverse autotransporter. The inverse autotransporters were reported to play a role in biofilm formation in *E. coli* and contributed to biofilm formation and virulence in *Yersinia ruckeri* (Martinez-Gil et al., 2017; Goh et al., 2019; Wrobel et al., 2020). However, deletion of *sinH* in the UPEC strain CFT073 did not impact the biofilm formation significantly, rather, the mutant displayed a significant fitness reduction during UTI in a murine model (Shea et al., 2022). Like *paa*, *sinH* appears to be conserved in *E. albertii*. BLASTn search of additional 57 complete *E. albertii* genomes deposited in GenBank as of April 2024 identified a *sinH* in all of them.

Like curli production, *E. albertii* strains differed greatly in biofilm formation on glass surfaces. Consistent with our previous report that, in STEC, curli fimbriae are important for biofilm formation on abiotic surfaces (Carter et al., 2016; Carter et al., 2019), all curli-producing *E. albertii* strains produced moderate or strong biofilms under the condition examined. Interestingly, some curli-positive strains produced visible biofilm following a 24-h incubation, while others did not produce biofilms until a 120-h incubation, implying a difference in biofilm development among the *E. albertii* strains. Furthermore, although no curli fimbriae were detected in chicken isolate 2014C-4356, strong biofilms were observed on day 5 of incubation, suggesting a role of other adhesins in biofilm formation. Strains 2014C-4356 and 07-3866 were the only strains carrying genes encoding the adhesive fimbriae that are located on a large plasmid commonly found in ETEC strains, such as pUMNK88_Hly (GenBank accession # NC_017643.1). It requires further investigation to determine if the plasmid-borne adhesive fimbriae are expressed in *E. albertii* and whether it contributes to biofilm formation in strains that do not produce curli fimbriae.

Biofilms of foodborne pathogens can enhance their survival and persistence in diverse ecological niches and serve as sources of contamination in food production environments and of infection in health-care environments. Adhesion is the first step in biofilm development and in establishing colonization in animal hosts. Strong adherence often implies enhanced surface attachment and biofilm formation, leading to increased fitness and pathogenic potential. Therefore, understanding the adhesion capability and the underlining factors in *E. albertii* would provide valuable information for development of effective control strategies. Our study revealed that curli fimbriae, Type 1 fimbriae, Sfm fimbriae, and HCP appear to be the common fimbrial adhesins in *E. albertii*, while adhesive fimbriae was a strain-specific trait. Among the numerous nonfimbrial adhesins identified in *E. albertii*, autotransporter adhesins EhaA, EhaB, EhaC, EhaG, and SinH, and the adherence factors EaeH and Paa are common, while Agn43, Cah, and Iha that are widespread in *E. coli*, are only associated with a few strains. *E. albertii* strains carry different combinations of fimbrial and nonfimbrial adhesins that may facilitate colonization of *E. albertii* in diverse niches. Our study further revealed great variations in expression of curli fimbriae and in biofilm production, suggesting complex regulation in expression of adhesins in *E. albertii*. Studies are needed to identify environmental cues that induce the adhesions expression and the receptors specifically interacted with each adhesin to gain insight into molecular basis of niche selection for *E. albertii*, an emerging human and avian pathogen.

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/ Table 1.

Author contributions

MC: Conceptualization, Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing. DC: Data curation, Methodology, Validation, Writing – review and editing. RL: Conceptualization, Methodology, Resources, Writing – review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of the article. This work was supported by the USDA-ARS CRIS project 2030-42000-052-000D and the Advanced Molecular Detection (AMD) Initiative grant number AMD-21 at the Centers for Disease Control and Prevention.

Acknowledgments

We wish to thank Nicole Laniohan for providing technical assistance in curli detection and biofilm assay.

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

Author disclaimer

The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention, the Department of Health and Human Services, or the United States government. Furthermore, the use of any product names, trade names, images, or commercial sources is for identification purposes only, and does not imply endorsement or government sanction by the U.S. Department of Health and Human Services.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2024.1456637/full#supplementary-material

SUPPLEMENTARY TABLE 1
E. albertii primary flagellar genes.

SUPPLEMENTARY TABLE 2
E. albertii secondary flagellar genes.

SUPPLEMENTARY TABLE 3
E. albertii fimbrial and non-fimbrial adhesin genes.

SUPPLEMENTARY TABLE 4
Statistical analyses of *E. albertii* biofilms.

References

Abbott, S. L., O'Connor, J., Robin, T., Zimmer, B. L., and Janda, J. M. (2003). Biochemical properties of a newly described *Escherichia species, Escherichia albertii*. *J. Clin. Microbiol.* 41, 4852–4854. doi: 10.1128/JCM.41.10.4852-4854.2003

Abraham, J. M., Freitag, C. S., Clements, J. R., and Eisenstein, B. I. (1985). An invertible element of DNA controls phase variation of type 1 fimbriae of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 82, 5724–5727. doi: 10.1073/pnas.82.17.5724

An, H., Fairbrother, J. M., Desautels, C., and Harel, J. (1999). Distribution of a novel locus called Paa (porcine attaching and effacing associated) among enteric *Escherichia coli*. *Adv. Exp. Med. Biol.* 473, 179–184. doi: 10.1007/978-1-4615-4143-1_17

Arai, S., Yamaya, S., Ohtsuka, K., Konishi, N., Obata, H., Ooka, T., et al. (2022). Detection of *Escherichia albertii* in Retail Oysters. *J. Food Prot.* 85, 173–179. doi: 10.4315/JFP-21-222

Barmettler, K., Biggel, M., Treier, A., Muchaamba, F., and Stephan, R. (2023). Livestock as possible reservoir of *Escherichia albertii* in Switzerland. *Schweiz. Arch. Tierheilkd.* 165, 299–306. doi: 10.17236/sat00393

Barnhart, M. M., and Chapman, M. R. (2006). Curli biogenesis and function. *Annu. Rev. Microbiol* 60, 131–147. doi: 10.1146/annurev.micro.60.080805.142106

Batisson, I., Guimond, M. P., Girard, F., An, H., Zhu, C., Oswald, E., et al. (2003). Characterization of the novel factor paa involved in the early steps of the adhesion mechanism of attaching and effacing *Escherichia coli*. *Infect. Immun* 71, 4516–4525. doi: 10.1128/IAI.71.8.4516-4525.2003

Bengtsson, R. J., Baker, K. S., Cunningham, A. A., Greig, D. R., John, S. K., Macgregor, S. K., et al. (2023). The genomic epidemiology of *Escherichia albertii* infecting humans and birds in Great Britain. *Nat. Commun.* 14:1707. doi: 10.1038/s41467-023-37312-3

Biscola, F. T., Abe, C. M., and Guth, B. E. (2011). Determination of adhesin gene sequences in, and biofilm formation by, O157 and non-O157 Shiga toxin-producing *Escherichia coli* strains isolated from different sources. *Appl. Environ. Microbiol.* 77, 2201–2208. doi: 10.1128/AEM.01920-10

Blomfield, I., and van der Woude, M. (2007). Regulation of fimbrial expression. *EcoSal Plus* 2:2. doi: 10.1128/ecosal.2.4.2.2

Bresolin, G., Trcek, J., Scherer, S., and Fuchs, T. M. (2008). Presence of a functional flagellar cluster Flag-2 and low-temperature expression of flagellar genes in *Yersinia enterocolitica* W22703. *Microbiology (Reading)* 154, 196–206. doi: 10.1099/mic.0.2007/008458-0

Carter, M. Q., Brandl, M. T., Louie, J. W., Kyle, J. L., Carychao, D. K., Cooley, M. B., et al. (2011). Distinct acid resistance and survival fitness displayed by Curli variants of enterohemorrhagic *Escherichia coli* O157:H7. *Appl. Environ. Microbiol.* 77, 3685–3695. doi: 10.1128/AEM.02315-10

Carter, M. Q., Feng, D., and Li, H. H. (2019). Curli fimbriae confer Shiga toxin-producing *Escherichia coli* a competitive trait in mixed biofilms. *Food Microbiol.* 82, 482–488. doi: 10.1016/j.fm.2019.03.024

Carter, M. Q., Louie, J. W., Feng, D., Zhong, W., and Brandl, M. T. (2016). Curli fimbriae are conditionally required in *Escherichia coli* O157:H7 for initial attachment and biofilm formation. *Food Microbiol.* 57, 81–89. doi: 10.1016/j.fm.2016.01.006

Carter, M. Q., Louie, J. W., Huynh, S., and Parker, C. T. (2014). Natural rpoS mutations contribute to population heterogeneity in *Escherichia coli* O157:H7 strains linked to the 2006 US spinach-associated outbreak. *Food Microbiol.* 44, 108–118. doi: 10.1016/j.fm.2014.05.021

Carter, M. Q., Parker, C. T., Louie, J. W., Huynh, S., Fagerquist, C. K., and Mandrell, R. E. (2012). RcsB contributes to the distinct stress fitness among *Escherichia coli* O157:H7 curli variants of the 1993 hamburger-associated outbreak strains. *Appl. Environ. Microbiol.* 78, 7706–7719. doi: 10.1128/AEM.02157-12

Carter, M. Q., Quinones, B., He, X., Pham, A., Carychao, D., Cooley, M. B., et al. (2023a). Genomic and phenotypic characterization of shiga toxin-producing *Escherichia albertii* strains isolated from wild birds in a major agricultural region in California. *Microorganisms* 11:2803. doi: 10.3390/microorganisms11112803

Carter, M. Q., Quinones, B., Laniohan, N., Carychao, D., Pham, A., He, X., et al. (2023b). Pathogenicity assessment of Shiga toxin-producing *Escherichia coli* strains isolated from wild birds in a major agricultural region in California. *Front. Microbiol.* 14:1214081. doi: 10.3389/fmicb.2023.1214081

Colin, R., Ni, B., Laganenka, L., and Sourjik, V. (2021). Multiple functions of flagellar motility and chemotaxis in bacterial physiology. *FEMS Microbiol. Rev.* 45:fuab038. doi: 10.1093/femsre/fuab038

Cookson, A. L., Cooley, W. A., and Woodward, M. J. (2002). The role of type 1 and curli fimbriae of Shiga toxin-producing *Escherichia coli* in adherence to abiotic surfaces. *Int. J. Med. Microbiol.* 292, 195–205. doi: 10.1078/1438-4221-00203

De Maayer, P., Pillay, T., and Coutinho, T. A. (2020). Flagella by numbers: Comparative genomic analysis of the supernumerary flagellar systems among the Enterobacterales. *BMC Genom.* 21:670. doi: 10.1186/s12864-020-07085-w

Dyer, J. G., Sriranganathan, N., Nickerson, S. C., and Elvinger, F. (2007). Curli production and genetic relationships among *Escherichia coli* from cases of bovine mastitis. *J. Dairy Sci.* 90, 193–201. doi: 10.3168/jds.S0022-0302(07)72620-6

Elpers, L., and Hensel, M. (2020). Expression and functional characterization of various chaperon-usher fimbriae, Curli Fimbriae, and Type 4 Pili of enterohemorrhagic *Escherichia coli* O157:H7 Sakai. *Front. Microbiol.* 11:378. doi: 10.3389/fmicb.2020.00378

Felfoldi, T., Heeger, Z., Vargha, M., and Marialigeti, K. (2010). Detection of potentially pathogenic bacteria in the drinking water distribution system of a hospital in Hungary. *Clin. Microbiol. Infect.* 16, 89–92. doi: 10.1111/j.1469-0691.2009.02795.x

Fink, R. C., Black, E. P., Hou, Z., Sugawara, M., Sadowsky, M. J., and Diez-Gonzalez, F. (2012). Transcriptional responses of *Escherichia coli* K-12 and O157:H7 associated with lettuce leaves. *Appl. Environ. Microbiol.* 78, 1752–1764. doi: 10.1128/AEM.07454-11

Goh, K. G. K., Moriel, D. G., Hancock, S. J., Phan, M. D., and Schembri, M. A. (2019). Bioinformatic and molecular analysis of inverse autotransporters from *Escherichia coli*. *mSphere* 4, e00572–19. doi: 10.1128/mSphere.00572-19

Gomes, T. A. T., Ooka, T., Hernandes, R. T., Yamamoto, D., and Hayashi, T. (2020). *Escherichia albertii* pathogenesis. *EcoSal Plus* 9:15. doi: 10.1128/ecosalplus.ESP-0015-2019

Gophna, U., Barlev, M., Seijffers, R., Oelschlager, T. A., Hacker, J., and Ron, E. Z. (2001). Curli fibers mediate internalization of *Escherichia coli* by eukaryotic cells. *Infect. Immun.* 69, 2659–2665. doi: 10.1128/IAI.69.4.2659-2665.2001

Hinenoya, A., Li, X. P., Zeng, X., Sahin, O., Moxley, R. A., Logue, C. M., et al. (2021). Isolation and characterization of *Escherichia albertii* in poultry at the pre-harvest level. *Zoon. Public Health* 68, 213–225. doi: 10.1111/zph.12812

Hinenoya, A., Wang, H., Patrick, E. M., Zeng, X., Cao, L., Li, X. P., et al. (2022). Longitudinal surveillance and comparative characterization of *Escherichia albertii* in wild raccoons in the United States. *Microbiol. Res.* 262:127109. doi: 10.1016/j.micres.2022.127109

Hirose, S., Konishi, N., Sato, M., Suzumura, K., Obata, H., Ohtsuka, K., et al. (2024). Growth and survival of *Escherichia albertii* in food and environmental water at various temperatures. *J. Food Prot.* 87:100249. doi: 10.1016/j.jfp.2024.100249

Iguchi, A., Takemura, T., Ogura, Y., Nguyen, T. T. H., Kikuchi, T., Okuno, M., et al. (2023). Genomic characterization of endemic diarrheagenic *Escherichia coli* and *Escherichia albertii* from infants with diarrhea in Vietnam. *PLoS Negl. Trop. Dis.* 17:e0011259. doi: 10.1371/journal.pntd.0011259

Iida, K., Mizunoe, Y., Wai, S. N., and Yoshida, S. (2001). Type 1 fimbriation and its phase switching in diarrheagenic *Escherichia coli* strains. *Clin. Diagn. Lab. Immunol.* 8, 489–495. doi: 10.1128/CDLI.8.3.489-495.2001

Ikeda, T., Shinagawa, T., Ito, T., Ohno, Y., Kubo, A., Nishi, J., et al. (2020). Hypoosmotic stress induces flagellar biosynthesis and swimming motility in *Escherichia albertii*. *Commun. Biol.* 3:87. doi: 10.1038/s42003-020-0816-5

Karatan, E., and Watnick, P. (2009). Signals, regulatory networks, and materials that build and break bacterial biofilms. *Microbiol. Mol. Biol. Rev.* 73, 310–347. doi: 10.1128/MMBR.00041-08

Khan, F., Tabassum, N., Pham, D. T. N., Oloketuyi, S. F., and Kim, Y. M. (2020). Molecules involved in motility regulation in *Escherichia coli* cells: A review. *Biofouling* 36, 889–908. doi: 10.1080/08927014.2020.1826939

Konno, T., Yatsuyanagi, J., Takahashi, S., Kumagai, Y., Wada, E., Chiba, M., et al. (2012). Isolation and identification of *Escherichia albertii* from a patient in an outbreak of gastroenteritis. *Jpn. J. Infect. Dis.* 65, 203–207. doi: 10.7883/yoken.65.203

Lima, M. P., Yamamoto, D., Santos, A. C. M., Ooka, T., Hernandes, R. T., Vieira, M. A. M., et al. (2019). Phenotypic characterization and virulence-related properties of *Escherichia albertii* strains isolated from children with diarrhea in Brazil. *Pathog. Dis.* 77:ftz014. doi: 10.1093/femspd/ftz014

Lindsey, R. L., Fedorka-Cray, P. J., Abley, M., Turpin, J. B., and Meinersmann, R. J. (2015). Evaluating the occurrence of *Escherichia albertii* in chicken carcass rinses by PCR, Vitek analysis, and sequencing of the rpoB gene. *Appl. Environ. Microbiol.* 81, 1727–1734. doi: 10.1128/AEM.03681-14

Maeda, E., Murakami, K., Sera, N., Ito, K., and Fujimoto, S. (2015). Detection of *Escherichia albertii* from chicken meat and giblets. *J. Vet. Med. Sci.* 77, 871–873. doi: 10.1292/jvms.14-0640

Maheux, A. F., Boudreau, D. K., Bergeron, M. G., and Rodriguez, M. J. (2014). Characterization of *Escherichia fergusonii* and *Escherichia albertii* isolated from water. *J. Appl. Microbiol.* 117, 597–609. doi: 10.1111/jam.12551

Martinez-Gil, M., Goh, K. G. K., Rackaityte, E., Sakamoto, C., Audrain, B., Moriel, D. G., et al. (2017). YeeJ is an inverse autotransporter from *Escherichia coli* that binds to peptidoglycan and promotes biofilm formation. *Sci. Rep.* 7:11326. doi: 10.1038/s41598-017-10902-0

Masuda, K., Ooka, T., Akita, H., Hiratsuka, T., Takao, S., Fukada, M., et al. (2020). Epidemiological aspects of *Escherichia albertii* Outbreaks in Japan and genetic characteristics of the causative pathogen. *Foodborne Pathog. Dis* 17, 144–150. doi: 10.1089/fpd.2019.2654

McWilliams, B. D., and Torres, A. G. (2014). Enterohemorrhagic *Escherichia coli* Adhesins. *Microbiol. Spectr.* 2:3. doi: 10.1128/microbiolspec.EHEC-0003-2013

Merino, S., Aquilini, E., Fulton, K. M., Twine, S. M., and Tomas, J. M. (2015). The polar and lateral flagella from *Plesiomonas shigelloides* are glycosylated with legionaminic acid. *Front. Microbiol.* 6:649. doi: 10.3389/fmicb.2015.00649

Moens, S., and Vanderleyden, J. (1996). Functions of bacterial flagella. *Crit. Rev. Microbiol.* 22, 67–100. doi: 10.3109/10408419609106456

Muchaamba, F., Barmettler, K., Treier, A., Houf, K., and Stephan, R. (2022). Microbiology and epidemiology of *Escherichia albertii*-an emerging elusive foodborne pathogen. *Microorganisms* 10:875. doi: 10.3390/microorganisms10050875

Murakami, K., Kimura, S., Nagafuchi, O., Sekizuka, T., Onozuka, D., Mizukoshi, F., et al. (2020). Flagellum expression and swimming activity by the zoonotic pathogen *Escherichia albertii*. *Environ. Microbiol. Rep.* 12, 92–96. doi: 10.1111/1758-2229.12818

Oaks, J. L., Besser, T. E., Walk, S. T., Gordon, D. M., Beckmen, K. B., Burek, K. A., et al. (2010). *Escherichia albertii* in wild and domestic birds. *Emerg. Infect. Dis.* 16, 638–646. doi: 10.3201/eid1604.090695

Ooka, T., Ogura, Y., Katsura, K., Seto, K., Kobayashi, H., Kawano, K., et al. (2015). Defining the genome features of *Escherichia albertii*, an emerging enteropathogen closely related to *Escherichia coli*. *Genome Biol. Evol.* 7, 3170–3179. doi: 10.1093/gbe/evv211

Ooka, T., Seto, K., Kawano, K., Kobayashi, H., Etoh, Y., Ichihara, S., et al. (2012). Clinical significance of *Escherichia albertii*. *Emerg. Infect. Dis.* 18, 488–492. doi: 10.3201/eid1803.111401

Ooka, T., Tokuoka, E., Furukawa, M., Nagamura, T., Ogura, Y., Arisawa, K., et al. (2013). Human gastroenteritis outbreak associated with *Escherichia albertii*, Japan. *Emerg. Infect. Dis.* 19, 144–146. doi: 10.3201/eid1901.120646

Ori, E. L., Takagi, E. H., Andrade, T. S., Miguel, B. T., Cergole-Novella, M. C., Guth, B. E. C., et al. (2018). Diarrhoeagenic *Escherichia coli* and *Escherichia albertii* in Brazil: Pathotypes and serotypes over a 6-year period of surveillance. *Epidemiol. Infect.* 147:e10. doi: 10.1017/S0950268818002595

O'Toole, G., Kaplan, H. B., and Kolter, R. (2000). Biofilm formation as microbial development. *Annu. Rev. Microbiol.* 54, 49–79. doi: 10.1146/annurev.micro.54.1.49

Pratt, L. A., and Kolter, R. (1998). Genetic analysis of *Escherichia coli* biofilm formation: Roles of flagella, motility, chemotaxis and type I pili. *Mol. Microbiol.* 30, 285–293. doi: 10.1046/j.1365-2958.1998.01061.x

Reisner, A., Haagensen, J. A., Schembri, M. A., Zechner, E. L., and Molin, S. (2003). Development and maturation of *Escherichia coli* K-12 biofilms. *Mol. Microbiol.* 48, 933–946. doi: 10.1046/j.1365-2958.2003.03490.x

Ren, C. P., Beatson, S. A., Parkhill, J., and Pallen, M. J. (2005). The Flag-2 locus, an ancestral gene cluster, is potentially associated with a novel flagellar system from *Escherichia coli*. *J. Bacteriol.* 187, 1430–1440. doi: 10.1128/JB.187.4.1430-1440.2005

Roe, A. J., Currie, C., Smith, D. G., and Gally, D. L. (2001). Analysis of type 1 fimbriae expression in verotoxigenic *Escherichia coli*: A comparison between serotypes O157 and O26. *Microbiology (Reading)* 147, 145–152. doi: 10.1099/00221287-147-1-145

Romling, U., Sierralta, W. D., Eriksson, K., and Normark, S. (1998). Multicellular and aggregative behaviour of *Salmonella typhimurium* strains is controlled by mutations in the agfD promoter. *Mol. Microbiol.* 28, 249–264. doi: 10.1046/j.1365-2958.1998.00791.x

Shea, A. E., Stocki, J. A., Himpsl, S. D., Smith, S. N., and Mobley, H. L. T. (2022). Loss of an intimin-like protein encoded on a uropathogenic *E. coli* pathogenicity island reduces inflammation and affects interactions with the urothelium. *Infect. Immun.* 90:e0027521. doi: 10.1128/IAI.00275-21

Sheikh, A., Luo, Q., Roy, K., Shabaan, S., Kumar, P., Qadri, F., et al. (2014). Contribution of the highly conserved EaeH surface protein to enterotoxigenic *Escherichia coli* pathogenesis. *Infect. Immun.* 82, 3657–3666. doi: 10.1128/IAI.01890-14

Snellings, N. J., Tall, B. D., and Venkatesan, M. M. (1997). Characterization of *Shigella* type 1 fimbriae: Expression, FimA sequence, and phase variation. *Infect. Immun.* 65, 2462–2467. doi: 10.1128/iai.65.6.2462-2467.1997

Uhlich, G. A., Keen, J. E., and Elder, R. O. (2002). Variations in the csgD promoter of *Escherichia coli* O157:H7 associated with increased virulence in mice and increased invasion of HEp-2 cells. *Infect. Immun.* 70, 395–399. doi: 10.1128/IAI.70.1.395-399.2002

Vogeleer, P., Tremblay, Y. D., Mafu, A. A., Jacques, M., and Harel, J. (2014). Life on the outside: Role of biofilms in environmental persistence of Shiga-toxin producing *Escherichia coli*. *Front. Microbiol.* 5:317. doi: 10.3389/fmicb.2014.00317

Walk, S. T., Alm, E. W., Gordon, D. M., Ram, J. L., Toranzos, G. A., Tiedje, J. M., et al. (2009). Cryptic lineages of the genus *Escherichia*. *Appl. Environ. Microbiol.* 75, 6534–6544. doi: 10.1128/AEM.01262-09

Wang, H., Li, Q., Bai, X., Xu, Y., Zhao, A., Sun, H., et al. (2016). Prevalence of eae-positive, lactose non-fermenting *Escherichia albertii* from retail raw meat in China. *Epidemiol. Infect.* 144, 45–52. doi: 10.1017/S0950268815001120

Wang, H., Zhang, L., Cao, L., Zeng, X., Gillespie, B., and Lin, J. (2022). Isolation and characterization of *Escherichia albertii* originated from the broiler farms in Mississippi and Alabama. *Vet. Microbiol.* 267:109379. doi: 10.1016/j.vetmic.2022.109379

Watnick, P., and Kolter, R. (2000). Biofilm, city of microbes. *J. Bacteriol.* 182, 2675–2679. doi: 10.1128/JB.182.10.2675-2679.2000

Wrobel, A., Saragliadis, A., Perez-Ortega, J., Sittman, C., Gottig, S., Liskiewicz, K., et al. (2020). The inverse autotransporters of Yersinia ruckeri, YrInv and YrIlm, contribute to biofilm formation and virulence. *Environ. Microbiol.* 22, 2939–2955. doi: 10.1111/1462-2920.15051

Xicohtencatl-Cortes, J., Monteiro-Neto, V., Ledesma, M. A., Jordan, D. M., Francetic, O., Kaper, J. B., et al. (2007). Intestinal adherence associated with type IV pili of enterohemorrhagic *Escherichia coli* O157:H7. *J. Clin. Invest.* 117, 3519–3529. doi: 10.1172/JCI30727

Xicohtencatl-Cortes, J., Monteiro-Neto, V., Saldana, Z., Ledesma, M. A., Puente, J. L., and Giron, J. A. (2009). The type 4 pili of enterohemorrhagic *Escherichia coli* O157:H7 are multipurpose structures with pathogenic attributes. *J. Bacteriol.* 191, 411–421. doi: 10.1128/JB.01306-08

Xu, B., Hatanaka, N., Awasthi, S. P., Hinenoya, A., and Yamasaki, S. (2024). Seasonality of detection rate of *Escherichia albertii* in wild raccoons (*Procyon lotor*) in Osaka, Japan. *J. Vet. Med. Sci.* 86, 180–183. doi: 10.1292/jvms.23-0372

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 15 October 2024
DOI 10.3389/fmicb.2024.1459840

OPEN ACCESS

EDITED BY
Xuanyu Tao,
University of Oklahoma, United States

REVIEWED BY
Hui Li,
University of Oklahoma, United States
Ricardo Oliveira,
National Institute for Agrarian and Veterinary Research (INIAV), Portugal
Jihong Li,
University of Pittsburgh, United States

*CORRESPONDENCE
Päivi Lahti
paivi.lahti@helsinki.fi

†These authors have contributed equally to this work

RECEIVED 05 July 2024
ACCEPTED 16 September 2024
PUBLISHED 15 October 2024

CITATION
Lahti P, Jaakkola K, Hörman A, Heikinheimo A, Sovijärvi A and Korkeala H (2024) Identification of a new *Clostridium perfringens* variant with a chromosomally encoded enterotoxin gene in a suspected persistent food poisoning outbreak in Eritrea. *Front. Microbiol.* 15:1459840. doi: 10.3389/fmicb.2024.1459840

Identification of a new *Clostridium perfringens* variant with a chromosomally encoded enterotoxin gene in a suspected persistent food poisoning outbreak in Eritrea

Päivi Lahti[1*†], Kaisa Jaakkola[1†], Ari Hörman[1,2], Annamari Heikinheimo[1,3], Ava Sovijärvi[2,4] and Hannu Korkeala[1]

[1]Department of Food Hygiene and Environmental Health, University of Helsinki, Helsinki, Finland, [2]Medical Services, Defence Command, The Finnish Defence Forces, Helsinki, Finland, [3]Microbiology Unit, Finnish Food Authority, Seinäjoki, Finland, [4]Retired, Turku, Finland

Clostridium perfringens is a causative agent of various human and animal enteric diseases including food poisoning. In this study, we describe an interesting case of a persistent food poisoning outbreak among Finnish peacekeepers in Eritrea, possibly caused by *Clostridium perfringens* carrying a new variant of the chromosomally encoded enterotoxin gene. *C. perfringens* strains causing food poisoning carry the enterotoxin gene, *cpe*, in its chromosome (c-*cpe*) or on a plasmid (p-*cpe*). PCR assays are widely used for toxinotype *C. perfringens* strains. The integration sites for the *cpe* gene are highly conserved, and PCR assays targeting the *cpe* gene and the adjacent IS elements (the IS*1470* in c-*cpe* and the IS*1470*-like or IS*1151* in p-*cpe* strains) are used to further determine the genetic location of the *cpe* gene. We sequenced nine enteropathogenic *C. perfringens* strains related to a persistent food poisoning outbreak among Finnish peacekeepers in Eritrea. Six of these strains produced non-typeable *cpe* results in the standard PCR assay due to divergence in the enterotoxin integration site. The gene order of the new variant of the chromosomal *cpe* insertion site with an additional IS1470 element impairing genotyping PCR assay for the location of *cpe* is described. In addition, variant c-*cpe* strains carried 58–81 copies of IS1470 in their genomes, compared to 9–23 copies found in previously described c-*cpe* strains. Thus, the present study represents an untraditional type of *C. perfringens* food poisoning caused by variant c-*cpe* strains, and the sequenced strains bring geographic variation to the existing strain collection of sequenced *C. perfringens*.

KEYWORDS

***Clostridium perfringens*, genotyping, variant strains, enterotoxin gene, food poisoning, persistent outbreak, IS element**

1 Introduction

Clostridium perfringens is an anaerobic, spore former causing gas gangrene, wound infections, and a variety of human and animal diseases involving the gastrointestinal (GI) system (Lindström et al., 2011; Kiu and Hall, 2018). Traditional *C. perfringens* food poisoning outbreaks are associated with temperature-abused food and typically affect many people at the same time. In addition,

cases of antibiotic-associated diarrhea (AAD) by *C. perfringens* may be transmitted through food (Lindström et al., 2011).

Both food poisoning and AAD cases of *C. perfringens* diarrhea are primarily caused by type F strains (strains carrying enterotoxin gene *cpe*), which produce pore-forming *C. perfringens* enterotoxin (CPE). The prevailing understanding is that all *C. perfringens* strains may carry the *cpe* gene, but only approximately 5% do and produce CPE (Miyamoto et al., 2006). The *cpe* allele encoding a 319 aa end product is located either in pCPF5603 or pCPF4969 plasmid (plasmid-mediated *cpe,* p-*cpe* strains) or in a transposable element Tn5565 integrated into the chromosome (chromosomal *cpe*, c-*cpe* strains) (Cornillot et al., 1995; Brynestad et al., 1997; Miyamoto et al., 2006; Li et al., 2010). The Tn5565 includes insertion sequences (IS) IS1470 and IS1469 directly adjacent to the *cpe* gene (Brynestad et al., 1997). Known *cpe*-carrying plasmids are all conjugative and horizontally transferable (Miyamoto et al., 2006). The plasmid *cpe* gene is adjacent to IS1151- or IS1470-like elements (Daube et al., 1993; Miyamoto et al., 2002).

Foodborne outbreaks are caused by both p-*cpe* and c-*cpe* strains (Lahti et al., 2008), but multiple studies have shown that most outbreaks are caused by chromosomal *cpe* strains. In addition, a 325 aa variant of the *cpe* gene of unknown clinical relevance has been described in the pCPBB1 plasmid (Miyamoto et al., 2011). Persistent longitudinal outbreaks of *C. perfringens* have been suggested in some studies (Kiu et al., 2019), and these outbreaks have been associated with plasmidial *cpe*-carrying strains.

Six toxin genes detected by PCR have been widely used for toxinotype strains of types A–G (Rood et al., 2018). The only invariably chromosomal gene in the toxinotyping scheme is alpha (*plc*), which is present in all *C. perfringens* strains, while the other toxinotyping toxins are carried on transposable elements or a family of conjugative plasmids (Li et al., 2013). Toxinotyping, therefore, does not reflect the phylogenetic lineage of strains. To address this, the availability of genomic sequences has led to the establishment of genetic lineages (Kiu et al., 2017; Feng et al., 2020; Jaakkola et al., 2021), and a virulence gene profile scheme including chromosomal genes has been proposed (Abdelrahim et al., 2019).

PCR genotyping is also used to detect the location and type of enterotoxin gene in enteropathogenic *C. perfringens*. Despite the transmissible nature of these genetic structures, the gene order on both chromosomal and plasmidial integration sites is so highly conserved that IS elements are utilized as PCR probe targets (Brynestad, 1997), and reported variant strains have been related to the presence of a larger variant of the enterotoxin gene itself (325 aa) (Heikinheimo et al., 2006; Li et al., 2007).

In this study, we describe a persistent food poisoning outbreak putatively caused by Eritrean *C. perfringens* strains. In addition, we describe six outbreak strains with chromosomal *cpe* that exhibit an additional IS1470 insertion sequence near the *cpe* gene, which affects the ability of the PCR assay to detect the *cpe* location and genotyping results.

2 Materials and methods

2.1 Outbreak and stool samples

The occurrence of gastroenteritis among Finnish peacekeepers deployed in the United Nations peacekeeping operation in Eritrea in Africa (UNMEE) increased during the summer and autumn of 2004. In the autumn of 2004, an investigation team from Finland was sent to Eritrea to investigate the possible outbreak, take measures, and give instructions to address the situation. Inspections of the camp food premises, water distribution system, and local water suppliers were conducted. Water samples were collected and analyzed for the total bacterial count, total coliforms, *E. coli*, parasites (*Giardia, Cryptosporidium,* and *Entamoeba histolytica*), and noroviruses using the standard drinking water methods. No *E. coli*, parasites, or noroviruses were detected in any of the 13 samples.

Of the 184 peacekeepers, 163 (91.3%) answered the epidemiological questionnaire and 98 (60.1%) of them had suffered from symptoms such as diarrhea (93.9%), flatulence (69.4%), abdominal pain (58.2%), nausea without vomiting (51.5%), lack of appetite (43.9%), abdominal distention (33.7%), fever (31.6%), vomiting (22.4%), muscular pain (17.3%), abdominal cramps and other symptoms (16.3%), and bloody diarrhea (7.1%). The illnesses occurred consistently from June to October, with no distinct peak in their occurrence. The analysis of questionnaires and medical reports revealed that on average, approximately five peacekeepers were sick each day from June to October.

Stool specimens were collected from 184 Finnish peacekeepers (sample numbers 1–184) and 38 local Eritrean staff (cooks and cleaners) members (sample numbers 200–237) and transported to the Laboratory of Helsinki University Central Hospital. *Salmonella*, *Shigella*, *Yersinia*, and *Campylobacter*, parasites (*Cryptosporidium*, *Cyclospora*, *Cystoisospora,* and *Dientamoeba*), norovirus and astrovirus, and antigens for *Giardia*, *Cryptosporidium*, and *Entamoeba histolytica* were investigated from the stool samples using standardized methods of the laboratory. The epidemiological and laboratory investigations revealed no specific bacteria, virus, or parasite connected to the illnesses. However, the symptoms resembled those of *Clostridium perfringens* food poisoning, and the fecal samples were taken for further examination.

2.2 Detection and isolation of *Clostridium perfringens*

Each fecal swab sample was dissolved into a tube that contained sterilized water, and the hydrophobic grid membrane filter-colony hybridization (HGMF-CH) (Heikinheimo et al., 2004) was used to detect and isolate *cpe*-carrying *C. perfringens* from the samples. The samples giving a positive signal in the HGMF-CH assay were considered positive for *cpe*-carrying *C. perfringens*. The probe-positive colonies were isolated from each sample to obtain *cpe*-carrying *C. perfringens* from a single sample and to further study the genetic relatedness of these isolates.

2.3 PCR and PFGE typing

We detected major toxins and *cpe* in the *C. perfringens* isolates with PCR, as previously described by Heikinheimo and Korkeala (2005) in 2004–2005. The current nomenclature was used in this article (Rood et al., 2018). *C. perfringens* strains NCTC 8239, ATCC 3626, CCUG 2036, CCUG 2037, and CCUG 44727 were used as positive controls. *C. perfringens* type F isolates were further studied by PCR to determine the *cpe* genotype based on the *cpe* insertion site

with IS elements. The total DNA was isolated by using Advamax beads (Edge Biosystems, Gaithersburg, MD, USA), according to the manufacturer's instructions. IS elements downstream of *cpe* that determine the *cpe* genotype (IS1151, IS1470-like, or IS1470) of each isolate were characterized by using PCR with the previously described primers (Daube et al., 1993; Brynestad, 1997; Brynestad et al., 1997; Miyamoto et al., 2002; Miyamoto et al., 2004) and protocols (Heikinheimo et al., 2006).

In pulsed-field gel electrophoresis (PFGE) analysis, the DNA was digested with ApaI (New England Biolabs, Beverly, MA, USA), and the genetic relationships between isolates were assessed using the previously described assay (Ridell et al., 1998), which was modified by adding thiourea to the electrophoresis running buffer (Leclair et al., 2006). The digital images of PFGE patterns were analyzed using Bionumerics software (version 4.6, Applied Maths, Sint-Martens-Latem, Belgium), and the similarity analysis of PFGE patterns was performed using the Dice coefficient (optimized 2%, tolerance 1.2%). Clustering and construction of dendrograms were performed by using the unweighted pair-group method with arithmetic averages.

2.4 Sequencing of selected *Clostridium perfringens* isolates and annotation

Nine strains were sequenced in 2022 and have been deposited in the GenBank. These nine strains were selected for sequencing based on the genotyping results, and they represented both chromosomal and variant chromosomal *cpe*-carrying isolates in the genotyping assay.

Genomic DNA of *C. perfringens* isolates was extracted, as described by Keto-Timonen et al. (2006), and whole-genome sequencing was performed using PacBio RSII (Institute of Biotechnology, Helsinki, Finland). Sequenced genomes were assembled using HGAP3 and checked for circularity using Gap4 (Staden et al., 2003; Chin et al., 2013). To improve the draft assembly, Illumina MiSeq reads and the Pilon tool were used for genome polishing (Walker et al., 2014). Both sequenced and downloaded genomes were annotated using Prokka (Seemann, 2014).

Selected reference strains (ATCC 13124, SM101, and Str. 13) and previously sequenced lineage IV strains were included in the cgMLST analysis (Supplementary Table S1). Genomes were downloaded from the Bacterial and Viral Bioinformatics Resource Center.[1]

2.5 Comparative genome analysis

Bacterial and Viral Bioinformatics Resource Center (see text footnote 1) was used to perform comparative genome analysis and to combine results with previous studies (Jaakkola et al., 2021). To determine the cgMLST target gene set and create a genome-wide gene-by-gene comparison, ChewBBACA (3.0.0) was used (Silva et al., 2018). A schema created by Abdel-Glil et al. (2021) available at https://www.cgmlst.org was used. The core genome of 63 *C. perfringens* genomes consisted of 1,236 genes. For cgMLST results, a minimal spanning tree was calculated by GrapeTree (Zhou et al., 2018) using MSTreeV2. Strains ($n = 63$) included in the comparative genome analysis are listed in Supplementary Table S1.

Sequenced genomes were queried for the presence of IS elements (IS1470, IS1469, and IS1151) and selected genetic features (Supplementary Table S2). For IS elements, the hits with >80% identity over 80% of length were considered a match. For other genetic features, an identity threshold of 90% was used.

2.6 Phylogenetic analysis

Single-nucleotide polymorphisms between sequenced strains and selected reference strain genomes were identified using Snippy 4.6. (Seemann et al., 2015), with strain SM101 (289380.15) as a reference. A phylogeny based on core-SNP alignment was created by IQ-TREE 2.3.0. (Nguyen et al., 2015). Bootstrap values for branches were approximated using ultrafast bootstrapping (-B 1000), and FigTree v1.4.4. was used to visualize the trees (Rambaut and Drummond, 2008). Strains ($n = 63$) included in phylogenetic analysis are listed in Supplementary Table S1.

3 Results

3.1 *cpe*-positive *Clostridium perfringens* samples

Altogether, 50 of 222 (22.5%) human stool samples gave a positive signal in the HGMF-CH and were considered positive for the presence of *cpe*-carrying *C. perfringens*. The prevalence of type F *C. perfringens* differed between the samples from peacekeepers and local staff members. Overall, 32 of 184 (17.4%) of the samples of the peacekeepers and 18 of 38 (47.3%) of the samples of local staff members were positive. Altogether, we isolated 96 *C. perfringens* isolates from 12 samples (Table 1). The isolates were regarded as *cpe*-positive since they gave a signal in the HGMF-CH.

Of the peacekeepers who gave *cpe*-positive or *cpe*-negative samples, 72 and 59%, respectively, had symptoms of intestinal disease between June and October 2004.

3.2 Genotyping results

Based on multiplex PCR, 94 (98%) of the 96 isolates were *cpe*-carrying type F (former type A). None of the 94 *cpe*-carrying isolates carried plasmidial *cpe*. PCR genotyping identified 49 (52%) of the 94 isolates as typical c-*cpe* isolates. The remaining 45 (48%) of the 94 isolates yielded PCR products of divergent size compared to c-*cpe*; thus, the genetic location of the *cpe* gene of these 45 isolates was interpreted as a variant. The PCR results are shown in Table 1.

We typed 78 (83%) of 94 *cpe*-positive isolates with PFGE. Isolates from four peacekeepers (samples 33, 44, 53, and 103) displayed four distinct patterns (Figure 1). Isolates from a fifth peacekeeper (sample 24) and four locally employed workers (203, 205, 206, and 233) displayed closely related patterns and included both c-*cpe* and variant isolates. Among this cluster, a subcluster of indistinguishable isolates from the fifth peacekeeper and three local staff members was selected for further analysis.

1 https://www.bv-brc.org/

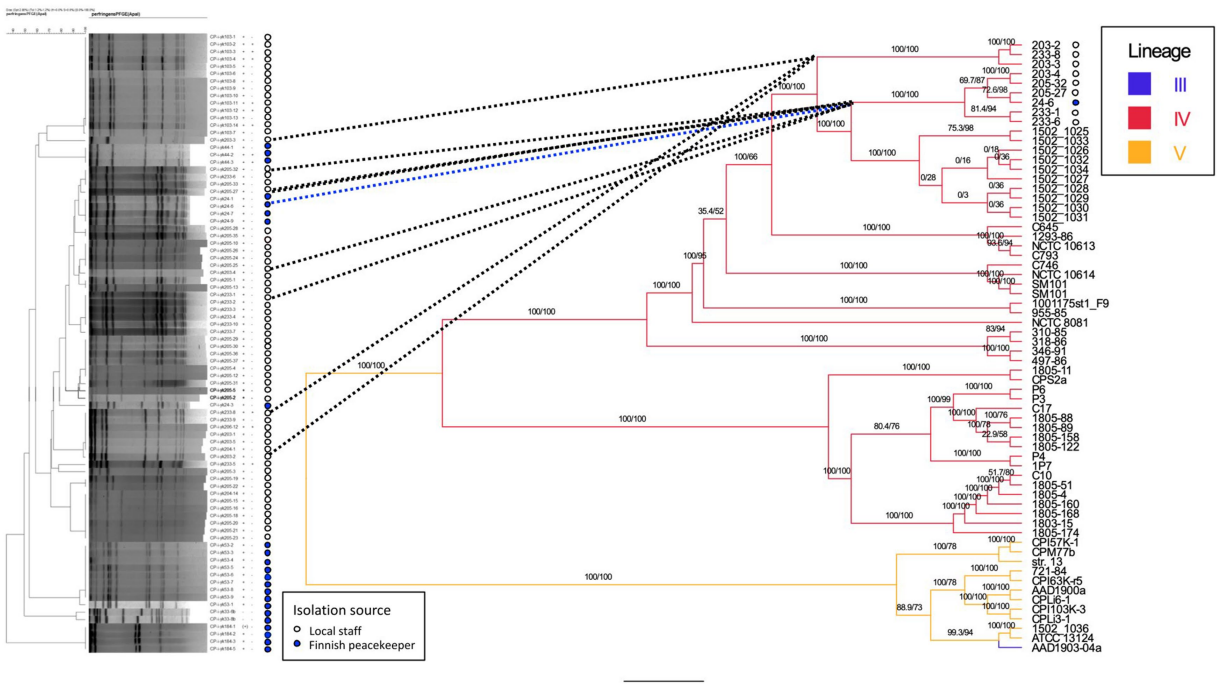


FIGURE 1
Phylogenetic tree of sequenced strains and previously published strains. Comparison is made between the dendrogram (Dice coefficient with 2%, UPGMA clustering, 84 isolates) of PFGE results (left) and the phylogenetic tree (IQ-TREE 2.3.0, visualized by FigTree v1.4.4., 63 isolates) based on whole-genome sequences (center, right). For the strains isolated in this study, the isolation source (local staff, peacekeeper) is annotated next to the PFGE tree, and bootstrapping values for tree nodes are given. For sequenced genomes and previously published genomes, the genetic lineages (III, IV, and V) of strains are indicated on the right. Nomenclature of strains: CP-i-yk was used as the study ID in the PFGE analysis. The numbers and hyphen (4–6 characters) following the study ID constitute the strain/genome ID. Reference strain = *C. perfringens* strain SM101.

3.3 Sequenced genomes

Nine isolates representing both c-*cpe* typed isolates and PCR variant *cpe* isolates were sequenced (Table 2). Sequencing revealed that all carried a chromosomally inserted *cpe* and they, despite the varied PCR results and the differences observed in PFGE, belonged to the same phylogenetic lineage IV and phylogenetically clustered together with c-*cpe* food poisoning isolates (Figure 2). Different c-*cpe* groups, 1 and 2, have been recently suggested (Jaakkola et al., 2021). C-*cpe* group 1 strains seem to be equipped for changing pH and acidic, high-temperature environments where iron uptake is competitive, and citrate utilization is beneficial, whereas strains of c-*cpe* group 2 lack these genes and operons. The isolates described here belonged to c-*cpe* group 1, among many well-researched food poisoning strains such as SM101, NCTC 8239, and NCTC 10613.

All sequenced *cpe* genes were highly conserved and encoded a 319 aa enterotoxin sequence. Genome analysis revealed that the six strains in Table 2 with variant *cpe* PCR results (24–6, 203–4, 205–27, 205–32, 233–1, and 233–6) carried an additional IS1470 sequence directly downstream the *cpe* gene (Figure 3), while the genomes typed as chromosomal in PCR (Table 2) had a typical gene order and genetic composition, as described by Miyamoto et al. (2006) around their chromosomally inserted *cpe* gene. For the sake of brevity, these six genes with new gene order for their *cpe* insertion site are called "variant c-*cpe* strains" in this article.

The additional IS1470 element in strains with variant gene order was 99% identical with the other, well-described IS1470 in the close vicinity of the *cpe* gene in chromosomal strains. Further analysis of IS1470 revealed that the variant c-*cpe* strains such as 24–6, 203–4, 205–27, 205–32, 233–6, and 233–1 had also accumulated other additional copies of IS1470 elements (Table 3). Variant c-*cpe* strains carried 58 to 81 copies of IS1470 in their genomes, while the majority of studied chromosomal strains carried 9 to 23 copies of IS1470. P-*cpe* strains carried under 10 copies of IS1470 (0 to 7) (Supplementary Table S2). All c-*cpe* strains carried 1 copy of IS1469 (Supplementary Table S2).

The traditional and variant c-*cpe* strains sequenced in this study had slight differences when selected genetic features were compared (Table 3). The traditional c-*cpe* strains lacked the arginine deiminase pathway Arc and the iron uptake system FeoAB, which were present in variant c-*cpe* strains. The traditional c-*cpe* strains sequenced in this study carried cellobiose metabolism operon and spore photoproduct lyase SplB, which were absent in the variant c-*cpe* strains. Sequenced strains were subjected to cgMLST analysis to characterize the strains and shed light on their epidemiological context. Analyzed *C. perfringens* strains (Supplementary Table S1) shared 1,236 genes in their core genome. Strains 24–6, 203–4,

10.3389/fmicb.2024.1459840

TABLE 1 *cpe*-positive samples yielded *cpe*-positive isolates, the isolation source of the samples, the number of *cpe*-positive isolates, and their *cpe* type (chromosomal/plasmidial).

Sample number	Isolation source	Number of *cpe*-positive isolates	Number of isolated *cpe* genotypes		
			c-*cpe* isolates IS1470	p-*cpe* isolates IS1470-like or IS1151	Variant isolates[a]
24	Peacekeeper	7	2	0	5
33	Peacekeeper	7 (+2 *cpe*-negative)	7	0	0
44	Peacekeeper	3	3	0	0
53	Peacekeeper	9	9	0	0
96	Peacekeeper	2	0	0	2
103	Peacekeeper	14	14	0	0
203	Local cook	5	4	0	1
204	Local cleaner	1	1	0	0
205	Local cleaner	32	2	0	30
206	Local cook	1	1	0	0
225	Local cook	3	3	0	0
233	Local cook	10	3	0	7

[a]Divergent size product in PCR detecting c-*cpe*.

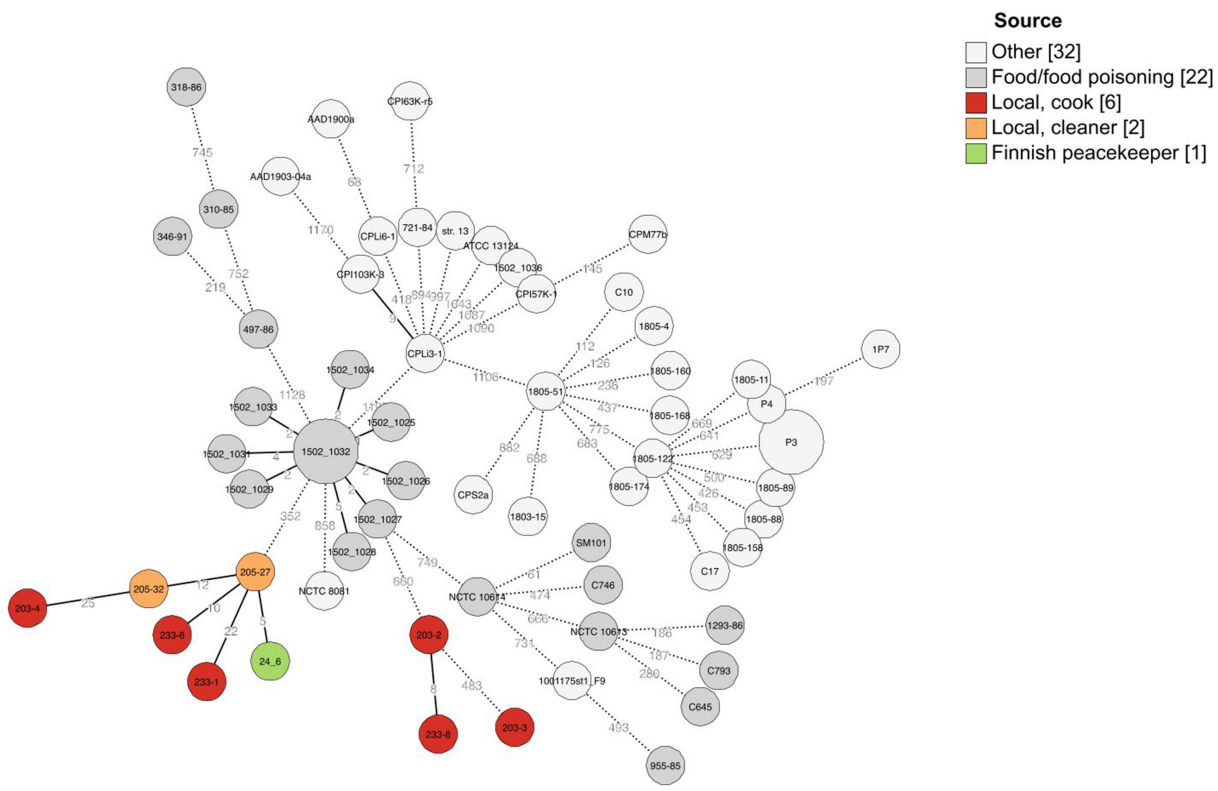


FIGURE 2
Minimal spanning tree of cgMLST results of 63 *C. perfringens* strains including here sequenced strains ($n = 9$) and previously published *C. perfringens* strains ($n = 54$). A scheme created by Abdel-Glil et al. (2021) was used for allele calling, and 1,236 loci belonging to the core genome were used to create the minimal spanning tree with GrapeTree (MSTreeV2). The isolation sources of strains have been added as a colored annotation, and branches below two allelic differences are collapsed.

205–27, 205–32, 233–1, and 233–6 formed a clade and shared highly similar cgMLST gene profiles (5 to 25 differences), suggesting genetic relatedness and a recent shared origin (Figure 1). These strains had been isolated from a peacekeeper (24–6), two local kitchen staff members (Cook #1203: 203–4 and Cook #2233: 233–1 and 233–6), and a local cleaner (205–27 and 205–32). Strains 203–2, 203–3, and 233–8 shared another clade with a genetic difference of 8 between 203–2 and 233–8 (Figure 2).

TABLE 2 Sequenced *C. perfringens* strains, their *cpe* type (chromosomal/plasmidial), initial PCR typing results, genetic lineage, isolation source, and BioSample accession number.

Strain	*cpe* type[a]	Initial PCR typing results	Genetic lineage[b]	Isolation source	BioSample accession number
203–2	Chromosomal	Chromosomal	IV	Cook #1	SAMN35541948
203–3	Chromosomal	Chromosomal	IV	Cook #1	SAMN35541949
203–4	Chromosomal	Variant	IV	Cook #1	SAMN35541950
205–27	Chromosomal	Variant	IV	Cleaner	SAMN35541951
205–32	Chromosomal	Variant	IV	Cleaner	SAMN35541952
233–1	Chromosomal	Variant	IV	Cook #2	SAMN35541953
233–6	Chromosomal	Variant	IV	Cook #2	SAMN35541954
233–8	Chromosomal	Chromosomal	IV	Cook #2	SAMN35541955
24–6	Chromosomal	Variant	IV	Peacekeeper	SAMN35541956

[a]Based on the sequenced genome.
[b]Feng et al. (2020) and Jaakkola et al. (2021).

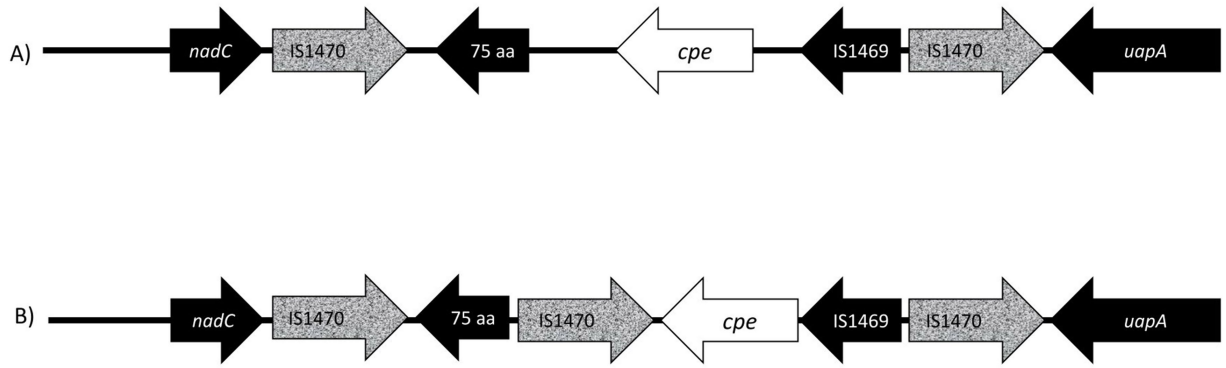


FIGURE 3
Presentation of gene order on chromosomal insertion sites of enterotoxin gene (*cpe*). **(A)** The well-described and conserved gene order (Brynestad et al., 1997) is commonly utilized for PCR genotyping, and probes target the IS1470 element downstream of the *cpe* gene and the *cpe* gene itself. In this study, six isolates (233–6, 233–1, 205–27, 203–4, 205–32, and 24–6) produced variant PCR results in PCR genotyping and had a new variant *cpe* loci **(B)** with a 1,053 aa IS1470 insertion element directly downstream of the *cpe* gene. The additional IS1470 element was 99% identical to the other, well-described IS1470 in the close vicinity of *the cpe* gene in chromosomal strains.

4 Discussion

In the present study, we found *C. perfringens* type F strains in the stool samples of peacekeepers and locally employed workers. Among the samples from locally employed workers and peacekeepers, 47 and 17%, respectively, tested positive for *cpe*-carrying *C. perfringens*. Altogether, 94 c-*cpe* strains were isolated from 12 people, including clinically healthy individuals (the locally employed workers and some peacekeepers) and those who were ill or recovering from intestinal disease (some peacekeepers). The isolation of c-*cpe* strains from multiple people suggests that there may have been one or more *C. perfringens* food poisoning outbreaks among peacekeepers with symptoms of disease as c-*cpe* strains are known to be a common cause of food poisoning (Lindström et al., 2011). However, since the sampling was conducted in October and the symptoms of the peacekeepers had occurred between June and October, CPE or large numbers of type F isolates were not found in the feces of affected people, and we could not definitively identify these symptoms being caused by *C. perfringens* type F. The local workers were not included in the epidemiological questionnaire or systematically interviewed, so it is unclear whether they had some symptoms of intestinal disease at the same time period as the peacekeepers. Interestingly, we obtained only c-*cpe* isolates, especially among locally employed workers. P-*cpe* strains are often present in the normal intestinal microbiota of many healthy people (Heikinheimo et al., 2006; Carman et al., 2008). Noteworthy, the isolation of *C. perfringens* strains was not successful from all *cpe*-positive samples, and whether p-*cpe* or c-*cpe* strains were present in these samples remains unknown.

In the present study, four peacekeepers had individual c-*cpe* strains according to the PFGE results. Those c-*cpe* strains were not found in anyone else. Based on the questionnaire, at least two of these peacekeepers had the symptoms of intestinal disease 1–2 months earlier, but not at the time of sampling. Moreover, one peacekeeper shared the same c-*cpe* strain with two local cooks and one local cleaner. According to the questionnaire, this peacekeeper had symptoms of intestinal disease 3 weeks earlier but not at the time of sampling. This may indicate that the peacekeepers had become carriers of the c-*cpe* strains after food poisoning.

While there is no clear evidence that c-*cpe* strains have been the cause of symptoms observed among the peacekeepers, the fact that we found genetically similar *C. perfringens* strains in samples from the locally employed workers and peacekeepers suggests that transmission

TABLE 3 Comparison of selected genetic features of the c-*cpe* strains sequenced in this study and some well-known *C. perfringens strains*.

cpe type	Strain	Lineage	Selected genes and genetic features																		
			N of IS1470	N of IS1469	N of IS1151	*pfoA*	*nanI*	Arc	NanJ	Fhu	FeoAB	Citrate	Cellobiose	Fucose	Ethanolamine	Myo-inositol	Biotin	CspLA	*splB*	SigV	*ssp4* HR
New variant of the c-*cpe* insertion site	203–4	IV	62	1	0	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	1
	205–27	IV	81	1	0	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	1
	205–32	IV	57	1	0	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	1
	233–1	IV	65	1	0	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	1
	233–6	IV	58	1	0	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	1
	24–6	IV	58	1	0	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	1
The known c-*cpe* insertion site	203–2	IV	23	1	0	0	0	0	0	1	0	1	1	0	0	0	0	0	1	0	1
	203–3	IV	31	1	0	0	0	0	0	1	0	1	1	0	0	0	0	0	1	0	1
	233–8	IV	23	1	0	0	0	0	0	1	0	1	1	0	0	0	0	0	1	0	1
	NCTC8081	IV	49	0	2	0	0	0	1	1	1	0	0	0	0	0	0	0	1	0	1
	NCTC8239	IV	23	1	0	0	0	1	0	1	1	1	1	0	0	0	0	0	0	0	1
	SM101	IV	16	1	0	0	0	1	0	1	1	1	1	0	0	0	0	0	0	1	1
	NCTC10240	IV	9	1	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0
Plasmidial *cpe*	CPE str. F4969	V	0	2	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	0
	AAD1903a	III	0	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0
cpe-negative	Str 13	V	0	0	0	1	1	1	1	1	0	1	0	0	1	1	1	1	1	1	0

Cpe type (New variant of the c-*cpe* insertion site/The known c-*cpe* insertion site/Plasmidial *cpe*/*cpe*-negative), the genetic lineage, number of IS elements (>80% identity, >80% length), and presence (1) and absence (0) of selected genes and genetic features [Arc = arginine deiminase pathway, Fhu = ferrichrome uptake system, FeoAB = iron uptake system, Citrate = citrate metabolism, Cellobiose = cellobiose metabolism operon, Fucose = fucose metabolism operon, Ethanolamine = ethanolamine utilization operon, Myo-inositol = myo-inositol utilization operon, Biotin = biotin biosynthesis, SplB = spore photoproduct lyase, SigV = sigma V, Ssp4 HR = gene allele with Asp at residue 36 and Asn at residue 72 associated with spore heat resistance (Li and McClane, 2008)] in sequenced *C. perfringens* strains. All IS1470 hits were located within the chromosome. Complete version of this table is available in Supplementary Table S2.

10.3389/fmicb.2024.1459840

of these strains has occurred among camp crew. The transmission has most likely been foodborne as this is the most common route of transmission for intestinal bacteria (Browne et al., 2017). This is further supported by the presence of the same strains in the kitchen staff. The high number of c-*cpe C. perfringens* among the local workers may even suggest humans as a reservoir of the c-*cpe* strains, and the persistence of c-*cpe* in peacekeepers several months after symptoms may even suggest that humans might become at least a transient reservoir of c-*cpe* strains in some cases. This subject needs to be further elucidated for better prevention of epidemics in the future. Reservoirs of the c-*cpe* strains are unclear, but humans as reservoirs have also been suggested previously (Heikinheimo et al., 2006).

In a previous study (Kiu et al., 2019), genetically highly similar p-*cpe C. perfringens* strains were associated with nine distinct care-home-associated outbreaks throughout a 5-year interval. Kiu et al. (2019) suggested that there was a common source linked to these outbreaks or transmission over time and space. Previous studies (Lahti et al., 2012; Jaakkola et al., 2021) have suggested that p-*cpe* strains and some c-*cpe* strains inhabit human intestines, so humans as carriers of food poisoning strains over time and space are possible. Since peacekeepers had symptoms evenly distributed during summer and autumn in 2004, it is possible that there had been a persistent food poisoning outbreak caused by the c-*cpe* strains carried by the kitchen workers.

An alternative explanation is that contaminated food ingredients have introduced the *C. perfringens* strains to the camp, where camp conditions may have facilitated the transmission of c-*cpe* strains among food consumers. C-*cpe* strains are known to enter the food chain and have been identified as the sole or predominant *cpe*-positive strains in retail meats from both the United States (Wen and McClane, 2004) and Turkey (Yibar et al., 2018). However, it is unclear how the c-*cpe* strains end up in the food chain, from food handlers or other sources. In the present study, the camp's food premises were adequate when inspected, and the UN procured food from international operators known for their high standards of food safety.

In the present study, the higher proportion of the c-*cpe C. perfringens* among the local workers than among the peacekeepers may also indicate a specific local source of c-*cpe C. perfringens*. The high numbers of these pathogenic isolates in the stool of one local cleaner (32 isolates) and a local cook (10 isolates) support this. The exposure to local foods outside the camp may have increased the risk of *C. perfringens* food poisoning.

The cgMLST scheme for *C. perfringens* has been used to type strains in several studies, and allele differences between these variable species are usually relatively large, with differences ranging from 200 to 1,000 alleles between strains (Abdel-Glil et al., 2021; Jaakkola et al., 2021). The allele differences between these sequenced strains varied between 5 and 25, and they formed a distinct cluster, supported by a separate clade in phylogenetic analysis (Figure 2). The differences between epidemiologically linked isolates are usually below 10 alleles in bacterial species (Schürch et al., 2018), but the cluster thresholds are dependent on the sequencing method and bacterial species, and larger cluster thresholds have been suggested for *C. perfringens* and its highly variable genome (Abdel-Glil et al., 2021). We suggested that the isolation of strains with less than 50 allele differences from a diverse group of people working on the same campsite suggests recent genetic relatedness, and considering the associated symptoms of gastrointestinal disease, there is a possible persistent outbreak among the residents at the camp. It is also possible that the chosen sequencing method has introduced some observed differences between the isolates.

The *cpe* insertion sites in the *C. perfringens* are well-described and conserved; therefore, they are widely used for strain typing. In our PCR typing, 45 of 94 c-*cpe* isolates were variants, and the sequencing revealed that the sequenced strains with variant results had an additional IS1470 element inserted directly downstream of the *cpe* gene. Variation in *cpe* loci arrangements has been previously reported by Abdelrahim et al. (2019), but this is the first time a *cpe* loci arrangement has been reported to affect PCR typing. Furthermore, the additional IS1470 element within *cpe* loci observed here has not been previously described.

The PCR genotyping probes in the *C. perfringens* assay target the IS1470 element downstream of the *cpe* gene and the *cpe* gene itself, and the presence of an additional IS1470 element downstream of the *cpe* gene impaired the standardized genotyping PCR assay. These variant c-*cpe* strains could be misidentified as strains with unknown *cpe* locations. Due to the variant c-*cpe,* the genotyping PCR primers need to be redesigned to detect reliably the variant c-*cpe* as well.

The traditional and variant c-*cpe* strains sequenced in this study all belonged to c-*cpe* group 1, which tolerate changing pH and acidic environments (Jaakkola et al., 2021). Interestingly, the variant c-*cpe* strains and traditional c-*cpe* strains had slight differences when selected genetic features were compared (Table 3). The variant c-*cpe* strains carried the arginine deiminase pathway Arc and the iron uptake system FeoAB, which were absent in traditional c-*cpe* strains and lacked the cellobiose metabolism operon, which was present in the traditional c-*cpe* strains. These possible differences in the metabolism between the c-*cpe* types may indicate that the variant c-*cpe* strains are better equipped for harsh conditions.

The variant c-*cpe* strains with an additional IS1470 element downstream of the *cpe* gene had a higher number of IS1470 elements in their entire genome compared to other *C. perfringens* strains. IS1470 element has been reported on chromosomes, never plasmids, and generally in moderate numbers (0–10 copies) (Brynestad et al., 1994). However, IS1470-like sequences are common on *C. perfringens* plasmids (Miyamoto et al., 2004). IS elements together with other transposable elements are important mutagenic agents enabling the host to adapt to new environmental challenges and colonize new niches. IS expansion has been linked to genome rearrangements, genome size reduction, and gene inactivation characteristic of the emergence of pathogenic strains (Parkhill et al., 2003; Siguier et al., 2006; Vandecraen et al., 2017), with famous examples of *Yersinia pseudotuberculosis* and *Yersinia pestis* (Parkhill et al., 2001; Chain et al., 2004), *Bordetella bronchiseptica*, and *Bordetella pertussis* (Parkhill et al., 2003).

In *C. perfringens*, the expansion of IS1470 elements has previously been reported in Darmbrand strain NCTC 8081 (Ma et al., 2012), and in other bacteria, the expansion of IS elements has been associated with an increase in virulence (Parkhill et al., 2003; Siguier et al., 2006). The relevance of IS1470 expansion for these Eritrean strains remains unknown, but the impact of IS1470 expansion on *C. perfringens* gene expression and host adaptation would be interesting topics for further research.

5 Conclusion

Our results suggest that persistent food poisoning outbreaks caused by *C. perfringens* type F strains can occur and that humans are a likely reservoir and carrier for enteropathogenic *C. perfringens*. We also conclude that the occurrence of additional IS1470 elements

in chromosomal *cpe*-carrying *C. perfringens* strains can impair the PCR typing, resulting in false-negative typing of these strains.

We present six new c-*cpe C. perfringens* genomes featuring an additional IS1470 element at the *cpe* insertion site and describe the organization of this new variant of the *cpe* locus. In addition, these variant c-*cpe* strains carry 58–81 copies of IS1470 in their genomes instead of 9–23 copies in previously described chromosomal *cpe* strains.

Our study contributes to the expansion of the pool of c-*cpe* strains by introducing Eritrean strains marking the first reported instances of c-*cpe* strains originating from Eastern Africa.

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 involving humans because in our study, the epidemiological investigation, including the literary questionnaire, sampling of the fecal specimens, and environmental samples, was done based on the Finnish legal obligations of food safety and health protection regulations. The epidemiologic investigation was carried out in 2004 by order of the Health Protection Authority of the Finnish Defence Forces. The order included all parts of the investigation (questionnaire, sampling and reporting). For the quick solving of the epidemics, there was a continuous right and an obligation to carry out the investigation ordered by the authorities of the Defence Forces. The research did not involve intervening in the physical integrity of research participants. The persons participated as study objects were volunteers who took the samples themselves and delivered them to the investigation group. No ethical concerns were identified, and no specific ethical approval was required. The studies were conducted in accordance with the local legislation and institutional requirements. 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 because persons participating as study objects in the investigation were beforehand informed volunteers, and they agreed that the investigation results could be reported anonymized. An informed consent procedure was used, but not a written informed consent procedure since the results were anonymized.

Author contributions

PL: Writing – review & editing, Writing – original draft, Investigation. KJ: Validation, Investigation, Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation. AHö: Resources, Writing – review & editing, Investigation, Conceptualization. AHe: Writing – review & editing, Methodology, Investigation. AS: Writing – review & editing, Resources, Investigation. HK: Writing – original draft, Supervision, Writing – review & editing, Resources, Project administration, Funding acquisition.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We thank Miia Lindström and Kirsi Ristkari from the University of Helsinki and Pia Laine from the Institute of Biotechnology for technical assistance and advice.

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.2024.1459840/full#supplementary-material

References

Abdel-Glil, M. Y. P., Thomas, J. L., Busch, A., Wieler, L. H., Neubauer, H., and Seyboldt, C. (2021). Comparative in Silico genome analysis of *Clostridium perfringens* unravels stable Phylogroups with different genome characteristics and pathogenic potential. *Sci. Rep.* 11, 1–15. doi: 10.1038/s41598-021-86148-8

Abdelrahim, M., Abakabir, N. R., Delannoy, S., Djellal, S., Le Négrate, M., Hadjab, K., et al. (2019). Large-scale genomic analyses and Toxinotyping of *Clostridium perfringens* implicated in foodborne outbreaks in France. *Front. Microbiol.* 10:777. doi: 10.3389/fmicb.2019.00777

Browne, H. P., Anne Neville, B., Forster, S. C., and Lawley, T. D. (2017). Transmission of the gut microbiota: spreading of health. *Nat. Rev. Microbiol.* 2017, 531–543. doi: 10.1038/nrmicro.2017.50

Brynestad, S. (1997). Genetic studies on *Clostridium perfringens* enterotoxin. Oslo, Norway: Norwegian College of Veterinary Medicine.

Brynestad, S., Iwanejko, L. A., Stewart, G. S. A. B., and Granum, P. E. (1994). A complex Array of Hpr consensus DNA recognition sequences proximal to the enterotoxin gene in *Clostridium perfringens* type a. *Microbiology* 140, 97–104. doi: 10.1099/13500872-140-1-97

Brynestad, S., Synstad, B., and Granum, P. E. (1997). The *Clostridium perfringens* enterotoxin gene is on a transposable element in type a human food poisoning strains. *Microbiology* 143, 2109–2115. doi: 10.1099/00221287-143-7-2109

Carman, R. J., Sayeed, S., Li, J., Genheimer, C. W., Hiltonsmith, M. F., Wilkins, T. D., et al. (2008). *Clostridium perfringens* toxin genotypes in the feces of healthy north Americans. *Anaerobe* 14, 102–108. doi: 10.1016/J.ANAEROBE.2008.01.003

Chain, P. S. G., Carniel, E., Larimer, F. W., Lamerdin, J., Stoutland, P. O., Regala, W. M., et al. (2004). Insights into the evolution of *Yersinia Pestis* through whole-genome comparison with *Yersinia Pseudotuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 13826–13831. doi: 10.1073/pnas.0404012101

Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474

Cornillot, E., Saint-Joanis, B., Daube, G., Katayama, S.-i.-i., Granum, P. E., Canard, B., et al. (1995). The enterotoxin gene (Cpe) of *Clostridium perfringens* can be chromosomal or plasmid-borne. *Mol. Microbiol.* 15, 639–647. doi: 10.1111/j.1365-2958.1995.tb02373.x

Daube, G., Simon, P., and Kaeckenbeeck, A. (1993). IS1151, an IS-like element of *Clostridium perfringens*. *Nucleic Acids Res.* 21:352. doi: 10.1093/NAR/21.2.352

Feng, Y., Fan, X., Zhu, L., Yang, X., Liu, Y., Gao, S., et al. (2020). Phylogenetic and genomic analysis reveals high genomic openness and genetic diversity of *Clostridium perfringens*. *Microbial Genomics* 6:e000441. doi: 10.1099/mgen.0.000441

Heikinheimo, A., and Korkeala, H. (2005). Multiplex PCR assay for Toxinotyping *Clostridium perfringens* isolates obtained from Finnish broiler chickens. *Lett. Appl. Microbiol.* 40, 407–411. doi: 10.1111/J.1472-765X.2005.01702.X

Heikinheimo, A., Lindström, M., Granum, P. E., and Korkeala, H. (2006). Humans as reservoir for enterotoxin gene-carrying *Clostridium perfringens* type A. *Emerg. Infect. Dis.* 12, 1724–1729. doi: 10.3201/eid1211.060478

Heikinheimo, A., Lindström, M., and Korkeala, H. (2004). Enumeration and isolation of Cpe-positive *Clostridium perfringens* spores from feces. *J. Clin. Microbiol.* 42, 3992–3997. doi: 10.1128/JCM.42.9.3992-3997.2004

Jaakkola, K., Virtanen, K., Lahti, P., Keto-Timonen, R., Lindström, M., and Korkeala, H. (2021). Comparative genome analysis and spore heat resistance assay reveal a new component to population structure and genome epidemiology within *Clostridium perfringens* enterotoxin-carrying isolates. *Front. Microbiol.* 12:2530. doi: 10.3389/FMICB.2021.717176/BIBTEX

Keto-Timonen, R., Heikinheimo, A., Eerola, E., and Korkeala, H. (2006). Identification of Clostridium species and DNA fingerprinting of *Clostridium perfringens* by amplified fragment length polymorphism analysis. *J. Clin. Microbiol.* 44, 4057–4065. doi: 10.1128/JCM.01275-06

Kiu, R., Caim, S., Alexander, S., Pachori, P., and Hall, L. J. (2017). Probing genomic aspects of the multi-host pathogen *Clostridium perfringens* reveals significant Pangenome diversity, and a diverse Array of virulence factors. *Front. Microbiol.* 8:2485. doi: 10.3389/fmicb.2017.02485

Kiu, R., Caim, S., Painset, A., Pickard, D., Swift, C., Dougan, G., et al. (2019). Phylogenomic analysis of gastroenteritis-associated *Clostridium perfringens* in England and Wales over a 7-year period indicates distribution of clonal toxigenic strains in multiple outbreaks and extensive involvement of enterotoxin-encoding (CPE) plasmids. *Microb. Genomics* 5:e000297. doi: 10.1099/mgen.0.000297

Kiu, R., and Hall, L. J. (2018). An update on the human and animal enteric pathogen *Clostridium perfringens*. *Emerg. Microbes Infect.* 7. doi: 10.1038/s41426-018-0144-8

Lahti, P., Heikinheimo, A., Johansson, T., and Korkeala, H. (2008). *Clostridium perfringens* type a strains carrying a plasmid-borne enterotoxin gene (genotype IS1151-Cpe or IS1470-like-Cpe) as a common cause of food poisoning. *J. Clin. Microbiol.* 46, 371–373. doi: 10.1128/JCM.01650-07

Lahti, P., Lindström, M., Somervuo, P., Heikinheimo, A., and Korkeala, H. (2012). Comparative genomic hybridization analysis shows different epidemiology of chromosomal and plasmid-borne Cpe-carrying *Clostridium perfringens* type a. *PLoS One* 7:e46162. doi: 10.1371/journal.pone.0046162

Leclair, D., Pagotto, F., Farber, J. M., Cadieux, B., and Austin, J. W. (2006). Comparison of DNA fingerprinting methods for use in investigation of type E botulism outbreaks in the Canadian Arctic. *J. Clin. Microbiol.* 44, 1635–1644. doi: 10.1128/JCM.44.5.1635-1644.2006

Li, J., Adams, V., Bannam, T. L., Miyamoto, K., Garcia, J. P., Uzal, F. A., et al. (2013). Toxin plasmids of *Clostridium perfringens*. *Microbiol. Mol. Biol. Rev.* 77, 208–233. doi: 10.1128/mmbr.00062-12

Li, J., and McClane, B. A. (2008). A novel small acid soluble protein variant is important for spore resistance of most *Clostridium perfringens* food poisoning isolates. *PLoS Pathog.* 4:e1000056. doi: 10.1371/journal.ppat.1000056

Li, J., Miyamoto, K., and McClane, B. A. (2007). Comparison of virulence plasmids among *Clostridium perfringens* type E isolates. *Infect. Immun.* 75, 1811–1819. doi: 10.1128/IAI.01981-06

Li, J., Miyamoto, K., Sayeed, S., and McClane, B. A. (2010). Organization of the Cpe Locus in CPE-positive *Clostridium perfringens* Type C and D isolates. *PLoS One* 5:e10932. doi: 10.1371/journal.pone.0010932

Lindström, M., Heikinheimo, A., Lahti, P., and Korkeala, H. (2011). Novel insights into the epidemiology of *Clostridium perfringens* type a food poisoning. *Food Microbiol.* 28, 192–198. doi: 10.1016/j.fm.2010.03.020

Ma, M., Li, J., and Mcclane, B. A. (2012). Genotypic and phenotypic characterization of *Clostridium perfringens* isolates from Darmbrand cases in post-world war ii Germany. *Infect. Immun.* 80, 4354–4363. doi: 10.1128/IAI.00818-12

Miyamoto, K., Chakrabarti, G., Morino, Y., and McClane, B. A. (2002). Organization of the Plasmid Cpe Locus in *Clostridium perfringens* type a isolates. *Infect. Immun.* 70, 4261–4272. doi: 10.1128/IAI.70.8.4261-4272.2002

Miyamoto, K., Fisher, D. J., Li, J., Sayeed, S., Akimoto, S., and McClane, B. A. (2006). Complete sequencing and diversity analysis of the enterotoxin-encoding plasmids in *Clostridium perfringens* type a non-food-borne human gastrointestinal disease isolates. *J. Bacteriol.* 188, 1585–1598. doi: 10.1128/JB.188.4.1585-1598.2006

Miyamoto, K., Wen, Q., and McClane, B. A. (2004). Multiplex PCR genotyping assay that distinguishes between isolates of *Clostridium perfringens* type a carrying a chromosomal enterotoxin gene (Cpe) locus, a plasmid Cpe locus with an IS1470-like sequence, or a plasmid Cpe locus with an IS1151 sequence. *J. Clin. Microbiol.* 42, 1552–1558. doi: 10.1128/JCM.42.4.1552-1558.2004

Miyamoto, K., Yumine, N., Mimura, K., Nagahama, M., Li, J., McClane, B. A., et al. (2011). Identification of novel *Clostridium perfringens* type E strains that carry an iota toxin plasmid with a functional enterotoxin gene. *PLoS One* 6:e20376. doi: 10.1371/journal.pone.0020376

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Parkhill, J., Sebaihia, M., Preston, A., Murphy, L. D., Thomson, N., Harris, D. E., et al. (2003). Comparative analysis of the genome sequences of *Bordetella Pertussis*, Bordetella Parapertussis and *Bordetella Bronchiseptica*. *Nat. Genet.* 35, 32–40. doi: 10.1038/NG1227

Parkhill, J., Wren, B. W., Thomson, N. R., Titball, R. W., Holden, M. T., Prentice, M. B., et al. (2001). Genome sequence of *Yersinia Pestis*, the causative agent of plague. *Nature* 413, 523–527. doi: 10.1038/35097083

Rambaut, A., and Drummond, A. (2008). FigTree: tree figure drawing tool, version 1.4. 4. Available at: http://tree.bio.ed.ac.uk/software/figtree/ (Accessed January 12, 2021).

Ridell, J., Björkroth, J., Eisgrüberg, H., Schalch, B., Stolle, A., and Korkeala, H. (1998). Prevalence of the enterotoxin gene and clonality of *Clostridium perfringens* strains associated with food-poisoning outbreaks. *J. Food Prot.* 61, 240–243. doi: 10.4315/0362-028X-61.2.240

Rood, J. I., Adams, V., Lacey, J., Lyras, D., McClane, B. A., Melville, S. B., et al. (2018). Expansion of the *Clostridium perfringens* toxin-based typing scheme. *Anaerobe* 53, 5–10. doi: 10.1016/j.anaerobe.2018.04.011

Schürch, A. C., Arredondo-Alonso, S., Willems, R. J. L., and Goering, R. V. (2018). Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene–based approaches. *Clin. Microbiol. Infect.* 24, 350–354. doi: 10.1016/j.cmi.2017.12.016

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

Seemann, T. (2015). Snippy: Fast bacterial variant calling from NGS reads. Available at: https://github.com/tseemann/snippy

Siguier, P., Filée, J., and Chandler, M. (2006). Insertion sequences in prokaryotic genomes. *Curr. Opin. Microbiol.* 9, 526–531. doi: 10.1016/J.MIB.2006.08.005

Silva, M., Machado, M. P., Silva, D. N., Rossi, M., Moran-Gilad, J., Santos, S., et al. (2018). *Microb. Genomics* 4:e000166. doi: 10.1099/mgen.0.000166

Staden, R., Judge, D. P., and Bonfield, J. K. (2003). "Managing sequencing projects in the GAP4 environment" in Introduction to bioinformatics. A Theoretical and Practical Approach. eds. S. A. Krawetz and D. D. Womble (Totawa, NJ 07512: Humana Press), 327–344.

Vandecraen, J., Chandler, M., Aertsen, A., and Van Houdt, R. (2017). The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit. Rev. Microbiol.* 43, 709–730. doi: 10.1080/1040841X.2017.1303661

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

Wen, Q., and McClane, B. A. (2004). Detection of Enterotoxigenic *Clostridium perfringens* type a isolates in American retail foods. *Appl. Environ. Microbiol.* 70, 2685–2691. doi: 10.1128/AEM.70.5.2685-2691.2004

Yibar, A., Cetin, E., Ata, Z., Erkose, E., and Tayar, M. (2018). *Clostridium perfringens* contamination in retail meat and meat-based products in Bursa, Turkey. *Foodborne Pathog Dis.* 15, 239–245. doi: 10.1089/FPD.2017.2350

Zhou, Z., Alikhan, N. F., Sergeant, M. J., Luhmann, N., Vaz, C., Francisco, A. P., et al. (2018). Grapetree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res.* 28, 1395–1404. doi: 10.1101/gr.232397.117

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 06 November 2024
DOI 10.3389/fmicb.2024.1483110

OPEN ACCESS

EDITED BY
Xuanyu Tao,
University of Oklahoma, United States

REVIEWED BY
Mu Peng,
Hubei Minzu University, China
Masataka Tsuda,
Tohoku University, Japan

*CORRESPONDENCE
Magdalena Szuplewska
✉ m.szuplewska@uw.edu.pl

RECEIVED 19 August 2024
ACCEPTED 14 October 2024
PUBLISHED 06 November 2024

CITATION
Szuplewska M, Sentkowska D, Lasek R, Decewicz P, Hałucha M, Funk Ł, Chmielowska C and Bartosik D (2024) Genome-wide comparative analysis of clinical and environmental strains of the opportunistic pathogen *Paracoccus yeei* (*Alphaproteobacteria*). *Front. Microbiol.* 15:1483110. doi: 10.3389/fmicb.2024.1483110

Genome-wide comparative analysis of clinical and environmental strains of the opportunistic pathogen *Paracoccus yeei* (*Alphaproteobacteria*)

Magdalena Szuplewska[1]*, Dorota Sentkowska[1], Robert Lasek[1], Przemysław Decewicz[2], Mateusz Hałucha[1], Łukasz Funk[1], Cora Chmielowska[1] and Dariusz Bartosik[1]

[1]Department of Bacterial Genetics, Institute of Microbiology, Faculty of Biology, University of Warsaw, Warsaw, Poland, [2]Department of Environmental Microbiology and Biotechnology, Institute of Microbiology, Faculty of Biology, University of Warsaw, Warsaw, Poland

Introduction: *Paracoccus yeei* is the first species in the genus *Paracoccus* to be implicated in opportunistic infections in humans. As a result, *P. yeei* strains provide a valuable model for exploring how bacteria shift from a saprophytic to a pathogenic lifestyle, as well as for investigating the role of horizontally transferred DNA in this transition. In order to gain deeper insights into the unique characteristics of this bacterium and the molecular mechanisms underlying its opportunistic behavior, a comparative physiological and genomic analysis of *P. yeei* strains was performed.

Results: Complete genomic sequences of 7 *P. yeei* isolates (both clinical and environmental) were obtained and analyzed. All genomes have a multipartite structure comprising numerous extrachromosomal replicons (59 different ECRs in total), including large chromids of the DnaA-like and RepB families. Within the mobile part of the *P. yeei* genomes (ECRs and transposable elements, TEs), a novel non-autonomous MITE-type element was identified. Detailed genus-wide comparative genomic analysis permitted the identification of *P. yeei*-specific genes, including several putative virulence determinants. One of these, the URE gene cluster, determines the ureolytic activity of *P. yeei* strains—a unique feature among *Paracoccus* spp. This activity is induced by the inclusion of urea in the growth medium and is dependent on the presence of an intact *nikR* regulatory gene, which presumably regulates expression of nickel (urease cofactor) transporter genes.

Discussion: This in-depth comparative analysis provides a detailed insight into the structure, composition and properties of *P. yeei* genomes. Several predicted virulence determinants (including URE gene clusters) were identified within ECRs, indicating an important role for the flexible genome in determining the opportunistic properties of this bacterium.

KEYWORDS

***Paracoccus yeei*, opportunistic pathogen, multipartite genome, chromid, evolution of pathogenic bacteria, transposable element, non-autonomous transposable element, urease**

1 Introduction

Bacteria of the genus *Paracoccus* (family *Paracoccaceae*, order *Rhodobacterales*, class *Alphaproteobacteria*) are known for their great metabolic diversity and flexibility. Many of them can switch between growth modes, using different carbon and energy sources, and employing various final electron acceptors (Czarnecki and Bartosik, 2019). There are currently 105 known species of the genus *Paracoccus* (NCBI Taxonomy, 17 July 2024), including *Paracoccus denitrificans*—the type strain of the genus (isolated in 1910) — commonly used as a model denitrifying organism (Beijerinck and Minkman, 1910; Bordel et al., 2024). A large number of these species were identified only recently and many have yet to be the subject of detailed studies.

Paracoccus spp. inhabit diverse marine and terrestrial environments, such as soil, sediments, sludge, brines or groundwater (e.g., Urakami et al., 1990; Siller et al., 1996; Tsubokura et al., 1999; Berry et al., 2003; Lee et al., 2004; Liu et al., 2006, 2008). In addition, some strains have been isolated from the rhizosphere or from the surface of other organisms, e.g., ticks, marine bryozoans and corals (Pukall et al., 2003; Ghosh et al., 2006; Carlos et al., 2017). The ubiquity of these bacteria is further demonstrated by their presence in house dust (Thompson et al., 2021).

Among these environmental bacteria, *Paracoccus yeei* (formerly classified as a eugonic oxidizer in group 2; strain EO-2) is of particular interest because it is the first species of the genus *Paracoccus* to be implicated in opportunistic infections in humans (Daneshvar et al., 2003). This bacterium (naturally occurring in soil) is not associated with a specific disease (Fosso et al., 2021). *P. yeei* isolates have been recovered from several clinical conditions, e.g., from the dialysate of a patient with peritonitis, myocarditis in a transplanted heart, corneal transplantation, bacteremia, keratitis, otitis and dermatological lesions (Lasek et al., 2018). The number of cases of infection by *P. yeei* appears to be increasing (Funke et al., 2004; Kanis et al., 2010; Schweiger et al., 2011; Courjaret et al., 2014; Arias and Clark, 2019; Aliste-Fernández et al., 2020; Fosso et al., 2021; Shifera et al., 2021; Bhikoo et al., 2022). However, this reported incidence is likely to be an underestimate since current common diagnostic tests do not detect this bacterium (Sack et al., 2017). Moreover, *P. yeei* cells can appear under-decolorized after Gram staining, so may be inadvertently reported or dismissed as gram-positive cocci (Dyer and Harris, 2020).

Analysis of a larger number of clinical cases suggests that immunocompromised patients are at increased risk of infection with this bacterium (Dyer and Harris, 2020). However, the natural reservoir of this opportunistic pathogen as well as the molecular bases of its pathogenicity remain unclear.

Numerous *P. yeei* strains have been isolated from everyday objects in hospital environments, which may play an important role in its transmission. This bacterium was found on the surface of medical equipment, such as operating tables, endoscopes or laryngoscopes (in direct contact with the mucosa, saliva and blood of patients) (Choi et al., 2017; Sartoretti et al., 2017; Pasquale et al., 2021), as well as on the surface of mobile phones of hospital employees and patients (Murgier et al., 2016; Cantais et al., 2020). Moreover, in non-hospital settings *P. yeei* strains were detected on the surface of benches, doors and walls of schools and kindergartens (Kruszewska et al., 2021), and on hand-drying devices in public toilets (Huesca-Espitia et al., 2018). Interestingly, this bacterium was also identified as a dominant member of microbial consortia contributing to the biodegradation of pre-Columbian canvasses in a museum collection (Pietrzak et al., 2017).

P. yeei represents a convenient model organism to study the switch from a saprophytic to a pathogenic lifestyle and to determine the role of horizontally acquired DNA in this process. Comparative genomic analyses of *P. yeei* and other members of the genus *Paracoccus* should yield a great deal of information on the mechanisms involved in this transition. Our previous analysis of the genomes of 3 *P. yeei* isolates (CCUG 32053, FDAARGOS_252 and TT13) gave the first insight into the genome composition, mobilome and metabolic potential of these bacteria (Lasek et al., 2018). However, the small number of strains analyzed was a major limitation of this study.

The main objective of the present study was to define the specific properties and genetic information of *P. yeei* species that are potentially involved in the process of pathogenesis, as well as to characterize mobile genetic elements—natural gene carriers that may be responsible for the horizontal transmission of virulence determinants. For 7 *P. yeei* strains available in our laboratory, we performed various physiological tests and detailed analyses of their mobile genetic elements. To identify *P. yeei*-specific genes, we conducted a comparative analysis of the 11 genomes of *P. yeei* strains available in the NCBI database (both clinical and environmental isolates) and four genomes of other members of the genus *Paracoccus*.

2 Materials and methods

2.1 Strains and culture conditions

P. yeei strains CCUG 13493, CCUG 17731, CCUG 32052, CCUG 32054, CCUG 46822, and CCUG 54214 were purchased from the Culture Collection of the University of Gothenburg (CCUG) (Sweden). Strain LM20 was isolated from black shale collected within the Lubin underground copper mine in Poland (Dziewit et al., 2015). *P. yeei* CCUG 13493R, CCUG 17731R, CCUG 32052R, CCUG 32054R, CCUG 46822R, CCUG 54214R, and LM20R, rifampicin-resistant derivatives of the respective wild-type strains, and *Escherichia coli* TG1 (Sambrook and Russell, 2001), were used as plasmid recipients. *E. coli* DH5α (Hanahan, 1983) was the host strain of helper plasmid pRK2013, used in triparental matings (Ditta et al., 1980) (Supplementary Table S1).

All strains were cultured in lysogeny broth (LB) medium (Sambrook and Russell, 2001) at 37°C for *E. coli* or 30°C for the other strains. Liquid cultures were grown under shaking conditions. When required, the medium was supplemented with 10% sucrose and the following concentrations of selective antibiotics: kanamycin, 50 μg/ml for all strains except *Achromobacter* sp. LM16R (300 μg/ml) (Dziewit et al., 2015); rifampicin, 50 μg/ml; tetracycline, 2 μg/ml for *P. yeei* and 20 μg/ml for *E. coli* strains.

2.2 Physiological analyses

The response to various growth conditions (temperature, pH and salinity) of *P. yeei* strains was determined as described previously by Dziewit et al. (2013). Motility was tested according to the method previously outlined by Dziewit et al. (2015). The hemolytic ability of different strains was tested by growth on blood agar plates that were incubated at 30 and 37°C under aerobic conditions—the results were read after 24 and 48 h of incubation. Siderophore production was assessed using the chrome azurol S (CAS) agar plate method, as described previously (Dziewit et al., 2015). The plates were incubated in the dark at 30°C for 72 h, after which halo formation around the colonies was assessed. Each isolate was phenotypically characterized using (i) the API 50 CH test system (to differentiate strains based on their ability to metabolize different carbohydrates) and (ii) the API ZYM test system (to detect and identify enzymatic activities); according to the recommendations of the supplier (bioMerieux, Marcy l'Etoile, France). Ureolytic activity was detected by cultivation on Christensen's urea medium, containing urea and a pH indicator (phenol red) (Christensen, 1946) or using rapid urease tests. *Proteus vulgaris* and/or *Klebsiella pneumoniae* strains (University of Warsaw collection) were used as positive controls, and *P. aminophilus* JCM 7686 and/or *P. aminovorans* JCM 7685 as negative controls (Urakami et al., 1990). The minimal inhibitory concentrations (MIC) of antimicrobial agents were determined for *P. yeei* strains by using Mueller Hinton II agar diffusion with E-tests as recommended by the supplier (bioMérieux). *E. coli* ATCC 25922 was used as a control. To confirm reproducibility, antimicrobial susceptibility testing was repeated at least twice. Since no interpretative guidelines for *P. yeei* susceptibility have been published, we used the criteria proposed by Arias and Clark (2019). MIC values for selected heavy metal ions were established on titration plates using a procedure described previously (Dziewit et al., 2013). The following heavy metal salts were used to prepare appropriate stock solutions in water: $NaAsO_2$; $3CdSO_4 \times 8H_2O$; $CoSO_4 \times 7H_2O$; $K_2Cr_2O_7$; $CuSO_4$; $HgCl_2$; $NiCl_2 \times 6H_2O$; NaO_3V; $ZnSO_4 \times 7H_2O$. Each microplate was monitored for growth using an automated microplate reader at 24-h intervals for 3 days. Isolates that demonstrated growth at the following minimum metal ion concentrations were classified as resistant using previously described criteria (Dziewit et al., 2015): (i) 20 mM V^{5+}, (ii) 1 mM As^{3+}, Cd^{2+}, Co^{2+}, Cu^{2+}, Ni^{2+}, Zn^{2+} or Cr^{6+}, and (iii) 0.1 mM Hg^{2+}.

2.3 DNA isolation, standard molecular biology procedures and PCR conditions

Plasmid DNA was isolated using the alkaline lysis procedure (Birnboim and Doly, 1979) or purified by CsCl-ethidium bromide gradient centrifugation (Sambrook and Russell, 2001). The visualization of mega-sized replicons was achieved by in-gel lysis and DNA electrophoresis (Eckhardt, 1978; Hynes and McGregor, 1990). Plasmid DNA was also isolated using a Plasmid Mini kit (A&A Biotechnology), GeneMATRIX Miniprep DNA Purification Kit (EUR_x) and GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific). Routine DNA manipulation was performed using standard methods (Sambrook and Russell, 2001). DNA amplification by PCR was performed in a Mastercycler (Eppendorf) using synthetic oligonucleotides (Supplementary Table S1), High Fidelity Taq DNA polymerase (Qiagen) or Phusion Hi-Fidelity DNA polymerase (Thermo Fisher Scientific), dNTPs and appropriate template DNAs, as described previously (Bartosik et al., 2003).

2.4 Introduction of plasmid DNA into bacterial cells

Chemical transformation of *E. coli* cells was performed by a standard method (Kushner, 1978). Plasmid DNA was introduced into *Paracoccus* spp. strains by triparental mating using helper *E. coli* strain DH5α carrying plasmid pRK2013 (containing the transfer system of plasmid RP4) (Ditta et al., 1980), as described previously (Lasek et al., 2018).

2.5 Plasmid host range testing

REP systems of extrachromosomal replicons of the analyzed strains were cloned in vector pABW1 (Bartosik et al., 1997) which cannot replicate in *Paracoccus* strains or other tested recipients (Supplementary Table S1). These shuttle plasmid constructs were introduced into the following strains: (i) class *Alphaproteobacteria—Agrobacterium tumefaciens* LBA 288R (Hooykaas et al., 1980), *Paracoccus aminophilus* JCM 7686R (without plasmids pAMI2, pAMI3) (Dziewit et al., 2014), *Paracoccus aminovorans* JCM 7686R (Czarnecki et al., 2017), *Paracoccus pantotrophus* KL100 (Jordan et al., 1997), (ii) class *Betaproteobacteria—Achromobacter* sp. LM16R (Dziewit et al., 2015) and (iii) class *Gammaproteobacteria—Cronobacter sakazakii* ATCC 29544 (Iversen et al., 2008) and *Pseudomonas* sp. LM6R (Dziewit et al., 2015).

2.6 Identification of functional transposable elements

Trap plasmids pMAT1 and pMEC1 (Bartosik et al., 2003; Szuplewska and Bartosik, 2009; Gay et al., 1985; Schneider et al., 2000) were used for the identification of functional TEs of *P. yeei* (Supplementary Table S1). These plasmids were transferred from *E. coli* TG1 to *P. yeei* strains by triparental mating. The "capture" of TEs was verified by PCR using primers specific to the selection cassettes, with DNA isolated from Tc^r and Suc^r mutants as the template (Supplementary Table S1). A total of 1,304 clones were tested this way. The amplified DNA fragments were subjected to electrophoretic analysis, which permitted the detection of DNA elements (TEs), ranging in size from ~0.8 to 1.5 kb, embedded in the cassettes. DNA sequencing of the termini of individual TEs was used to identify the complete elements within the genome sequences of their parental strains. All trapped TEs were sequenced and compared with the ISfinder database (Siguier et al., 2006).

2.7 Mutational analysis of URE modules

To confirm that the identified URE modules are responsible for ureolytic activity, mutational analysis was performed using the gene replacement method. Mutations (insertion of a Km^r cassette) were introduced into URE type 1 modules, either in the 5′ region of the Ureα subunit gene or within the *nikR* gene. Both mutations were created in the strain CCUG 32053 (which contains only URE type 1) and the strain CCUG 13493 (which also contains a URE type 2 module). The following procedure was used to obtain *ureα*::Km^r and *nikR*::Km^r mutants: (i) DNA fragments containing *ureα-ureβ-ureγ* or part of the region encoding NikR and nickel transporters were amplified by PCR using CCUG 32053 DNA as the template, (ii) these DNA fragments were cloned into the SacI and SphI sites of the mobilizable suicide vector pDS132 (Philippe et al., 2004), (iii) the Km^r cassette from plasmid pDIY-Km (Dziewit et al., 2011) was inserted into the AleI site within these fragments to obtain constructs pDS132-*ureα* and pDS132-*nikR* (Supplementary Table S1); (iv) these plasmids were then introduced into strains CCUG 32053R and CCUG 13493R by conjugation. Mutant clones *ureα*::Km^r and *nikR*::Km^r were selected on medium supplemented with kanamycin (for recombinant selection) and sucrose (for counter-selection). The introduced mutations were confirmed by DNA sequencing. Ureolytic activity of the obtained mutants was tested following growth on minimal medium supplemented with 1% urea, using rapid urease tests. Complementation of the mutants was performed using plasmids pBBR-*ureα* or pBBR-*nikR*, which were prepared by cloning PCR fragments containing *ureα-ureβ-ureγ* or part of the region encoding NikR and nickel transporters into vector pBBR-MCS5 (Gm^r) (Kovach et al., 1995) (Supplementary Table S1).

2.8 Genome sequencing

The *P. yeei* genomes were sequenced using a combination of Oxford Nanopore and Illumina technologies, as described previously (Lasek et al., 2018).

2.9 Bioinformatic analyses

Sequence annotation and bioinformatic analyses (identification of tRNA genes and rRNA operons, relaxases—MOB, toxin-antitoxin modules, transposable elements—TE) were performed as described previously (Lasek et al., 2018). Potential ICE and IME elements were predicted using ICEfinder (Liu et al., 2019).

Categories for clusters of orthologous groups (COGs) were assigned to each protein by performing a local RPS-BLAST search against the COG database (last updated on January 22, 2015). A threshold e-value of 1e-5 was used, and only the top BLAST hits were taken into account (Tatusov et al., 2003).

The core genome of *Paracoccus* spp. and *P. yeei* species was defined based on the complete genomic sequences of *P. aminophilus* JCM 7686 (Dziewit et al., 2014), *P. aminovorans* JCM 7685 (Czarnecki et al., 2017), *P. contaminans* RKI16-01929T (Aurass et al., 2017) and *P. denitrificans* PD1222 (NC_008686-8), and eleven *P. yeei* strains: CCUG 32053 (Lasek et al., 2018), FDAARGOS_252 (NZ_CP020440–47), FDAARGOS_643 (NZ_CP044078-82), TT13 (Lim et al., 2018) and CCUG 13493, CCUG 17731, CCUG 32052, CCUG 32054, CCUG 46822, CCUG 54214, LM20 (Dziewit et al., 2015). Proteins encoded within the genomes were used in all-against-all BLASTp searches utilized by a stand-alone version of OrthoVenn2 (Xu et al., 2019), using an e-value of 1e-15 as the threshold and inflation level of 1.5. Based on this analysis, all proteins were clustered into groups reflecting similarity and designated as core proteins when encoded by all eleven genomes or as singletons when the respective gene was identified in only a single genome. A similar approach with a different manner of grouping was applied to identify strain-specific gene clusters.

Protein datasets from the Virulence Factor Database VFDB (Chen et al., 2016), the Comprehensive Antibiotic Resistance Database CARD (Alcock et al., 2020) and Virulence Factors Database VICTORS (Sayers et al., 2019), all downloaded on April 1 2021, were used for BLASTp searches (e-value threshold of 1e-30 and 80% of query coverage per HSP) with the proteomes of *P. yeei* and four other *Paracoccus* spp. strains listed above to identify putative *P. yeei* species virulence determinants.

EasyFig (Sullivan et al., 2011) was used to perform comparative genomic analyses (including genomes and plasmidomes comparisons) and visualize the results. RNA secondary structures were predicted by *in silico* folding using Mfold software (Zuker, 2003).

Phylogenetic analysis of the genus *Paracoccus* was based on alignment of concatenated nucleotide sequences of selected core genes: *atpD, dnaA, dnaK, gyrB, recA, rpoB*, and *thrC* (homologous genes of *Roseobacter denitrificans* OCh 114 were used as an outgroup), as described previously (Lasek et al., 2018). Nucleotide alignments were obtained with Mafft (Katoh et al., 2019). Then, concatenated genes were analyzed with ModelTest-NG (Darriba et al., 2020), checking all models to select the best-fit nucleotide substitution model. The selected substitution model was applied in RaxML-NG (Kozlov et al., 2019) with 2,000 regular bootstrap replicates performed on the best Maximum Likelihood (ML) tree selected from 100 independently generated ML starting trees.

2.10 Nucleotide sequence accession numbers

The nucleotide sequences of the *P. yeei* chromosomes and extrachromosomal replicons were deposited in GenBank (NCBI), with the following accession numbers: (i) CCUG 13493—CP038080 (chromosome), CP038073 (plasmid pYEE13493P1), CP038074 (pYEE13493P2), CP038075 (pYEE13493P3), CP038076 (pYEE13493P4), CP038077 (pYEE13493P5), CP038078 (pYEE13493P7), CP038079 (pYEE13493P8), CP038081 (pYEE13493P6); (ii) CCUG 17731—CP038042 (chromosome), CP038035 (pYEE17731P1), CP038036 (pYEE17731P2), CP038037 (pYEE17731P3), CP038038 (pYEE17731P4), CP038039 (pYEE17731P5), CP038040 (pYEE17731P6), CP038041 (pYEE17731P7); (iii) CCUG 32052—CP038090 (chromosome),

CP038082 (pYEE32052P1), CP038083 (pYEE32052P2), CP038084 (pYEE32052P3), CP038085 (pYEE32052P4), CP038086 (pYEE32052P5), CP038087 (pYEE32052P6), CP038088 (pYEE32052P7), CP038089 (pYEE32052P8); (iv) CCUG 32054—CP038095 (chromosome), CP038091 (pYEE32054P6), CP038092 (pYEE32054P7), CP038093 (pYEE32054P8), CP038094 (pYEE32054P9), CP038096 (pYEE32054P1), CP038097 (pYEE32054P2), CP038098 (pYEE32054P3), CP038099 (pYEE32054P4), CP038100 (pYEE32054P5); (v) CCUG 46822—CP038056 (chromosome), CP038043 (pYEE46822P1), CP038044 (pYEE46822P2), CP038045 (pYEE46822P3), CP038046 (pYEE46822P4), CP038047 (pYEE46822P5), CP038048 (pYEE46822P6), CP038049 (pYEE46822P7), CP038050 (pYEE46822P8), CP038051 (pYEE46822P9), CP038052 (pYEE46822P10), CP038053 (pYEE46822P11), CP038054 (pYEE46822P12), CP038055 (pYEE46822P13); (vi) CCUG 54214—CP038061 (chromosome), CP038057 (pYEE51214P1), CP038058 (pYEE51214P2), CP038059 (pYEE51214P3), CP038060 (pYEE51214P4); (vii) LM20—CP038072 (chromosome), CP038062 (pLM20P1), CP038063 (pLM20P2), CP038064 (pLM20P3), CP038065 (pLM20P4), CP038066 (pLM20P5), CP038067 (pLM20P6), CP038068 (pLM20P7), CP038069 (pLM20P8), CP038070 (pLM20P9), CP038071 (pLM20P10). The nucleotide sequences of newly identified ISs (IS*Pye*73—IS*Pye*80) and MITE*Pye1* were deposited in the ISfinder database (Siguier et al., 2006).

3 Results

3.1 *P. yeei* strains selected for characterization

Seven (five clinical and two environmental) isolates of *P. yeei* were selected for detailed characterization (Table 1). Clinical strains were sourced from two continents—they were isolated in different geographical locations, at different times and from different clinical cases (Table 1). The other strains were isolated either from an anthropogenic environment (CCUG 54214; a metal surface), or from the natural environment (LM20; organic-rich black shale from the Lubin copper mine in Poland; Dziewit et al., 2015). The clinical isolate CCUG 32053, characterized in a previous study (Lasek et al., 2018), was used as a reference strain.

3.1.1 Physiological and phenotypic characterization

P. yeei is a non-motile, oxidase- and catalase-positive gram-negative bacterium. The analyzed strains were facultative aerobes capable of nitrate reduction and denitrification (except for CCUG 32052). The strains were mesophilic, capable of growth at temperatures ranging from 21 to 37°C (except LM20, which grew from 15 to 37°C) (Table 2). All strains grew in LB medium at pH values close to 7 (ranging from 5 to 8 or 9), typical for neutrophilic bacteria. Salinity tolerance testing showed that the isolates could tolerate NaCl concentrations of 4–6%. These bacteria were non-hemolytic, did not produce bacteriocins, and exhibited weak adhesion to polystyrene. However, all produced siderophores, as determined by the chrome azurol S (CAS) agar plate assay. They demonstrated the ability to degrade 27 carbohydrates (out of 49 tested), with identical profiles for all strains except LM20, which is unable to ferment mannitol, sorbitol or arabitol (Table 2). The enzymatic capabilities of these strains were assessed by testing for 19 hydrolytic enzymes from various groups. All strains produced alkaline phosphatase (except CCUG 13493), esterase, lipase esterase, leucine arylamidase, acid phosphatase, naphthol-AS-BI phosphohydrolase, and α-glucosidase, with only CCUG 17731 producing β-galactosidase (Table 2). The strains also displayed ureolytic activity, which is a unique feature among *Paracoccus* spp.

MIC values for selected antibiotics (ampicillin, ciprofloxacin, erythromycin, gentamicin, tetracycline, and vancomycin) were determined for each strain (Supplementary Table S2). The data were analogous to those from previous studies of clinical isolates of *P. yeei*, showing sensitivity to beta-lactams, especially aminopenicillins, as well as macrolides and aminoglycosides (Funke et al., 2004; Wallet et al., 2010; Schweiger et al., 2011; Sastre et al., 2016; Arias and Clark, 2019; Aliste-Fernández et al., 2020). Nevertheless, we observed variations in the susceptibility of the isolates, e.g., CCUG 13493 was most sensitive to the tested antibiotics, CCUG 32054 was the most tolerant, and some strains (CCUG 32052, CCUG 32054, CCUG 46822, LM20) were less sensitive to ciprofloxacin (Supplementary Table S2).

MIC values for selected heavy metal ions were similar for all strains, with no major differences between clinical and environmental isolates. CCUG 54214 showed higher tolerance to As^{3+}, Ni^{2+}, V^{5+}, and Zn^{2+}, while LM20 showed higher tolerance to As^{3+}, Cd^{2+}, and Hg^{2+}, reflecting the adaptation of these strains to the environments from which they were isolated (Supplementary Table S2).

3.2 Genomic features

The genomes of the seven strains of *P. yeei* were fully sequenced and characterized. Their genome sizes range from 4,423,927 to 4,829,807 bp (average 4,610,003 bp). All are multipartite, containing from 4 (CCUG 54214) to 13 (CCUG 46822) extrachromosomal replicons (ECRs), that range in size from 3.5 kb (pYEE46822P1 of CCUG 46822) to 485 kb (pYEE17731P7 of CCUG 17731) (Table 3; Figure 1B; Supplementary Figure S1). In total, the strains contain 59 ECRs (11,665,529 bp), which on average constitute ~25% of the size of each genome (Table 3; Figure 1B). The physicochemical parameters of the genome sequences were determined, including the %GC content (average 67%), the number of protein-coding sequences (CDSs) (average 4,200) and pseudogenes (average 190). The GC content of the ECRs ranges from 53% (pYEE46822P4, CCUG 46822) to 69% (pYEE13493P5, CCUG 13493). The number of CDSs present in ECRs ranges from 849 (CCUG 54214), accounting for 20% of the entire genome, to 1,270 (CCUG 17731) (28%) (Supplementary Table S3). The rRNA operons and most of the tRNA genes are localized within the chromosomes, although some additional tRNA genes (2–3) were identified within ECRs (Table 3).

TABLE 1 *Paracoccus yeei* strains analyzed in this study.

P. yeei strain	Source of isolation			References
	Specimen	Locality	Year	
Clinical isolates (human)				
CCUG[a] 13493	Foot wound	USA (Virginia)	1980	–
CCUG 17731	Water, hemodialysis	France (Grenoble)	1985	–
CCUG 32052	Cerebrospinal fluid	USA (Puerto Rico)	1983	–
CCUG 32053[b]	Eye	USA (Missouri)	1981	Lasek et al., 2018
CCUG 32054	Facial sinus	USA (Washington)	1985	–
CCUG 46822	Abdominal dialysate	USA (Pennsylvania)	1988	Daneshvar et al., 2003
Environmental isolates				
CCUG 54214	Metal product, industry	Sweden	2007	–
LM20	Black shale (copper mine)	Poland (Lubin)	2015	Dziewit et al., 2015

[a]*P. yeei* strains with the designation CCUG were purchased from the Culture Collection of the University of Gothenburg (CCUG) (Sweden).
[b]Reference strain.

3.3 Extrachromosomal replicons

The analyzed *P. yeei* strains contain numerous ECRs (59 in total)—from 4 (CCUG 54214) to 13 (CCUG 46822) (Figure 1B; Supplementary Figure S1). These are mainly large DNA molecules, constituting a significant proportion of their host strain genomes—from 22% (CCUG 54214) to >28% (CCUG 13493) (Figure 1B). The replication system (REP) and the gene encoding the replication initiation protein (Rep) within each ECR were identified. The REP modules (and entire plasmids) were classified based on the presence of specific amino acid (aa) sequence signatures in the Rep proteins (Figure 2A). Replicons of the DnaA-like, RepB (RepB1 and RepB2), RepABC, RepC, Rep_3 and HTH_36 families were identified, all of which are common among *Paracoccus* spp. and other *Alphaproteobacteria* (Figures 1C, 2).

The ECRs carrying REPs of a given type are generally similar in size, with the exception of RepB1 and RepB2 replicons, in which greater size variation is observed (Figure 2B). The REPs of the DnaA-like, RepB and RepABC types (occurring in each *P. yeei* strain) are characteristic for large replicons, which is consistent with previous observations (e.g., Lasek et al., 2018). Detailed characterization of individual ECRs, including their size, nucleotide sequence GC content, coding capacity and tRNA gene distribution, is presented in Supplementary Table S3.

To test the host range of individual ECRs representing the different replicon families, their REP regions were cloned in mobilizable shuttle plasmids and introduced by triparental mating into recipient strains from different classes of *Proteobacteria* (listed in Section 2). This analysis revealed that all of the analyzed REPs function solely in *Alphaproteobacteria*, so have a relatively narrow host range.

The *P. yeei* ECRs contain numerous *loci* enabling their stable maintenance in bacterial cells and populations. The DnaA-like, RepB, and RepABC replicons all carry type I partitioning systems, encoding Walker-type ATPases (Thomas, 1981) (Figure 2A). The vast majority of ECRs also have class II toxin-antitoxin (TA) systems, mainly encoding toxins of the VapC family (18 toxins), as well as toxins of the RelE/ParE (14), PhD/YefM (14) or parDE (8) families. None of the ECRs encode a complete type IV secretion system (T4SS), suggesting that these replicons are not self-transmissible. However, 16 of them contain genes encoding predicted relaxases of the Mob_C (14), Mob_Q (1), and Mob_{HEN} (1) families (Garcillán-Barcia et al., 2009), which are typical components of genetic modules, enabling plasmid mobilization for conjugative transfer (MOB).

Further analysis was conducted to verify whether the *P. yeei* genomes contain ECRs that meet the criteria for chromids (secondary chromosomes)—essential replicons of plasmid origin. This revealed that all the DnaA-like, RepB1 and RepB2 replicons can be considered chromids (Figure 1C), since they (i) carry plasmid-type REP modules, (ii) have a nucleotide sequence composition that is close to that of the chromosome (<1% difference in GC content), and (iii) possess a set of core genes (in one copy in the genome) characteristic for the entire genus (Harrison et al., 2010). The genes carried by these chromids participate in several important metabolic pathways, e.g., encoding acetyl-CoA C-acyltransferase, and proteins likely to be involved in biosynthesis of the lipoyl cofactor, in oxidative metabolism, leucine metabolism, and electron transport via the respiratory chain (complex I).

The Rep proteins of the DnaA-like, RepB1 and RepB2 chromids display a high level of aa sequence identity within a given replicon group. These replicons are also more highly conserved in terms of their structure and genetic load (Supplementary Table S5). In contrast, the replication initiators of the RepABC, Rep_3 and HTH_36 families (and their entire replicons) show much greater sequence variability. Many of these replicons appear to be unique among *P. yeei* strains, suggesting their recent horizontal acquisition (Supplementary Table S5).

TABLE 2 Phenotypic and physiological characteristics of *P. yeei* strains.

Characteristic(s)	*P. yeei* clinical isolates					Environmental isolates	
	CCUG 13493	CCUG 17731	CCUG 32052	CCUG 32054	CCUG 46822	CCUG 54214	LM20[a]
	General features						
Temperature range for growth (°C, optimum)	21–37 (30)	21–37 (30)	21–37 (30)	21–37 (30)	21–37 (30)	21–37 (30)	15–37 (30)
NaCl tolerance (%)	6	4	4	6	4	6	4
pH range for growth	5–9	5–8	5–9	5–8	5–8	5–9	4–9
Motility	□	□	□	□	□	□	□
Hemolysis	□	□	□	□	□	□	□
Siderophore production	■	■	■	■	■	■	■
	Nitrate metabolism						
Nitrate reduction	■	■	■	■	■	■	■
Gas from nitrate	■	■	□	■	■	■	■
	Carbohydrate fermentation (API 50 CH test)						
D-Arabinose, L-arabinose, D-ribose, D-xylose, L-xylose, D-galactose, D-glucose, D-fructose, D-mannose, L-sorbose, L-rhamnose, dulcitol, inositol, D-cellibiose, D-lactose, D-melibiose, xylitol, gentibiose, D-lyxose, D-tagatose, D-fucose, L-fucose, L-arabitol	■	■	■	■	■	■	■
D-Adonitol	■	■	■	■	■	■	■
D-Mannitol, D-sorbitol, D-arabitiol	■	■	■	■	■	■	□
Glycerol, erythritol, methyl-βD-xylopyranoside, methyl-αD-mannopyranoside, methyl-αD-glucopyranoside, N-acetylglucosamine, amygdaline, arbutine, esculine, salicin, D-maltose, D-sacharose, D-trehalose, inulin, D-melezitose, D-raffinose, amidon, glycogen, D-turanose, potassium gluconate, potassium 2 ketogluconate, potassium 5 ketogluconate	□	□	□	□	□	□	□
	Enzyme activities (including API ZYM test)						
Esterase, esterase lipase, leucine arylamidase, acid phosphatase, naphthol-AS-BI-phosphohydrolase,α-glukosidase	■	■	■	■	■	■	■
Alkaline phosphatase	□	■	■	■	■	■	■
β-Galactosidase	□	■	□	□	□	□	□
Lipase, valine arylamidase, cystine arylamidase, trypsin, α-chymotrypsin, α-galactosidase, β-glucuronidase, β-glucosidase, N-acetyl-β-glucosamidase, α-mannosidase, α-fucosidase	□	□	□	□	□	□	□
Catalase	■	■	■	■	■	■	■
Oxidase	■	■	■	■	■	■	■
Urease activity	■	■	■	■	■	■	■

[a]Growth conditions (temp. range, pH range, NaCl tolerance) for strain LM20 were as described previously (Dziewit et al., 2015).

Filled square ■ indicates a positive result, empty square □ indicates a negative result.

10.3389/fmicb.2024.1483110

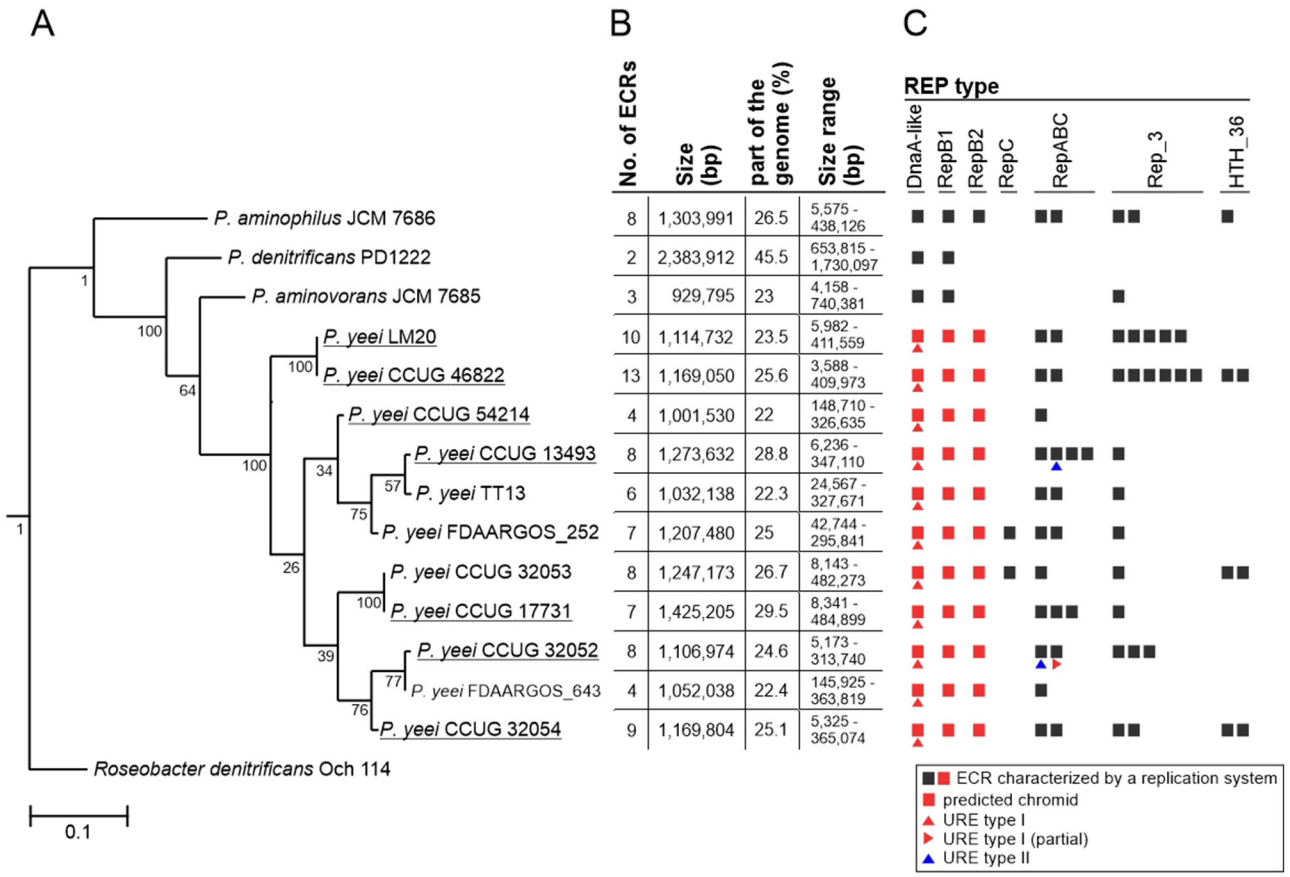


FIGURE 1
Distribution of ECRs in *P. yeei* genomes. **(A)** Phylogenetic tree of *P. yeei* constructed using core gene nucleotide sequences. Three other members of the genus *Paracoccus*—*P. aminophilus* JCM 7686, *P. aminovorans* JCM 7685, and *P. denitrificans* PD1222—as well as *Roseobacter denitrificans* Och 114 (as an outgroup) were included in the analysis. Bar represents 0.1 nucleotide substitutions per position. **(B)** Summary of data for ECRs of the analyzed *Paracoccus* spp. strains. **(C)** Distribution of different REP module types among ECRs of *Paracoccus* spp.

3.4 Comparative analysis of *P. yeei* genomes

Phylogenetic analysis showed that the two environmental isolates (CCUG 54214 and LM20) are situated on different branches of the *P. yeei* phylogenetic tree and are closely related to the clinical isolates. Strains LM20 and CCUG 46822 form a separate cluster, while the other five strains can be divided into two phylogenetic groups: (i) CCUG 54214 and CCUG 13493, and (ii) CCUG 17731, CCUG 32052 and CCUG 32054 (Figure 1A).

Comparative genomic analysis of the *P. yeei* strains (including clinical isolates CCUG 32053, FDAARGOS_643, FDAARGOS_252, and TT13) revealed a conserved genome structure, preserved ECR composition and synteny of gene arrangement in this species (Supplementary Table S5). There is a particularly high degree of synteny between the genomes of the strains LM20 and CCUG 46822 (Figure 3). The ECRs of these strains are nearly identical, although they display minor rearrangements (Supplementary Table S5).

Our analysis identified two chromosome structure variants (one found in strains CCUG 17731, CCUG 32052, CCUG 32053, CCUG 32054, CCUG 54214, FDAARGOS_643, TT13, and the other in CCUG 13493, CCUG 46822, LM20, FDAARGOS_252), which correlates with the phylogeny of the strains (Figure 3; Supplementary Table S5). The differences result mainly from inversions or translocations of fairly large DNA segments (Figure 3).

The DnaA-like replicons show a high degree of conservation, although with some rearrangements, e.g., translocation of DNA segments in strains CCUG 32053 and CCUG 13493, sometimes with inversion of the transferred DNA segment (Supplementary Table S5).

3.5 Genetic load of *P. yeei* genomes

Analysis of the proteome of *P. yeei* determined *in silico* permitted the assignment of individual proteins to functional COG groups. For all genomes a similar proportion of the encoded proteins belong to each category, indicating a conserved genome structure and limited variability within the species. Interestingly, significant similarity in the distribution of COG groups was found

TABLE 3 General features of *P. yeei* genomes.

General features	Clinical isolates					Environmental isolates	
	CCUG 13493	CCUG 17731	CCUG 32052	CCUG 32054	CCUG 46822	CCUG 54214	LM20
Genome size (bp)	4,423,927	4,829,807	4,505,223	4,652,732	4,562,823	4,554,217	4,741,293
ECRs	8	7	8	9	13	4	10
Total ECR size (bp) (% of entire genome)	1,273,632 (28.8)	1,425,205 (29.5)	1,106,974 (24.6)	1,169,804 (25.1)	1,169,050 (25.6)	1,001,530 (22)	1,114,732 (23.5)
G + C content (%)							
Genome	67.6	67.3	67.4	67.5	67.4	67.5	67.3
ECRs (range)	62.8–69	56.9–68.3	58.5–68.4	60.6–68.5	53.5–68.9	64.5–68.5	57.8–68.9
CDSs							
Genome	4,066	4,465	4,168	4,284	4,152	4,162	4,353
ECRs (% of entire genome)	1,117 (27.5)	1,270 (28.4)	964 (23.1)	1,021 (23.8)	998 (24)	849 (20.4)	957 (22)
Pseudogenes							
Genome	195	205	179	163	208	188	217
ECRs (%)	96 (49.2)	100 (48.8)	78 (43.6)	74 (45.4)	98 (47.1)	67 (35.6)	95 (43.8)
RNA genes							
tRNA genes in genome	50	51	51	51	50	50	50
tRNA genes in ECRs	2	3	2	3	2	2	2
rRNA operons[a]	3	3	3	3	3	3	3
Integrative elements							
Predicted prophage regions[b]	2	5	5	4	3	6	7
Predicted ICE/IME elements[c]	2	3	0	2	0	2	1
Transposase genes[d]							
Genome	119	145	150	84	136	125	150
ECRs (% of entire genome)	57 (48)	64 (44)	66 (44)	32 (38)	56 (41)	40 (32)	52 (35)

[a] 16S-23S-5S rRNA operons.
[b] Phastest predicted prophage regions (complete) (Supplementary Table S4).
[c] ICEfinder predicted ICE/IME elements (Supplementary Table S4).
[d] Including truncated forms.

between the chromosomes and ECRs, suggesting that ECRs may contain a large amount of genetic information of chromosomal origin (Figure 4).

In all cases, a considerable number of genes are of unknown function (R and S categories) or could not be classified into any COG group: 27–31% for chromosomes and 27–32% for ECRs. The largest group of classified proteins comprises those related to amino acid transport and metabolism (category E; 9–10%), followed by carbohydrate metabolism (G; 5–6%), transcription (K; 6%), transport and metabolism of inorganic ions (P; 5–6%), energy acquisition processes (C; 5%), and the synthesis of cell wall, membrane, and envelope components, including capsules (M; 5%) (Supplementary Figure S2). In the case of ECRs, the largest fraction of classified proteins was assigned to categories E (7–10%), G (8–11%), P (6–8%), and K (6–8%) (Figure 4B). There are no significant differences between clinical and environmental isolates in the distribution of COG categories.

Transport-related genes constitute a dominant group of functional genes in the *P. yeei* genomes—from 530 (CCUG 32052) to over 600 CDSs (CCUG 17731). These genes encode various types of transporters: ABC, MSF (Major Facilitator Superfamily), P-type, RND (Resistance-Nodulation-Division), TRAP (Tripartite ATP-independent Periplasmic), and TTT (Tripartite Tricarboxylate Transporters) (Table 4; Supplementary Table S6). ECRs (mainly DnaA-like and RepB chromids) contain over 35% of the ABC transporter genes. Less common in ECRs are MFS, RND and TTT transporter genes. Only a few genes encoding proteins with similarity to P-type ATPase integral membrane transporters were found in the analyzed genomes.

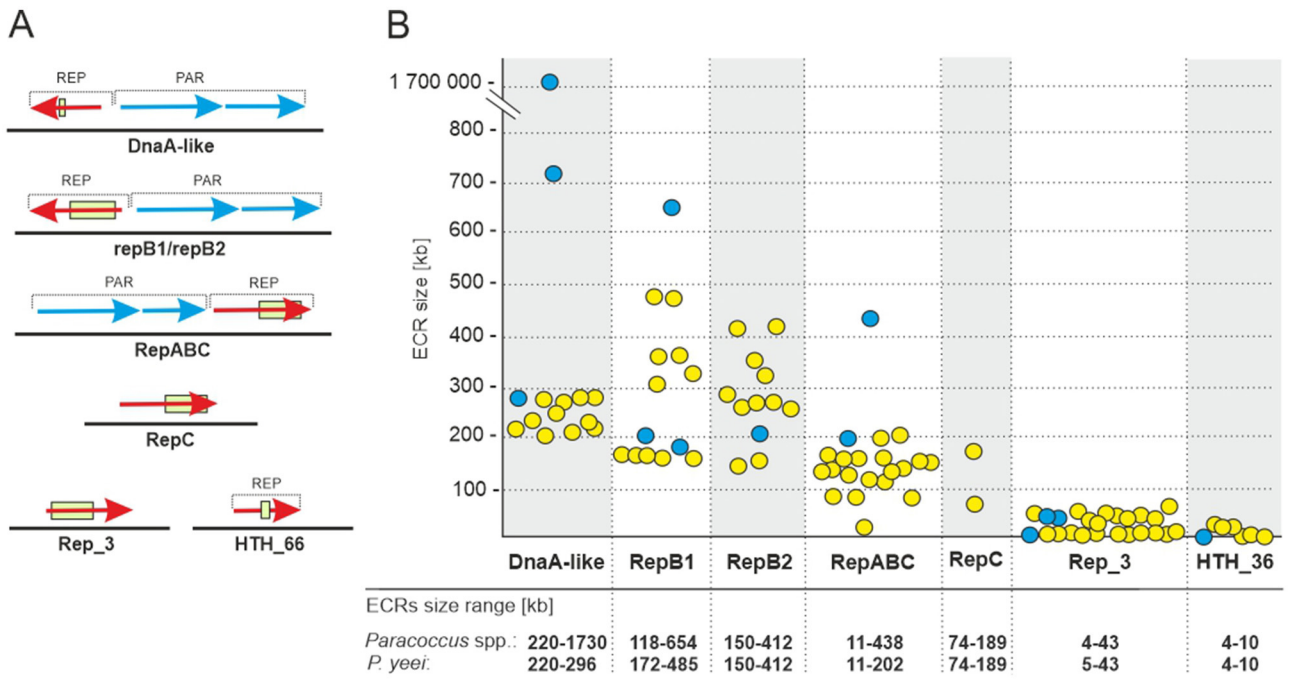


FIGURE 2
ECRs identified in the genomes of *P. yeei* strains. **(A)** Genetic organization of the replication (REP) and partitioning modules of the ECRs. The box indicates the location of sequence motifs specific to each type of replication initiation protein. **(B)** Size range of *P. yeei* replicons carrying a given REP module type (yellow). The blue color indicates replicons present in the genomes of *P. aminophilus* JCM 7686, *P. aminovorans* JCM 7685, and *P. denitrificans* PD1222.

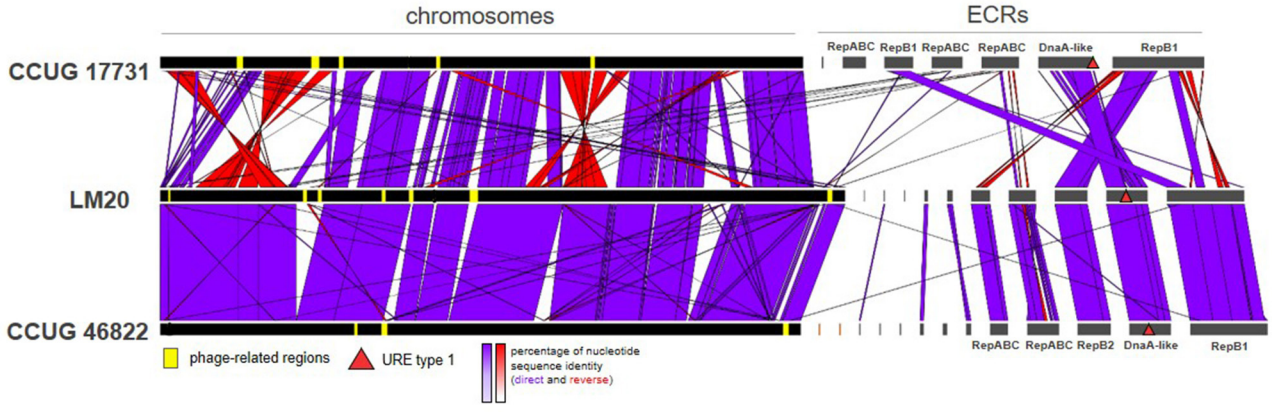


FIGURE 3
Comparative genomic analysis of three *P. yeei* strains, CCUG 17731, LM20, and CCUG 46822, illustrating the two chromosome structural variants identified in this species.

The ECRs also carry numerous genes that may increase the adaptability of their host strains in the environment. All types of chromids and RepABC plasmids encode potential iron or carbohydrate transporters. Siderophore receptor genes were identified in all RepB2 plasmids, including the FhuADCB ferric hydroxamate transporter, which is essential for the uptake of Fe^{3+}-aerobactin. All RepB1 replicons carry a *mauMGJCBDEAF* gene cluster involved in the utilization of methylamine (Baker et al., 1998). Organisms that utilize C1 compounds as their sole carbon and energy sources are described as methylotrophs and this trait is common in bacteria of the genus *Paracoccus* (Czarnecki and Bartosik, 2019). Methylotrophs play an important role in global carbon, nitrogen and sulfur cycling, and they have been successfully employed in the bioremediation of contaminated soils (Chistoserdova, 2011).

The RepB1 replicons of three phylogenetically closely related *P. yeei* strains (CCUG 17731, CCUG 32052, CCUG 32054) (Figure 1A) contain a putative operon (*phnCDEFGHIJKLMNOP*)

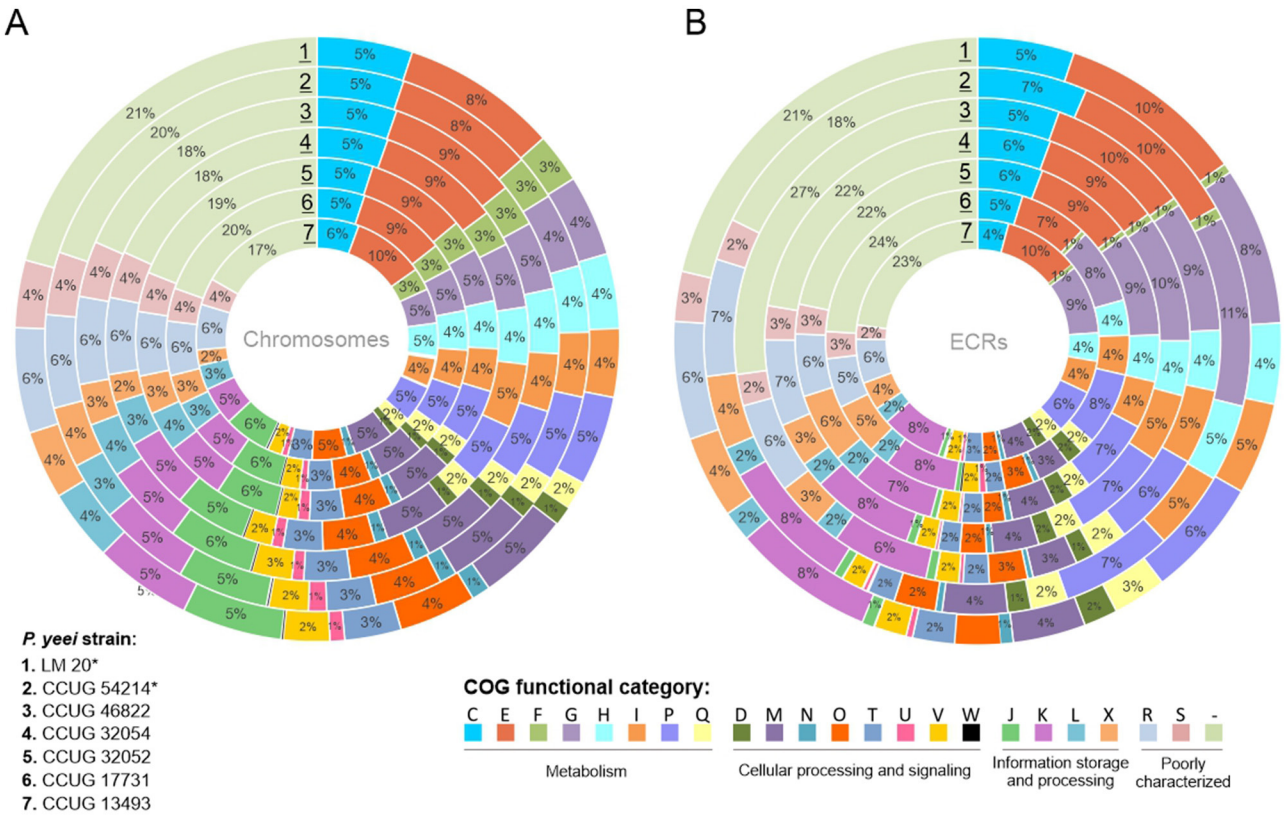


FIGURE 4
Distribution of genes encoding proteins belonging to different COG functional categories within **(A)** the chromosomes and **(B)** ECRs of the analyzed *P. yeei* strains (*environmental isolates). Each colored segment indicates the relative contribution of a functional category as a percentage of total COGs. Each ring represents a different strain of *P. yeei*. COG functional categories: C, energy production and conversion; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; P, inorganic ion transport and metabolism; Q, secondary metabolite biosynthesis, transport and catabolism; D, cell cycle control, cell division, chromosome partitioning; M, cell wall/membrane/envelope biogenesis; N, cell motility; O, posttranslational modification, protein turnover, chaperones; T, signal transduction mechanisms; U, intracellular trafficking, secretion and vesicular transport; V, defense mechanisms; W, extracellular structures; Z, cytoskeleton; J, translation, ribosomal structure and biogenesis; K, transcription; L, replication, recombination and repair; X, mobilome, i.e., prophages, transposons; R, general function, prediction only; S, function unknown, COG not assigned.

encoding carbon-phosphorus lyase, which permits the utilization of phosphorus from a wide range of stable phosphonate compounds containing a C-P bond (Makino et al., 1991; Horsman and Zechel, 2017; Amstrup et al., 2023). In numerous environments, inorganic phosphate, a vital nutrient, is available in very limited amounts, compelling microorganisms to rely on alternative phosphorus sources for survival (Podzelinska et al., 2009). Therefore, the ability to uptake and breakdown phosphonates would be advantageous.

Enzymes catalyzing rhamnose biosynthesis (RfbABCD) are encoded by all RepB2 replicons (CCUG 32052 has an additional copy of the *rfbABCD* operon in Rep_3 plasmid pYEE32052P3). Rhamnose is present in the O-antigens of many gram-negative bacteria, forming part of the lipopolysaccharide (LPS). This deoxy-hexose sugar is also present in capsular polysaccharides, covalently bound to the cell wall, and in exopolysaccharides that are loosely associated with the cell wall (Marolda and Valvano, 1995). The *P. yeei* strains analyzed in this study display characteristic mucoid growth, probably related to the presence of a polysaccharide capsule (data not shown).

The ECRs also carry genes for proteins involved in (i) the cobalamin (vitamin B12) biosynthetic pathway (*cobAWNGHIJKLMBF*; DnaA-like chromids), (ii) ectoine biosynthesis (*ehuABCD*; RepB1 chromids), (iii) propionate catabolism (*mmgE/prpD*; several RepB1, RepB2, and RepABC replicons), (iv) disulfide bond formation (*dsbA*, *dsbB*, and *dsbE*), and (v) resistance to arsenic compounds (permease Arc3, arsenate reductase ArsC, flavoprotein ArsH, repressor ArsR, and transporter ArsJ). Notably, all RepB1 chromids carry a complete conserved set of *crt* genes (*crtXYIBZE-idi*) for carotenoid synthesis (Maj et al., 2013). However, colonies of the strains with these chromids lack the characteristic color associated with carotenoid pigment production and these compounds could not be detected by HPLC analysis (data not shown).

In summary, the DnaA-like, RepB1 and RepB2 chromids and the RepABC plasmids carry the largest number of genes of adaptive potential. The genetic load of smaller replicons (HTH_36 or Rep_3) consists mainly of genes of unknown function. The distribution of selected genes in individual ECRs is presented in Supplementary Table S3.

TABLE 4 Distribution and characterization of transport-related genes in *P. yeei* genomes.

Transporter type	*P. yeei* clinical isolates					Environmental isolates	
	CCUG 13493	CCUG 17731	CCUG 32052	CCUG 32054	CCUG 46822	CCUG 54214	LM20
ABC No. of genes	366	390	370	384	341	342	342
Chromosome (%)	219 (60%)	235 (60%)	241 (65%)	241 (63%)	225 (66%)	213 (62%)	226 (66%)
ECRs (%)	147 (40%)	155 (40%)	129 (35%)	143 (37%)	116 (34%)	129 (38%)	116 (34%)
MSF	31	34	29	34	30	28	29
Chromosome	22 (71%)	23 (68%)	23 (79%)	28 (82%)	22 (73%)	20 (71%)	21 (72%)
ECRs	9 (29%)	11 (32%)	6 (21%)	6 (18%)	8 (27%)	8 (29%)	8 (28%)
P-type	4	7	4	5	3	8	6
Chromosome	3 (75%)	4 (57%)	3 (75%)	4 (80%)	3 (100%)	8 (100%)	6 (100%)
ECRs	1 (25%)	3 (43%)	1 (25%)	1 (20%)	0	0	0
RND	21	15	15	17	19	20	20
Chromosome	10 (48%)	10 (67%)	10 (67%)	14 (82%)	11 (58%)	15 (75%)	11 (55%)
ECRs	11 (52%)	5 (33%)	5 (33%)	3 (18%)	9 (42%)	5 (25%)	9 (45%)
TRAP	19	17	19	25	19	20	19
Chromosome	11 (58%)	11 (65%)	13 (68%)	15 (60%)	10 (53%)	14 (70%)	10 (53%)
ECRs	8 (42%)	6 (35%)	6 (32%)	10 (40%)	9 (47%)	6 (30%)	9 (47%)
TTT	16	15	13	16	19	27	16
Chromosome	16 (100%)	13 (87%)	10 (77%)	13 (81%)	16 (84%)	19 (70%)	13 (81%)
ECRs	0	2 (13%)	3 (23%)	3 (19%)	3 (16%)	8 (30%)	3 (19%)
Others/not defined	106	125	107	103	108	114	113
Chromosome	84 (79%)	82 (66%)	81 (76%)	73 (71%)	82 (76%)	89 (78%)	87 (77%)
ECRs	22 (21%)	43 (34%)	26 (24%)	30 (29%)	26 (24%)	25 (22%)	26 (23%)
Total	**563**	**604**	**530**	**557**	**540**	**559**	**545**

3.6 Transposable elements

A search for transposable elements (TEs) in the *P. yeei* genomes revealed the presence of over 900 *tnp* genes (complete or partial) encoding transposases from 18 families of insertion sequences (ISs): IS*3*, IS*5*, IS*6*, IS*21*, IS*30*, IS*66*, IS*91*, IS*110*, IS*256*, IS*481*, IS*630*, IS*701*, IS*1182*, IS*1202*, IS*1380*, IS*1595*, IS*L3*, and ISNCY (Figure 5; Supplementary Table S7). The number of *tnp* genes in individual genomes ranges from 84 (CCUG 32054) to 150 (CCUG 32052 and LM20). The dominant elements are representatives of the IS*3*, IS*5*, IS*110*, and IS*256* families. Several IS families were identified in only one or a few isolates, e.g., IS*6*—CCUG 17731 and CCUG 32052, IS*91*—CCUG 46822, IS*630*—CCUG 32054 and LM20, IS*1202*—CCUG 46822, IS*1380*—CCUG 46822, IS*1595*—CCUG 13493, CCUG 46822 and LM20, and ISNCY—CCUG 13493 and CCUG 32054 (Supplementary Table S7).

Numerous ISs are localized within ECRs. Surprisingly, most of the largest ECRs (DnaA-like and RepB chromids), except for those from CCUG 46822 and LM20, contain relatively few ISs, indicating a more conserved replicon structure and the presence of essential genes. The replicons richest in these elements are RepABC plasmids, which in most cases carry almost 50% of the *tnp* genes of a given strain (Supplementary Table S7).

In the chromosomes and ECRs of strains CCUG 46822 and LM20 several repetitive IS arrays were identified with a structure resembling composite transposon (Supplementary Table S7). These DNA regions are flanked by divergently oriented complete isoforms of IS*Pye46* (1361 bp; IS*110* family) and contained within them are *tnp* genes of the IS*256* family (related to IS*Pye43*) and/or the IS*5* family (related to IS*Pye12*). Each copy of IS*Pye46* is flanked by identical DRs (5′-TG-3′), so it is not clear whether these IS*110*-IS*5*/IS*256*-IS*110* arrays constitute functional TEs or are the result of preferential transposition of IS*Pye46* into ISs representing the IS*256* or IS*5* families. Interestingly, IS*Pye46* is dominant in the analyzed *P. yeei* genomes (47 copies in total); however, most of the copies are present in strains CCUG 46822 (19 copies) and LM20 (24 copies) (Figure 5; Supplementary Table S7). No related IS*110*-IS*5*/IS*256*-IS*110* arrays were identified in the genomes of other bacteria by BLASTn analysis.

Interestingly, a 159-bp-long sequence element was identified in most chromosomes of the analyzed *P. yeei* strains (Supplementary Table S7). These elements show high sequence similarity (96–100% identity), lack any open reading frames, occur in intergenic regions at different locations, and are flanked by short (4 bp) direct repeats (DRs), suggesting their acquisition through transposition (Supplementary Table S7). Additionally, they contain

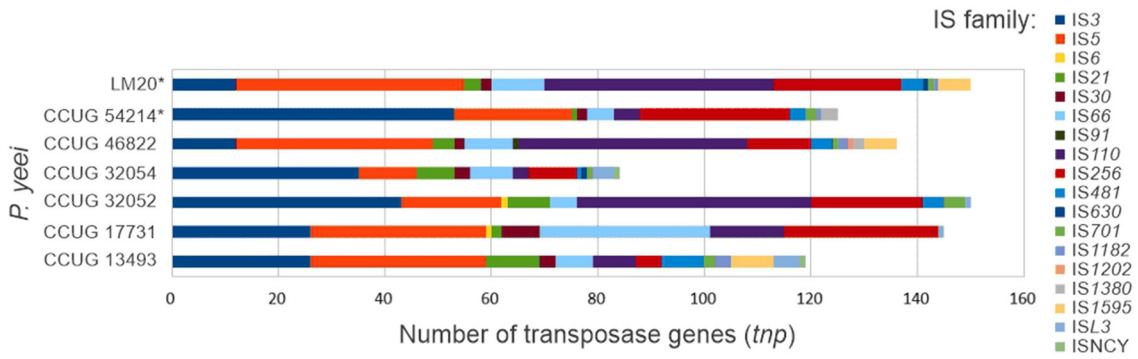


FIGURE 5
Distribution of insertion sequences in *P. yeei* genomes (*environmental isolates).

inverted repeats at both termini (TIRs; IRL and IRR) (18 bp) with significant similarity to the TIRs of insertion sequences of the IS*1182* family, including IS*Pye18*, previously identified in the CCUG 32053 genome (Lasek et al., 2018) (Figure 6A). Analysis of the distribution of IS*Pye18* in *P. yeei* genomes revealed the presence of intact isoforms of this element in strains FDAARGOS_252 (1 copy; plasmid 4) and CCUG 32053 (1 copy; plasmid pYEE3). However, related ISs from the IS*1182* family were also identified in the genomes of CCUG 13493, CCUG 46822, CCUG 54214, and LM20 (Supplementary Figure S3).

Considering (i) the potential mobility of these DNA regions, (ii) their small size, (iii) the presence of TIRs, (iv) DRs, and (v) the absence of a transposase gene, it seems likely that they constitute a novel group of MITE-type non-autonomous TEs (Delihas, 2011), that are most probably mobilized for transposition by ISs of the IS*1182* family. This element was named MITE*Pye1* (ISfinder).

The activity of TEs of *P. yeei* was investigated *in vivo* using trap plasmids that allow positive selection of transposition events. Two trap plasmids were used, pMEC1 and pMAT1, containing the *cI-tetA* and *sacB* selection cassettes, respectively. Transposition of TEs into these cassettes resulted in the appearance of new phenotypes—the cells carrying a mutated plasmid become resistant to tetracycline (pMEC1) or to sucrose (pMAT1) (see Section 2 for details).

Despite the presence of a very large number of transposase genes in the *P. yeei* genomes, use of the entrapment vectors led to the identification of only six distinct elements: (i) three ISs previously identified *in silico* in the CCUG 32053 genome—IS*Pye2* (IS*5* family, IS*5* group), IS*Pye38* (IS*5* family) and IS*Pye41* (IS*5* family, IS*903* group), (ii) an isoform of IS*1247* of *Xanthobacter autotrophicus* (IS*1380* family), and (iii) two novel elements, designated IS*Pye79* and IS*Pye80*, representing the IS*5* and IS*427* groups within the IS*5* family, respectively.

The distribution of every defined IS of *P. yeei* (ISfinder and this study) across all *P. yeei* genomes is shown in Supplementary Figure S3. The most abundant elements are (i) IS*Pye46*—47 copies in 5 strains (most copies present in CCUG 46822 and LM20), (ii) IS*Pye53*—45 copies in seven strains, and (iii) IS*Pye41*—30 copies in six strains (most copies present in CCUG 17731 and CCUG 32053). Many of the ISs are present in only 1 copy, e.g., (i) IS*Pye13* (chromosome FDAARGOS_252), (ii) IS*Pye63* (chromosome of CCUG 32052), and (iii) IS*1247*a (chromosome of CCUG 46822).

3.7 *P. yeei* strain- and species-specific genes

The core genomes of *Paracoccus* spp. and of *P. yeei* were defined based on the complete genomic sequences of *P. aminophilus* JCM 7686 (Dziewit et al., 2014), *P. aminovorans* JCM 7685 (Czarnecki et al., 2017), *P. contaminans* RKI16-01929T (Aurass et al., 2017), *P. denitrificans* PD1222, plus eleven *P. yeei* strains. Comparative genomic analysis was used to define (i) the core genome of bacteria of the genus *Paracoccus* spp., (ii) genes specific to *P. yeei* (not found in other *Paracoccus* spp. with complete genomes), and (iii) genes that are unique to individual *P. yeei* strains (singletons). The collected data, also showing the distribution of these genes in the individual replicons of each *P. yeei* strain, are presented in Supplementary Table S8.

This analysis revealed that the core genome of *Paracoccus* spp. consists of 1,647 protein-encoding genes, of which 116 are located within ECRs of *P. yeei*. The total number of *P. yeei*-specific genes is 250 and the number of singletons ranges from 138 (CCUG 46822) to 332 (FDAARGOS 252) (Supplementary Table S8). In some strains, the majority of singletons occur in the chromosome, while in others (CCUG 13493, CCUG 17731) they are predominantly in ECRs. The vast majority of singletons encode proteins of unknown function, although a predicted role in adaptation could be assigned to some genes from the *P. yeei* environmental isolates. The chromosome of LM20 encodes (i) the membrane protein YeeE, which mediates the uptake of thiosulfate as an inorganic sulfur source for cysteine synthesis (Tanaka et al., 2020), (ii) mercury transporter Mer, and (iii) the sensor kinase KdpD, which senses potassium levels, and the *kdpABC* operon, which encodes a high-affinity potassium uptake system crucial for bacterial survival in low potassium environments (Laermann et al., 2013; Ali et al., 2017). In the CCUG 54214 genome, many transport-related genes (ABC, RND, TTT types) were distinguished among the singletons, as well as genes for toxin-antitoxin systems (e.g., HipA-, PrlF-, Yha-family), similarly to clinical strains.

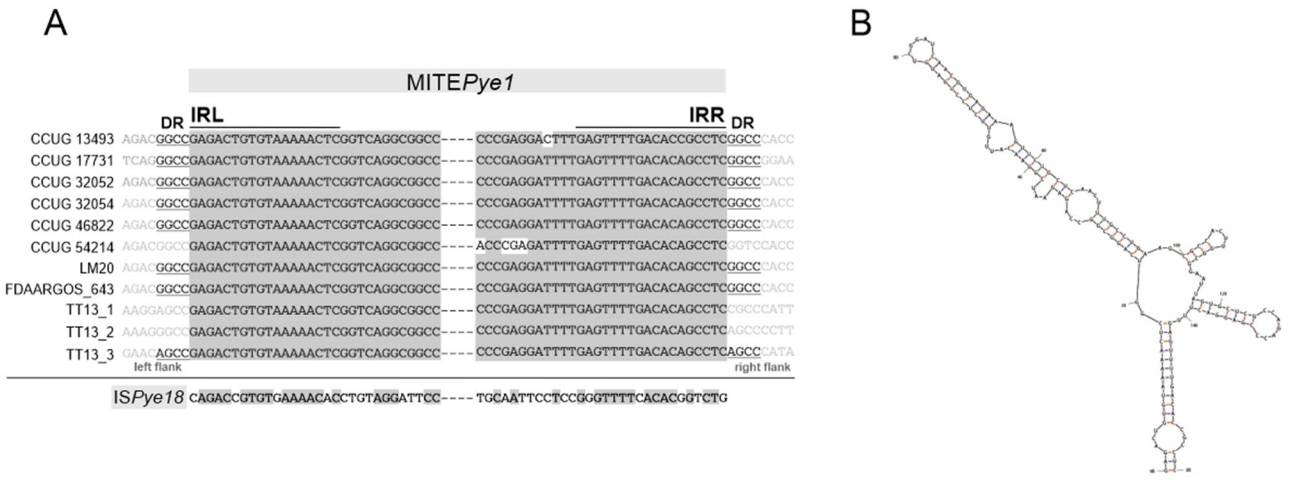


FIGURE 6
Non-autonomous transposable element MITE*Pye1* identified in *P. yeei* strains. **(A)** Alignment of terminal inverted repeat nucleotide sequences (IRL, left IR; IRR, right IR) of complete MITE*Pye1* elements identified in *P. yeei* genomes and IS*Pye18*. Identical residues are indicated by gray shading. Nucleotide sequences of direct repeats (DRs) generated by MITE*Pye1* elements during transposition are indicated by underlining. **(B)** RNA secondary structures predicted by *in silico* folding using Mfold software for MITE*Pye1*. The minimum folding energy (ΔG −47.93) of the predicted secondary structures was calculated by Mfold.

The largest group of *P. yeei*-specific genes (found in chromosomes and DnaA-like, RepB1, RepB2 and RepABC replicons) encode transposases from different IS families. The second largest group of genes encodes transporters, mostly of the ABC type, which are located in chromosomes and in DnaA-like and RepB1 chromids. Other abundant genes, found mainly in RepB1 chromids, encode transcription regulators of the LuxR, CopG, TetR/AcrR, MarR, LacI, AraC, TetR, DeoR/GlpR, and NikR families (Supplementary Table S8).

Examples of proteins encoded by other *P. yeei*-specific genes include a urease enzyme, components of the trehalose, rhamnose, and cobalamin biosynthetic pathways, chaperone proteins, and factors possibly involved in the cell's response to stress conditions (Supplementary Table S8).

3.8 Putative virulence determinants

The proteomes of a number of *Paracoccus* strains were analyzed in order to identify potential virulence factors: (i) the seven *P. yeei* strains described in this study, (ii) other *P. yeei* strains whose complete genomic sequences are available (CCUG 32053, TT13, FDAARGOS_262, FDAARGOS_643), as well as (iii) several environmental isolates of the genus *Paracoccus*, including the type strain *P. denitrificans* PD1222. These protein sets were compared against databases containing bacterial virulence factors and antibiotic resistance genes (see Section 2 for details). Relatively few of the identified determinants were unique to *P. yeei* (all clinical isolates). The pool of identified genes includes those encoding six types of protein, that were previously mentioned by Lasek et al. (2018), i.e., (i) methionine sulfoxide reductase (*msrA1*, *msrA2*, *msrB*), (ii) diguanylate cyclase, (iii) superoxide dismutase, (iv) sugar transferase, (v) type IV secretion system components, and (vi) urease (Supplementary Tables S3, S9).

The chromosomally-located *msr* genes of CCUG 32052 and CCUG 46822 encode predicted proteins sharing 61–100% aa sequence similarity, while the homologous proteins encoded by other *P. yeei* strains are less well-conserved (38–55% aa identity). Msr proteins participate in the detoxification of reactive oxygen intermediates (both prokaryotic and eukaryotic cells lacking Msr are sensitive to oxidative stress) (Denkel et al., 2011). Moreover, MsrA has been shown to be critical for the survival of *Erwinia chrysanthemi* (El Hassouni et al., 1999), *Mycoplasma genitalium* (Das et al., 2012), and *Helicobacter pylori* (Alamuri and Maier, 2006) in their infected hosts. Genes encoding components of a putative type IV secretion system (T4SS) were identified only in CCUG 17731 (30–50% aa identity) and in CCUG 13493 (25–50% aa identity)—within RepABC plasmids in both cases. However, these were not complete modules, unlike the one previously identified in CCUG 32053 (Lasek et al., 2018).

The remaining virulence factors, present in all analyzed genomes, include the gene encoding diguanylate cyclase, an enzyme synthesizing cyclic diguanylate (c-di-GMP)—an important bacterial second messenger. This signaling molecule is involved in the regulation of a number of complex physiological processes, including biofilm formation and motility, which affect the pathogenesis of many bacteria (Tamayo et al., 2007).

Genes encoding several other putative virulence factors were identified within ECRs, although they were not unique to *P. yeei*. These include proteins participating in the following: (i) numerous ABC and RND transport systems (DnaA-like, RepB, RepABC replicons), (ii) siderophore synthesis (RepB2 replicons, chromosomes), (iii) polysaccharide capsule production (RepB2 replicons), and (iv) trehalose synthesis (DnaA-like type replicons).

All *P. yeei* strains contain a set of genes (URE) involved in the synthesis of the metalloenzyme urease, which hydrolyzes urea to ammonia and CO_2. These genes are located within essential replicons—DnaA-like chromids. They display synteny and a high degree of nucleotide sequence identity (98%), indicating their

biological importance. The role of urease in bacterial pathogenesis is well-documented (e.g., Mobley, 1996; Dupuy et al., 1997; Burne and Chen, 2000; Berutti et al., 2014; Rutherford, 2014; Graham and Miftahussurur, 2018; Zhou et al., 2019; Minami et al., 2021).

3.9 URE gene clusters

Besides genes for the main urease subunits (UreA—γ subunit, UreB—β subunit, and UreC—α subunit) and accessory proteins (*ureD*, *ureE*, *ureF*, and *ureG*), the URE gene cluster also contains the *nikR* gene (encoding a predicted nickel-dependent transcription factor) and numerous ABC transporter genes, likely to be involved in the uptake of Ni^{2+} (an essential cofactor for urease; Rutherford, 2014). It is probable that NikR regulates the expression of the transporter genes (Figure 7). Interestingly, in *P. yeei* strain CCUG 32052, part of the URE gene cluster, containing the *ureABC* and *ureD* genes, has been duplicated and transferred to another ECR (RepABC; pYEE32052P4) (99% nucleotide sequence identity to the corresponding URE genes) (Figure 7).

The URE gene cluster of DnaA-like replicons (designated URE type 1) (Figure 7) is unique to and characteristic of *P. yeei* strains. However, in two strains (CCUG 13493 and CCUG 32052) another set of *ure* genes (URE type 2; Figure 7), possibly associated with urease synthesis, was identified. These genes are present in RepABC plasmids in both strains (pYEE13493P4 and pYEE32052P5, respectively) and their genetic organization is different from that of URE type 1. In the URE type 2 cluster, the γ and β urease subunits are encoded by a single fusion gene *ureBC*, the order and transcriptional orientation of the accessory genes *ureFEDG* is different, there are fewer transporter genes and the *nikR* gene is absent (Figure 7).

URE type 2 represents a novel type of ureolytic gene cluster which has a different evolutionary origin from URE type 1. The two clusters share some sequence similarity between the genes encoding the main urease subunits, but no similarity was found in case of the accessory genes and their arrangement is different.

The ureolytic activity of three *P. yeei* strains (CCUG 32053, CCUG 32052—containing the duplicated part of URE type 1, and CCUG 32052—also carrying URE type 2; Figure 7) was examined using Christensen's differential medium without (control) and with the addition of urea (1%) (see Section 2 for details). All the strains displayed ureolytic activity, which was induced by the presence of urea in the medium.

To unequivocally confirm the role of the type 1 URE gene cluster in the observed ureolytic activity, mutations were introduced by insertion of a kanamycin resistance cassette into the *ureC* (encoding the α subunit of urease) or *nikR* genes of CCUG 32053 and CCUG 13493. These mutations completely abolished ureolytic activity in these strains, indicating the importance of both *ureC* and *nikR* in determining this phenotype. The lack of ureolytic activity in mutated strain CCUG 13493, which also contains URE type 2, indicates that this second cluster does not produce an active urease under the tested conditions and its genes are unable to complement mutations introduced in URE type 1.

4 Discussion

These comprehensive comparative genomic analyses of a pool of strains of the opportunistic bacterium *P. yeei*, have provided valuable information on (i) the structural diversity of genomes within this species, (ii) the components of the mobilome, their impact on genome structure and properties, and (iii) the genetic information determining the opportunistic properties of this species. Importantly, the analyzed strains originated from different environments (clinical and natural) and were isolated at different times. Therefore, the collected data better reflects the extent of strain variability within this species, and also permits the identification of genes conserved during evolution, including those involved in the process of pathogenesis (Table 1).

The physiological and phenotypic properties of the analyzed strains were found to be very similar. High levels of similarity were also evident when the structure and genetic content of the genomes of clinical and environmental isolates were compared. The relatedness of these strains is reflected in the topology of the *P. yeei* phylogenetic tree (Figure 1A).

All strains contain numerous ECRs of varying sizes, ranging from 4 to 485 kbp. It should be noted that strain CCUG 46822 carries the largest number of ECRs (13) among *Paracoccus* spp. strains analyzed so far. The main ECRs are large replicons representing the DnaA-like, RepB (RepB1 and RepB2 subgroups) and RepABC families, which is consistent with our previous observations (Lasek et al., 2018) (Figures 1B, C). All of the DnaA-like, RepB1 and RepB2 replicons (which meet the criteria of chromids) contain numerous adaptive genes (facilitating survival in challenging environmental conditions) plus a set of conserved genes of the core *Paracoccus* spp. genome. However, the genetic load of these replicons is not equal. DnaA-like replicons can be considered major chromids due to (i) the presence of the largest number of core genes, (ii) the relatively low variability in their structure and size (especially in relation to DnaA-like replicons of other *Paracoccus* species) (Figure 3; Supplementary Table S5), and (iii) the presence of the URE gene cluster, which is chromosomally located in other pathogenic bacteria.

DnaA-like chromids are characteristic for *Paracoccus* spp. and play an important role in the biology of these bacteria. Related essential replicons were previously identified and analyzed e.g., in *P. denitrificans* PD1222, *P. aminophilus* JCM 7686 and *P. aminovorans* JCM 7685 (Dziewit et al., 2014; Czarnecki et al., 2017). The removal of DnaA-like and RepB chromids from *P. aminophilus* cells was attempted (Dziewit et al., 2014), but this was only successful in the latter case. The strain lacking RepB replicons grew much more slowly on complete medium and was unable to grow at all on minimal media (regardless of the type of carbon source). Based on these observations, two classes of chromids were proposed: obligatorily essential primary chromids (e.g., DnaA-like) and facultatively essential secondary chromids (e.g., RepB).

It should be noted that DnaA-like chromids have yet to be identified in other taxonomic groups of bacteria, which is consistent with the assertion of Harrison et al. (2010) that particular chromid types are characteristic for a specific taxonomic group of bacteria (mainly genus). This claim seems reasonable given that these types of replicons are likely to have arisen from plasmids

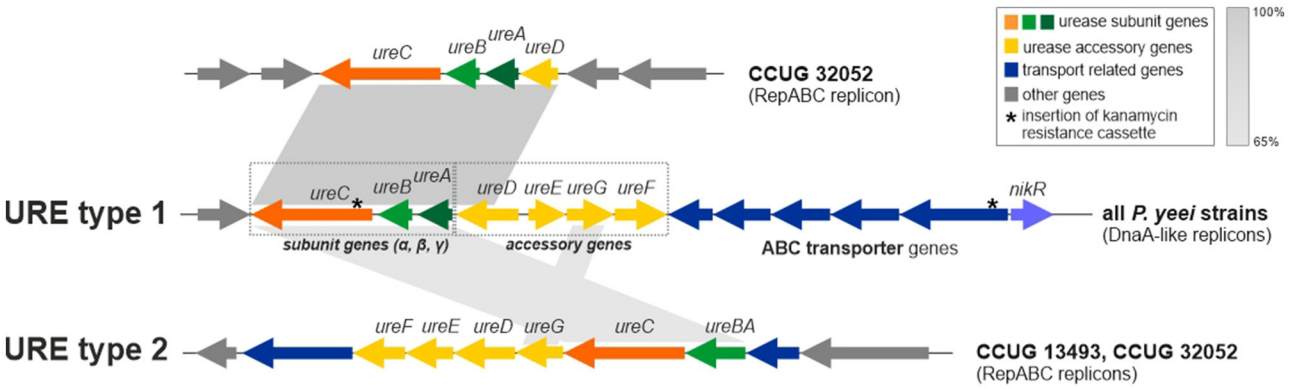


FIGURE 7
Genetic organization of urease gene clusters (URE) located in *P. yeei* ECRs—URE type 1 (present in all DnaA-like chromids) and URE type 2 (present in two RepABC plasmids of *P. yeei* CCUG 13493 and CCUG 32052). Shaded areas connect homologous DNA regions.

(potentially with a narrow host range) specific to a particular group of bacteria. DnaA-like replicons therefore have some aptitude to generate chromids. However, it has yet to be explained why, for example, RepABC replicons, which are the main chromids of e.g.. *Rhizobium* spp. and *Agrobacterium* spp. (Landeta et al., 2011; diCenzo et al., 2013; Döhlemann et al., 2017), despite also being common in *Paracoccus* spp., do not play the same role in these bacteria.

The flexible genome of *P. yeei* also comprises transposable elements. A previous study examining strain CCUG 32053 highlighted an unusual diversity of insertion sequences, not seen in other *Paracoccus* spp. strains (Lasek et al., 2018). This analysis resulted in the examination of the prevalence of 80 ISs (IS*Pye*1–IS*Pye*80) sequences in the genomes of *P. yeei*, that had not been previously characterized in the transposable mobilome of *Paracoccus* spp. (Dziewit et al., 2012). In this study, we searched for putative TEs in all available complete genomic sequences of *P. yeei*. The results of these analyses showed (i) the widespread presence of ISs in this species (although there is great variability in their number in individual strains), (ii) the predominance of ISs belonging to the IS*5*, IS*3*, IS*110*, and IS*256* families (Figure 5; Supplementary Figure S3), and the absence of (iii) *P. yeei*-specific elements conserved in all genomes or (iv) transposons that could be linked with potential determinants of pathogenesis. A more global comparative analysis using data collected in the ISfinder database showed that the overall distribution pattern of IS families in the *P. yeei* genomes corresponds to their distribution across the genus *Paracoccus*. This also highlighted the ubiquity and diversity of these elements in *Paracoccus* spp. in comparison to other genera of *Alphaproteobacteria* (Supplementary Figure S4).

The presence of multiple copies of particular ISs in *P. yeei* genomes may promote homologous recombination, resulting in genomic structural rearrangements. It is likely that such events led to the generation of the two chromosome structural variants identified in this study (Figure 3). Notably, the large number of TEs predicted *in silico* did not correlate with the number of active elements identified by the application of trap plasmids. Only a few elements that were the most dynamic in the process of transposition were captured. The low activity of TEs may be due to the presence of multiple regulatory systems keeping this process in check to reduce the likelihood of lethality caused by insertional inactivation of host housekeeping genes (Lipszyc et al., 2022).

Analysis of the *P. yeei* genomes also identified a novel group of non-autonomous TEs (MITE-type; MITE*Pye1*). These elements contain TIRs and are bordered by DRs, so resemble functional TEs in this respect (Figure 6A). Their TIRs share significant sequence similarity with terminal sequences of IS*1182*-family elements, which suggests that they originate from defective ISs and their transposition may be trans-activated by compatible transposases. In the chromosomes, these MITE-type elements occur in low numbers (1–3) and are preferentially located within intergenic regions (Supplementary Table S7). Mfold analysis revealed that these elements are able to fold into long stem-loop structures at the RNA level (Figure 6B). Therefore, their co-transcription with upstream genes, may influence the conformation and stability of the resulting transcripts. The altered expression may result in various phenotypes depending on the specific gene function (Szuplewska and Bartosik, 2009; Dziewit et al., 2012; Szuplewska et al., 2014).

Particular attention was given to the identification of genes responsible for the opportunistic phenotype of *P. yeei*. Collections of *P. yeei* species-specific genes were searched for potential virulence factors. A noteworthy finding of this analysis was the very large number of transporter genes, which is a unique feature among *Paracoccus* spp. Although it is difficult to directly link their presence to the ability of these bacteria to cause opportunistic infections, they could potentially play an important role in pathogenesis.

The URE gene clusters of *P. yeei* identified in the screen for potential pathogenicity determinants were examined in more detail. These clusters are involved in the synthesis of urease, an enzyme that acts as a virulence factor in many pathogenic bacteria, including *Staphylococcus* spp. (90% of methicillin-resistant *S. aureus* display ureolytic activity), *Helicobacter pylori*, *Mycobacterium tuberculosis*, *Mycobacterium bovis*, anaerobic *Clostridium perfringens* and *Vibrio parahaemolyticus* (Dupuy et al., 1997; Berutti et al., 2014; Graham and Miftahussurur, 2018; Zhou et al., 2019; Minami et al., 2021). In the case of *H. pylori*, urease is one of the major virulence factors, since strains unable to produce

this enzyme cannot colonize the gastric mucosa. The ammonia released by urease action affects the tissue, causing damage to epithelial cells. Urease activity also enables pathogens to survive and proliferate in macrophages (Fu et al., 2018). This enzyme elicits a response from the human immune system by stimulating the production of antibodies (Konieczna et al., 2012).

Two diverse URE gene clusters were identified in the *P. yeei* genomes: type 1—species specific, and type 2—present only in two RepABC plasmids. We showed that URE type 1 determined the urease activity, which was induced by the presence of urea in the growth medium. Activity of the URE type 2 cluster could not be demonstrated. It may be speculated that these URE module variants require different substrates/activators to initiate the synthesis of urease. Nevertheless, it is highly probable that the presence of additional URE gene clusters may enhance the ureolytic capacity of the bacteria, which highlights the biological importance of this phenotype. It is important to note that urea is present in large quantities in the human body, mainly in the kidneys, but is also found in the stomach, blood serum, sweat and milk (Fu et al., 2018; Graham and Miftahussurur, 2018; Schimmel et al., 2021).

Mutations introduced into the *ureC* (encoding a subunit of urease) or *nikR* (encoding a putative regulator of nickel transporter gene expression) genes of URE type 1 completely abolished urease activity. This effect was previously observed in analogous mutants of *H. pylori*. One of the main phenotypes of an *H. pylori nikR* mutant was the absence of nickel-responsive induction of urease expression (Van Vliet et al., 2002). Further studies showed that NikR binds to the *ureA* promoter (in a nickel-dependent manner), which results in nickel-induced transcription and expression of a ureolytic phenotype (Ernst et al., 2005). We hypothesize that the *P. yeei ure* genes may be regulated in a similar way.

As the conserved URE type 1 gene cluster of *P. yeei* is unique among *Paracoccus* spp., it is highly probable that urease plays an important role in the pathogenesis of this species. The verification of this hypothesis is an immediate goal of our future studies.

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

MS: Conceptualization, Formal analysis, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. DS: Formal analysis, Investigation, Visualization, Writing – review & editing. RL: Data curation, Formal analysis, Visualization, Writing – review & editing. PD: Data curation, Formal analysis, Methodology, Software, Writing – review & editing. MH: Formal analysis, Investigation, Writing – review & editing. ŁF: Investigation, Writing – review & editing. CC: Formal analysis, Investigation, Writing – review & editing. DB: Conceptualization, Formal analysis, Methodology, 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. This research was funded by the National Science Center (NCN), Poland (grant no. 2016/21/D/NZ2/03824).

Acknowledgments

We would like to thank Aleksandra Pawłot for her experimental assistance with physiological characterization of bacteria, Albert Roethel for his help with the identification of transposable elements, and Anna Rokowska and Kamil Krakowski for their contribution to genomic data 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.2024.1483110/full#supplementary-material

References

Alamuri, P., and Maier, R. J. (2006). Methionine sulfoxide reductase in *Helicobacter pylori*: interaction with methionine-rich proteins and stress-induced expression. *J. Bacteriol.* 188:5839. doi: 10.1128/JB.00430-06

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

Ali, M. K., Li, X., Tang, Q., Liu, X., Chen, F., Xiao, J., et al. (2017). Regulation of inducible potassium transporter KdpFABC by the KdpD/KdpE two-component system in *Mycobacterium smegmatis*. *Front. Microbiol.* 8:570. doi: 10.3389/fmicb.2017.00570

Aliste-Fernández, M., Sanfeliu-Sala, I., and Sánchez-Delgado, J. (2020). Bacteremia caused by *Paracoccus yeei* in patient with compensated cirrhosis of the liver. *Enferm. Infecc. Microbiol. Clin.* 38, 451–452. doi: 10.1016/j.eimc.2020.01.013

Amstrup, S. K., Ong, S. C., Sofos, N., Karlsen, J. L., Skjerning, R. B., Boesen, T., et al. (2023). Structural remodelling of the carbon–phosphorus lyase machinery by a dual ABC ATPase. *Nat. Commun.* 14, 1–12. doi: 10.1038/s41467-023-36604-y

Arias, M. A., and Clark, J. (2019). *Paracoccus yeei* as a cause of peritoneal dialysis peritonitis in the United Kingdom. *IDCases* 15:e00486. doi: 10.1016/j.idcr.2019.e00486

Aurass, P., Karste, S., Trost, E., Glaeser, S. P., Kämpfer, P., and Flieger, A. (2017). Genome sequence of *Paracoccus contaminans* LMG 29738T, isolated from a water microcosm. *Genome Announc.* 5, 487–504. doi: 10.1128/genomeA.00487-17

Baker, S. C., Ferguson, S. J., Ludwig, B., Page, M. D., Richter, O.-M. H., and van Spanning, R. J. M. (1998). Molecular genetics of the genus *Paracoccus*: metabolically versatile bacteria with bioenergetic flexibility. *Microbiol. Mol. Biol. Rev.* 62, 1046–1078. doi: 10.1128/MMBR.62.4.1046-1078.1998

Bartosik, D., Sochacka, M., and Baj, J. (2003). Identification and characterization of transposable elements of *Paracoccus pantotrophus*. *J. Bacteriol.* 185:3753. doi: 10.1128/JB.185.13.3753-3763.2003

Bartosik, D., Wlodarczyk, M., and Thomas, C. M. (1997). Complete nucleotide sequence of the replicator region of *Paracoccus (Thiobacillus) versutus* pTAV1 plasmid and its correlation to several plasmids of *Agrobacterium* and *Rhizobium* species. *Plasmid* 38, 53–59. doi: 10.1006/plas.1997.1295

Beijerinck, M. W., and Minkman, D. C. J. (1910). Bildung und verbrauch von stickoxydul durch bakterien. *Zentbl. Bakteriol. Parasitenkd. Infektionskr. Hyg. Abt. II* 25, 30–63.

Berry, A., Janssens, D., Hümbelin, M., Jore, J. P. M., Hoste, B., Cleenwerck, I., et al. (2003). *Paracoccus zeaxanthinifaciens* sp. nov., a zeaxanthin-producing bacterium. *Int. J. Syst. Evol. Microbiol.* 53, 231–238. doi: 10.1099/ijs.0.02368-0

Berutti, T. R., Williams, R. E., Shen, S., Taylor, M. M., and Grimes, D. J. (2014). Prevalence of urease in *Vibrio parahaemolyticus* from the Mississippi Sound. *Lett. Appl. Microbiol.* 58, 624–628. doi: 10.1111/lam.12237

Bhikoo, R., Chong, S., McLeod, K., Heyworth, P., and McAllister, I. L. (2022). *Paracoccus yeei*, a novel bacterial cause of endophthalmitis following intravitreal injection. *Ocul. Immunol. Inflamm.* 30, 995–997. doi: 10.1080/09273948.2020.1836233

Birnboim, H. C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* 7, 1513–1523. doi: 10.1093/nar/7.6.1513

Bordel, S., Martín-González, D., Börner, T., Muñoz, R., and Santos-Beneit, F. (2024). Genome-scale metabolic model of the versatile bacterium *Paracoccus denitrificans* Pd1222. *mSystems* 9:e0107723. doi: 10.1128/msystems.01077-23

Burne, R. A., and Chen, Y. Y. M. (2000). Bacterial ureases in infectious diseases. *Microbes Infect.* 2, 533–542. doi: 10.1016/S1286-4579(00)00312-9

Cantais, A., Grattard, F., Gagnaire, J., Mory, O., Plat, A., Lleres-Vadeboin, M., et al. (2020). Longitudinal study of viral and bacterial contamination of hospital pediatricians' mobile phones. *Microorganisms* 8, 1–12. doi: 10.3390/microorganisms8122011

Carlos, C., Pereira, L. B., and Ottoboni, L. M. M. (2017). Comparative genomics of *Paracoccus* sp. SM22M-07 isolated from coral mucus: insights into bacteria-host interactions. *Curr. Genet.* 63, 509–518. doi: 10.1007/s00294-016-0658-3

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

Chistoserdova, L. (2011). Modularity of methylotrophy, revisited. *Environ. Microbiol.* 13, 2603–2622. doi: 10.1111/j.1462-2920.2011.02464.x

Choi, J. H., Cho, Y. S., Lee, J. W., Shin, H. B., and Lee, I. K. (2017). Bacterial contamination and disinfection status of laryngoscopes stored in emergency crash carts. *J. Prev. Med. Public Health* 50:158. doi: 10.3961/jpmph.17.013

Christensen, W. B. (1946). Urea decomposition as a means of differentiating proteus and paracolon cultures from each other and from *Salmonella* and *Shigella* Types. *J. Bacteriol.* 52, 461–466. doi: 10.1128/jb.52.4.461-466.1946

Courjaret, J. C., Drancourt, M., and Hoffart, L. (2014). *Paracoccus yeei* keratitis in a contact lens wearer. *Eye Contact Lens* 40, e21–e22. doi: 10.1097/ICL.0b013e31829e8fc7

Czarnecki, J., and Bartosik, D. (2019). Diversity of methylotrophy pathways in the genus *Paracoccus* (*Alphaproteobacteria*). *Curr. Issues Mol. Biol.* 33, 117–131. doi: 10.21775/cimb.033.117

Czarnecki, J., Dziewit, L., Puzyna, M., Prochwicz, E., Tudek, A., Wibberg, D., et al. (2017). Lifestyle-determining extrachromosomal replicon pAMV1 and its contribution to the carbon metabolism of the methylotrophic bacterium *Paracoccus aminovorans* JCM 7685. *Environ. Microbiol.* 19, 4536–4550. doi: 10.1111/1462-2920.13901

Daneshvar, M. I., Hollis, D. G., Weyant, R. S., Steigerwalt, A. G., Whitney, A. M., Douglas, M. P., et al. (2003). *Paracoccus yeeii* sp. nov. (formerly CDC Group EO-2), a novel bacterial species associated with human infection. *J. Clin. Microbiol.* 41:1289. doi: 10.1128/JCM.41.3.1289-1294.2003

Darriba, D.i., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., and Flouri, T. (2020). ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* 37, 291–294. doi: 10.1093/molbev/msz189

Das, K., de la Garza, G., Maffi, S., Saikolappan, S., and Dhandayuthapani, S. (2012). Methionine sulfoxide reductase A (MsrA) deficient *Mycoplasma* genitalium shows decreased interactions with host cells. *PLoS ONE* 7:e0036247. doi: 10.1371/journal.pone.0036247

Delihas, N. (2011). Impact of small repeat sequences on bacterial genome evolution. *Genome Biol. Evol.* 3:959. doi: 10.1093/gbe/evr077

Denkel, L. A., Horst, S. A., Rouf, S. F., Kitowski, V., Böhm, O. M., Rhen, M., et al. (2011). Methionine sulfoxide reductases are essential for virulence of *Salmonella typhimurium*. *PLoS ONE* 6:e0026974. doi: 10.1371/journal.pone.0026974

diCenzo, G., Milunovic, B., Cheng, J., and Finan, T. M. (2013). The tRNAarg gene and engA are essential genes on the 1.7-Mb pSymB megaplasmid of *Sinorhizobium meliloti* and were translocated together from the chromosome in an ancestral strain. *J. Bacteriol.* 195, 202–212. doi: 10.1128/JB.01758-12

Ditta, G., Stanfield, S., Corbin, D., and Helinski, D. R. (1980). Broad host range DNA cloning system for gram-negative bacteria: construction of a gene bank of *Rhizobium meliloti*. *Proc. Natl. Acad. Sci. U. S. A.* 77, 7347–7351. doi: 10.1073/pnas.77.12.7347

Döhlemann, J., Wagner, M., Happel, C., Carrillo, M., Sobetzko, P., Erb, T. J., et al. (2017). A family of single copy repABC-type shuttle vectors stably maintained in the alpha-proteobacterium *Sinorhizobium meliloti*. *ACS Synth. Biol.* 6, 968–984. doi: 10.1021/acssynbio.6b00320

Dupuy, B., Daube, G., Popoff, M. R., and Cole, S. T. (1997). *Clostridium perfringens* urease genes are plasmid borne. *Infect. Immun.* 65:2313. doi: 10.1128/iai.65.6.2313-2320.1997

Dyer, J., and Harris, P. (2020). *Paracoccus yeei* – An emerging pathogen or incidental finding? *Pathology* 52:S123. doi: 10.1016/j.pathol.2020.01.419

Dziewit, L., Adamczuk, M., Szuplewska, M., and Bartosik, D. (2011). DIY series of genetic cassettes useful in construction of versatile vectors specific for *Alphaproteobacteria*. *J. Microbiol. Methods* 86, 166–174. doi: 10.1016/j.mimet.2011.04.016

Dziewit, L., Baj, J., Szuplewska, M., Maj, A., Tabin, M., Czyzkowska, A., et al. (2012). Insights into the transposable mobilome of *Paracoccus* spp. (*Alphaproteobacteria*). *PLoS ONE* 7:e0032277. doi: 10.1371/journal.pone.0032277

Dziewit, L., Czarnecki, J., Wibberg, D., Radlinska, M., Mrozek, P., Szymczak, M., et al. (2014). Architecture and functions of a multipartite genome of the methylotrophic bacterium *Paracoccus aminophilus* JCM 7686, containing primary and secondary chromids. *BMC Genom.* 15, 1–16. doi: 10.1186/1471-2164-15-124

Dziewit, L., Pyzik, A., Matlakowska, R., Baj, J., Szuplewska, M., and Bartosik, D. (2013). Characterization of *Halomonas* sp. ZM3 isolated from the Zelazny Most post-flotation waste reservoir, with a special focus on its mobile DNA. *BMC Microbiol.* 13:59. doi: 10.1186/1471-2180-13-59

Dziewit, L., Pyzik, A., Szuplewska, M., Matlakowska, R., Mielnicki, S., Wibberg, D., et al. (2015). Diversity and role of plasmids in adaptation of bacteria inhabiting the Lubin copper mine in Poland, an environment rich in heavy metals. *Front. Microbiol.* 6:152. doi: 10.3389/fmicb.2015.00152

Eckhardt, T. (1978). A rapid method for the identification of plasmid desoxyribonucleic acid in bacteria. *Plasmid* 1, 584–588. doi: 10.1016/0147-619X(78)90016-1

El Hassouni, M., Chambost, J. P., Expert, D., Van Gijsegem, F., and Barras, F. (1999). The minimal gene set member msrA, encoding peptide methionine sulfoxide reductase, is a virulence determinant of the plant pathogen *Erwinia chrysanthemi*. *Proc. Natl. Acad. Sci. U. S. A.* 96, 887–892. doi: 10.1073/pnas.96.3.887

Ernst, F. D., Kuipers, E. J., Heijens, A., Sarwari, R., Stoof, J., Penn, C. W., et al. (2005). The nickel-responsive regulator NikR controls activation and repression of gene transcription in *Helicobacter pylori*. *Infect. Immun.* 73, 7252–7258. doi: 10.1128/IAI.73.11.7252-7258.2005

Fosso, C., Maillart, E., Beun, B., Touzani, F., Mahadeb, B., and Clevenbergh, P. (2021). Opportunistic peritonitis in peritoneal dialysis: the example of *Paracoccus yeei*. *Clin. Case Rep.* 9:4176. doi: 10.1002/ccr3.4176

Fu, M. S., Coelho, C., De Leon-Rodriguez, C. M., Rossi, D. C. P., Camacho, E., Jung, E. H., et al. (2018). *Cryptococcus neoformans* urease affects the outcome of intracellular pathogenesis by modulating phagolysosomal pH. *PLoS Pathog.* 14:7144. doi: 10.1371/journal.ppat.1007144

Funke, G., Frod, R., and Sommer, H. (2004). First comprehensively documented case of *Paracoccus yeei* infection in a human. *J. Clin. Microbiol.* 42, 3366–3368. doi: 10.1128/JCM.42.7.3366-3368.2004

Garcillán-Barcia, M. P., Francia, M. V., and De La Cruz, F. (2009). The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol. Rev.* 33, 657–687. doi: 10.1111/j.1574-6976.2009.00168.x

Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T., and Kado, C. I. (1985). Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J. Bacteriol.* 164:918. doi: 10.1128/JB.164.2.918-921.1985

Ghosh, W., Mandal, S., and Roy, P. (2006). *Paracoccus bengalensis* sp. nov., a novel sulfur-oxidizing chemolithoautotroph from the rhizospheric soil of an Indian tropical leguminous plant. *Syst. Appl. Microbiol.* 29, 396–403. doi: 10.1016/j.syapm.2005.10.004

Graham, D. Y., and Miftahussurur, M. (2018). *Helicobacter pylori* urease for diagnosis of *Helicobacter pylori* infection: a mini review. *J. Adv. Res.* 13, 51–57. doi: 10.1016/j.jare.2018.01.006

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166, 557–580. doi: 10.1016/S0022-2836(83)80284-8

Harrison, P. W., Lower, R. P. J., Kim, N. K. D., and Young, J. P. W. (2010). Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends Microbiol.* 18, 141–148. doi: 10.1016/j.tim.2009.12.010

Hooykaas, P. J. J., den Dulk-Ras, H., and Schilperoort, R. A. (1980). Molecular mechanism of Ti plasmid mobilization by R plasmids: isolation of Ti plasmids with transposon-insertions in *Agrobacterium tumefaciens*. *Plasmid* 4, 64–75. doi: 10.1016/0147-619X(80)90083-9

Horsman, G. P., and Zechel, D. L. (2017). Phosphonate biochemistry. *Chem. Rev.* 117, 5704–5783. doi: 10.1021/acs.chemrev.6b00536

Huesca-Espitia, L., del, C., Aslanzadeh, J., Feinn, R., Joseph, G., Murray, T. S., et al. (2018). Deposition of bacteria and bacterial spores by bathroom hot-air hand dryers. *Appl. Environ. Microbiol.* 84, e00044-18. doi: 10.1128/AEM.00044-18

Hynes, M. F., and McGregor, N. F. (1990). Two plasmids other than the nodulation plasmid are necessary for formation of nitrogen-fixing nodules by *Rhizobium leguminosarum*. *Mol. Microbiol.* 4, 567–574. doi: 10.1111/j.1365-2958.1990.tb00625.x

Iversen, C., Mullane, N., McCardell, B., Tall, B. D., Lehner, A., Fanning, S., et al. (2008). Cronobacter gen. nov., a new genus to accommodate the biogroups of *Enterobacter sakazakii*, and proposal of *Cronobacter sakazakii* gen. nov., comb. nov., *Cronobacter malonaticus* sp. nov., *Cronobacter turicensis* sp. nov., *Cronobacter muytjensii* sp. nov., *Cronobacter dublinensis* sp. nov., *Cronobacter genomospecies* 1, and of three subspecies, *Cronobacter dublinensis* subsp. dublinensis subsp. nov., *Cronobacter dublinensis* subsp. lausannensis subsp. nov. and *Cronobacter dublinensis* subsp. lactaridi sub. *Int. J. Syst. Evol. Microbiol.* 58, 1442–1447. doi: 10.1099/ijs.0.65577-0

Jordan, S. L., McDonald, I. R., Kraczkiewicz-Dowjat, A. J., Kelly, D. P., Rainey, F. A., Murrell, J. C., et al. (1997). Autotrophic growth on carbon disulfide is a property of novel strains of *Paracoccus denitrificans*. *Arch. Microbiol.* 168, 225–236. doi: 10.1007/s002030050492

Kanis, M. J., Oosterheert, J. J., Lin, S., Boel, C. H. E., and Ekkelenkamp, M. B. (2010). Corneal graft rejection complicated by *Paracoccus yeei* infection in a patient who had undergone a penetrating keratoplasty. *J. Clin. Microbiol.* 48:323. doi: 10.1128/JCM.01798-09

Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20, 1160–1166. doi: 10.1093/bib/bbx108

Konieczna, I., Żarnowiec, P., Kwinkowski, M., Kolesińska, B., Fraczyk, J., Kamiński, Z., et al. (2012). Bacterial urease and its role in long-lasting human diseases. *Curr. Protein Pept. Sci.* 13:789. doi: 10.2174/138920312804871094

Kovach, M. E., Elzer, P. H., Steven Hill, D., Robertson, G. T., Farris, M. A., Roop, R. M., et al. (1995). Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* 166, 175–176. doi: 10.1016/0378-1119(95)00584-1

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35, 4453–4455. doi: 10.1093/bioinformatics/btz305

Kruszewska, E., Grze,ś, H., Czupryna, P., Pancewicz, S., Groth, M., Wondim, M., et al. (2021). Fogging with peracetic acid in schools and kindergartens. *Front. Public Health* 9:697917. doi: 10.3389/fpubh.2021.697917

Kushner, S. (1978). "Genetic engineering," in *Proc. Symp. on Genetic Engineering*, eds. H. Boyer, and S. Nicosia (Amsterdam: Elsevier).

Laermann, V., Cudić, E., Kipschull, K., Zimmann, P., and Altendorf, K. (2013). The sensor kinase KdpD of *Escherichia coli* senses external K+. *Mol. Microbiol.* 88, 1194–1204. doi: 10.1111/mmi.12251

Landeta, C., Dávalos, A., Cevallos, M. Á., Geiger, O., Brom, S., and Romero, D. (2011). Plasmids with a chromosome-like role in rhizobia. *J. Bacteriol.* 193, 1317–1326. doi: 10.1128/JB.01184-10

Lasek, R., Szuplewska, M., Mitura, M., Decewicz, P., Chmielowska, C., Pawlot, A., et al. (2018). Genome structure of the opportunistic pathogen *Paracoccus yeei* (*Alphaproteobacteria*) and identification of putative virulence factors. *Front. Microbiol.* 9:2553. doi: 10.3389/fmicb.2018.02553

Lee, J. H., Kim, Y. S., Choi, T. J., Lee, W. J., and Kim, Y. T. (2004). *Paracoccus haeundaensis* sp. nov., a Gram-negative, halophilic, astaxanthin-producing bacterium. *Int. J. Syst. Evol. Microbiol.* 54, 1699–1702. doi: 10.1099/ijs.0.63146-0

Lim, J. Y., Hwang, I., Ganzorig, M., Pokhriyal, S., Singh, R., and Lee, K. (2018). Complete genome sequence of *Paracoccus yeei* TT13, isolated from human skin. *Genome Announc.* 6, e01514-17. doi: 10.1128/genomeA.01514-17

Lipszyc, A., Szuplewska, M., and Bartosik, D. (2022). How do transposable elements activate expression of transcriptionally silent antibiotic resistance genes? *Int. J. Mol. Sci.* 23:63. doi: 10.3390/ijms23158063

Liu, M., Li, X., Xie, Y., Bi, D., Sun, J., Li, J., et al. (2019). ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucl. Acids Res.* 47, D660–D665. doi: 10.1093/nar/gky1123

Liu, X. Y., Wang, B. J., Jiang, C. Y., Liu, S. J., and Liu, S. J. (2006). *Paracoccus sulfuroxidans* sp. nov., a sulfur oxidizer from activated sludge. *Int. J. Syst. Evol. Microbiol.* 56, 2693–2695. doi: 10.1099/ijs.0.64548-0

Liu, Z. P., Wang, B. J., Liu, X. Y., Dai, X., Liu, Y. H., and Liu, S. J. (2008). *Paracoccus halophilus* sp. nov., isolated from marine sediment of the South China Sea, China, and emended description of genus *Paracoccus* Davis 1969. *Int. J. Syst. Evol. Microbiol.* 58, 257–261. doi: 10.1099/ijs.0.65237-0

Maj, A., Dziewit, L., Czarnecki, J., Wlodarczyk, M., Baj, J., Skrzypczyk, G., et al. (2013). Plasmids of carotenoid-producing *Paracoccus* spp. (*Alphaproteobacteria*) - structure, diversity and evolution. *PLoS ONE* 8:e0080258. doi: 10.1371/journal.pone.0080258

Makino, K., Kim, S. K., Shinagawa, H., Amemura, M., and Nakata, A. (1991). Molecular analysis of the cryptic and functional phn operons for phosphonate use in *Escherichia coli* K-12. *J. Bacteriol.* 173, 2665–2672. doi: 10.1128/jb.173.8.2665-2672.1991

Marolda, C. L., and Valvano, M. A. (1995). Genetic analysis of the dTDP-rhamnose biosynthesis region of the *Escherichia coli* VW187 (O7:K1) *rfb* gene cluster: identification of functional homologs of *rfbB* and *rfbA* in the *rff* cluster and correct location of the *rffE* gene. *J. Bacteriol.* 177, 5539–5546. doi: 10.1128/jb.177.19.5539-5546.1995

Minami, M., Hashikawa, S., Ando, T., Kobayashi, H., Goto, H., and Ohta, M. (2021). Involvement of CO_2 generated by urease in multiplication of *Helicobacter pylori*. *GSC Adv. Res. Rev.* 7, 045–053. doi: 10.30574/gscarr.2021.7.3.0123

Mobley, H. L. T. (1996). The role of *Helicobacter pylori* urease in the pathogenesis of gastritis and peptic ulceration. *Aliment. Pharmacol. Ther.* 10(Suppl. 1), 57–64. doi: 10.1046/j.1365-2036.1996.22164006.x

Murgier, J., Coste, J. F., Cavaignac, E., Bayle-Iniguez, X., Chiron, P., Bonnevialle, P., et al. (2016). Microbial flora on cell-phones in an orthopedic surgery room before and after decontamination. *Orthop. Traumatol. Surg. Res.* 102, 1093–1096. doi: 10.1016/j.otsr.2016.09.014

Pasquale, L., Maurano, A., Cengia, G., Da Massa Carrara, P., German,à, B., Graziani, M. G., et al. (2021). Infection prevention in endoscopy practice: comparative evaluation of re-usable vs single-use endoscopic valves. *Infect. Prev. Pract.* 3:100123. doi: 10.1016/j.infpip.2021.100123

Philippe, N., Alcaraz, J. P., Coursange, E., Geiselmann, J., and Schneider, D. (2004). Improvement of pCVD442, a suicide plasmid for gene allele exchange in bacteria. *Plasmid* 51, 246–255. doi: 10.1016/j.plasmid.2004.02.003

Pietrzak, K., Puchalski, M., Otlewska, A., Wrzosek, H., Guiamet, P., Piotrowska, M., et al. (2017). Microbial diversity of pre-Columbian archaeological textiles and the effect of silver nanoparticles misting disinfection. *J. Cult. Herit.* 23, 138–147. doi: 10.1016/j.culher.2016.07.007

Podzelinska, K., He, S. M., Wathier, M., Yakunin, A., Proudfoot, M., Hove-Jensen, B., et al. (2009). Structure of PhnP, a phosphodiesterase of the carbon-phosphorus lyase pathway for phosphonate degradation. *J. Biol. Chem.* 284:17216. doi: 10.1074/jbc.M808392200

Pukall, R., Laroche, M., Kroppenstedt, R. M., Schumann, P., Stackebrandt, E., and Ulber, R. (2003). *Paracoccus seriniphilus* sp. nov., an L-serine-dehydratase-producing coccus isolated from the marine bryozoan *Bugula plumosa*. *Int. J. Syst. Evol. Microbiol.* 53, 443–447. doi: 10.1099/ijs.0.02352-0

Rutherford, J. C. (2014). The emerging role of urease as a general microbial virulence factor. *PLoS Pathog.* 10:e1004062. doi: 10.1371/journal.ppat.1004062

Sack, J., Peaper, D. R., Mistry, P., and Malinis, M. (2017). Clinical implications of *Paracoccus yeeii* bacteremia in a patient with decompensated cirrhosis. *IDCases* 7:9. doi: 10.1016/j.idcr.2016.11.008

Sambrook, J., and Russell, D. (2001). *Molecular Cloning: A Laboratory Manual, 3rd Edn*. New York, NY: Cold Spring Harbor Laboratory Press.

Sartoretti, T., Sartoretti, E., Bucher, C., Doert, A., Binkert, C., Hergan, K., et al. (2017). Bacterial contamination of ultrasound probes in different radiological institutions before and after specific hygiene training: do we have a general hygienical problem? *Eur. Radiol.* 27, 4181–4187. doi: 10.1007/s00330-017-4812-1

Sastre, A., González-Arregoces, J., Romainoik, I., Mariño, S., Lucas, C., Monfá, E., et al. (2016). *Paracoccus yeei* peritonitis in peritoneal dialysis. *Nefrología* 36, 445–446. doi: 10.1016/j.nefroe.2016.09.011

Sayers, S., Li, L., Ong, E., Deng, S., Fu, G., Lin, Y., et al. (2019). Victors: a web-based knowledge base of virulence factors in human and animal pathogens. *Nucleic Acids Res.* 47:D693. doi: 10.1093/nar/gky999

Schimmel, P., Kleinjans, L., Bongers, R. S., Knol, J., and Belzer, C. (2021). Breast milk urea as a nitrogen source for urease positive *Bifidobacterium infantis*. *FEMS Microbiol. Ecol.* 97:fiab019. doi: 10.1093/femsec/fiab019

Schneider, D., Faure, D., Noirclerc-Savoye, M., Barrière, A. C., Coursange, E., and Blot, M. (2000). A broad-host-range plasmid for isolating mobile genetic elements in gram-negative bacteria. *Plasmid.* 44, 201–207. doi: 10.1006/PLAS.2000.1483

Schweiger, M., Stiegler, P., Scarpatetti, M., Wasler, A., Sereinigg, M., Prenner, G., et al. (2011). Case of *Paracoccus yeei* infection documented in a transplanted heart. *Transplant Infectious Disease* 13, 200–203. doi: 10.1111/j.1399-3062.2010.00571.x

Shifera, A. S., Pockrandt, C., Rincon, N., Ge, Y., Lu, J., Varabyou, A., et al. (2021). Identification of microbial agents in tissue specimens of ocular and periocular sarcoidosis using a metagenomics approach. *F1000Res* 10:820. doi: 10.12688/f1000research.55090.1

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

Siller, H., Rainey, F. A., Stackebrandt, E., and Winter, J. (1996). Isolation and characterization of a new gram-negative, acetone-degrading, nitrate-reducing bacterium from soil, *Paracoccus solventivorans* sp. nov. *Int. J. Syst. Bacteriol.* 46, 1125–1130. doi: 10.1099/00207713-46-4-1125

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

Szuplewska, M., and Bartosik, D. (2009). Identification of a mosaic transposable element of *Paracoccus* marcusii composed of insertion sequence IS*Pmar4* (IS*As1* family) and an IS*1247*a-driven transposable module (TMo). *FEMS Microbiol. Lett.* 292, 216–221. doi: 10.1111/j.1574-6968.2009.01495.x

Szuplewska, M., Ludwiczak, M., Lyzwa, K., Czarnecki, J., and Bartosik, D. (2014). Mobility and generation of mosaic non-autonomous transposons by Tn*3*-derived inverted-repeat miniature elements (TIMEs). *PLoS ONE* 9:0105010. doi: 10.1371/journal.pone.0105010

Tamayo, R., Pratt, J. T., and Camilli, A. (2007). Roles of cyclic diguanylate in the regulation of bacterial pathogenesis. *Annu. Rev. Microbiol.* 61, 131–148. doi: 10.1146/annurev.micro.61.080706.093426

Tanaka, Y., Yoshikaie, K., Takeuchi, A., Ichikawa, M., Mori, T., Uchino, S., et al. (2020). Crystal structure of a YeeE/YedE family protein engaged in thiosulfate uptake. *Sci. Adv.* 6:aba7637. doi: 10.1126/sciadv.aba7637

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., et al. (2003). The COG database: an updated vesion includes eukaryotes. *BMC Bioinf.* 4, 1–14. doi: 10.1186/1471-2105-4-41

Thomas, C. M. (1981). Complementation analysis of replication and maintenance functions of broad host range plasmids RK2 and RP1. *Plasmid* 5, 277–291. doi: 10.1016/0147-619X(81)90005-6

Thompson, J. R., Argyraki, A., Bashton, M., Bramwell, L., Crown, M., Hursthouse, A. S., et al. (2021). Bacterial diversity in house dust: characterization of a core indoor microbiome. *Front. Environ. Sci.* 9:754657. doi: 10.3389/fenvs.2021.754657

Tsubokura, A., Yoneda, H., and Mizuta, H. (1999). *Paracoccus carotinifaciens* sp. nov., a new aerobic Gram-negative astaxanthin-producing bacterium. *Int. J. Syst. Bacteriol.* 49, 277–282. doi: 10.1099/00207713-49-1-277

Urakami, T., Araki, H., Oyanagi, H., Suzuki, K. I., and Komagata, K. (1990). *Paracoccus aminophilus* sp. nov. and *Paracoccus aminovorans* sp. nov., which utilize N,N-dimethylformamide. *Int. J. Syst. Bacteriol.* 40, 287–291. doi: 10.1099/00207713-40-3-287

Van Vliet, A. H. M., Poppelaars, S. W., Davies, B. J., Stoof, J., Bereswill, S., Kist, M., et al. (2002). NikR mediates nickel-responsive transcriptional induction of urease expression in *Helicobacter pylori*. *Infect. Immun.* 70:2846. doi: 10.1128/IAI.70.6.2846-2852.2002

Wallet, F., Blondiaux, N., de Sainte Foy, C. L., Loïez, C., Armand, S., Pagniez, D., et al. (2010). *Paracoccus yeei*: a new unusual opportunistic bacterium in ambulatory peritoneal dialysis. *Int. J. Infect. Dis.* 14:30. doi: 10.1016/j.ijid.2009.03.030

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., et al. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucl. Acids Res.* 47, W52–W58. doi: 10.1093/nar/gkz333

Zhou, C., Bhinderwala, F., Lehman, M. K., Thomas, V. C., Chaudhari, S. S., Yamada, K. J., et al. (2019). Urease is an essential component of the acid response network of *Staphylococcus aureus* and is required for a persistent murine kidney infection. *PLoS Pathog.* 15:1007538. doi: 10.1371/journal.ppat.1007538

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.* 31, 3406–3415. doi: 10.1093/nar/gkg595

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 11 December 2024
DOI 10.3389/fmicb.2024.1496223

OPEN ACCESS

EDITED BY
Qun Gao,
Beijing Normal University, China

REVIEWED BY
Anastasia D. Gazi,
Institut Pasteur, France
Ashish Kumar Singh,
Center of Innovative and Applied Bioprocessing (CIAB), India
Xianglilan Zhang,
State Key Laboratory of Pathogen and Biosafety of China, China

*CORRESPONDENCE
Carlos A. Santiviago
csantiviago@ciq.uchile.cl
David Pezoa
dpezoa@udla.cl

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

RECEIVED 13 September 2024
ACCEPTED 12 November 2024
PUBLISHED 11 December 2024

CITATION
Amaya FA, Blondel CJ, Reyes-Méndez F, Rivera D, Moreno-Switt A, Toro M, Badilla C, Santiviago CA and Pezoa D (2024) Genomic analysis of *Salmonella* isolated from surface water and animal sources in Chile reveals new T6SS effector protein candidates. *Front. Microbiol.* 15:1496223. doi: 10.3389/fmicb.2024.1496223

Genomic analysis of *Salmonella* isolated from surface water and animal sources in Chile reveals new T6SS effector protein candidates

Fernando A. Amaya[1†], Carlos J. Blondel[2†], Felipe Reyes-Méndez[3], Dácil Rivera[4], Andrea Moreno-Switt[4], Magaly Toro[5,6], Consuelo Badilla[3], Carlos A. Santiviago[1]* and David Pezoa[3,7]*

[1]Laboratorio de Microbiología, Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santiago, Chile, [2]Facultad de Medicina y Facultad de Ciencias de la Vida, Instituto de Ciencias Biomédicas, Universidad Andrés Bello, Santiago, Chile, [3]Núcleo de Investigación en One Health, Facultad de Medicina Veterinaria y Agronomía, Universidad de Las Américas, Santiago, Chile, [4]Escuela de Medicina Veterinaria, Facultad de Agronomía e Ingeniería Forestal, Facultad de Ciencias Biológicas y Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile, [5]Joint Institute for Food Safety and Applied Nutrition (JIFSAN), University of Maryland, College Park, MD, United States, [6]Instituto de Nutrición y Tecnología de los Alimentos (INTA), Universidad de Chile, Santiago, Chile, [7]Departamento de Ciencias Químicas y Biológicas, Universidad Bernardo O'Higgins, Santiago, Chile

Type VI Secretion Systems (T6SS), widely distributed in Gram-negative bacteria, contribute to interbacterial competition and pathogenesis through the translocation of effector proteins to target cells. *Salmonella* harbor 5 pathogenicity islands encoding T6SS (SPI-6, SPI-19, SPI-20, SPI-21 and SPI-22), in which a limited number of effector proteins have been identified. Previous analyses by our group focused on the identification of candidate T6SS effectors and cognate immunity proteins in *Salmonella* genomes deposited in public databases. In this study, the analysis was centered on *Salmonella* isolates obtained from environmental sources in Chile. To this end, bioinformatics and comparative genomics analyses were performed using 695 genomes of *Salmonella* isolates representing 44 serotypes obtained from surface water and animal sources in Chile to identify new T6SS effector proteins. First, T6SS gene clusters were identified using the SecreT6 server. This analysis revealed that most isolates carry the SPI-6 T6SS gene cluster, whereas the SPI-19 and SPI-21 T6SS gene clusters were detected in isolates from a limited number of serotypes. In contrast, the SPI-20 and SPI-22 T6SS gene clusters were not detected. Subsequently, each ORF in the T6SS gene clusters identified was analyzed using bioinformatics tools for effector prediction, identification of immunity proteins and functional biochemical prediction. This analysis detected 20 of the 37 T6SS effector proteins previously reported in *Salmonella*. In addition, 4 new effector proteins with potential antibacterial activity were identified in SPI-6: 2 Rhs effectors with potential DNase activity (PAAR-RhsA-NucA_B and PAAR-RhsA-GH-E) and 2 effectors with potential RNase activity (PAAR-RhsA-CdiA and RhsA-CdiA). Interestingly, the repertoire of SPI-6 T6SS effectors varies among isolates of the same serotype. In SPI-19, no new effector protein was detected. Of note, some Rhs effectors of SPI-19 and SPI-6 present C-terminal ends with unknown function. The presence of cognate immunity proteins carrying domains present in *bona fide* immunity proteins suggests that these effectors have antibacterial activity. Finally, two new effectors were identified in SPI-21: one with potential peptidoglycan hydrolase activity and another with potential membrane pore-forming activity. Altogether, our work broadens the repertoire of *Salmonella* T6SS

effector proteins and provides evidence that SPI-6, SPI-19 and SPI-21 T6SS gene clusters harbor a vast array of antibacterial effectors.

KEYWORDS

Salmonella, T6SS, Chile, effector, immunity protein

Introduction

The type VI secretion system (T6SS) is an apparatus composed of 13 structural proteins and several accessory proteins that deliver protein effectors into target cells by means of a contractile mechanism (Coulthurst, 2019; Cherrak et al., 2019). The T6SS needle is composed of an inner tube made of a stack of Hcp hexamer rings that is tipped by a trimer of VgrG and a proline-alanine–alanine-arginine repeat (PAAR) protein. This internal structure is surrounded by a contractile sheath of polymerized TssB/TssC subunits assembled in an extended, metastable conformation (Silverman et al., 2013; Cherrak et al., 2019). Contraction of the sheath propels the needle complex toward the target cell (Brackmann et al., 2017). T6SS effector proteins are classified as either cargo or specialized effectors. Cargo effectors are transported by non-covalent interaction with some core components (Coulthurst, 2019), while specialized effectors are VgrG, Hcp or PAAR proteins carrying additional domains (Durand et al., 2014; Whitney et al., 2014; Diniz and Coulthurst, 2015; Ma et al., 2017; Pissaridou et al., 2018).

T6SS effector proteins can target prokaryotic and/or eukaryotic cells (Coulthurst, 2019; Monjarás Feria and Valvano, 2020). Among the anti-bacterial effector proteins, some target the peptidic or glycosidic bonds of the peptidoglycan (Ma and Mekalanos, 2010; Russell et al., 2012; Srikannathasan et al., 2013; Whitney et al., 2013; Berni et al., 2019; Wood et al., 2019), or the FtsZ cell division ring (Ting et al., 2018). These anti-bacterial effectors are usually encoded in bi-cistronic elements with their cognate immunity proteins (E/I pairs) in order to avoid self-intoxication and killing of sibling cells (Russell et al., 2012). Other T6SS effectors target eukaryotic cells, such as those disrupting the actin or microtubule cytoskeleton networks (Monjarás Feria and Valvano, 2020), while trans-kingdom effectors target both bacterial and eukaryotic cells (Jiang et al., 2014). These effectors include those forming pores in membranes or targeting conserved molecules such as NAD^+ and $NADP^+$, and macromolecules such as DNA, RNA and phospholipids (Whitney et al., 2015; Tang et al., 2018; Ahmad et al., 2019). In many enteric pathogens (e.g., *Salmonella*, *Shigella* and *Vibrio*), the T6SS contributes to colonization of the intestinal tract of infected hosts (Sana et al., 2016; Chassaing and Cascales, 2018). On the other hand, strains of the gut commensal *Bacteroides fragilis* use their T6SSs for competition against other Bacteroidales species (Coyne and Comstock, 2019). Hence, the T6SS is a key player in bacterial warfare.

The *Salmonella* genus includes more than 2,600 serotypes distributed between species *S. enterica* and *S. bongori* (Issenhuth-Jeanjean et al., 2014), which differ in clinical signs and host range (Uzzau et al., 2000). In *Salmonella*, five T6SS gene clusters have been identified within *Salmonella* Pathogenicity Islands (SPIs) SPI-6, SPI-19, SPI-20, SPI-21, and SPI-22 (Blondel et al., 2009; Fookes et al., 2011; Bao et al., 2019). These T6SS gene clusters are distributed in 4 different evolutionary lineages: The SPI-6 T6SS gene cluster belongs to subtype i3, SPI-19 T6SS gene cluster to subtype i1, SPI-22 T6SS gene cluster to subtype i4a, and both SPI-20 and SPI-21 T6SS gene clusters to subtype i2 (Bao et al., 2019). Besides their distinct evolutionary origin, these five T6SS gene clusters are differentially distributed among distinct serotypes, subspecies, and species of *Salmonella* (Blondel et al., 2009; Bao et al., 2019).

In *Salmonella*, only a few studies have addressed the role played by the T6SSs in interbacterial and eukaryotic relationships, and most of our understanding regarding the contribution of T6SSs to *Salmonella* infection cycle, virulence and pathogenesis comes from studies of $T6SS_{SPI-6}$ in *S.* Typhimurium and $T6SS_{SPI-19}$ in *S.* Dublin (Mulder et al., 2012; Pezoa et al., 2013; Pezoa et al., 2014; Sana et al., 2016; Sibinelli-Sousa et al., 2022; Xian et al., 2020; Blondel et al., 2010; Hespanhol et al., 2022). Furthermore, knowledge of the presence and distribution of T6SS effector proteins is derived from studies using strains representing a limited number of serotypes (Russell et al., 2012; Benz et al., 2013; Sana et al., 2016; Whitney et al., 2013; Sibinelli-Sousa et al., 2020; Lorente-Cobo et al., 2022; Koskiniemi et al., 2014; Amaya et al., 2022; Jurėnas et al., 2022; Blondel et al., 2023). Consequently, information regarding *Salmonella* T6SS effector proteins is still scarce. Indeed, only 37 T6SS effectors and candidate effectors that target different bacterial molecules such as peptidoglycan, nucleic acids and bacterial ribosomes have been currently identified in a few serotypes (Blondel et al., 2009; Russell et al., 2012; Benz et al., 2013; Whitney et al., 2013; Koskiniemi et al., 2014; Sana et al., 2016; Ho et al., 2017; Sibinelli-Sousa et al., 2020; Amaya et al., 2022; Jurėnas et al., 2022; Lorente-Cobo et al., 2022; Hespanhol et al., 2022; Blondel et al., 2023). This is an important knowledge gap as the T6SS effector proteins are the ultimate mediators of the T6SS activity and thus, their identification and characterization are pivotal for a better understanding of *Salmonella* infectious cycle and in its contribution to environmental fitness and pathogenic potential.

Nowadays, there is increasing evidence that *Salmonella enterica* can persist in diverse environments such as aquatic ecosystems, maintaining a reservoir in surface waters and becoming a serious risk to public health and animal production systems. It is conceivable that the T6SS could mediate in part this persistence since it has been shown that *S.* Typhimurium requires the $T6SS_{SPI-6}$ to survive intracellularly in environmental amoebas such as *Dictyostelium discoideum* (Riquelme et al., 2016). Interestingly, in Chile some serotypes such as *S.* Infantis, *S.* Newport and *S.* Typhimurium have been frequently isolated in surface waters during the last decade, imposing a significant threat to human and animal health since these serotypes usually carry an arsenal of antimicrobial resistance genes (Chen et al., 2024a,b). These Chilean isolates could be an untapped reservoir of new T6SS effector proteins. Importantly, *Salmonella* strains isolated from surface waters in Chile will shed light not only on the vast arsenal of T6SS effector repertoire but could also provide insight into geographic adaptation of *Salmonella*.

In this study, we performed bioinformatic and comparative genomic analyses of a dataset of 695 *S. enterica* genomes representing 44 serotypes isolated from different environmental sources in Chile, mostly surface waters. Our analysis revealed that most genomes only harbor the SPI-6 T6SS gene cluster, and that within its variable region 3 (VR3) we found four new candidate T6SS effectors with predicted nuclease activity. Noteworthy, many putative SPI-6 rearrangement

hotspot (Rhs) effectors identified in this study harbor C-terminal extensions with unknown function. Overall, the diversity and distribution of T6SS effector proteins in Chilean *Salmonella* isolates suggest that different combinations of these proteins may contribute to the environmental fitness and pathogenic potential.

Materials and methods

Environmental samples and *Salmonella* isolation

Water samples were collected as part of a previous study (Toro et al., 2022) from sites in the Maipo, Mapocho, Claro and Lontué watersheds from the rivers themselves and connected tributaries, such as canals. Animal samples were collected as part of a previous study (Rivera et al., 2021) from industrial dairy farms, backyard systems and wild animals in the Región de Coquimbo, Región de Valparaíso, Región Metropolitana and Región del Libertador General Bernardo O'Higgins, Chile. A detailed description of sampling procedures, geographical location of samples and the procedure employed for *Salmonella* isolation from water an animal samples can be found elsewhere (Rivera et al., 2021; Toro et al., 2022).

Whole genome sequencing, assembly, and quality control

For sequencing, each isolate was grown overnight at 37°C in tryptic soy broth and 1 mL of culture was used to purify DNA with the DNeasy Blood and Tissue Qiagen kit (Qiagen, CA, United States). Ratios of absorbance at 260 nm and 230 nm were obtained using a MaestroNano spectrophotometer (Maestro, Korea) and a QUBIT fluorimeter (Life Technologies, CA, United States). Libraries were prepared with the Illumina DNA Prep kit (Illumina, CA, United States) on the Sciclone G3 NGSx iQ Workstation (Perkin Elmer, MA, United States), and sequencing was performed on the Illumina NextSeq 2000 using the NextSeq 1000/2000 P2 reagents 300 cycles with the 150 paired-end chemistry (Illumina, CA, United States). Reads were examined for quality using FastQC (Galaxy version 0.69) (Wingett and Andrews, 2018) and trimmed using Trimmomatic (Galaxy version 0.36.4), with a minimum required quality of 20, averaging across 4 bases (Bolger et al., 2014). Processed reads were assembled using SPAdes (Galaxy version 3.11.1) with kmer sizes of 99 and 127, and careful correction (Bankevich et al., 2012). Assemblies were checked for quality using QUAST (Galaxy version 4.6.3) (Gurevich et al., 2013) and finally deposited in the NCBI Bioproject 560,080.[1]

In silico serotyping was carried out using SeqSero (Galaxy version 2.0.1) (Zhang et al., 2015) and SISTR (Galaxy version 1.0.2) (Yoshida et al., 2016). Finally, a single-nucleotide polymorphism (SNP) analysis was performed to identify clonality among isolates from the same sample. Clones were defined as isolates with genomes having 20 or fewer SNPs, as described by Pightling et al. (2018). According to this criterion, genome sequences from non-clonal isolates obtained from the same sample were selected for subsequent analysis. Thus, the genome sequence dataset analyzed in this study includes 695 *S. enterica* genomes from 44 distinct serotypes (Supplementary Table S1).

Identification of T6SS gene clusters

The T6SS prediction tool from the Secret6 web server[2] was used to identify T6SS gene clusters encoding the minimal 13 core components of a T6SS in each genome (Zhang et al., 2023). For selection of positive matches, a BLASTp 2.10.1+ identity threshold for T6SS prediction >30% and an E-value <0.0001 were used. These threshold values have been successfully used to identify T6SS gene clusters in *Salmonella* genomes (Amaya et al., 2022; Blondel et al., 2023).

Identification of candidate T6SS effectors

To identify putative T6SS effectors encoded within the *Salmonella* genomes analyzed, each ORF encoded within the T6SS gene clusters identified was analyzed with the Bastion6 pipeline[3] (Wang et al., 2018) excluding the 13 T6SS core components. ORFs presenting a Bastion6 score ≥ 0.7 were considered as candidate T6SS effectors. It is worth mentioning that a Bastion6 score ≥ 0.5 is routinely used as default setting for detection of T6SS effectors. However, we decided to use a score ≥ 0.7 to perform a more strict analysis. Each Bastion6 prediction was further analyzed using tools implemented in the Operon-Mapper web server[4] (Taboada et al., 2018) to determine whether it was part of a single transcriptional unit that also encoded a putative immunity protein [i.e., a small protein with potential signal peptides (SignalP 6.0) and/or transmembrane domains (TMHMM 2.0)]. Conserved functional domains and motifs in the candidate T6SS effectors were identified using the PROSITE, NCBI-CDD, Motif-finder, and Pfam databases (Kanehisa et al., 2002; Sigrist et al., 2013; Finn et al., 2014; Lu et al., 2019) implemented in the GenomeNet search engine.[5] An E-value cutoff score of 0.01 was used. In addition, for each putative effector and immunity protein identified, a biochemical functional prediction was performed by HMM homology searches using the HHpred HMM-HMM comparison tool[6] (Zimmermann et al., 2017). Finally, a candidate T6SS effector was defined as "new" when it meets two criteria: (i) it includes at least one domain previously linked to antibacterial activity, and (ii) this domain has not been described as part of a T6SS effector in publicly available databases.

Hierarchical clustering analysis of the new T6SS effectors

For hierarchical clustering analysis, a presence/absence matrix of each T6SS effector and candidate effector was constructed for each

1 https://www.ncbi.nlm.nih.gov/bioproject/560080

2 https://bioinfo-mml.sjtu.edu.cn/SecReT6/t6ss_prediction.php

3 https://bastion6.erc.monash.edu

4 https://biocomputo.ibt.unam.mx/operon_mapper

5 https://www.genome.jp

6 https://toolkit.tuebingen.mpg.de/tools/hhpred

bacterial genome by means of BLASTn analyses and manual curation of the data (Supplementary Table S2). A 90% identity and 90% sequence coverage threshold was used to select positive matches, as done in previous analyses conducted by our group (Amaya et al., 2022; Blondel et al., 2023). The matrix generated was uploaded as a csv file to the online server MORPHEUS[7] using default parameters (i.e., one minus Pearson's correlation and average linkage method).

Phylogenetic analyses of *Salmonella* T6SS gene clusters

TssC aminoacid sequences encoded in T6SS gene clusters from 605 *Salmonella* genomes were concatenated and aligned with ClustalW using the Molecular Evolutionary Genetics Analysis (MEGA) software version 7.0 (Kumar et al., 2016). A phylogenetic tree was built from the alignments obtained from MEGA by performing a bootstrap test of phylogeny (1,000 replications) using the maximum-likelihood method with a Jones-Taylor-Thornton correction model.

Analysis of T6SS effectors distribution

The DNA sequence encoding each T6SS effector identified in this study was subjected to tBLASTx analyses to find orthologs in all *Salmonella* genome sequences deposited in the NCBI database (March, 2024) (Supplementary Tables S3, S4). For selection of positive matches, a 90% identity and 90% sequence coverage threshold was used. Conservation of sequences was determined by independent multiple sequence alignments using T-Coffee Expresso[8] (Notredame et al., 2000), MAFFT[9] (Katoh et al., 2017), and ESPript 3[10] (Robert and Gouet, 2014). Comparative genomic analyses of T6SS gene clusters were performed using Mauve version 2.3.1[11] (Darling et al., 2004) and EasyFig version 2.2.5[12] (Sullivan et al., 2011). Nucleotide sequences were analyzed using Artemis version 18[13] (Rutherford et al., 2000).

Results

T6SS gene clusters are widely distributed among Chilean *Salmonella* isolates

Previous analyses performed by our group have aimed in the identification of candidate T6SS effectors and cognate immunity proteins in *Salmonella* genomes deposited in public databases (Amaya et al., 2022; Blondel et al., 2023). In the present study, the analysis focused on genome sequences of *Salmonella* isolates obtained from different environmental sources in Chile, in order to shed light on the repertoire of T6SS candidate effectors present in *Salmonella* inhabiting our local geography. To this end, we analyzed a database of 695 high-quality sequenced *Salmonella* genomes from strains isolated from surface water and animal sources. Most isolates in this collection come from surface waters (674 isolates representing 34 serotypes), while 21 isolates representing only 8 serotypes were obtained from animal sources (14 in chicken, 3 in pigeon, 2 in pig and 2 in duck). Interestingly, the most frequently isolated serotypes were *S.* Infantis ($n = 169$), *S.* Agona ($n = 71$) and *S.* Newport ($n = 11$).

To identify T6SS gene clusters we used the T6SS prediction tool from the SecreT6 web server (see text footnote 2), which identified 622 putative T6SS gene clusters in 608 *Salmonella* genomes (Table 1; Supplementary Table S1). A more in-depth analysis revealed that these T6SS gene clusters correspond to those encoded in SPI-6, SPI-19 and SPI-21 (Table 1; Supplementary Figure S1). We could not identify T6SS gene clusters encoded in SPI-20 or SPI-22 in the genome of any isolate from our database. The SPI-6 T6SS gene cluster is widely distributed in 518 of the 695 genomes analyzed (74.5%), while the SPI-19 and SPI-21 T6SS gene clusters were only detected in 89 (12.8%) and 14 (2%) genomes, respectively (Table 1). Most isolates carried a unique T6SS gene cluster in SPI-6, SPI-19 or SPI-21, while a group of isolates belonging to serotype *S.* Livingstone harbors both SPI-6 and SPI-19 T6SS gene clusters. In contrast, no complete T6SS gene cluster was detected in isolates belonging to serotypes *S.* Enteritidis and *S.* Stanley.

To identify high-confidence putative effectors encoded within every T6SS gene cluster detected, each ORF within these gene clusters was analyzed based on four criteria: (i) identification of candidate effectors through Bastion6 analysis (a bioinformatic tool that predicts T6SS effectors based on amino acid sequence, evolutionary information, and physicochemical properties); (ii) identification of putative immunity proteins by operon prediction (Operon-mapper; Taboada et al., 2018) and detection of signal peptides (SignalP 6.0) and transmembrane domains (TMHMM 2.0); (iii) identification of conserved functional domains associated with *bona fide* T6SS effectors (INTERPROSCAN, PROSITE, NCBI-CDD, MOTIF, and Pfam); and (iv) functional biochemical prediction using the HHpred HMM-HMM server. In addition, we further analyzed these T6SS gene clusters to identify potential unannotated ORFs that could encode putative effectors and cognate immunity proteins. Thus, our analysis revealed the presence of 6 new effector candidates encoded within the SPI-6 (4 effectors) and SPI-21 (2 effectors) T6SS gene clusters.

The VR3 within the SPI-6 T6SS gene cluster of isolates from surface waters harbor four candidate T6SS effector proteins

Most T6SS effector proteins identified in *Salmonella* are encoded within three variable regions (VR1-3) of SPI-6 (Blondel et al., 2023). We have previously shown that the VR3 of SPI-6, located downstream of the *tssI* gene, exhibits the greatest diversity of *Salmonella* T6SS effectors (Blondel et al., 2023). This is mainly due to the presence of a variable number of Rhs effector proteins that harbor C-terminal extensions encoding endonuclease domains, such as DNases, RNases, and deaminases, as well as ADP-ribosyltransferases (Blondel et al., 2023).

Our analysis identified 4 new putative effector proteins and cognate immunity proteins (Table 2; Figure 1) encoded in the VR3 of SPI-6 distributed in isolates of serotypes *S.* Braenderup, *S.* Albany,

7 https://software.broadinstitute.org/morpheus
8 https://tcoffee.crg.eu/apps/tcoffee/do:expresso
9 https://mafft.cbrc.jp/alignment/server/index.html
10 https://espript.ibcp.fr/ESPript/ESPript/
11 https://darlinglab.org/mauve/mauve.html
12 https://mjsull.github.io/Easyfig/files.html
13 https://sanger-pathogens.github.io/Artemis/Artemis/

10.3389/fmicb.2024.1496223

TABLE 1 T6SS effectors and cognate immunity proteins encoded in T6SS gene clusters in Chilean *Salmonella* isolates.

Source of sample (Number of isolates)	T6SS gene cluster	T6SS effector[a]	Serotypes (Number of isolates with the corresponding T6SS effector)
Water (510), Chicken (5), Duck (2), Pig (2), Pigeon (3)	SPI-6	Tae2	*S.* Adelaide (4), *S.* Albany (1), *S.* Anatum (12), *S.* Bovismorbificans (37), *S.* Braenderup (4), *S.* Brandenburg (4), *S.* Cerro (12), *S.* Corvallis (9), *S.* Derby (1), *S.* Edinburgh (13), *S.* Give (4), *S.* Hadar (2), *S.* Heidelberg (1), *S.* Infantis (152), *S.* I -:b:1,5 (2), *S.* I 1,4,[5],12:d:- (1), *S.* I 1,4,[5],12:i:- (1), *S.* Johannesburg (1), *S.* Kentucky (1), *S.* Montevideo (2), *S.* Muenchen (5), *S.* Newport (1), *S.* Oranienburg (5), *S.* Panama (15), *S.* Paratyphi B (2), *S.* Sandiego (3), *S.* Santiago (4), *S.* Senftenberg (35), *S.* Soerenga (3), *S.* Tennessee (2), *S.* Thompson (10), *S.* Typhimurium (46), *S.* Worthington (4)
		Tae4	*S.* Adelaide (4), *S.* Albany (1), *S.* Anatum (12), *S.* Bovismorbificans (38), *S.* Braenderup (3), *S.* Cerro (11), *S.* Corvallis (10), *S.* Derby (1), *S.* Edinburgh (13), *S.* Give (4), *S.* Goldcoast (11), *S.* Hadar (2), *S.* Heidelberg (1), *S.* Infantis (151), *S.* I -:b:1,5 (3), *S.* I 1,4,[5],12:d:- (1), *S.* I 1,4,[5],12:i:- (1), *S.* Kentucky (1), *S.* Livingstone (23), *S.* Mbandaka (4), *S.* Montevideo (2), *S.* Muenchen (7), *S.* Newport (44), *S.* Oranienburg (5), *S.* Panama (15), *S.* Paratyphi B (2), *S.* Sandiego (3), *S.* Santiago (4), *S.* Senftenberg (33), *S.* Soerenga (3), *S.* Tennessee (2), *S.* Thompson (10), *S.* Typhimurium (46), *S.* Worthington (4)
		Tge2P	*S.* Adelaide (4), *S.* Bovismorbificans (38), *S.* Braenderup (3), *S.* Corvallis (10), *S.* Give (1), *S.* Hadar (2), *S.* Heidelberg (1), *S.* Infantis (152), *S.* I 1,4,[5],12:d:- (1), *S.* Johannesburg (1), *S.* Kentucky (1), *S.* Livingstone (7), *S.* Mbandaka (4), *S.* Muenchen (7), *S.* Newport (23), *S.* Sandiego (3), *S.* Senftenberg (35), *S.* Soerenga (3), *S.* Tennessee (2), *S.* Thompson (10), *S.* Typhimurium (1), *S.* Worthington (4)
		Tlde1	*S.* Adelaide (4), *S.* Albany (1), *S.* Anatum (12), *S.* Bovismorbificans (38), *S.* Braenderup (3), *S.* Brandenburg (4), *S.* Cerro (13), *S.* Corvallis (10), *S.* Derby (1), *S.* Goldcoast (11), *S.* Hadar (2), *S.* Heidelberg (1), *S.* Infantis (152), *S.* I 1,4,[5],12:d:- (1), *S.* I 1,4,[5],12:i:- (1), *S.* Johannesburg (1), *S.* Kentucky (1), *S.* Livingstone (25), *S.* Mbandaka (4), *S.* Muenchen (7), *S.* Newport (35), *S.* Paratyphi B (2), *S.* Sandiego (3), *S.* Senftenberg (1), *S.* Soerenga (3), *S.* Tennessee (2), *S.* Thompson (10), *S.* Typhimurium (46), *S.* Worthington (4)
		L-Ala, D-Glu endopeptidase	*S.* Bovismorbificans (37), *S.* Braenderup (1), *S.* Brandenburg (4), *S.* Edinburgh (13), *S.* Give (4), *S.* I -:b:1,5 (4), *S.* Johannesburg (1), *S.* Mbandaka (4), *S.* Montevideo (2), *S.* Newport (18), *S.* Oranienburg (5), *S.* Panama (15), *S.* Sandiego (3), *S.* Worthington (4)
		PgP	*S.* Braenderup (2)
		TseH-like	*S.* Edinburgh (13), *S.* I -:b:1,5 (6), *S.* Panama (15)
		Peptidase_M64	*S.* Braenderup (2), *S.* Give (4), *S.* Montevideo (2), *S.* Senftenberg (34), *S.* Tennessee (2)
		RhsA-HNHc	*S.* Tennessee (2)
		RhsA-Ntox47	*S.* Brandenburg (2), *S.* I 1,4,[5],12:i:- (1), *S.* Typhimurium (44)
		RhsA-Tox-HNH-EHHH	*S.* Braenderup (2), *S.* Derby (1)
		PAAR-RhsA-HNHc	*S.* Anatum (1), *S.* Edinburgh (1), *S.* Infantis (132), *S.* Kentucky (1), *S.* Senftenberg (1)
		PAAR-RhsA-Ntox47	*S.* Give (3), *S.* Livingstone (8), *S.* Muenchen (7), *S.* Newport (14), *S.* Panama (15), *S.* Sandiego (2)
		PAAR-RhsA-Tox-HNH-EHHH	*S.* Johannesburg (1), *S.* Tennessee (2)
		PAAR-RhsA-AHH	*S.* Goldcoast (11)
		PAAR-RhsA-GIY-YIG	*S.* Livingstone (8)
		RhsA-Tox-ART-HYD1	*S.* Thompson (7)
		PAAR-RhsA-Tox-ART-HYD1	*S.* Johannesburg (1)
		Rhs_{main}	*S.* Typhimurium (36)
		PAAR-RhsA-NucA_B	*S.* Braenderup (1)
		PAAR-RhsA-GH-E	*S.* Albany (1)
		PAAR-RhsA-CdiA	*S.* Tennesse (2)
		RhsA-CdiA	*S.* Derby (1)
		PAAR-RhsA-CT	*S.* Adelaide (4), *S.* Braenderup (3), *S.* Brandenburg (2), *S.* Cerro (12), *S.* Derby (1), *S.* Edinburgh (5), *S.* Give (1), *S.* Hadar (2), *S.* Heidelberg (1), *S.* I 1,4,[5],12:i:- (1), *S.* Mbandaka (1), *S.* Montevideo (2), *S.* Newport (21), *S.* Paratyphi B (2), *S.* Sandiego (3), *S.* Soerenga (3), *S.* Thompson (10), *S.* Typhimurium (1), *S.* Worthington (4)
		RhsA-CT	*S.* Braenderup (2), *S.* Cerro (12), *S.* Edinburgh (5), *S.* Give (1), *S.* Hadar (2), *S.* Johannesburg (1), *S.* Thompson (9), *S.* Typhimurium (1)

(Continued)

TABLE 1 (Continued)

Source of sample (Number of isolates)	T6SS gene cluster	T6SS effector[a]	Serotypes (Number of isolates with the corresponding T6SS effector)
Water (66)	SPI-19	**PAAR-RhsA-CT**	*S.* Agona (65), *S.* I 4:f,g,s:1,2 (1)
Water (13)	SPI-21	VgrG-PyocinS-HNHc	*S.* IIIb 35:i:z (1), *S.* IIIb 48:i:z (7)
		Glucosaminidase	*S.* IIIb 35:i:z (1), *S.* IIIb 48:i:z (11)
		BTH_I2691	*S.* IIIb 35:i:z (1)

[a]T6SS effectors and immunity proteins are designated according their formal name (in the case of those previously reported in the literature) or indicating the functional domains present in the predicted proteins (in the case of those having no formal names). New T6SS candidate effectors identified in this study are highlighted in **bold type**.

TABLE 2 New putative T6SS effectors and cognate immunity proteins encoded in the SPI-6 T6SS gene cluster of Chilean *Salmonella* isolates.

T6SS effector genes					Cognate T6SS immunity protein genes	
ORF(s)	Size (aa)	Serotype-isolate	Variable Region	Predicted activity/ Domain	ORF(s)	TM or signal peptide/ Domain[a]
Effectors targeting nucleic acids						
FA1083_3621	1,498	*S.* Braenderup FA1083	3	DNase/PAAR-RhsA-NucA_B	FA1083_3620	No/DUF6707
FA1443_1959	1,566	*S.* Albany FA1443	3	DNase/PAAR-RhsA-GH-E	FA1443_1960	No/TPR
FA1455_4074	1,560	*S.* Tennessee FA1455	3	RNase/PAAR-RhsA-CdiA	FA1455_4073	No/MafI
CFSAN035156_3316		*S.* Tennessee CFSAN035156			CFSAN035156_3317	
FA1451_3438	372	*S.* Derby FA1451	3	RNase/RhsA-CdiA	FA1451_3439	No/AntA

[a]Presence or absence of transmembrane domains (TM) or a signal peptide, and protein domains present in the putative immunity protein genes.

S. Tennessee and *S.* Derby. Three of these candidates are specialized Rhs effector proteins with predicted nuclease activity, including 2 DNases and 1 RNase, while only one is a cargo Rhs effector with putative RNase activity (Table 2). The first putative effector (FA1083_3621 in *S.* Braenderup FA1083) is a large 1,498 amino acid Rhs protein that harbors an N-terminal PAAR domain and a C-terminal Nuclease A/Nuclease B (NucA_B) domain with predicted DNase activity (Table 2; Figure 1). It should be noted that *FA1083_3621* is predicted to be encoded in a bi-cistronic unit with *FA1083_3620* (Table 2). This latter ORF encodes a 204 amino acid protein with a DUF6707 domain that may correspond to the cognate immunity protein of FA1083_3621. The second candidate effector (FA1443_1959 in *S.* Albany FA1443) with predicted DNase activity also corresponds to a 1,566 amino acid Rhs protein that harbors an N-terminal PAAR domain and the putative GH-E domain in its C-terminal end (Table 2; Figure 1). The GH-E domain is found in members of the HNH/ENDO VII superfamily nuclease with conserved glycine, histidine and glutamate residues. This putative effector was also predicted to be co-transcribed with its respective putative immunity protein gene that encodes a tetratricopeptide repeat (TPR)-containing protein (FA1443_1960 in *S.* Albany FA1443). The third candidate effector (FA1455_4074 in *S.* Tennessee FA1455) is a 1,560 amino acid Rhs protein with a predicted N-terminal PAAR domain and a C-terminal contact-dependent growth inhibition protein A (CdiA) domain with putative RNase activity (Table 2; Figure 1). The gene encoding this candidate effector is predicted to be part of a bi-cistronic unit with *FA1455_4073*, encoding its putative immunity protein (Table 2; Figure 1). Of note, FA1455_4073 harbors a multiple adhesin family I (MafI) domain that is frequently found in cognate immunity proteins of bacterial toxin systems (Zhang et al., 2012). The fourth new candidate effector identified in this study is a 372 amino acid Rhs protein with a predicted CdiA domain in its C-terminal end (FA1451_3438 in *S.* Derby FA1451) (Table 2; Figure 1). *FA1451_3438* is predicted to be co-transcribed with *FA1451_3439*, encoding its cognate immunity protein (Table 2; Figure 1). FA1451_3439 harbors an anti-repressor A (AntA) domain usually found in phage anti-repressor proteins (Sandt et al., 2002). It is worth mentioning that the CdiA domain found in candidate effectors FA1455_4074 and FA1451_3438 has not been previously associated with any Rhs effector protein in *Salmonella*.

The genetic structure and repertoire of effector proteins encoded in the SPI-6 T6SS gene cluster vary considerably among *Salmonella* isolates of the same serotype

It has been reported that the genetic structure of the T6SS gene clusters and the repertoire of effector proteins varies between different serotypes of *Salmonella* (Amaya et al., 2022; Blondel et al., 2023). Therefore we analyzed the genetic structure of SPI-6 and the distribution of previously identified effector proteins (Table 1; Supplementary Table S2). We identified 19 out of the 32 previously reported effectors encoded in the SPI-6 T6SS gene cluster. The three

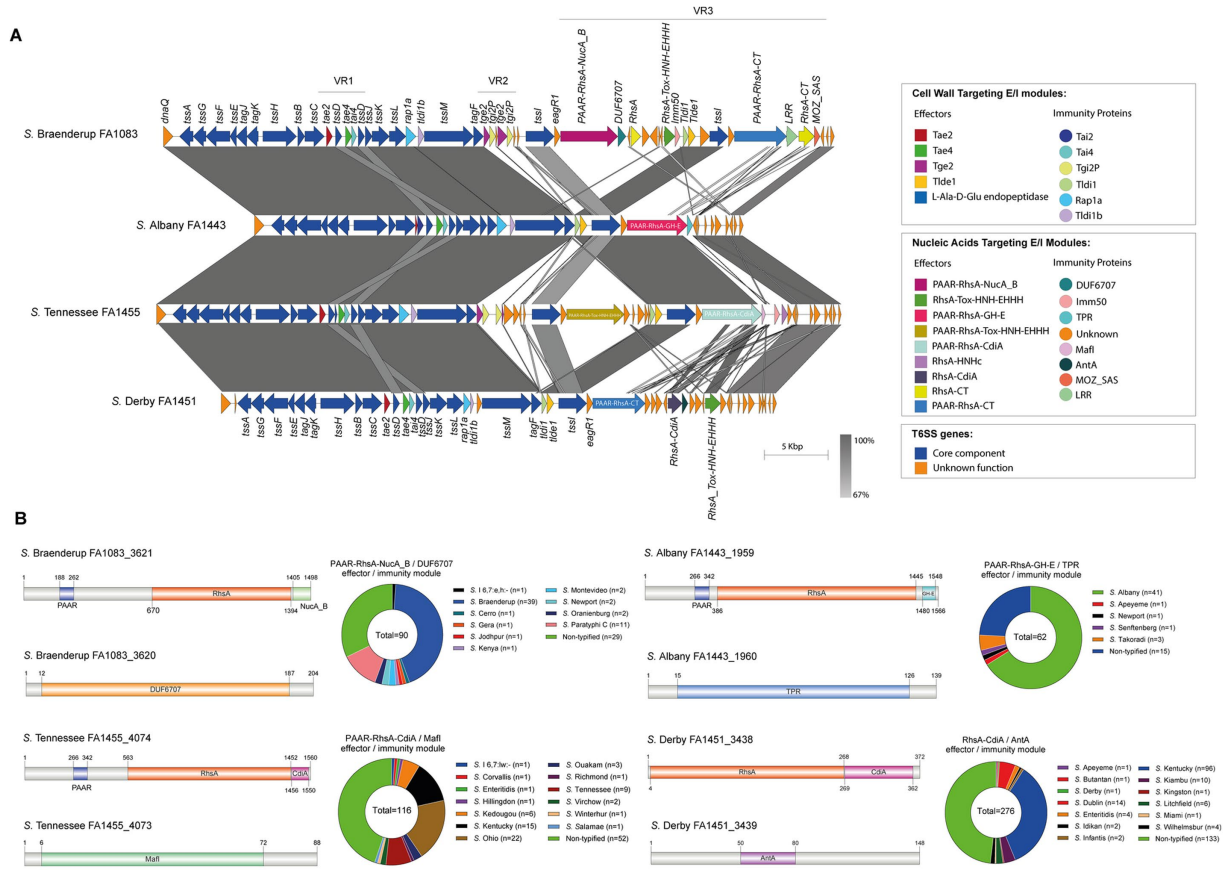


FIGURE 1
The SPI-6 T6SS gene cluster encodes new putative T6SS effector proteins. **(A)** Comparative genomic analysis of the SPI-6 T6SS cluster of *S.* Braenderup FA1083, *S.* Albany FA1443, *S.* Tennessee FA1455 and *S.* Derby FA1451. BLASTn sequence alignment was performed and visualized using EasyFig (Sullivan et al., 2011). **(B)** Schematic representation and distribution among *Salmonella* genomes of each new effector and immunity protein identified. ORFs encoding new E/I modules are highlighted in different colors according to the predicted functions. Homologs for each component were identified by BLASTn analyses as described in Materials and Methods.

most frequently distributed T6SS effectors are encoded in VR1-2 of SPI-6. These effector proteins were Tae4 (34/36), Tae2 (32/36) and Tlde1 (29/36). In VR3, the region showing the greatest diversity of *Salmonella* T6SS effectors, the most prevalent effector proteins were PAAR-RhsA-Ntox47 (6/36) and PAAR-RhsA-HNHc (5/36).

Next, we performed a hierarchical clustering analysis to shed lights into the distribution of effectors and candidate effectors encoded in the SPI-6 T6SS gene cluster identified (Supplementary Table S1). As illustrated in Figure 2, the four *bona fide* effectors encoded within VR1-2 (Tae2, Tae4, Tge2 and Tlde1) were the most conserved across the genomes of isolates representing 29 to 34 *Salmonella* serotypes. However, some of these effectors are missing from the genomes of all isolates from a few *Salmonella* serotypes. In VR3, the most prevalent effector protein was PAAR-RhsA-Ntox47, while PAAR-RhsA-AHH, PAAR-RhsA-GIY-YIG, PAAR-RhsA-Tox-ART-HYD1, RhsA-Tox-ART-HYD1 and RhsA-HNHc were the least prevalent. It is worth mentioning that a greater diversity of VR3-encoded effectors is observed in those serotypes that lack some of the more conserved VR1-2-encoded effectors (Figure 2).

Analysis of genetic structure variation of the SPI-6 T6SS gene cluster between serotypes and between isolates of the same serotype revealed interesting observations. First, we identified a variable number of *tssI-eagR-rhs* gene modules encoded in VR3. A number of isolates from serotypes *S.* Braenderup, *S.* Kentucky, *S.* Sandiego and *S.* Tennessee harbor two *tssI-eagR-rhs* modules (Figure 3), while most isolates from serotypes carrying the SPI-6 T6SS gene cluster only harbor one *tssI-eagR-rhs* module (Figure 4). Remarkably, in *S.* Braenderup the genetic structure of SPI-6 differs between isolates CFSAN43223, FA0982 and FA1083. CFSAN43223 has only one *tssI-eagR-rhs* module, while FA0982 and FA1083 have two of these modules, as previously reported in *S.* Tennessee isolate CFSAN070645 (Blondel et al., 2023) (Figure 3; Supplementary Figure S2). Isolates FA0982 and FA1083 encode the RhsA-Tox-HNH-EHHH effector, as well as two other effectors harboring C-terminal ends with unknown function (PAAR-RhsA-CT and RhsA-CT). Additionally, isolate FA1083 encodes a new PAAR-RhsA-NucA_B effector with putative DNase activity, as described above (Figures 1, 3). It is important to note that isolate CFSAN43223 has an internal deletion within VR2 in comparison to isolates FA0982 and FA1083, and encodes only the Tlde1 effector. In contrast, isolates FA0982 and FA1083 encode two copies of the Tge2 effector in VR2 (Supplementary Figure S2). In *S.* Kentucky, our analysis of the single isolate present in the database (CFSAN035145) identified two *tssI-eagR-rhs* modules in VR3. These modules encode the PAAR-RhsA and PAAR-RhsA-HNHc effector proteins, respectively (Figure 3). Notably,

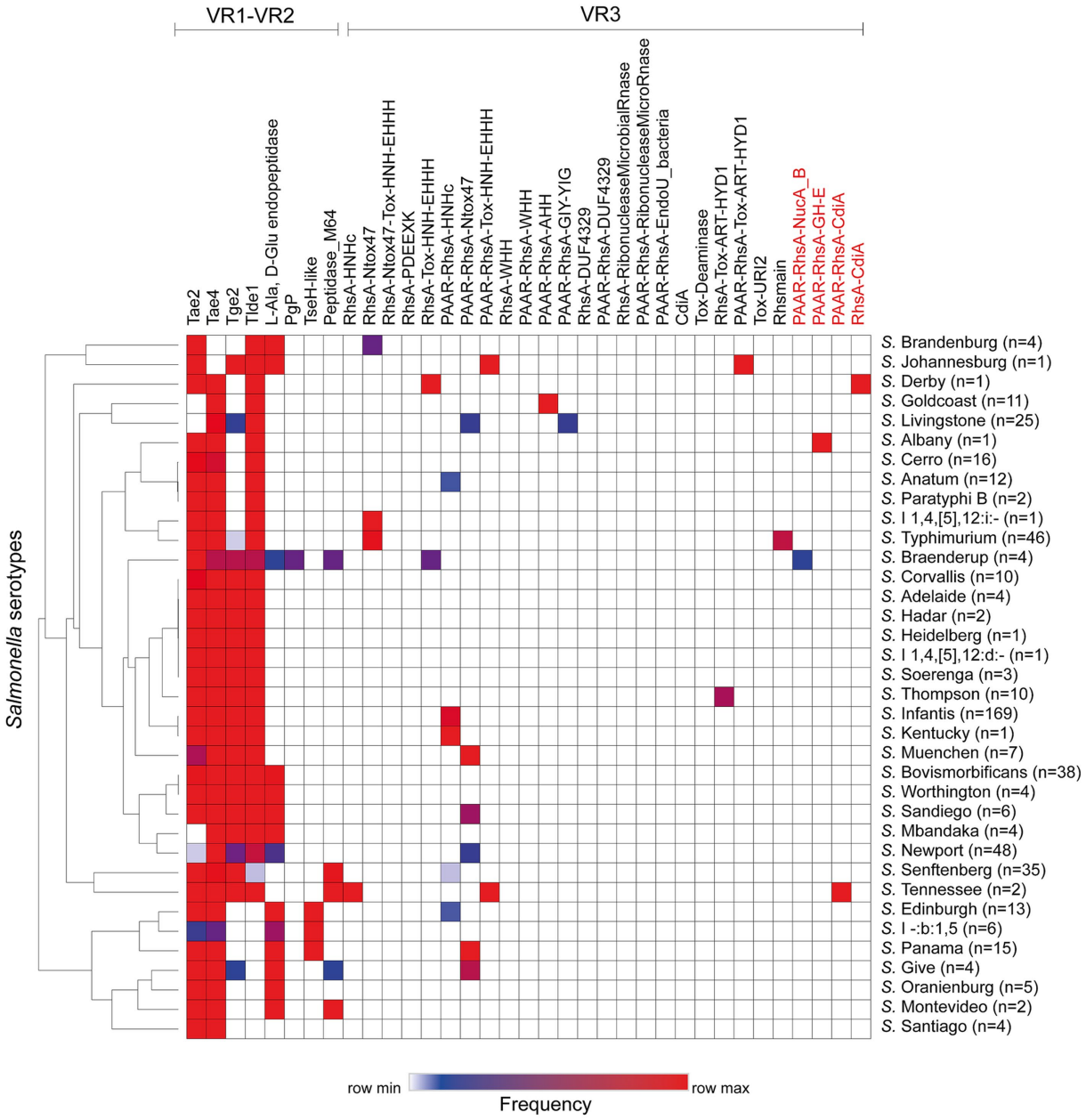


FIGURE 2
Prevalence of ORFs encoding T6SS effectors and candidate effectors in the SPI-6 T6SS gene cluster of Chilean *Salmonella* isolates. A hierarchical clustering analysis was conducted using MORPHEUS, as detailed in the Materials and Methods section. The color code in the heatmap indicates the frequency of a given ORF among all isolates of a particular *Salmonella* serotype. The names of new T6SS candidate effectors identified in this study are highlighted in red.

the first *tssI-eagR-rhs* module has a high sequence identity with only one gene module previously reported in *S.* Tennessee CFSAN070645 (Blondel et al., 2023). Similarly, the second *tssI-eagR-rhs* module of *S.* Kentucky CFSAN035145 shows high sequence identity with the corresponding module encoded in VR3 of *S.* Typhimurium 14028s. Furthermore, *S.* Kentucky CFSAN035145 harbors an ORF with a predicted DUF4056 domain encoded in a bi-cistronic unit in VR2 never reported in *Salmonella*, which may constitute a new T6SS candidate effector (Figure 3). In *S.* Sandiego, the genetic structure of the SPI-6 T6SS gene cluster is conserved between isolates FA0894 and CFSAN105324, that harbor two *tssI-eagR-rhs* gene modules encoding a PAAR-RhsA-CT (C-terminal end with unknown function) and the PAAR-RhsA-Ntox47 effector proteins, respectively (Figure 3). A genomic comparative analysis of this latter effector with the corresponding T6SS effector in *S.* Typhimurium 14028s suggest that in isolates of serotype *S.* Sandiego the Rhs_{main} and RhsA-Ntox47 were at some point a single ORF that was later split due to the accumulation of nonsense mutations (Figure 3). Similar to *S.* Kentucky, the two *tssI-eagR-rhs* gene modules encoded in SPI-6 of *S.* Sandiego share high sequence identity with the corresponding gene modules encoded in *S.* Tennessee CFSAN070645 and *S.* Typhimurium 14028s, respectively (Figure 3). It is worth mentioning that Chilean *S.* Sandiego isolates harbor the Tae2 and Tae4 effector proteins encoded in VR1, as well as Tge2 and Tlde1 effectors encoded in VR2. Finally, in *S.* Tennessee, the

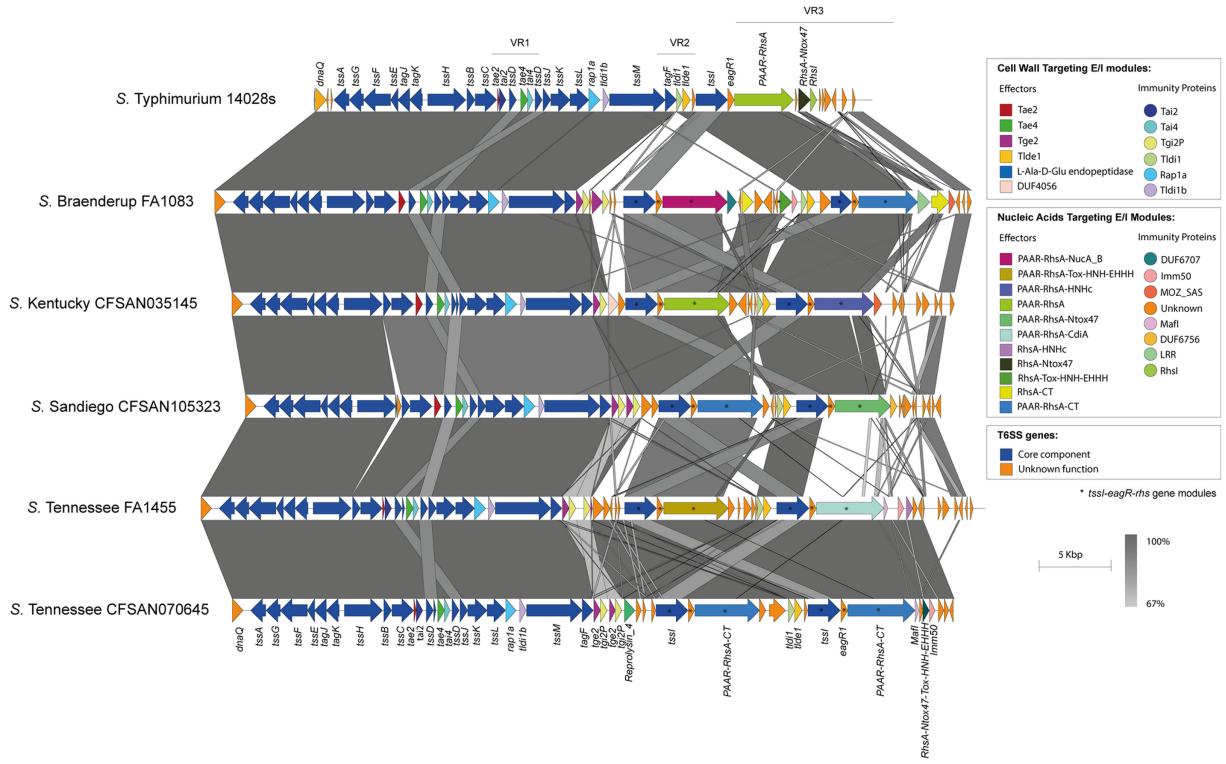


FIGURE 3
The SPI-6 T6SS gene cluster in a number of Chilean *Salmonella* isolates includes two *tssI-eagR-rhs* gene modules in VR3. Comparative genomic analysis of the SPI-6 T6SS cluster of *S.* Braenderup FA1083, *S.* Kentucky CFSAN035145, *S.* Sandiego CFSAN105323 and *S.* Tennessee FA1455 and CFSAN070645. BLASTn sequence alignment was performed and visualized using EasyFig (Sullivan et al., 2011). ORFs encoding E/I modules are highlighted in different colors according to the confirmed or predicted functions. The *tssI-eagR-rhs* gene modules of the SPI-6 T6SS gene cluster are demarked by asterisks. Grayscale represents the percentage of identity between nucleotide sequences. The SPI-6 T6SS gene cluster from *S.* Typhimurium 14028s was used for comparative purposes.

genomic organization of the T6SS gene cluster encoded in SPI-6 is highly conserved not only among Chilean isolates but also among previously reported *S.* Tennessee isolates (Blondel et al., 2023) (Figure 3). Isolates of this serotype harbor two *tssI-eagR-rhs* gene modules encoding a PAAR-RhsA-Tox-HNH-EHHH and a PAAR-RhsA-CdiA T6SS effector proteins, respectively. Interestingly, unlike the other serotypes described above, these two *tssI-eagR-rhs* gene modules do not share any sequence identity with the corresponding module in *S.* Typhimurium 14028s. Altogether, these results suggest a distinct evolutionary origin of *tssI-eagR-rhs* gene modules within the SPI-6 T6SS gene cluster.

On the other hand, the isolates belonging to the remaining 32 serotypes only contain one *tssI-eagR-rhs* gene module encoded in the SPI-6 T6SS gene cluster. In these isolates, the distribution of known and new candidate effectors varies considerably, even among representatives of the same serotype. This is the case of *S.* Livingstone, where two groups of isolates are distinguished. In the first group, the VR3 encodes the PAAR-RhsA-Ntox47 effector, while isolates in the second group harbor the PAAR-RhsA-GIY-YIG effector (Figure 5). In addition, the VR2 in the first group encodes the Tge2 and Tlde1 effector proteins, while in the second group only encodes Tlde1 (Figure 5; Supplementary Table S1). Remarkably, the first group only harbor the SPI-6 T6SS gene cluster while the second group also encodes the SPI-19 T6SS gene cluster. Furthermore, the genetic structure of the SPI-6 T6SS cluster in the first group differs more with the T6SS gene cluster of *S.* Typhimurium 14028s when compared to the second group (Figure 5).

In isolates of serotype *S.* Give, the SPI-6 T6SS gene cluster shows structural differences in VR2 and VR3. In VR2, the isolate CFSAN043231 encodes the Tge2 and Peptidase M64 effector proteins, while other isolates (CFSAN119452, CFSAN119453, and CFSAN119454) carry a bi-cistronic unit encoding proteins with unknown function (Supplementary Figure S3). The putative immunity protein encoding-gene of this bi-cistronic unit harbors a DUF4229 domain found in integral membrane proteins (Wang et al., 2023). Another intriguing structural difference exists in VR3, where isolates CFSAN119452, CFSAN119453, and CFSAN119454 encode a PAAR-RhsA-Ntox47 effector protein, while isolate CFSAN043231 encodes a PAAR-RhsA-CT and an RhsA-CT, both harboring C-terminal ends with unknown functions (Supplementary Figure S3). Notably, the putative immunity protein encoding-gene of the RhsA-CT candidate effector harbors the Imm9 domain, which is frequently found in cognate immunity proteins of bacterial toxin systems with RNase activity (Zhang et al., 2012). Thus, the presence of the Imm9 domain in the putative immunity protein-encoding gene suggests that the C-terminal end of the RhsA-CT candidate effector has RNase activity.

The genetic organization of the SPI-6 T6SS gene cluster in *S.* Newport varies between two groups of isolates. In the first group, the isolates encode the PAAR-RhsA-Ntox47 effector in VR3 and the Tge2 effector in VR2. Furthermore, in VR3, these isolates also contain

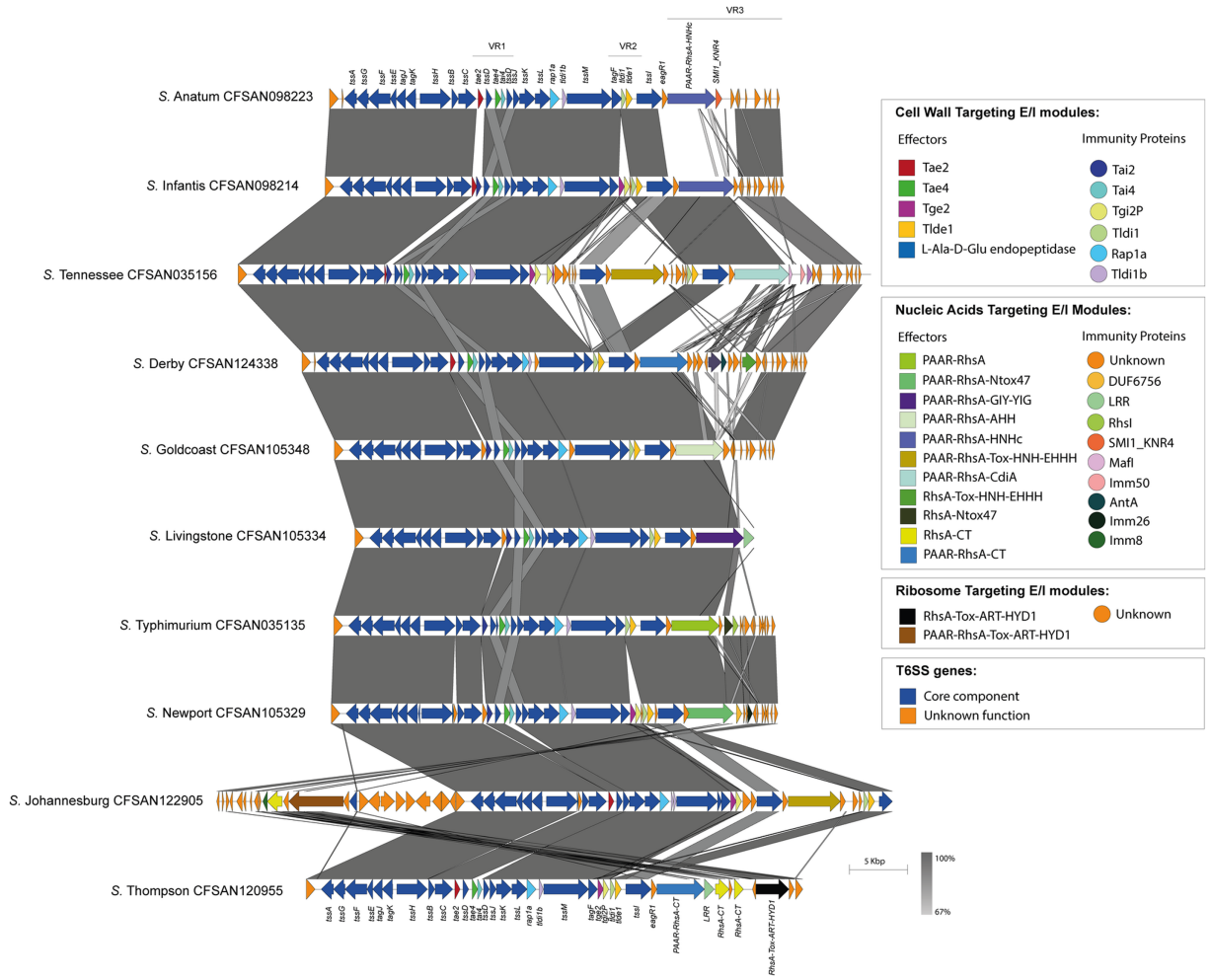


FIGURE 4
The T6SS$_{SPI-6}$ effector repertoire varies among Chilean *Salmonella* isolates. Comparative genomic analysis of the SPI-6 T6SS cluster of selected *Salmonella* isolates representing different serotypes. BLASTn sequence alignment was performed and visualized using EasyFig (Sullivan et al., 2011). ORFs encoding T6SS core components are shown in blue. ORFs encoding E/I modules are highlighted in different colors according to the confirmed or predicted functions. Grayscale represents the percentage of identity between nucleotide sequences.

an ORF with a predicted DUF6769 domain encoded in a bi-cistronic unit with an ORF harboring an Imm26 domain, which is typically found in cognate immunity proteins of bacterial toxin systems with RNase activity (Zhang et al., 2012). The presence of the Imm26 domain in this ORF suggests that the DUF6769-containing protein is a candidate effector with RNase activity. On the other hand, isolates in the second group encode the PAAR-RhsA-CT effector in VR3 and do not encode the Tge2 effector in VR2 (Supplementary Figure S4). Of note, there is no sequence identity between the Rhs elements of both groups of isolates, suggesting a different origin. In addition, the sequence of the C-terminal end of the PAAR-RhsA-CT effector encoded in these isolates shows high sequence similarity with the Rhs element of *S.* Typhi CT18 (Supplementary Figure S4).

Similar findings were also identified in *S.* Edinburgh, where two groups of isolates were distinguished. In VR3, isolates in the first group encode the PAAR-RhsA-HNHc effector protein, while isolates in the second group encode the PAAR-RhsA-CT and RhsA-CT effectors with C-terminal ends with unknown function (Supplementary Figure S5). Notably, S. Edinburgh is one of the three serotypes in which the TseH-like effector is predicted to be encoded in VR2 (Supplementary Figure S5; Supplementary Table S1).

Finally, the SPI-6 T6SS gene cluster in the remaining 32 serotypes is highly conserved among isolates within the same serotype. However, the T6SS effector repertoire and its distribution varies considerably among these 32 serotypes (Figure 4). Notably, in VR3 these serotypes encode several T6SS effector proteins with different anti-bacterial activities, including putative DNases such as PAAR-RhsA-HNHc (*S.* Anatum, *S.* Edinburgh, *S.* Infantis, *S.* Kentucky, *S.* Senftenberg), RhsA-HNHc (*S.* Tennessee), RhsA-Tox-HNH-EHHH (*S.* Braenderup, *S.* Derby), PAAR-RhsA-Tox-HNH-EHHH (*S.* Johannesburg, *S.* Tennessee), PAAR-RhsA-AHH (*S.* Goldcoast) and PAAR-RhsA-GIY-YIG (*S.* Livingstone); putative RNases such as RhsA-Ntox47 (*S.* Brandenburg, *S.* I 1,4,[5],12:i:-, *S.* Typhimurium), PAAR-RhsA-Ntox47 (*S.* Give, *S.* Livingstone, *S.* Muenchen, *S.* Newport, *S.* Panama, *S.* Sandiego) and DUF4329 (*S.* Anatum); and putative ADP-ribosyltransferases such as PAAR-RhsA-Tox-ART-HYD1 (*S.* Johannesburg), RhsA-Tox-ART-HYD1 (*S.* Thompson) and RhsA$_{main}$ (*S.* Typhimurium). Notably, 19 out of these 32 serotypes encode PAAR-RhsA-CT and RhsA-CT effectors harboring

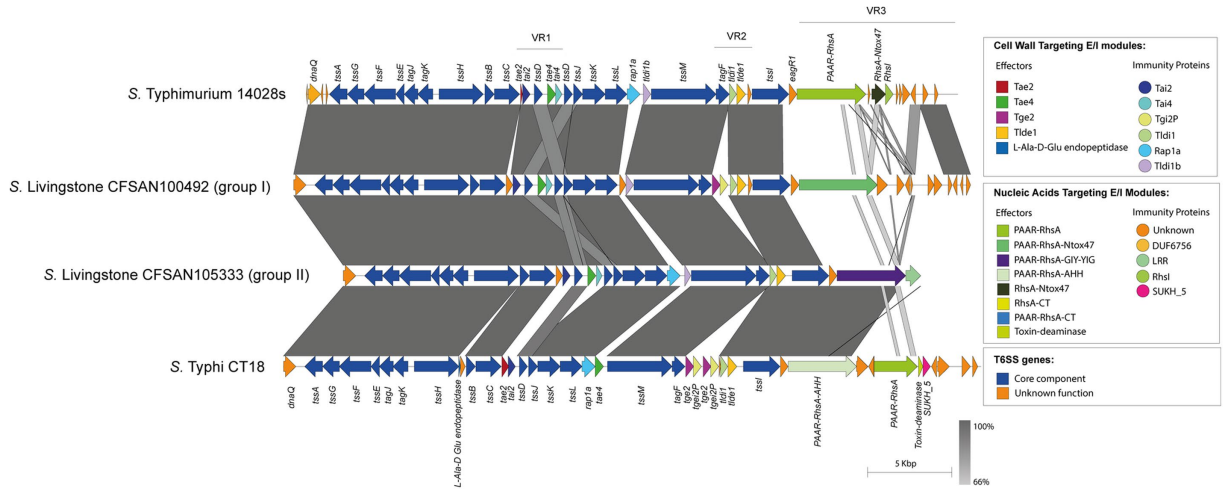


FIGURE 5
The genetic structure and repertoire of effector proteins encoded in the SPI-6 T6SS gene cluster vary among isolates of serotype *S.* Livingstone. Comparative genomic analysis of the SPI-6 T6SS cluster in *S.* Livingstone isolates. BLASTn sequence alignment was performed and visualized using EasyFig (Sullivan et al., 2011). ORFs encoding E/I modules are highlighted in different colors according to the confirmed or predicted functions. SPI-6 T6SS gene clusters from *S.* Typhimurium 14028s and *S.* Typhi CT18 were used for comparative purposes.

C-terminal ends with unknown function (Table 1; Figure 4). For instance, *S.* Johannesburg isolate CFSAN 122905 encodes an RhsA-CT candidate effector, along with a putative immunity protein harboring an Imm8 domain, which is commonly found in immunity proteins of bacterial toxin systems with RNase activity (Zhang et al., 2012). This result suggests that the C-terminal end of the RhsA-CT candidate effector has RNase activity.

The SPI-19 Rhs effectors of Chilean *Salmonella* serotypes harbor C-terminal ends with protein domains of unknown function

The SPI-19 encodes a T6SS gene cluster present in some of the most prevalent *Salmonella* serotypes worldwide, such as *S.* Dublin, *S.* Agona, *S.* Weltevreden and *S.* Gallinarum, among others. Despite its contribution to intestinal colonization, antibacterial activity and cytotoxicity against macrophages (Blondel et al., 2013; Blondel et al., 2010; Pezoa et al., 2013, 2014; Schroll et al., 2019; Xian et al., 2020) no effector protein of this T6SS has been experimentally validated and tested. This is an important knowledge gap as infections triggered by these serotypes cause major economic problems in animal production and public health issues.

Our analysis identified the SPI-19 T6SS gene cluster in isolates representing 4 out of the 42 serotypes encoding T6SS. Of note, the genetic structure of this T6SS gene cluster differs among isolates of these 4 serotypes (Figure 6). In *S.* Agona, there are two groups of isolates that encode a PAAR-RhsA-CT effector and differ in the putative cognate immunity protein. The first group encodes a putative immunity protein with a predicted TPR domain, while in the second group this protein harbors an Imm40 domain that is frequently found in cognate immunity proteins of bacterial toxin systems with RNase activity (Zhang et al., 2012) (Figure 6). Therefore, the presence of the Imm40 domain in the putative immunity protein-encoding gene suggests that the C-terminal end of the PAAR-RhA-CT candidate effector has RNase activity. Of note, a single *S.* Agona isolate (CFSAN100497) lacks the SPI-19 T6SS gene cluster and harbors that encoded in SPI-6, which encodes the effector RhsA-Ntox47. This SPI-6 T6SS gene cluster exhibits high homology to the corresponding cluster in *S.* Typhimurium 14028s (Supplementary Figure S6).

In the case of the only isolate of serotype *S.* I 4:f,g,s:1,2 analyzed, the SPI-19 T6SS gene cluster exhibits high sequence conservation between the *tssK* and *tssI* core component genes with those encoded in the corresponding cluster of *S.* Dublin and *S.* Gallinarum (Figure 6). However, this serotype encodes a PAAR-RhsA-CT effector that has a different origin from the corresponding effector of *S.* Dublin and *S.* Gallinarum. Furthermore, the cognate immunity protein of this PAAR-RhsA-CT effector harbors an Imm40 domain (Zhang et al., 2012) (Figure 6), suggesting that the C-terminal end of PAAR-RhsA-CT has RNase activity.

Although we were not able to identify new effector candidates in the SPI-19 T6SS gene cluster of isolates belonging to serotypes *S.* IV 43:z4,z23:- and *S.* Livingstone, we found some features worth mentioning. In the case of serotype *S.* IV 43:z4,z23:-, the SPI-19 T6SS gene cluster is highly conserved among the 3 isolates analyzed. However, it shares lower degree of sequence identity with the corresponding gene cluster of *S.* Dublin and *S.* Gallinarum (Figure 6). The same was true for the group of 14 *S.* Livingstone isolates carrying both SPI-6 and SPI-19 T6SS gene clusters described above (Figure 6).

The SPI-21 T6SS gene cluster from *S. enterica* subspecies *arizonae* and *diarizonae* encodes two candidate effectors

To date there is very limited information regarding the effector proteins encoded in the SPI-21 T6SS gene cluster. Only one candidate effector has been described in *S. enterica* subsp. *arizonae* serotype 62:z4,z23:- reference strain RSK2980, which corresponds to a specialized VgrG protein with a C-terminal extension including a pyocin domain

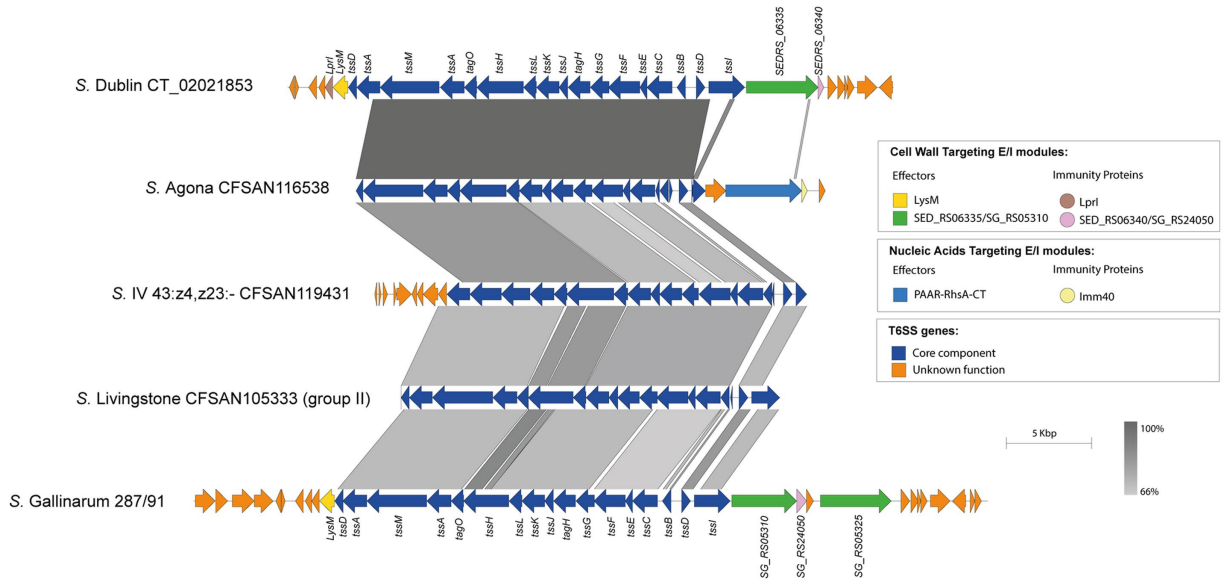


FIGURE 6
The SPI-19 T6SS gene cluster differs among Chilean *Salmonella* isolates and encodes putative T6SS Rhs effector proteins harboring C-terminal ends with domains of unknown function. Comparative genomic analysis of the SPI-19 T6SS gene cluster of *S.* Agona CFSAN116538, *S.* IV 43:z4,z23:- CFSAN119431 and *S.* Livingstone CFSAN105333. BLASTn sequence alignment was performed and visualized using EasyFig (Sullivan et al., 2011). ORFs encoding T6SS core components are shown in blue. ORFs encoding E/I modules are highlighted in different colors according to the confirmed or predicted functions. SPI-19 T6SS gene clusters from *S.* Dublin CT_02021853 (top) and *S.* Gallinarum 287/91 (bottom) were used for comparative purposes. Grayscale represents the percentage of identity between nucleotide sequences.

(S Type) (Blondel et al., 2009; Ho et al., 2017). Indeed, our bioinformatic analysis identified the VgrG-PyocinS-HNHc effector in most isolates of *S. enterica* subsp. *diarizonae* serotypes 48:i:z and 35:i:z (*S.* IIIb 48:i:z and *S.* IIIb 35:i:z, respectively) analyzed (Table 1; Figure 7A). The predicted cognate immunity protein of this candidate effector includes a inhibitory immunity protein of colicin DNase and pyocins (Col_Imm_like) domain, frequently present in immunity proteins of bacterial toxin systems (Zhang et al., 2012) (Figure 7A). Noteworthy, the SPI-21 T6SS gene cluster in all isolates of *S. enterica* subsp. *diarizonae* analyzed encodes a new candidate effector including a glucosaminidase domain with predicted peptidoglycan hydrolase activity (Table 3; Figure 7B). The predicted cognate immunity protein carries the domain with no name (DWNN). Furthermore, the SPI-21 T6SS gene cluster in the only isolate of *S.* IIIb 35:i:z analyzed (CFSAN111176) encodes a second new candidate effector with a predicted BTH_I2691 domain (Table 3; Figure 7B). Of note, BTH_I2691 is a T6SS effector protein originally described in *B. thailandensis* (Russell et al., 2012), which exhibits structural homology to colicin Ia (Parret et al., 2003). This suggests that the BTH_I2691 candidate effector protein may have membrane pore-forming activity. Finally, the SPI-21 T6SS gene cluster in all isolates of *S. enterica* subsp. *diarizonae* analyzed exhibit a relatively low degree of sequence identity with the corresponding gene cluster in *S. enterica* subsp. *arizonae* RSK2980 (Figure 7A).

Global genome-wide distribution analysis of the new candidate effectors identified in SPI-6 and SPI-21 T6SS gene clusters

The identification of 6 new candidate T6SS effectors, harboring protein domains frequently found in bacterial toxin systems, prompted us to determine their distribution across *Salmonella*. To this end, the nucleotide sequence corresponding to the ORF encoding each candidate effector was used in tBLASTx searches in publicly available *Salmonella* genome sequences deposited in the NCBI database (March, 2024) and the distribution of each effector was determined. Our analysis revealed that the new candidate effectors are distributed in a limited number of serotypes (Figures 1B, 7B). Indeed, effectors PAAR-RhsA-NucA_B, PAAR-RhsA-CdiA and RhsA-CdiA (encoded in the SPI-6 T6SS gene cluster) are distributed in 10 to 13 serotypes, while effector PAAR-RhsA-GH-E is distributed only in 5 serotypes (Figure 1B). In the case of the two candidate effectors encoded in the SPI-21 T6SS gene cluster, they are restricted to isolates of *S. enterica* subsp. *arizonae* and *S. enterica* subsp. *diarizonae* (Figure 7B).

Discussion

The T6SS has emerged as a significant virulence and environmental fitness factor for Gram-negative bacteria. The T6SS is a versatile machine that delivers a wide range of effector proteins to bacterial and/or eukaryotic cells. As a result, it has become an essential weapon for mediating interbacterial competition and host-cell interactions for many bacterial pathogens. In *Salmonella*, five T6SS gene clusters have been identified within pathogenicity islands SPI-6, SPI-19, SPI-20, SPI-21, and SPI-22 (Blondel et al., 2009; Fookes et al., 2011) which belong to 4 different evolutionary lineages. However, information regarding the presence and distribution of T6SS gene clusters and their effector proteins is still limited, partly because most analyses have focused on a limited number of strains of a few serotypes.

In this study, to expand our knowledge regarding the distribution of T6SS gene clusters and the repertoire of T6SS effector proteins in *Salmonella*, we performed bioinformatic and comparative genomic analyses of a dataset including 695 *S. enterica* genomes, representing

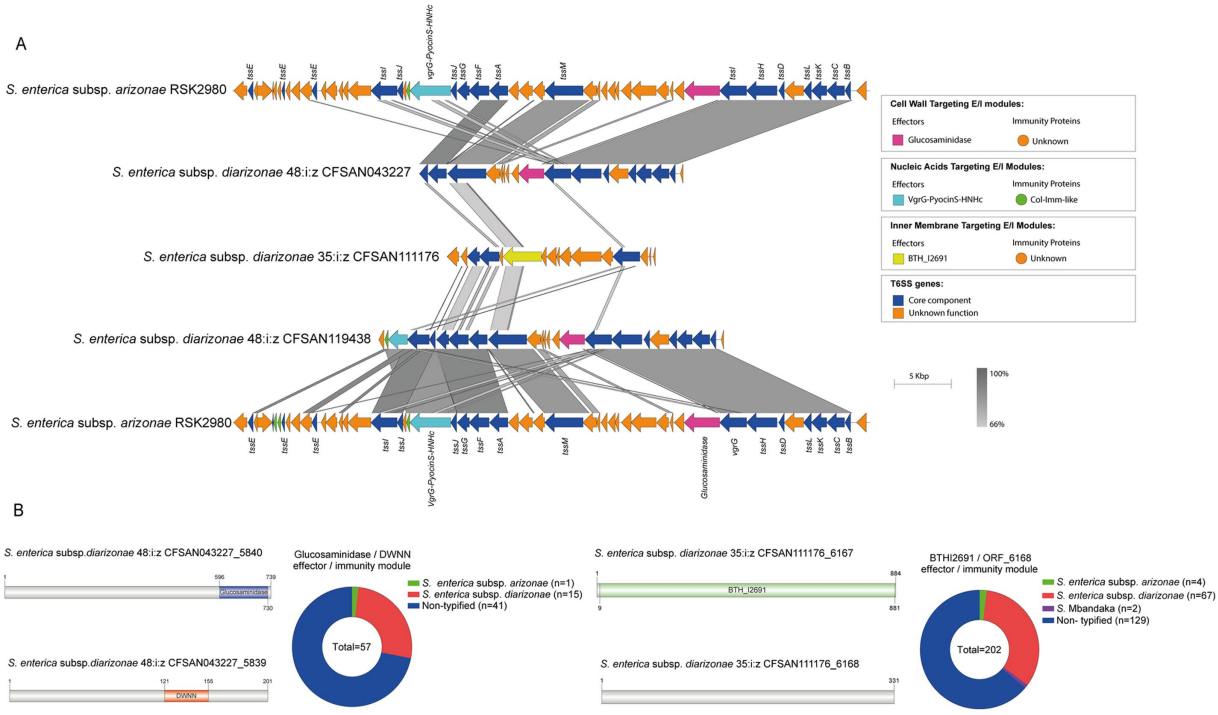


FIGURE 7
The SPI-21 T6SS gene cluster encodes new putative T6SS effector proteins. **(A)** Comparative genomic analysis of the SPI-21 T6SS cluster of *S. enterica* subsp. *diarizonae* 48:i:z CFSAN043227, *S. enterica* subsp. *diarizonae* 35:i:z CFSAN111176 and *S. enterica* subsp. *diarizonae* 48:i:z CFSAN119408. BLASTn sequence alignment was performed and visualized using EasyFig (Sullivan et al., 2011). SPI-21 T6SS gene cluster from *S. enterica* subsp. *arizonae* RSK2980 was used for comparative purposes. **(B)** Schematic representation and distribution among *Salmonella* genomes of each new effector and immunity protein identified. ORFs encoding new E/I modules are highlighted in different colors according to the predicted functions. Homologs for each component were identified by BLASTn analyses, as described in Materials and Methods.

TABLE 3 New putative T6SS effectors and cognate immunity proteins encoded in the SPI-21 T6SS gene cluster of Chilean *Salmonella* isolates.

T6SS effector genes				Cognate T6SS immunity protein genes	
ORF(s)	Size (aa)	Serotype-isolate	Predicted activity/ Domain	ORF(s)	TM or signal peptide/Domain[a]
Effectors targeting peptidoglycan					
CFSAN043227_5840	739	*S.* IIIb 48:i:z CFSAN043227	Peptidoglycan hydrolase/ Glucosaminidase	CFSAN043227_5839	No/DWNN
CFSAN119438_4687		*S.* IIIb 48:i:z CFSAN119438		CFSAN119438_4688	
Effectors targeting inner membrane					
CFSAN111176_6167	884	*S.* IIIb 35:i:z	Membrane-pore forming/ BTH_I2691	CFSAN111176_6166	No/No

[a]Presence or absence of transmembrane domains (TM) or a signal peptide, and protein domains present in the putative immunity protein genes.

44 serotypes isolated in Chile from different sources including surface waters, backyard systems and wildlife, among others. As expected, the SPI-6 T6SS gene cluster was the most prevalent in isolates of 36 different serotypes (87.48% of total *Salmonella* isolates), suggesting that the $T6SS_{SPI-6}$ is one of the most critical molecular toolboxes for *Salmonella* pathogenicity and environmental fitness. Our analysis also confirmed previous observations suggesting that the $T6SS_{SPI-19}$ is prevalent only in a subset of *Salmonella* serotypes, perhaps reflecting a contribution to *Salmonella* fitness in specialized environments and/or hosts (Blondel et al., 2009; Bao et al., 2019). Interestingly, we provide the first report on the presence of both SPI-6 and SPI-19 T6SS gene clusters in isolates of serotype *S.* Livingstone, as previously reported only in serotypes *S.* Dublin and *S.* Weltevreden (Blondel et al., 2009; Bao et al., 2019). Since the presence of multiple T6SSs in the same isolate is not common among *Salmonella* serotypes, it is still unclear how such multiplicity contributes to their environmental adaptation and/or pathogenic potential. Other T6SS gene clusters are restricted to specific serotypes. For instance, we identified the SPI-21 T6SS gene cluster only in isolates belonging to *S. enterica* subsp. *arizonae* and *S. enterica* subsp. *diarizonae,* as previously reported (Blondel et al., 2009; Bao et al., 2019). Regarding the repertoire of T6SS effector proteins of the Chilean *Salmonella* isolates, we identified 20 out of the 37 effectors previously identified in *Salmonella* (Blondel et al., 2009; Russell et al., 2012; Benz et al., 2013; Whitney et al., 2013; Koskiniemi et al., 2014; Sana et al., 2016; Ho et al., 2017; Sibinelli-Sousa et al., 2020; Amaya et al., 2022; Jurėnas et al., 2022; Lorente-Cobo et al.,

2022; Hespanhol et al., 2022; Blondel et al., 2023). These effector proteins are distributed across 42 serotypes. It is notable that the content and distribution of T6SS effector proteins in local *Salmonella* isolates differs from previous reports (Blondel et al., 2023) and show differences between isolates of the same serotype. It is therefore tempting to speculate that diverse combinations of these proteins may have different effects on the environmental fitness, which could differentially contribute to geographic adaptations and/or pathogenic potential of *Salmonella* strains. Further experimental work is required to confirm this hypothesis.

One of these differences is exemplified by the variable number of *tssI-eagR-rhs* gene modules within the VR3 of the SPI-6 T6SS gene cluster. All these modules encode different T6SS effectors and candidate effectors. In *Salmonella*, 23 T6SS effector proteins with putative nuclease activity targeting DNA and RNA have been identified so far encoded in VR3 (Blondel et al., 2009; Koskiniemi et al., 2014; Ho et al., 2017; Amaya et al., 2022; Hespanhol et al., 2022; Blondel et al., 2023). In this work, we identified 4 new candidate effector proteins with potential nuclease activity within VR3 in SPI-6. This expands our knowledge regarding the versatility of the *Salmonella* T6SS effectors in targeting bacterial nucleic acids and highlights how they are one of the main bacterial targets of *Salmonella* T6SS effector proteins. Most of these effector proteins correspond to Rhs proteins with C-terminal ends including domains with predicted antibacterial activities, thus contributing to the diversification of the molecular targets of T6SSs in *Salmonella*. This was expected, given that previous studies have demonstrated that the VR3 of the SPI-6 T6SS gene cluster encodes a variable number of Rhs elements (Blondel et al., 2009; Amaya et al., 2022; Blondel et al., 2023) and that several Rhs proteins carry C-terminal polymorphic endonuclease domains, which are associated with T6SS effectors in *Salmonella* and other bacteria (Zhang et al., 2012; Koskiniemi et al., 2014; Amaya et al., 2022; Blondel et al., 2023).

Another exciting observation is that many of the putative SPI-6 and SPI-19 Rhs effectors identified in this study harbor C-terminal extensions with unknown function. However, the presence of putative immunity proteins encoded next to these Rhs proteins suggests that these effectors have an antibacterial activity. Thus, it is tempting to speculate that the arsenal of *Salmonella* T6SS effectors harbors a diverse array of protein domains with yet-to-be-discovered activities and bacterial targets.

Regarding the SPI-19 T6SS gene cluster, we could not identify new T6SS candidate effectors encoded in the genome of the local isolates analyzed. Of note, the previously identified T6SS candidate effectors, SED_RS06235 and SED_RS06335, encoded in the SPI-19 T6SS gene cluster of *S.* Dublin CT_02021853 harbor the LysM and metallopeptidase M91 domains, respectively (Amaya et al., 2022), both of which target the peptidoglycan layer.

The only known T6SS effector encoded in the SPI-21 T6SS gene cluster corresponds to VgrG-PyocinS-HNHc, which harbors putative nuclease activity and was previously identified in *S. enterica* subsp. *arizonae* 62:z4,z23:-s reference strain RSK2980 (Blondel et al., 2009; Ho et al., 2017). Noteworthy, the SPI-21 T6SS gene cluster from our local *Salmonella* isolates encodes two new candidate effector proteins. The first one includes a glucosaminidase domain with peptidoglycan hydrolase activity, while the second one harbors the BTH_I2691 domain with predicted membrane-pore forming activity. This is the first report of a T6SS candidate effector harboring the BTH_I2691 domain present in the *Salmonella* genus, which expands our knowledge on the molecules targeted by T6SS in competing bacteria. Furthermore, this BTH_I2691 domain exhibits predicted structural homology to colicin Ia, a bactericidal protein that forms a voltage-dependent channel in the inner membrane of target cells (Parret et al., 2003). These findings suggest that $T6SS_{SPI-21}$ attacks different bacterial targets (i.e., nucleic acids, peptidoglycan and inner membrane), contributing to the fitness and virulence of both *S. enterica* subsp. *arizonae* and *S. enterica* subsp. *diarizonae*.

Finally, the distribution analysis of the six new T6SS candidate effectors identified in this study in *Salmonella* genomes from the NCBI database revealed that they are distributed in a limited number of serotypes, in contrast to the distribution previously reported for other T6SS candidate effectors in *Salmonella* (Blondel et al., 2023).

Altogether, our work broadens the repertoire of *Salmonella* T6SS effector proteins and provides evidence that the SPI-6, SPI-19 and SPI-21 T6SS gene clusters harbor a vast array of potential antibacterial effectors. This diversity is particularly evident in the VR3 of the SPI-6 T6SS gene cluster in our local *Salmonella* isolates, especially in those serotypes that lack some of the most conserved T6SS effectors encoded in VR2 (Figure 6). Finally, although this study increases the number of putative *Salmonella* antibacterial effectors against competing bacteria, it cannot be ruled out that those new candidate effectors targeting nucleic acids and cellular membranes may also affect eukaryotic cells. This represents a significant gap in our current understanding of the roles played by T6SS in host-pathogen interaction. In fact, no T6SS effector protein identified to date in *Salmonella* has been confirmed to target eukaryotic organisms, despite the clear contribution of *Salmonella* T6SSs to intracellular replication, survival and cytotoxicity inside the host immune cells (Mulder et al., 2012; Blondel et al., 2013; Schroll et al., 2019). Further research is required to address this issue.

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

FA: Conceptualization, Formal analysis, Validation, Writing-original draft, Writing review and editing, Resources, Project administration, Funding acquisition. CaB: Conceptualization, Formal analysis, Validation, Writing-original draft, Writing review and editing, Resources, Project administration, Funding acquisition, Methodolgy, Investigation, Visualization. FR-M: Conceptualization, Formal analysis, Validation, Writing-original draft, Writing review and editing, Resources, Project administration, Funding acquisition. DR: Conceptualization, Formal analysis, Validation, Writing-original draft, Writing review and editing, Resources, Project administration, Funding acquisition. AM-S: Conceptualization, Formal analysis, Validation, Writing-original draft, Writing review and editing, Resources, Project administration, Funding acquisition. MT: Conceptualization, Formal analysis, Validation, Writing-original draft, Writing review and editing, Resources, Project administration, Funding acquisition. CoB: Conceptualization, Formal analysis, Validation, Writing-original draft, Writing review and editing, Resources, Project administration, Funding acquisition. CS:

Conceptualization, Formal analysis, Validation, Writing-original draft, Writing review and editing, Resources, Project administration, Funding acquisition, Supervision. DP: Conceptualization, Formal analysis, Validation, Writing-original draft, Writing review and editing, Resources, Project administration, Funding acquisition, Methodolgy, Investigation, Visualization, Supervision.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. DP was supported by FONDECYT grant 11240160 and Fondo Concursable Proyectos de Investigación Regulares UDLA 2023 DI-13/23. CS was supported by FONDECYT grant 1212075. CaB was supported by FONDECYT grant 1241637, ECOS-ANID ECOS200037, and HHMI-Gulbenkian International Research Scholar Grant #55008749. AM-S was supported by FONDECYT grant 1231082. MT was supported by the FDA of the U.S. Department of Health and Human Services (HHS) as part of financial assistance award U01FDU001418. FA was supported by CONICYT/ANID fellowship 21191925.

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.2024.1496223/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Phylogenetic analysis and E/I module composition of T6SS gene clusters in Chilean *Salmonella* isolates. Concatenated TssC aminoacid sequences encoded in the genome of 605 Chilean *Salmonella* isolates were aligned with ClustalW using MEGA version 7.0. Next, a maximum-likelihood phylogenetic tree was built from the alignment using a bootstrap test of phylogeny (1,000 replications) with a Jones-Taylor-Thornton correction model.

SUPPLEMENTARY FIGURE S2
The genetic structure and repertoire of effector proteins encoded in the SPI-6 T6SS gene cluster vary among isolates of serotype *S.* Braenderup. Comparative genomic analysis of the SPI-6 T6SS gene cluster in isolates of *S.* Braenderup. BLASTn sequence alignment was performed and visualized using EasyFig version 2.2.5 (Sullivan et al., 2011). ORFs encoding E/I modules are highlighted in different colors according to the confirmed or predicted functions. SPI-6 T6SS gene clusters from *S.* Typhimurium 14028s and *S.* Tennessee CFSAN070645 were used for comparative purposes.

SUPPLEMENTARY FIGURE S3
The genetic structure and repertoire of effector proteins encoded in the SPI-6 T6SS gene cluster vary among isolates of serotype *S.* Give. Comparative genomic analysis of the SPI-6 T6SS cluster in isolates of *S.* Give. BLASTn sequence alignment was performed and visualized using EasyFig version 2.2.5 (Sullivan et al., 2011). ORFs encoding E/I modules are highlighted in different colors according to the confirmed or predicted functions. SPI-6 T6SS gene clusters from *S.* Typhimurium 14028s and *S.* Typhi CT18 were used for comparative purposes.

SUPPLEMENTARY FIGURE S4
The genetic structure and repertoire of effector proteins encoded in the SPI-6 T6SS gene cluster vary among isolates of serotype *S.* Newport. Comparative genomic analysis of the SPI-6 T6SS cluster in isolates of *S.* Newport. BLASTn sequence alignment was performed and visualized using EasyFig version 2.2.5 (Sullivan et al., 2011). ORFs encoding E/I modules are highlighted in different colors according to the confirmed or predicted functions. SPI-6 T6SS gene cluster from *S.* Newport SL254 was used for comparative purposes.

SUPPLEMENTARY FIGURE S5
The genetic structure and repertoire of effector proteins encoded in the SPI-6 T6SS gene cluster vary among isolates of serotype *S.* Edinburgh. Comparative genomic analysis of the SPI-6 T6SS cluster in isolates of *S.* Edinburgh. BLASTn sequence alignment was performed and visualized using EasyFig version 2.2.5 (Sullivan et al., 2011). ORFs encoding E/I modules are highlighted in different colors according to the confirmed or predicted functions. SPI-6 T6SS gene clusters from *S.* Typhimurium 14028s and *S.* Typhi CT18 were used for comparative purposes.

SUPPLEMENTARY FIGURE S6
The SPI-6 T6SS gene cluster from *S.* Agona CFSAN100497 and *S.* Typhimurium 14028s share high sequence identity. Comparative genomic analysis of the SPI-6 T6SS gene cluster of S. Agona CFSAN100497 and *S.* Typhimurium 14028s. BLASTn sequence alignment was performed and visualized using EasyFig version 2.2.5 (Sullivan et al., 2011). ORFs encoding E/I modules are highlighted in different colors according to the confirmed or predicted functions.

SUPPLEMENTARY TABLE S1
Dataset of *Salmonella* genomes retrieved from Bioproject 560080 (https://www.ncbi.nlm.nih.gov/bioproject/560080).

SUPPLEMENTARY TABLE S2
Frequency of *Salmonella* isolates of a particular serotype harboring each effector and candidate effector encoded in SPI-6.

SUPPLEMENTARY TABLE S3
Distribution of SPI-6 T6SS effectors and candidate effectors in *Salmonella* genomes. The DNA sequence encoding each T6SS effector identified in this study was subjected to tBLASTx analyses to find orthologs in all *Salmonella* genome sequences deposited in the NCBI database (March 2024).

SUPPLEMENTARY TABLE S4
Distribution of SPI-21 T6SS effectors and candidate effectors in *Salmonella* genomes. The DNA sequence encoding each T6SS effector identified in this study was subjected to tBLASTx analyses to find orthologs in all *Salmonella* genome sequences deposited in the NCBI database (March 2024).

References

Ahmad, S., Wang, B., Walker, M. D., Tran, H.-K. R., Stogios, P. J., Savchenko, A., et al. (2019). An interbacterial toxin inhibits target cell growth by synthesizing (p)ppApp. *Nature* 575, 674–678. doi: 10.1038/s41586-019-1735-9

Amaya, F. A., Blondel, C. J., Barros-Infante, M. F., Rivera, D., Moreno-Switt, A. I., Santiviago, C. A., et al. (2022). Identification of type VI secretion systems effector proteins that contribute to interbacterial competition in *Salmonella* Dublin. *Front. Microbiol.* 13:811932. doi: 10.3389/fmicb.2022.811932

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

Bao, H., Zhao, J.-H., Zhu, S., Wang, S., Zhang, J., Wang, X.-Y., et al. (2019). Genetic diversity and evolutionary features of type VI secretion systems in *Salmonella*. *Future Microbiol.* 14, 139–154. doi: 10.2217/fmb-2018-0260

Benz, J., Reinstein, J., and Meinhart, A. (2013). Structural insights into the effector – immunity system Tae4/Tai4 from *Salmonella typhimurium*. *PLoS One* 8:e67362. doi: 10.1371/journal.pone.0067362

Berni, B., Soscia, C., Djermoun, S., Ize, B., and Bleves, S. (2019). A type VI secretion system trans-kingdom effector is required for the delivery of a novel antibacterial toxin in *Pseudomonas aeruginosa*. *Front. Microbiol.* 10:1218. doi: 10.3389/fmicb.2019.01218

Blondel, C. J., Amaya, F. A., Bustamante, P., Santiviago, C. A., and Pezoa, D. (2023). Identification and distribution of new candidate T6SS effectors encoded in *Salmonella* Pathogenicity Island 6. *Front. Microbiol.* 14:1252344. doi: 10.3389/fmicb.2023.1252344

Blondel, C. J., Jiménez, J. C., Contreras, I., and Santiviago, C. A. (2009). Comparative genomic analysis uncovers 3 novel loci encoding type six secretion systems differentially distributed in *Salmonella* serotypes. *BMC Genomics* 10:354. doi: 10.1186/1471-2164-10-354

Blondel, C. J., Jiménez, J. C., Leiva, L. E., Alvarez, S. A., Pinto, B. I., Contreras, F., et al. (2013). The type VI secretion system encoded in *Salmonella* Pathogenicity Island 19 is required for *Salmonella enterica* serotype Gallinarum survival within infected macrophages. *Infect. Immun.* 81, 1207–1220. doi: 10.1128/iai.01165-12

Blondel, C. J., Yang, H.-J., Castro, B., Chiang, S., Toro, C. S., Zaldívar, M., et al. (2010). Contribution of the type VI secretion system encoded in SPI-19 to chicken colonization by *Salmonella enterica* serotypes Gallinarum and Enteritidis. *PLoS One* 5:e11724. doi: 10.1371/journal.pone.0011724

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

Brackmann, M., Nazarov, S., Wang, J., and Basler, M. (2017). Using force to punch holes: mechanics of contractile nanomachines. *Trends Cell Biol.* 27, 623–632. doi: 10.1016/j.tcb.2017.05.003

Chassaing, B., and Cascales, E. (2018). Antibacterial weapons: targeted destruction in the microbiota. *Trends Microbiol.* 26, 329–338. doi: 10.1016/j.tim.2018.01.006

Chen, Z., Moreno-Switt, A. I., Reyes-Jara, A., Delgado-Suarez, E., Adell, A. D., Oliveira, C. J. B., et al. (2024a). A multicenter genomic epidemiological investigation in Brazil, Chile, and Mexico reveals the diversity and persistence of *Salmonella* populations in surface waters. *mBio* 15:e0077724. doi: 10.1128/mbio.00777-24

Chen, Z., Toro, M., Moreno-Switt, A. I., Adell, A. D., Delgado-Suárez, E. J., Bonelli, R. R., et al. (2024b). Unveiling the genomic landscape of *Salmonella enterica* serotypes Typhimurium, Newport, and Infantis in Latin American surface waters: a comparative analysis. *Microbiol. Spectr.* 12:e0004724. doi: 10.1128/spectrum.00047-24

Cherrak, Y., Flaugnatti, N., Durand, E., Journet, L., and Cascales, E. (2019). Structure and activity of the type VI secretion system. *Microbiol. Spectr.* 7:10-1128. doi: 10.1128/microbiolspec.PSIB-0031-2019

Coulthurst, S. (2019). The type VI secretion system: a versatile bacterial weapon. *Microbiology* 165, 503–515. doi: 10.1099/mic.0.000789

Coyne, M. J., and Comstock, L. E. (2019). Type VI secretion systems and the gut microbiota. *Microbiol. Spectr.* 7:10-1128. doi: 10.1128/microbiolspec.PSIB-0009-2018

Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704

Diniz, J. A., and Coulthurst, S. J. (2015). Intraspecies competition in *Serratia marcescens* is mediated by type VI-secreted Rhs effectors and a conserved effector-associated accessory protein. *J. Bacteriol.* 197, 2350–2360. doi: 10.1128/jb.00199-15

Durand, E., Cambillau, C., Cascales, E., and Journet, L. (2014). VgrG, Tae, Tle, and beyond: the versatile arsenal of type VI secretion effectors. *Trends Microbiol.* 22, 498–507. doi: 10.1016/j.tim.2014.06.004

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223

Fookes, M., Schroeder, G. N., Langridge, G. C., Blondel, C. J., Mammina, C., Connor, T. R., et al. (2011). *Salmonella bongori* provides insights into the evolution of the salmonellae. *PLoS Pathog.* 7:e1002191. doi: 10.1371/journal.ppat.1002191

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

Hespanhol, J. T., Sanchez-Limache, D. E., Nicastro, G. G., Mead, L., Llontop, E. E., Chagas-Santos, G., et al. (2022). Antibacterial T6SS effectors with a VRR-Nuc domain are structure-specific nucleases. *eLife* 11:e82437. doi: 10.7554/eLife.82437

Ho, B. T., Fu, Y., Dong, T. G., and Mekalanos, J. J. (2017). *Vibrio cholerae* type 6 secretion system effector trafficking in target bacterial cells. *Proc. Natl. Acad. Sci. USA* 114, 9427–9432. doi: 10.1073/pnas.1711219114

Issenhuth-Jeanjean, S., Roggentin, P., Mikoleit, M., Guibourdenche, M., De Pinna, E., Nair, S., et al. (2014). Supplement 2008–2010 (no. 48) to the white–Kauffmann–Le minor scheme. *Res. Microbiol.* 165, 526–530. doi: 10.1016/j.resmic.2014.07.004

Jiang, F., Waterfield, N. R., Yang, J., Yang, G., and Jin, Q. (2014). A *Pseudomonas aeruginosa* type VI secretion phospholipase D effector targets both prokaryotic and eukaryotic cells. *Cell Host Microbe* 15, 600–610. doi: 10.1016/j.chom.2014.04.010

Jurėnas, D., Rey, M., Byrne, D., Chamot-Rooke, J., Terradot, L., and Cascales, E. (2022). *Salmonella* antibacterial Rhs polymorphic toxin inhibits translation through ADP-ribosylation of EF-Tu P-loop. *Nucleic Acids Res.* 50, 13114–13127. doi: 10.1093/nar/gkac1162

Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucleic Acids Res.* 30, 42–46. doi: 10.1093/nar/30.1.42

Katoh, K., Rozewicki, J., and Yamada, K. D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20, 1160–1166. doi: 10.1093/bib/bbx108

Koskiniemi, S., Garza-Sánchez, F., Sandegren, L., Webb, J. S., Braaten, B. A., Poole, S. J., et al. (2014). Selection of orphan Rhs toxin expression in evolved *Salmonella enterica* serovar Typhimurium. *PLoS Genet.* 10:e1004255. doi: 10.1371/journal.pgen.1004255

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Lorente-Cobo, N., Sibinelli-Sousa, S., Biboy, J., Vollmer, W., Bayer-Santos, E., and Prehna, G. (2022). Molecular characterization of the type VI secretion system effector Tlde1a reveals a structurally altered LD-transpeptidase fold. *J. Biol. Chem.* 298:102556. doi: 10.1016/j.jbc.2022.102556

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., et al. (2019). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 48, D265–D268. doi: 10.1093/nar/gkz991

Ma, A. T., and Mekalanos, J. J. (2010). In vivo actin cross-linking induced by *Vibrio cholerae* type VI secretion system is associated with intestinal inflammation. *Proc. Natl. Acad. Sci.* 107, 4365–4370. doi: 10.1073/pnas.0915156107

Ma, J., Sun, M., Dong, W., Pan, Z., Lu, C., and Yao, H. (2017). PAAR-Rhs proteins harbor various C-terminal toxins to diversify the antibacterial pathways of type VI secretion systems. *Environ. Microbiol.* 19, 345–360. doi: 10.1111/1462-2920.13621

Monjarás Feria, J., and Valvano, M. A. (2020). An overview of anti-eukaryotic T6SS effectors. *Front. Cell. Infect. Microbiol.* 10:584751. doi: 10.3389/fcimb.2020.584751

Mulder, D. T., Cooper, C. A., and Coombes, B. K. (2012). Type VI secretion system-associated gene clusters contribute to pathogenesis of *Salmonella enterica* serovar Typhimurium. *Infect. Immun.* 80, 1996–2007. doi: 10.1128/iai.06205-11

Notredame, C., Higgins, D. G., and Heringa, J. (2000). T-coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205–217. doi: 10.1006/jmbi.2000.4042

Parret, A. H., Schoofs, G., Proost, P., and De Mot, R. (2003). Plant lectin-like bacteriocin from a rhizosphere-colonizing *Pseudomonas* isolate. *J. Bacteriol.* 185, 897–908. doi: 10.1128/JB.185.3.897-908.2003

Pezoa, D., Blondel, C. J., Silva, C. A., Yang, H.-J., Andrews-Polymenis, H., Santiviago, C. A., et al. (2014). Only one of the two type VI secretion systems encoded in the *Salmonella enterica* serotype Dublin genome is involved in colonization of the avian and murine hosts. *Vet. Res.* 45:2. doi: 10.1186/1297-9716-45-2

Pezoa, D., Yang, H.-J., Blondel, C. J., Santiviago, C. A., Andrews-Polymenis, H. L., and Contreras, I. (2013). The type VI secretion system encoded in SPI-6 plays a role in gastrointestinal colonization and systemic spread of *Salmonella enterica* serovar Typhimurium in the chicken. *PLoS One* 8:e63917. doi: 10.1371/journal.pone.0063917

Pightling, A. W., Pettengill, J. B., Luo, Y., Baugher, J. D., Rand, H., and Strain, E. (2018). Interpreting whole-genome sequence analyses of foodborne bacteria for regulatory applications and outbreak investigations. *Front. Microbiol.* 9:1482. doi: 10.3389/fmicb.2018.01482

Pissaridou, P., Allsopp, L. P., Wettstadt, S., Howard, S. A., Mavridou, D. A. I., and Filloux, A. (2018). The *Pseudomonas aeruginosa* T6SS-VgrG1b spike is topped by a PAAR protein eliciting DNA damage to bacterial competitors. *Proc. Natl. Acad. Sci.* 115, 12519–12524. doi: 10.1073/pnas.1814181115

Riquelme, S., Varas, M., Valenzuela, C., Velozo, P., Chahin, N., Aguilera, P., et al. (2016). Relevant genes linked to virulence are required for *Salmonella* Typhimurium to survive intracellularly in the social amoeba *Dictyostelium discoideum*. *Front. Microbiol.* 7:1305. doi: 10.3389/fmicb.2016.01305

Rivera, D., Allel, K., Dueñas, F., Tardone, R., Soza, P., Hamilton-West, C., et al. (2021). Screening the presence of non-typhoidal *Salmonella* in different animal systems and the assessment of antimicrobial resistance. *Animals (Basel)* 11:1532. doi: 10.3390/ani11061532

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–W324. doi: 10.1093/nar/gku316

Russell, A. B., Singh, P., Brittnacher, M., Bui, N. K., Hood, R. D., Carl, M. A., et al. (2012). A widespread bacterial type VI secretion effector superfamily identified using a heuristic approach. *Cell Host Microbe* 11, 538–549. doi: 10.1016/j.chom.2012.04.007

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., et al. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944–945. doi: 10.1093/bioinformatics/16.10.944

Sana, T. G., Flaugnatti, N., Lugo, K. A., Lam, L. H., Jacobson, A., Baylot, V., et al. (2016). *Salmonella* Typhimurium utilizes a T6SS-mediated antibacterial weapon to establish in the host gut. *Proc. Natl. Acad. Sci.* 113, E5044–E5051. doi: 10.1073/pnas.1608858113

Sandt, C. H., Hopper, J. E., and Hill, C. W. (2002). Activation of prophage *eib* genes for immunoglobulin-binding proteins by genes from the IbrAB genetic island of

Escherichia coli ECOR-9. *J. Bacteriol.* 184, 3640–3648. doi: 10.1128/JB.184.13.3640-3648.2002

Schroll, C., Huang, K., Ahmed, S., Kristensen, B. M., Pors, S. E., Jelsbak, L., et al. (2019). The SPI-19 encoded type-six secretion-systems (T6SS) of *Salmonella enterica* serovars Gallinarum and Dublin play different roles during infection. *Vet. Microbiol.* 230, 23–31. doi: 10.1016/j.vetmic.2019.01.006

Sibinelli-Sousa, S., de Araújo-Silva, A. L., Hespanhol, J. T., and Bayer-Santos, E. (2022). Revisiting the steps of *Salmonella* gut infection with a focus on antagonistic interbacterial interactions. *FEBS J.* 289, 4192–4211. doi: 10.1111/febs.16211

Sibinelli-Sousa, S., Hespanhol, J. T., Nicastro, G. G., Matsuyama, B. Y., Mesnage, S., Patel, A., et al. (2020). A family of T6SS antibacterial effectors related to l,d-transpeptidases targets the peptidoglycan. *Cell Rep.* 31:107813. doi: 10.1016/j.celrep.2020.107813

Sigrist, C. J. A., de Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., et al. (2013). New and continuing developments at PROSITE. *Nucleic Acids Res.* 41, D344–D347. doi: 10.1093/nar/gks1067

Silverman, J. M., Agnello, D. M., Zheng, H., Andrews, B. T., Li, M., Catalano, C. E., et al. (2013). Haemolysin coregulated protein is an exported receptor and chaperone of type VI secretion substrates. *Mol. Cell* 51, 584–593. doi: 10.1016/j.molcel.2013.07.025

Srikannathasan, V., English, G., Bui, N. K., Trunk, K., O'Rourke, P. E. F., Rao, V. A., et al. (2013). Structural basis for type VI secreted peptidoglycan dl-endopeptidase function, specificity and neutralization in *Serratia marcescens*. *Acta Crystallogr. D Biol. Crystallogr.* 69, 2468–2482. doi: 10.1107/s0907444913022725

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

Taboada, B., Estrada, K., Ciria, R., and Merino, E. (2018). Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics* 34, 4118–4120. doi: 10.1093/bioinformatics/bty496

Tang, J. Y., Bullen, N. P., Ahmad, S., and Whitney, J. C. (2018). Diverse NADase effector families mediate interbacterial antagonism via the type VI secretion system. *J. Biol. Chem.* 293, 1504–1514. doi: 10.1074/jbc.ra117.000178

Ting, S.-Y., Bosch, D. E., Mangiameli, S. M., Radey, M. C., Huang, S., Park, Y.-J., et al. (2018). Bifunctional immunity proteins protect bacteria against FtsZ-targeting ADP-ribosylating toxins. *Cell* 175, 1380–1392.e14. doi: 10.1016/j.cell.2018.09.037

Toro, M., Weller, D., Ramos, R., Diaz, L., Alvarez, F. P., Reyes-Jara, A., et al. (2022). Environmental and anthropogenic factors associated with the likelihood of detecting *Salmonella* in agricultural watersheds. *Environ. Pollut.* 306:119298. doi: 10.1016/j.envpol.2022.119298

Uzzau, S., Brown, D. J., Wallis, T., Rubino, S., Leori, G., Bernard, S., et al. (2000). Host adapted serotypes of *Salmonella enterica*. *Epidemiol. Infect.* 125, 229–255. doi: 10.1017/s0950268899004379

Wang, J., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., Gwadz, M., Lu, S., et al. (2023). The conserved domain database in 2023. *Nucleic Acids Res.* 51, D384–D388. doi: 10.1093/nar/gkac1096

Wang, J., Yang, B., Leier, A., Marquez-Lago, T. T., Hayashida, M., Rocker, A., et al. (2018). Bastion6: a bioinformatics approach for accurate prediction of type VI secreted effectors. *Bioinformatics* 34, 2546–2555. doi: 10.1093/bioinformatics/bty155

Whitney, J. C., Beck, C. M., Goo, Y. A., Russell, A. B., Harding, B. N., Leon, J. A. D., et al. (2014). Genetically distinct pathways guide effector export through the type VI secretion system. *Mol. Microbiol.* 92, 529–542. doi: 10.1111/mmi.12571

Whitney, J. C., Chou, S., Russell, A. B., Biboy, J., Gardiner, T. E., Ferrin, M. A., et al. (2013). Identification, structure, and function of a novel type VI secretion peptidoglycan glycoside hydrolase effector-immunity pair. *J. Biol. Chem.* 288, 26616–26624. doi: 10.1074/jbc.m113.488320

Whitney, J. C., Quentin, D., Sawai, S., LeRoux, M., Harding, B. N., Ledvina, H. E., et al. (2015). An interbacterial $NAD(P)^+$ glycohydrolase toxin requires elongation factor Tu for delivery to target cells. *Cell* 163, 607–619. doi: 10.1016/j.cell.2015.09.027

Wingett, S. W., and Andrews, S. (2018). FastQ screen: a tool for multi-genome mapping and quality control. *F1000Res* 7:1338. doi: 10.12688/f1000research.15931.2

Wood, T. E., Howard, S. A., Förster, A., Nolan, L. M., Manoli, E., Bullen, N. P., et al. (2019). The *Pseudomonas aeruginosa* T6SS delivers a periplasmic toxin that disrupts bacterial cell morphology. *Cell Rep.* 29, 187–201.e7. doi: 10.1016/j.celrep.2019.08.094

Xian, H., Yuan, Y., Yin, C., Wang, Z., Ji, R., Chu, C., et al. (2020). The SPI-19 encoded T6SS is required for *Salmonella* Pullorum survival within avian macrophages and initial colonization in chicken dependent on inhibition of host immune response. *Vet. Microbiol.* 250:108867. doi: 10.1016/j.vetmic.2020.108867

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

Zhang, D., de Souza, R. F., Anantharaman, V., Iyer, L. M., and Aravind, L. (2012). Polymorphic toxin systems: comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol. Direct* 7:18. doi: 10.1186/1745-6150-7-18

Zhang, J., Guan, J., Wang, M., Li, G., Djordjevic, M., Tai, C., et al. (2023). SecReT6 update: a comprehensive resource of bacterial type VI secretion systems. *Sci. China Life Sci.* 66, 626–634. doi: 10.1007/s11427-022-2172-x

Zhang, S., Yin, Y., Jones, M. B., Zhang, Z., Deatherage Kaiser, B. L., Dinsmore, B. A., et al. (2015). *Salmonella* serotype determination utilizing high-throughput genome sequencing data. *J. Clin. Microbiol.* 53, 1685–1692. doi: 10.1128/JCM.00323-15

Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., et al. (2017). A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* 430, 2237–2243. doi: 10.1016/j.jmb.2017.12.007

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 05 February 2025
DOI 10.3389/fmicb.2025.1527049

OPEN ACCESS

EDITED BY
Renmao "Tim" Tian,
Illinois Institute of Technology, United States

REVIEWED BY
Juan Campos Guillén,
Autonomous University of Queretaro, Mexico
Taya Forde,
University of Glasgow, United Kingdom
Atmika Paudel,
GenEndeavor LLC, United States
Susanna J. Sabin,
Centers for Disease Control and Prevention (CDC), United States

*CORRESPONDENCE
Thuto Gomolemo Magome
magomethuto22@gmail.com

RECEIVED 12 November 2024
ACCEPTED 15 January 2025
PUBLISHED 05 February 2025

CITATION
Magome TG, Surleac M, Hassim A, Bezuidenhout CC, van Heerden H and Lekota KE (2025) Decoding the anomalies: a genome-based analysis of *Bacillus cereus* group strains closely related to *Bacillus anthracis*.
Front. Microbiol. 16:1527049.
doi: 10.3389/fmicb.2025.1527049

Decoding the anomalies: a genome-based analysis of *Bacillus cereus* group strains closely related to *Bacillus anthracis*

Thuto Gomolemo Magome[1]*, Marius Surleac[2,3], Ayesha Hassim[4], Cornelius Carlos Bezuidenhout[1], Henriette van Heerden[4] and Kgaugelo Edward Lekota[1]

[1]Unit for Environmental Sciences and Management, Microbiology, North-West University, Potchefstroom, South Africa, [2]The Research Institute of the University of Bucharest, Bucharest, Romania, [3]National Institute for Infectious Diseases "Matei Balș", Bucharest, Romania, [4]Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Onderstepoort, South Africa

Introduction: The *Bacillus cereus* group encompasses a complex group of closely related pathogenic and non-pathogenic bacterial species. Key members include *B. anthracis*, *B. cereus*, and *B. thuringiensis* organisms that, despite genetic proximity, diverge significantly in morphology and pathogenic potential. Taxonomic challenges persist due to inconsistent classification methods, particularly for *B. cereus* isolates that resemble *B. anthracis* in genetic clustering.

Methods: This study investigated *B. cereus* group isolates from blood smears of animal carcasses in Kruger National Park, uncovering an unusual isolate with *B. cereus* features based on classical microbiological tests yet *B. anthracis*-like genomic similarities with an Average Nucleotide Identity (ANI) of ≥95%. Using comparative genomics, pan-genomics and whole genome Single Nucleotide Polymorphism (wgSNP) analysis, a total of 103 *B. cereus* group genomes were analyzed, including nine newly sequenced isolates from South Africa and a collection of isolates that showed some classification discrepancies, thus classified as "anomalous."

Results and discussion: Of the 36 strains identified as *B. anthracis* in GenBank, 26 clustered phylogenetically with the four confirmed *B. anthracis* isolates from South Africa and shared 99% ANI. Isolates with less than 99% ANI alignment to *B. anthracis* exhibited characteristics consistent with *B. cereus* and/or *B. thuringiensis*, possessing diverse genetic profiles, insertion elements, resistance genes, and virulence genes features, contrasting with the genetic uniformity of typical *B. anthracis*. The findings underscore a recurrent acquisition of mobile genetic elements within *B. cereus* and *B. thuringiensis*, a process infrequent in *B. anthracis*.

Conclusion: This study highlights the pressing need for standardized taxonomic criteria in *B. cereus* group classification, especially as anomalous isolates emerge. This study supports the existing nomenclature framework which offers an effective solution for classifying species into genomospecies groups. We recommend isolates with ANI ≥99% to standard reference *B. anthracis* be designated as typical *B. anthracis* in GenBank to maintain taxonomic clarity and precision.

KEYWORDS

***Bacillus cereus* group, *Bacillus anthracis*, whole genome sequencing, pan-genomics, average nucleotide identity**

1 Introduction

The *Bacillus cereus sensu lato* (*s. l*) also referred as *B. cereus* group, is a complex cluster of Gram-positive, spore-forming, rod-shaped bacteria comprising both pathogenic and non-pathogenic species (Okinaka and Keim, 2016). Over 18 members have been classified as part of the *B. cereus* group; *B. albus*, *B. anthracis*, *B. cereus*, *B. cytotoxicus*, *B. luti*, *B. mobilis*, *B. mycoides*, *B. nitratireducens B. pacificus*, *B. paranthracis*, *B. paramycoides*, *B. proteolyticus*, *B. pseudomycoides*, *B. toyonensis*, *B. thuringiensis*, *B. tropicus*, *B. wiedmannii*, and *B. weihenstephanensis* (Carroll et al., 2020a,b). The most prominent pathogenic species include *B. anthracis*, *B. cereus*, and *B. thuringiensis*, each displaying unique phenotypic and virulence traits. For instance, *B. anthracis* presents as non-motile, encapsulated, non-hemolytic in sheep blood agar, and sensitive to both γ-phage and penicillin (WHO, 2008), while *B. cereus* and *B. thuringiensis* are hemolytic, motile, non-encapsulated, and exhibits γ-phage and penicillin resistance (Vilas-Bôas et al., 2007; Kolstø et al., 2009).

Bacillus anthracis is a zoonotic pathogen which causes anthrax disease (Turnbull, 2002). The disease primarily affects livestock, wildlife and humans and may present as cutaneous anthrax, acquired through contact with contaminated food or meat of animal with the disease, and inhalation anthrax, from breathing in airborne anthrax spores (WHO, 2008). Pathogenicity in *B. anthracis* is primarily attributed to plasmids pXO1 and pXO2 which harbor the anthrax toxin genes that include protective antigen (*pagA*), lethal factor (*lef*) and edema factor (*cya*) and the poly-γ-D-glutamic acid (PGA) capsule genes (*capABCDE*), respectively (Turnbull, 2002; Pena-Gonzalez et al., 2018). In contrast, *Bacillus cereus* acts as an opportunistic pathogen capable of causing gastrointestinal and non-gastrointestinal infections (Kotiranta et al., 2000; Bottone, 2010). Gastrointestinal disease may manifest as diarrheal (linked to toxins such as hemolysin BL, non-hemolytic enterotoxin, and cytotoxin K) or emetic illness (due to the cereulide toxin encoded by *cesABCDPTH*) (Schoeni and Wong, 2005; Owusu-Kwarteng et al., 2017; Ehling-Schulz et al., 2006). On the other hand, *Bacillus thuringiensis*, often employed as a natural insecticide which is not harmful to humans, contains insecticidal crystal protein genes (*cry* and/or *cyt*), which may lead to its misidentification as *B. cereus* in the absence of these genes (Kolstø et al., 2009).

Due to the clinical, agricultural, and economic importance of *B. cereus s.l*, especially the pathogenic species, substantial research has focused on their classification and taxonomy (Baldwin, 2020; Carroll et al., 2020a,b). The classification of these species was historically based on characteristics such as hemolytic activity, colony morphology, γ-phage and penicillin activity and the detection of virulence markers that were alleged to be species-specific (Rasko et al., 2005; Kamar et al., 2013; Ehling-Schulz et al., 2019). However, the emergence of exceptional genomes, such as *B. thuringiensis* isolates carrying *cry* genes that phylogenetically grouped with *B. anthracis* (Kolstø et al., 2009), along with atypical *B. cereus* and *B. cereus* biovar *anthracis* strains carrying the pBCXO1 and pBCXO2 plasmids, which contain anthrax virulence factors (Antonation et al., 2016; Baldwin, 2020), and the loss of virulence plasmids in *B. anthracis* isolates (Marston et al., 2005), suggested that the identification and classification of *B. cereus* group isolates should not rely solely on phenotypes and virulence factors (Baldwin, 2020; Brézillon et al., 2015). In some instances, molecular analysis targeting virulence determinants of *B. anthracis* revealed the presence of homologous PGA that synthesizes capsular genes (*capABCDE*) of *B. anthracis* in other *Bacillus* species (*pgsABCDE*) (Lekota et al., 2016; Lekota et al., 2018). Moreover, varying phenotypic characteristics by expressed bacterial isolates and/or limited genomic databases may have led to nomenclatural discrepancies observed in the *B. cereus* group (Afshinnekoo et al., 2015; Abdelli et al., 2023), which could have profound implications for clinical and public health settings (Carroll et al., 2020a,b; Carroll et al., 2022a,b,c). To date, the phenotypic characteristics (i.e., hemolysis activity, capsule presence/absence), plasmids and virulence factors still play a critical role in routine diagnosis and surveillance of potential disease outbreaks associated with pathogenic members of the *B. cereus* group (Mogaji et al., 2024).

Genetic techniques such as DNA–DNA hybridization (DDH) (Goris et al., 2007), multiple-locus variable-number tandem repeat analysis (MLVA) (Marston et al., 2006), amplified fragment length polymorphisms (Guinebretière et al., 2008), 16S RNA (Braun et al., 2021) and 23S RNA (Sacchi et al., 2002) were entirely instrumental methods used to differentiate members of the *B. cereus* group. The growing scheme of technology and rising novel species saw an increase in the application of more gene-specific target sequencing, such as single- and multi-locus sequence typing (SLST and MLST) methods, which became and remained some of the essential tools used for the identification of the *B. cereus* group members (Guinebretière et al., 2008). The pantoate-ß-alanine ligase (*panC*) emerged as a popular target locus used in SLST to assign members of the *B. cereus* group into seven distinct phylogenetic groups (group I-VII) based on sequence variations providing superior resolution compared to other traditional methods such the 16S RNA (Guinebretière et al., 2008). The seven-group assignment was later updated to an eight-group *panC* group assignment/genomospecies (Group I-VIII) with species names as representative names for some of the assigned groups (Carroll et al., 2022a,b,c). Group I contains species closely related to *B. pseudomycoides*. Group II (mosaicus/luti) includes certain strains that exhibit unique characteristics but are less commonly identified. Group III (mosaicus) is notably the most diverse group containing strains with emetic toxin production (cereulide), anthrax virulence genes production, and traditional *B. anthracis*, *B. cereus* and *B. thuringiensis* species closely related to traditional *B. anthracis*. Group IV (*B. cereus sensu stricto*) is a group that contains strains closely related to *B. cereus sensu stricto*, frequently identified in clinical and food-related sources. *Bacillus thuringiensis* species are also located in Group IV. Group V (*B. toyonensis*) contains strains that have been less frequently characterized but are part of the broader *Bacillus cereus* group. Group VI (*B. mycoides/paramycoides*) primarily consists of psychrotolerant species, indicating their ability to grow at lower temperatures (Liu et al., 2018). Group VII (*B. cytotoxicus*) contains specific strains with cytotoxic effects related to *cytotoxin K-1* (Fagerlund et al., 2004) and Group VIII (*B. mycoides*), which contains species closely related to *B. mycoides*.

The study of Liu et al. (2015) applied an integrated approach utilizing molecular techniques, including digital DNA–DNA hybridization (dDDH) with a ≥70% similarity threshold (Goris et al., 2007), MLST, and 16S RNA analysis. This approach clustered 224 genomes of *B. cereus s.l* into 30 clusters. Notably, dDDH analysis revealed that 20 genomes initially classified as *B. cereus* or *B. thuringiensis* exhibited unique genomic characteristics and were subsequently identified as *B. anthracis*, referred to as 'anomalous *B. anthracis*' to distinguish them from traditional *B. anthracis* isolates (Liu et al., 2015).

In a separate study, 2,231 *B. cereus* group genomes were analyzed using Fast ANI, which found that approximately 66.2% of the genomes belong to several genomospecies at an ANI threshold of 95% (Carroll et al., 2020a,b). Furthermore, additional anomalous *B. cereus* group isolates were identified, including isolates with phenotypic features typical of *B. anthracis*, *B. cereus*, and *B. thuringiensis*, which exhibited an ANI ≥ 95% with canonical *B. anthracis*, yet lacked anthrax-specific virulence genes. This work also proposed a nomenclatural framework which integrated virulence detection, MLST, ANI and *panC* group assignment for the *B. cereus* group to harmonize phenotypic and genomic classification, aiming to reduce misclassification and misinterpretation of species in clinical and industrial settings, ultimately addressing public health implications (Carroll et al., 2020a,b).

The estimated pan-genome of the *B. cereus* group encompasses approximately 60,000 genes, with around 600 core genes shared by 99% of analyzed strains (Bazinet, 2017). This estimate, however, is likely evolving with the classification of new members and the availability of additional genomes. Pan-genome analysis has been instrumental in distinguishing essential core and accessory genes, revealing the functional versatility of the *B. cereus* group and the unique adaptive capacities of individual strains (Kim et al., 2017). While pan-genome analysis illuminates the broader genetic landscape, providing insights into environmental adaptability and metabolic functionality, whole-genome single nucleotide polymorphism (wgSNP) analysis offers a high-resolution for delineating differentiating genetic diversity within closely related strains (Bogaerts et al., 2023). Single nucleotide polymorphisms are evolutionarily stable markers that contribute to elucidating deep phylogenetic relationships among global strains (Pearson et al., 2004; Girault et al., 2014; Lekota et al., 2024). Given the monomorphic nature of *B. anthracis*, wgSNP analysis has proven valuable in differentiating traditional *B. anthracis* strains from other closely related *B. cereus* group members.

This study investigated the genomic cohesion among "anomalous" *B. cereus* group isolates, prompted by the identification of an isolate with phenotypic features consistent with *B. cereus* yet a phylogenetic profile aligning with typical *B. anthracis*. A genome-based comparative approach was employed, encompassing pan-genomic and wgSNP analyses of previously reported "anomalous" *B. cereus* group strains from the studies of Liu et al. (2015) and Carroll et al. (2020a,b), as well as newly sequenced isolates displaying typical *B. anthracis* and *B. cereus* phenotypes. The isolates investigated in this study were obtained from archival animal blood smears collected in anthrax endemic regions of Kruger National Park, South Africa, as part of their anthrax outbreak surveillance program. The isolates in this study were initially screened for anthrax virulence markers (*pagA*, *lef*, and *capB*) and the chromosomal marker Ba-1, with phenotypic characteristics of the isolated recorded as detailed in Ochai et al. (2024). The collaborative work of the surveillance program is put in place to monitor anthrax outbreaks caused by traditional *B. anthracis* and/or any other potential isolates that may cause anthrax disease.

2 Materials and methods

2.1 Sample collection and screening

Bacterial cultures of samples collected between 2012 and 2015 (Table 1), were isolated from blood smears obtained from animal carcasses from the anthrax endemic regions of Kruger National Park (KNP) as described by Ochai et al. (2024). Briefly, blood smears on microscope slides were sterilely scrapped into 1.5 mL centrifuge tubes. Two hundred microliters of phosphate-buffered saline (PBS; Thermo Scientific, MA, United States) was added into the tube and half of the aliquot was spread plated on 5% sheep blood agar and incubated overnight at 37°C. All isolates with different colony morphologies were treated as different isolates and were sub-cultured on 5% sheep blood agar to obtain pure colonies. All the isolates were primarily subjected to classical methods that included microscopy, morphology, motility, hemolysin activity, γ-phage and penicillin sensitivity tests as described by the World Health Organization (WHO, 2008). The isolates were screened for the presence/absence of anthrax-toxin markers (*pagA*, *lef*, *capB*) and the *B. anthracis* chromosomal marker Ba-1 using SYBR green as prescribed by WHO (2008) and Taqman probe qPCR-based method as prescribed by Zincke et al. (2020). This was done in order to investigate any other *Bacillus* isolates which may carry anthrax virulence genes. The results are summarized in Supplementary Table S1, extracted from the study of Ochai et al. (2024).

2.2 Genomic extractions and sequencing

Genomic DNA of the isolates ($n = 9$) was extracted from overnight pure cultures using the Pure link Genomic DNA kit (Thermo Fisher Scientific, United States) following the manufacturer's protocol. The concentration and quality of the DNA were determined using the Qubit 2.0 fluorometer (Thermofisher-Scientific, United States). The DNA was sent to the Agricultural Research Council-Biotechnology Platform (ARC-BTP) for whole genome sequencing. Sequence

TABLE 1 *Bacillus anthracis* and *B. cereus* isolates from South Africa were cultured from blood smears from animal carcasses in the Kruger National Park, South Africa.

Strain	Genus	Animal source	Genus/ species	Ranger section
AX2012-121	*Bacillus*	Wildebeest	*Connochaetes*	Mooiplaas
AX2013-496	*Bacillus*	Rhinoceros	*Diceros bicornis*	Houtboschrand
AX2014-912	*Bacillus*	Impala	*Aepyceros melampus*	Lower Sabie
AX2014-949	*Bacillus*	Rhinoceros	*Diceros bicornis*	Houtboschrand
AX2016-1771Ac	*Bacillus*	Zebra	*Equus quagga*	Pafuri
AX2015-1136	*Bacillus*	Impala	*Aepyceros melampus*	Pafuri
AX2015-1152	*Bacillus*	Nyala	*Tragelaphus angasii*	Pafuri
AX2015-1270	*Bacillus*	Zebra	*Equus quagga*	Pafuri
AX2015-1277A	*Bacillus*	African Buffalo	*Syncerus caffer*	Mooiplaas

libraries of the isolates were constructed using the MGIEasy FS DNA Prep Kit (BGI, China) according to the manufacturer's protocol. The prepared libraries were sequenced using the BGI MGISEQ-2000 platform (BGI Shenzhen, China) using the paired-end 2 × 150 bp to generate reads. Four genomic sequences (AX2015-1136, AX2015-1152, AX2015-1270, and AX2015-1277A) were described and published in our previous study (Magome et al., 2024). For this study, we included these four isolates as they form part of an anthrax surveillance study in KNP from the same culture collection of the *B. cereus* group.

2.3 Genome assembly and annotation

De novo sequencing assembly, adapter trimming and polishing were performed using Shovill v4.6.0 (Seemann et al., 2020). Transeq from EMBOSS was used for the translation of nucleotide sequences into amino acids (Rice et al., 2000). Diamond software was used to compare the contig sequences against the protein databases by BlastX (the following command option was used: diamond blastx --max-target-seqs 0 --more-sensitive --id 70 -p 8 --subject-cover 90) (Buchfink et al., 2015; Buchfink et al., 2023). In-house AWK, Python and Excel scripts were further used to filter the resulting data. The annotation of the selected isolates was first performed with Prokka v1.14 using the command option: prokka --cpus 8 --gcode 11 --rnammer --compliant --center XXX (Seemann, 2014), an updated annotation tool Bakta (Schwengers et al., 2021) was then used with the following command argument: bakta --db db --verbose --threads 8. The resulting output from annotations was further used as input for Roary 3.13.0 (the following options have been used: "*-g 80000 -e --mafft -p 8 -r -qc -r -z -f*") (Page et al., 2015). Gene ontology classification of the pangenome data, was performed using DeepNOG (Feldbauer et al., 2020) and COG Classifier tools (Shimoyama, 2022). Whole genome SNP analysis was performed using Snippy v4.6.0[1] (Seemann, 2015). The clean alignment of the sequences was subjected to bootstrap =100 phylogenetic tree built with RaXML (Stamatakis, 2014).

2.4 Genome identification and phylogenetic placement

The sequenced genomes were first identified using the Pub-MLST species-ID search tool[2] (Jolley et al., 2018) and the Genome Taxonomy Database (GTDB) v1.7.0 which incorporates the Fast Average Nucleotide Identity (ANI) on KBase app (Arkin et al., 2018). Pan-genomic placement together with ANI comparison using the sequenced genomes ($n = 9$) and the 18 recognized members of the *B. cereus* group obtained from GenBank was constructed using the integrated prokaryotes genome and pan-genome analysis web services (IPGA) v1.09[3] (Liu et al., 2022). A comprehensive comparative genomics analysis was conducted on 88 anomalous genomes of the *B. cereus* group, which were retrieved from GenBank and classified into three major species: *B. anthracis* ($n = 36$), *B. cereus* ($n = 37$), and *B. thuringiensis* ($n = 15$). This analysis included five *B. cereus* biovar *anthracis* strains (CAM, CAR, CI, DRC, and UFBc0001) sourced from animal samples in Cameroon, the Central African Republic, Côte d'Ivoire, and the Democratic Republic of the Congo. These strains have reportedly caused anthrax-like diseases in their hosts (Antonation et al., 2016; Baldwin, 2020). Additionally, the atypical *B. cereus* strain BC-AK, isolated from a kangaroo in China, was included, along with nine South African isolates from this study (Supplementary Tables S2, S3). In total, this analysis examined 103 genomes. Moreover, further description of the genomes was conducted using BTyper3 (Carroll et al., 2020a,b) that involved *panC* group assignment, PubMLST genome identification, GTDB identification and ANI comparison within the IPGA (Liu et al., 2022). The pan-genome trees were based on gene presence/absence and were visualized using tvBOT incorporated in the web-based tool ChiPlot[4] (Xie et al., 2023).

2.5 Identification of mobile elements, virulence factors, and resistance genes

Various nucleotide/protein sequence databases, such as comprehensive antibiotic resistance database (CARD) (Alcock et al., 2023), ResFinder (Feldgarden et al., 2019), BacMet (Pal et al., 2014), BacAnt (Hua et al., 2021), MobileElementFinder (Johansson et al., 2021), PAIDB (Yoon et al., 2015), Iceberg2 (Liu M. et al., 2019; Liu B. et al., 2019) were used for mobile genetic elements and resistance gene predictions. Specifically, antibiotic resistance determinants were identified in each assembled genome using the ResFinder [−db ResFinder] (Feldgarden et al., 2019) with the minimum identity and coverage thresholds of 75 (−minid 75) and 50% (−mincov 50), respectively. Virulence factors in the sequenced genomes were mined using the Virulence Factor Database [−db vfdb] (Chen et al., 2016; Liu M. et al., 2019; Liu B. et al., 2019), using minimum identity and coverage thresholds of 75 (−minid 75) and 50% (−mincov 50), respectively. Further mining and annotation of antibiotic resistance genes (ARG), integrons and transposable elements was conducted using the BacAnt web-based tool (Hua et al., 2021), whereas BacMet was used to mine for metal resistance genes (Pal et al., 2014).

3 Results

3.1 Phenotypic and molecular characteristics

Microscopic examination of the nine bacterial isolates revealed that all were Gram-positive rod-shaped cells. Notably, four isolates (AX2015-1136, AX2015-1152, AX2015-1270, and AX2015-1277A) displayed long rod chains, while the remaining five isolates (AX2012-121, AX2013-496, AX2014-912, AX2014-949, and AX2016-1771Ac) were characterized by shorter rod-shaped cells, presumptively identified as *B. cereus*. Among these, the latter five were β-hemolytic,

1 https://github.com/tseemann/snippy

2 https://pubmlst.org/species-id

3 https://nmdc.cn/ipga/

4 https://www.chiplot.online/index.html

motile, and showed γ-phage and penicillin resistance. Conversely, the four isolates previously identified as *B. anthracis* (AX2015-1136, AX2015-1152, AX2015-1270, and AX2015-1277A) were non-hemolytic, γ-phage and penicillin sensitive (Ochai et al., 2024). Molecular screening of these isolates for anthrax virulence genes indicated the presence of *pagA*, *lef*, and the chromosomal marker Ba-1 in *B. anthracis* isolates. The other five isolates, AX2012-121, AX2013-496, AX2014-912, AX2014-949, and AX2016-1771Ac, tested positive for the *lef* gene. The *B. cereus* isolate AX2014-912 tested positive for anthrax chromosomal marker Ba-1, while isolate AX2016-1771Ac amplified for the *capB*, *lef*, and *pagA* marker (Supplementary Table S1).

3.2 Genome metrics and identification of the *Bacillus cereus* group isolates

The genome features of the sequenced isolates are presented in Table 2. Quality assessment of the assembled genomes showed an average of 99.43% completeness for all the genomes ($n = 9$). The genomes were initially identified using Pub-MLST and GTDB-tk v1.7.0, which incorporates Fast average nucleotide identity (ANI) by matching the query genomes against the closest reference strains incorporated in the database. The isolates AX2015-1136, AX2015-1152, AX2015-1270, and AX2015-1277A were identified as *B. anthracis* based on PubMLST species identifier and shared an ANI of ≥99% score when matched with the *B. anthracis* Vollum reference isolate (GCA_000007825.1) based on GTDB-tk v1.7.0. The isolates AX2012-121, AX2013-496, AX2014-912, and AX2014-949 were classified as *B. cereus*, based on PubMLST species identifier and each shared ≥98 ANI score when matched with *B. cereus* ATCC 14579 (GCA_000008725.1) using GTDB-tk v1.7.0. The genome size of the *B. cereus* genomes ranged from 5.34 Mb to 5.52 Mb, with a GC content ranging from 35.0 to 35.2%. Isolate AX2016-1771Ac was identified as *B. cereus* on the PubMLST species identifier. However, based on GTDB-tk v1.7.0, the closest reference genome to isolate AX2016-1771Ac was *B. anthracis* Vollum (GCA_000007825.1). The two genomes shared ≥97 ANI, which is well above the 95% ANI threshold typically used for prokaryotic species delineation (Jain et al., 2018).

The genome size of AX2016-1771Ac was 5.33 Mb with a GC content of 35.1% (Table 2). Phylogenetic placement using pan-genome and ANI analysis of the 18 recognized *B. cereus* group isolates in IPGA v1.09 (Figure 1) showed that isolates identified as *B. cereus* (AX2012-121, AX2013-496, AX2014-912, and AX2014-949) grouped with *B. cereus* ATCC 14579. The *B. anthracis* genomes (AX2015-1136, AX2015-1152, AX2015-1270, and AX2015-1277A) grouped more closely with *B. anthracis* Ames TYPE-STRAIN (GCA_000007845.1). However, isolate AX2016-1771Ac grouped closely with the *B. anthracis* isolates than with *B. cereus* isolates (Figure 1).

3.3 Phylogenetic placement of the *Bacillus cereus* group genomes using various databases

The pan-genome and wgSNP analysis revealed that four isolates (AX2015-1136, AX2015-1152, AX2015-1270, and AX2015-1277A) from South Africa identified as *B. anthracis* based on phenotypic characteristics, GTDB-tk and PubMLST classification, clustered with 26 additional genomes identified as *B. anthracis* on GenBank (PR01, PR02, PR05, PR06, PR07, PR08, PR09-1, PR09-4, PR10-4, Parent1, Parent2, Sterne, deltaSterne, BA_V770-NP1-R-ATCC 14185, Gmb1,

TABLE 2 Genome features of the nine South African sequenced *Bacillus cereus* group isolates.

Isolate	PubMLST	GTDB-tk and BTyper3-PubMLST	Contigs	Largest contig	Genome size (bp)	N50	G + C Content %	Coding sequences	RNAs
AX2015-1136	*B. anthracis*	*B. anthracis*	72	548,990	5,359,201	157,399	35.1	5,943	77
AX2015-1152	*B. anthracis*	*B. anthracis*	76	641,194	5,459,155	186,582	35.1	5,861	58
AX2015-1270	*B. anthracis*	*B. anthracis*	59	450,267	5,445,999	221,866	35.1	5,847	49
AX2015-1277A	*B. anthracis*	*B. anthracis*	43	451,964	5,451,556	538,273	35.1	5,848	52
AX2012-121	*B. cereus*	*B. cereus*	90	833,493	5,524,290	194,899	35.0	5,598	58
AX2013-496	*B. cereus*	*B. cereus*	143	443,994	5,489,039	85,188	35.1	5,617	77
AX2014-912	*B. cereus*	*B. cereus*	126	813,864	5,466,335	113,038	35.1	5,593	69
AX2014-949	*B. cereus*	*B. cereus*	98	430,535	5,431,140	211,170	35.2	5,494	74
AX2016-1771Ac	*B. cereus*	*B. anthracis*	44	616,581	5,338,489	304,668	35.1	5,568	54

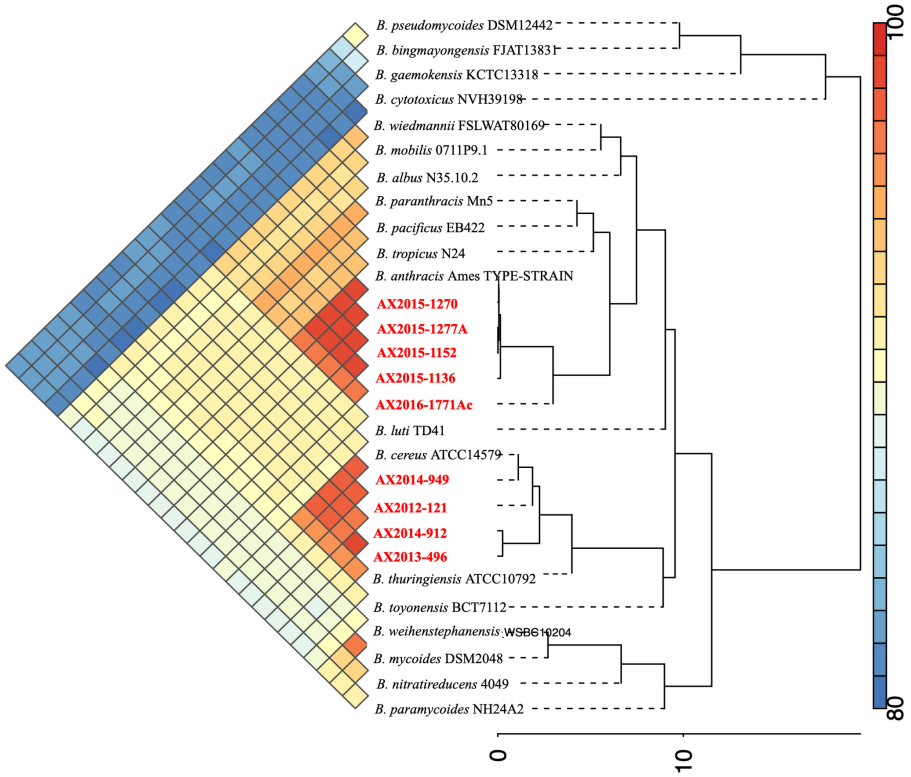


FIGURE 1
Phylogenetic placement and Average Nucleotide identity comparison using the 18 recognized members of the *B. cereus* group obtained from Carroll et al. (2020a,b), including the South African genomes (n = 9) from this study (marked in red).

Sen3, Sen2Col2, UT308, BAP417, BA781, Smith1013, Pasteur, A46, A1055, 2000031021, and 2000031052). This collective group, totalling 30 genomes, was consistently classified as *B. anthracis* based on GTDB-tk and BTyper3-PubMLST analysis and clustered in the mosaicus *panC* group III, sharing a ≥ 99% ANI (Figure 2; Supplementary Tables S2–S5). Given the genetic monomorphism characteristic of traditional *B. anthracis* strains, the wgSNP analysis (Figure 2B) supports the inclusion of these isolates within the *B. anthracis* lineage. Thus the 30 genomes are proposed to be accepted as part of the traditional or typical *B. anthracis*. This group of strains displayed diverse isolation sources, including laboratory/vaccine strains (n = 17), environment-soil (n = 1), animals (n = 9), and unknown source type (n = 3) from locations including the United States of America (USA), Senegal, Gambia, Pakistan and South Africa (Figure 2).

Ten of the genomes classified on GenBank as *B. anthracis* (Ames TYPE-STRAIN, AFS072084, PFAB2, L19, MCC1A01412, AFS081271, MCC1A02161, N1ZF-2, RIT375, and F34) did not co-cluster with the previously mentioned 30 *B. anthracis* genomes based on pan-genome and wgSNP analyses (Figure 2). Furthermore, ANI analysis showed that the 10 genomes shared ≤98 ANI identity with typical *B. anthracis* genomes (AX2015-1136, AX2015-1152, AX2015-1270, and AX2015-1277A) (Supplementary Table S5). These genomes originated from diverse sources, such as hot spring (India, strain PFAB2), salt lake (Algeria, strain F34), sediment (South China sea: strains MCCC1A02161, MCC1A01412, N1ZF-2, and L19), soybean plant (strain AFS081271), corn plant (strain AFS072084), stem tissue of a *Chamaecostus cuspidatus* plant (Puerto Rico, strain RIT375) and laboratory/vaccine (Ames TYPE-STRAIN). Despite being classified as *B. anthracis* under BTyper3-PubMLST and as *panC* group III within the mosaicus group, one genome (AFS081271) was classified as *B. thuringiensis* by GTDB-tk in IPGA, aligning with isolate B4082 classified as *B. cereus* in GenBank isolated from a food source (pea soup in the Netherlands) (Figure 2). The genomes AFS081271 and B4082 were classified as *B. thuringiensis* based on GTDB-tk in IPGA v1.09 and classified as *B. anthracis* based on Btyper3 PubMLST belonging to *panC* group II of mosaicus/luti (strain B4082) and in *panC* group III of the mosaicus group (strain AFS081271) (Figure 2). The genomes AFS081271 and B4082 shared a ≥ 99% ANI when compared against each other, suggesting they are the same species. Moreover they clustered with four South African isolates (AX2012-121, AX2013-496, AX2014-912, and AX2014-949) identified as *B. cereus* under GTDB-tk and PubMLST (Figure 2).

The four *B. cereus* genomes, isolated from animal blood smears in South Africa, grouped under *panC* group IV in the *B. cereus sensu stricto* group based on Btyper3, showing 97–98% ANI among themselves and ≤ 91% ANI compared to the GenBank *B. cereus* genomes (Figure 2; Supplementary Tables S4, S5). Thirty-five GenBank *B. cereus* genomes were identified as *panC* group III of the mosaicus group and two GenBank *B. cereus* genomes (B4082 and MOD1_Bc20), isolated from food sources were identified as *panC* group II of the mosaicus/luti group, however they were classified as *B. anthracis* in BTyper3 (Figure 2). The results above highlight the discrepancies that exist within the various genomic identification

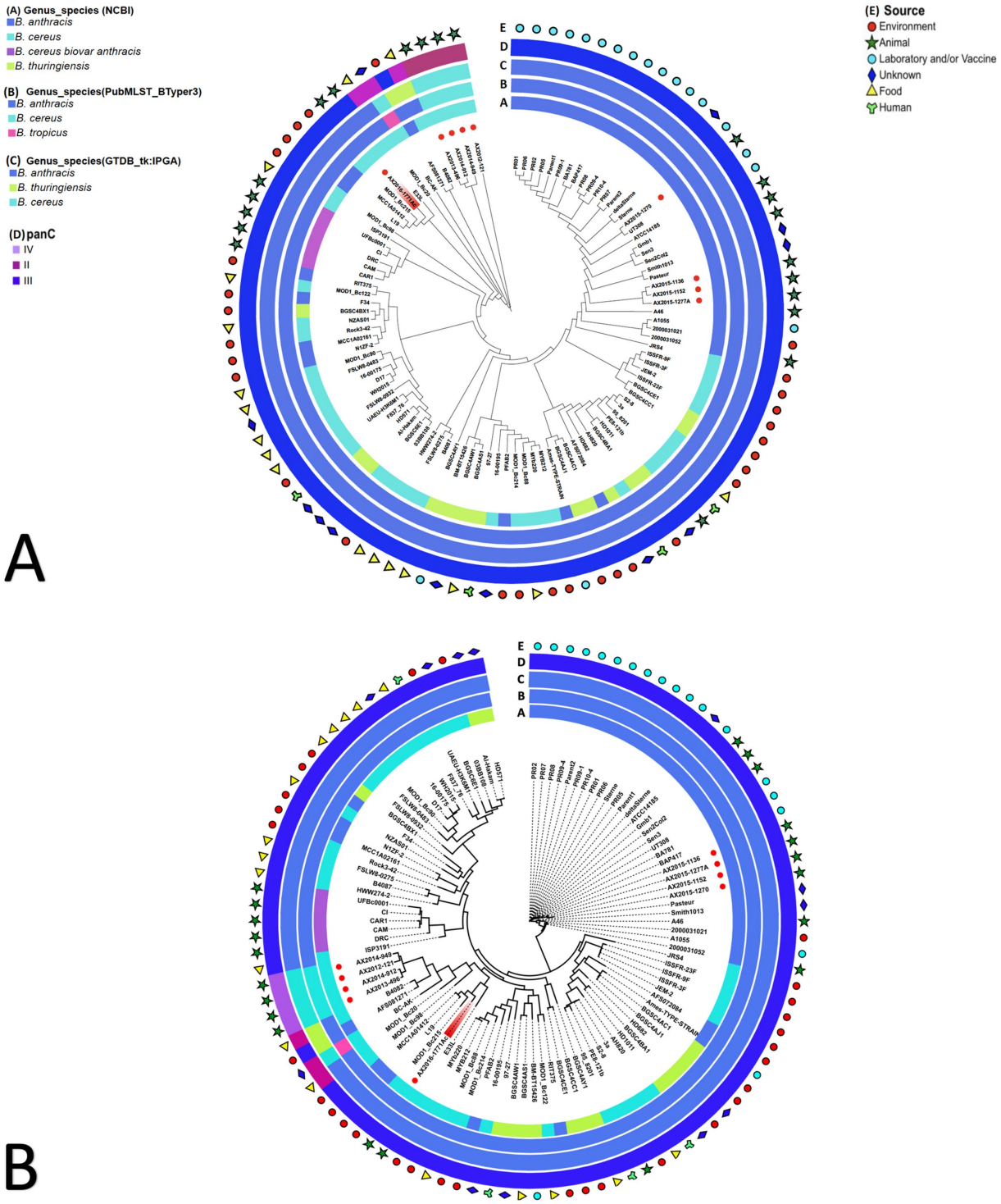


FIGURE 2
(A) Pan-genome and **(B)** whole genome single nucleotide polymorphism (wgSNP) analysis of 103 *Bacillus cereus* group genomes analyzed. The red dots in the inner circle indicate the South African isolates from this study. The AX2016-1771Ac highlighted in red. The rings surrounding the phylogenetic tree represent the following: **(A)** current genus species assignment as submitted on GenBank database; (B) BTyper3 PubMLST genome species assignment; (C) GTDB-tk-IPGA species assignment; (D) *panC* group assignment; (E) general source from which samples were obtained from.

databases which further complicate taxonomic classification. Additionally, an anomalous AX2016-1771Ac isolate, originating from a zebra blood smear, sequenced in this study, grouped with *B. cereus* (E33L and MOD1_Bc215) and *B. anthracis* (L19 and MCC1A01412) genomes using pan-genome and wgSNP analysis (Figure 2). These isolates, identified as *B. anthracis* under GTDB-tk and BTyper3-PubMLST, aligned with *panC* group III in the mosaicus group. The anomalous isolate AX2016-1771Ac shared 97–98% ANI with similar

genomes from the South China Sea (strains L19 and MCC1A01412), United States baby wipes (MOD1_Bc215), and a Namibian zebra (strain E33L). Two *B. thuringiensis* strains (HD571 and Al-Hakam) that form part of the anomalous clustered with *B. cereus* strains BGSC6E1, 03BB108, F837_76, and UAEU-H3K6M1 as *panC* group III species of the mosaicus group (Figure 2). GenBank classified *B. cereus* strains (FLSW8-0483, WH2015, D17, and MOD1_Bc90) isolated from food in the United States and strain 16-00175 from France formed part of this group. Meanwhile, the *B. anthracis* Ames TYPE-STRAIN and AFS072084 genomes formed close associations with *B. thuringiensis* strains (BGSC4AJ1 from Mexico and BGSC4AC1 from India), with a notable ≥99% ANI score shared between *B. anthracis* Ames TYPE-STRAIN and *B. thuringiensis* BGSC4AJ1 (Supplementary Table S5).

The anomalous *B. cereus* ISP3191 genome isolated from food source, clustered closely with the *B. cereus* biovar *anthracis* strains (CAM, CAR, CI, DRC, and UFBc0001) isolated from animals (gorilla, chimpanzees, goat, and colobus monkey) in Western Africa, sharing ≥99% ANI. This cluster displayed a 97% ANI similarity with 82 other GenBank genomes, including the four typical *B. anthracis* strains (AX2015-1136, AX2015-1152, AX2015-1270, and AX2015-1277A) from this study. A subset of anomalous isolates (AFS081271, L19, B4082, MOD1_Bc20, MOD1_Bc98, and MOD1_Bc215) exhibited ≤96% ANI with *B. cereus* biovar *anthracis* isolates. Moreover, the *B. cereus* (AX2012-121, AX2013-496, AX2014-912, and AX2014-949) isolates from this study, shared <91% ANI with *B. cereus* biovar *anthracis*. The AX2015-1771Ac isolate shared ≥96% ANI score with *B. cereus* biovar *anthracis* isolates. However, AX2015-1771Ac shared <94% ANI with atypical *B. cereus* strain BC-AK. These results indicate that *B. cereus* group isolates may belong to different species clusters when a threshold ≥95% is adopted for all species in the same group.

3.4 Comparative annotation and functional analysis of *Bacillus cereus* group genomes

To predict gene content and perform pan-genomic analysis, 103 genomes within the *Bacillus cereus* group were annotated using two distinct tools: Prokka and Bakta. Initially, the annotation based on Prokka identified a total of 38,079 genes, with 26,173 (68. 73%) classified as hypothetical proteins. In contrast, Bakta annotation assigned 37,678 genes, only 5,391 (14.31%) were predicted to be hypothetical proteins. This suggests that Bakta may offer improved functional predictions for gene products within the *B. cereus* group, potentially reducing ambiguity around hypothetical proteins (Figure 3).

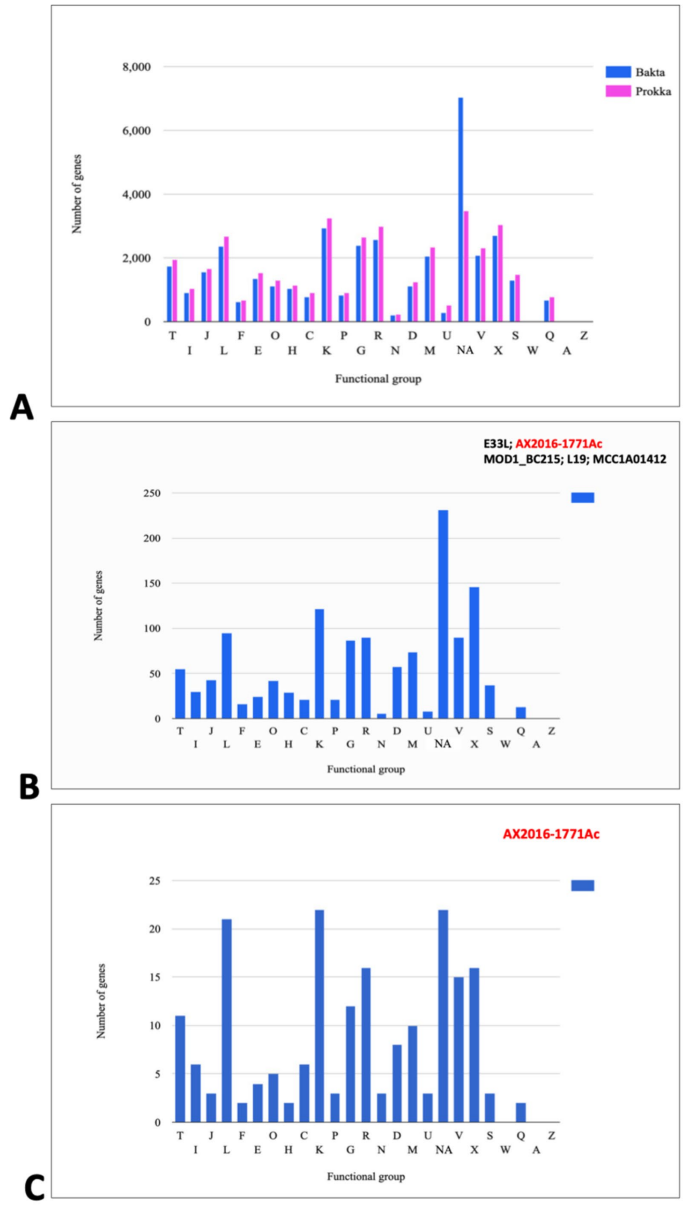
In the Bakta annotation, 1,338 cloud genes were solely found in five anomalous isolates AX2016-1771Ac, *B. anthracis* strains (L19 and MCC1A01412) as well as *B. cereus* strains (E33L and MOD1_Bc215). Within this subset, 271 of these genes were assigned as hypothetical proteins. Additionally, 14 genes with distinct functional roles were identified exclusively in these five isolates. These genes included transport-related proteins (e.g., ABC transporter permease), structural proteins (e.g., spore coat protein), regulatory proteins (e.g., transcriptional regulators such as LysR and YdeE), metabolic enzymes (e.g., putative ubiquinone/menaquinone methyltransferase and NAD-dependent epimerase/dehydratase family protein), and other specific proteins (e.g., FAD-dependent oxidoreductase and DUF-domain proteins like DUF998, DUF952, DUF4430, and DUF3380). These unique genes may indicate specialized functionalities within these isolates. Functional categorization based on the cluster of orthologous groups (COG) analysis revealed that from the 1,338 clouds genes of the five compared strains, 146 genes (10.9%) were associated with mobilomes, including prophages and transposons functional group (X), followed by the transcription functional group (K) with 122 (9.1%) gene count, and the replication, recombination and repair functional group (L) with 95 (7.1%) gene counts (Figure 3B; Supplementary Table S6). Notably, the AX2016-1771Ac isolate contained 195 unique genes, which were not detected in other *B. cereus* group isolates, based on the COGs (Figure 3C).

3.5 Antibiotic resistance genes in *Bacillus cereus* group

A total of 16 antibiotic resistance genes (ARGs) were distributed across 103 *Bacillus cereus* group genomes analyzed in this study (Figure 4). The most abundant ARGs identified across all genomes (n = 103) were the beta-lactamase genes *Bla1* (96%), *BcII* (95%), and *Bla2* (78%). A distinct class A beta-lactamase gene, *BcI*, was exclusively detected in four *B. cereus* genomes as (AX2012-121, AX2013-496, AX2014-912, and AX2014-949) and the anomalous GenBank classified *B. thuringiensis* strain BM-BT15426. Additionally, the beta-lactamase gene, $Bla_{TEM\text{-}116}$ was solely detected in the anomalous *B. cereus* strain FSLW8-0932, which was isolated from a food source in the United States. Another resistant gene, *satA* (encoding streptothricin N-acetyltransferase) was found in 48 genomes, but was absent in the *B. cereus sensu stricto* isolates (AX2012-121, AX2013-496, AX2014-912, and AX2014-949). The Fosfomycin resistance gene *FosB2* was identified in 30 genomes, all suggested in this study to represent traditional/typical *B. anthracis* with high genomic similarity (≥99 ANI) (Figure 2). Notably, the 30 *B. anthracis* isolates, exhibited a consistent ARG profile that included *BcII*, *Bla1*, *Bla2*, *FosB2*, and *satA* (Figure 4), a typical profile for *B. anthracis* isolates.

The fosfomycin gene *FosB1* was detected in 23 anomalous genomes made up of 12 *B. cereus* isolates (16-00195, 3a, 95_8201, D17, E33L, MOD1_Bc20, MOD1_Bc214, MOD1_Bc98, PE8-121b, Rock3-42, S2-8, and BGSC6E1), four *B. anthracis* isolates (RIT375, PFAB2, NIZF-2, and MCC1A02161), two *B. thuringiensis* isolates (HD10111 and BGSC4BA1), including the atypical *B. cereus* BC-AK strain and the four *B. cereus sensu stricto* isolates (AX2012-121, AX2013-496, AX2014-912, and AX2014-949) from this study (Figure 4). Furthermore, the vancomycin resistance gene, *vanR-A*, was detected in 15 genomes that included *B. cereus* classified genomes (AX2012-121, AX2013-496, AX2014-912, AX2016-1771AC, 03BB108, B4082, F837_76, MOD1_Bc20, MOD1_Bc215, and BGSC6E1), *B. anthracis* genomes (AFS081271, MCC1A01412, and L19) and *B. thuringiensis* (HD571 and Al-Hakam). The gene *rpoB* known to confer resistance to rifampicin was detected in the South African *B. cereus* strains (AX2012-121, AX2013-496, and AX2014-912), the atypical *B. cereus* BC-AK strain associated with anthrax-like disease, as well as in the anomalous *B. anthracis* F34 strain isolated from Algeria. The *vanS-Pt* gene that is associated with ruminant microbiota in *Paeniballicus* strains (Guardabassi et al., 2005) and vancomycin resistance was only detected in two *B. cereus* genomes AX2013-496 and AX2014-912 sequenced in this study. The tetracycline efflux pump *tet(45)* was

10.3389/fmicb.2025.1527049



Functional group	Description
T	Signal transduction mechanisms
I	Lipid transport and metabolism
J	Translation, ribosomal structure and biogenesis
L	Replication, recombination and repair
F	Nucleotide transport and metabolism
E	Amino acid transport and metabolism
O	Posttranslational modification, protein turnover, chaperones
H	Coenzyme transport and metabolism
C	Energy production and conversion
K	Transcription
P	Inorganic ion transport and metabolism
G	Carbohydrate transport and metabolism
R	General function prediction only
N	Cell motility
D	Cell cycle control, cell division, chromosome partitioning
M	Cell wall/membrane/envelope biogenesis
U	Intracellular trafficking, secretion, and vesicular transport
NA	Not_Assigned
V	Defense mechanisms
X	Mobilome: prophages, transposons
S	Function unknown
W	Extracellular structures
Q	Secondary metabolites biosynthesis, transport and catabolism
A	RNA processing and modification
Z	Cytoskeleton

FIGURE 3
Cluster of orthologous groups (COGs) analysis indicating annotated genes in different functional groups **(A)** COG gene count comparison of the two annotation tools Prokka (pink) and Bakta (blue). **(B)** COG shared between the branch containing the five genomes *B. cereus* (AX2020-1771Ac, E33L, and MOD1_BC215) and *B. anthracis* (L19 and MCC1A01412) based on Bakta annotation. **(C)** COG corresponds to the set of genes that are unique to the anomalous sequenced isolate AX2016-1771Ac sequenced in this study. Each bar height corresponds to the total number of genes in the compartment that were assigned to the COG functional group. NA (Not Assigned compartment represents the genes that could not be assigned in a functional compartment).

detected in anomalous *B. anthracis* F34 and *B. thuringiensis* BGSC4BX1 isolates. The tetracycline efflux pump *tet(L)* was detected in the anomalous *B. cereus* isolates BGSC6E1 and MOD1_Bc98. The sequenced typical *B. cereus* isolates AX2013-496 and AX2014-912 each presented seven ARGs namely the *BcI*, *BcII*, *Bla1*, *rpoB*, *FosB1*, *vanR-A*, and *vanS-Pt*. The *B. cereus* AX2012-121 strain contained six ARGs namely the *BcI*, *BcII*, *Bla1*, *rpoB*, *FosB1*, and *vanR-A*. The *B. cereus* AX2014-949 strain contained only four ARGs, which were identified as *BcI*, *BcII*, *Bla1*, and *FosB1*. The anomalous *B. cereus* strain AX2016-1771Ac also contained four ARGs namely *BcII*, *Bla1*, *Bla2*, and *vanR-A*. *Bacillus cereus* and *B. thuringiensis* isolates indicated a potential to carry varying resistance gene profiles from the typical *B. anthracis* isolates.

3.6 Virulence factors in *Bacillus cereus* group

In this analysis, 117 virulence factor genes were found across 103 genomes analyzed in this study, notably genes which form part of a cluster were counted as individual genes. It has already been established that the 88 GenBank isolates *B. anthracis* ($n = 36$), *B. cereus*

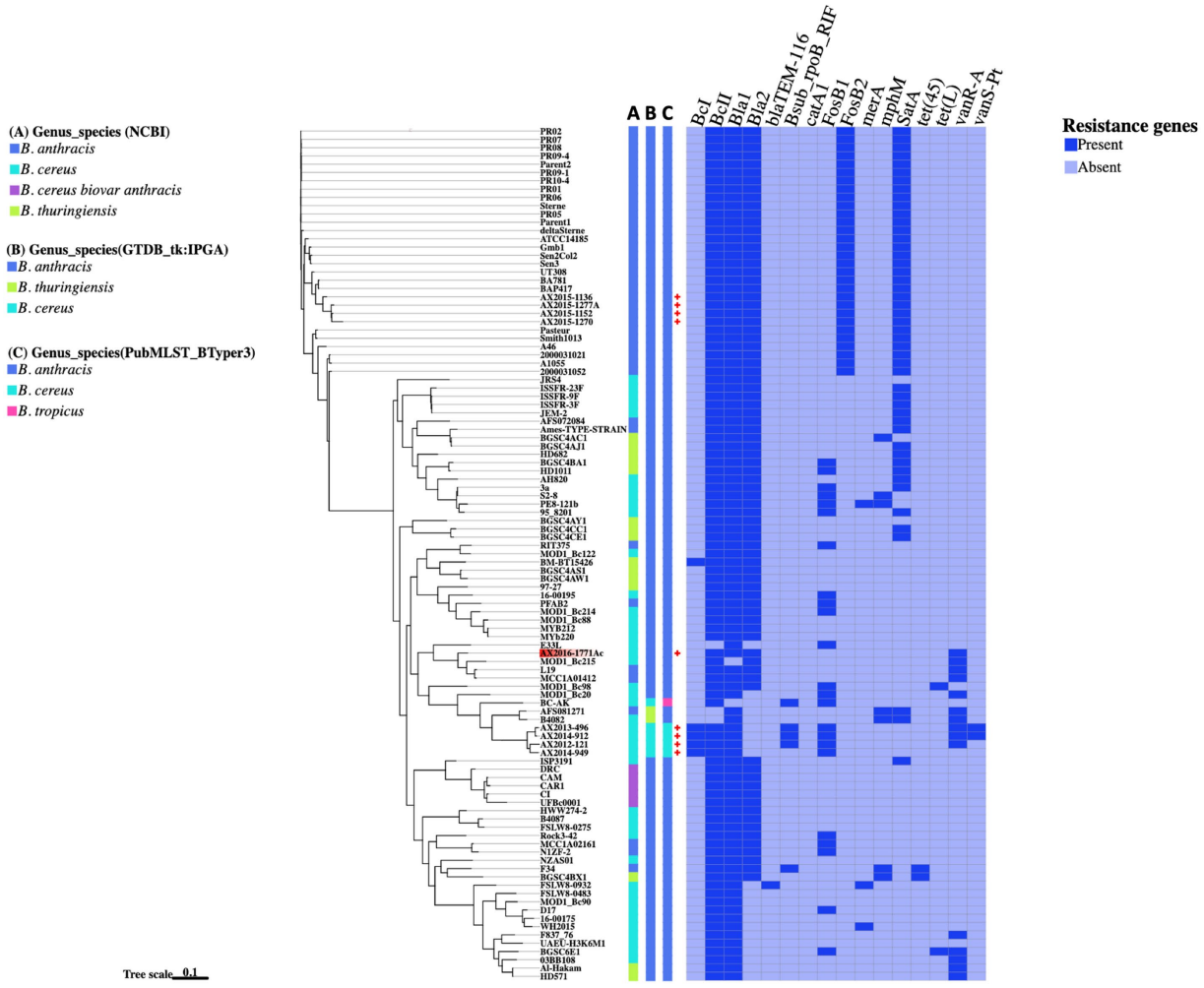


FIGURE 4
The presence and absence of resistance genes comparison between the *Bacillus* genomes. The phylogenetic tree is based on the wgSNP analysis of the 103 *Bacillus* genomes. The South African genomes from this study are marked with red crosses. A–C Blocks represents names assigned to the genomes on different genome identification databases: (A) Current genus species assigned to the genomes on NCBI; (B) Genus species assigned by the Genome taxonomy database in IPGA v1.09 (GTDB-tk); (C) Genus species assigned by the PubMLST in BTyper3.

($n = 37$), and *B. thuringiensis* ($n = 15$) do not contain the anthrax virulence genes: edema factor- *cya*, lethal factor- *lef*, the protective antigen -*pagA* located on the pXO1 plasmid (Carroll et al., 2020a,b). In this study, the virulence genes cya, *lef* and *pagA* were found present in the confirmed *B. anthracis* isolates AX2015-1136, AX2015-1152, AX2015-1270, and AX2015-1277A, atypical *B. cereus* BC-AK isolate and the *B. cereus* biovar *anthracis* isolates (UFBc0001, CAR, CAM, and CI). The anomalous isolate AX2016-1771Ac did not contain the anthrax virulence genes *cya*, *lef* and *pagA*. All the capsular genes *capABCDE* and the capsule synthesis transcriptional regulator genes *acpA* and *acpB* were detected in the identified and classified traditional *B. anthracis* isolates (deltaSterne, AX2015-1152, AX2015-1270, AX2015-1277A, A1055, 2000031021, 2000013052, A46, Smith1013, and Pasteur), *B. cereus* biovar *anthracis* isolates (UFBc0001, CI, CAM, CAR, and DRC) and the *B. cereus* BC-AK isolate. The *B. cereus* 03BB108 isolated from the environment (dust particles) in the United States and the *B. anthracis* isolate N1ZF-2 from sediments in China both contained the capsule synthesis regulator gene *acpB* and the capsule genes *capA* and *cap*C. The virulence regulator gene of *B. anthracis* (*atxA*), and the *has-ACB* gene cluster that encodes for the hyaluronic acid capsule was detected in sequenced *B. anthracis* isolates (AX2015-1136, AX2015-1152, AX2015-1270, AX2015-1277A), and anomalous *B. anthracis* strains BAP417 and BA781, the *B. cereus* BC-AK isolate including the *B. cereus* biovar *anthracis* isolates UFBc0001, CI, CAM, and CAR (Figure 5).

The insecticidal crystalline delta-endotoxin gene *cryIII* was detected in the *B. anthracis* Ames TYPE-STRAIN and *B. thuringiensis* BGSC4AC1 and BGSC4AJ1 strains that shared a common ancestor. At least two or more genes belonging to the gene clusters *asb-ABDEF* encoding for the biosynthetic machinery for petrobactin and the *dhb-ABCEF* bacillibactin which are siderophores involved in iron acquisition of *Bacillus* species were detected in all the genomes. Although *B. anthracis* is non-motile, gene clusters related to flagellum synthesis (*flg*, *flh*, and *fli*) were present in all the genomes. The non-hemolytic enterotoxin *nhe-ABC* cluster was present in over 99% percent of the genomes. Whereas genes of the hemolytic enterotoxin gene complex *hbl-ACD* were detected in 24 genomes which included anomalous *B. cereus*, *B. anthracis*, and *B. thuringiensis*. The cytotoxin

gene *cytK* was present in 49.6% of the genomes classified as *B. anthracis*, *B. cereus*, and *B. thuringiensis*, however our sequenced *B. anthracis* strains lack this gene. The hydrolase *cesH* gene which forms part of the cereulide operon was detected in six anomalous *B. cereus* strains (3a, 95_8201, BA087, BGSC6E1, PE8-121b, and S2-8). A more detailed description of the virulence genes is included in Supplementary Table S7.

3.7 Determination of the insertion sequences on the 103 genomes

A total of 50 insertion sequences (IS) and five mobile insertion cassettes (MIC) were detected among the 103 genomes (Figure 6). The 50 IS were distributed as follows: IS231 (n = 15), ISBce (n = 15), ISBth (n = 16), ISBwe (n = 2), ISBt (1), ISBsp8 (n = 1), and IS232 (n = 1). Mobile insertion cassettes included MICBan1 (n = 1), MICBce (3) and MICBth (n = 1). The insertion sequence IS231L was the most prevalent, as it was detected in all 103 genomes analyzed in this study. Among the 30 genomes classified as *B. anthracis* in this study, insertion sequences IS231 (L and S), ISBt (2 and 7) and the insertion cassette MICBan1 were found in AX2015-1136, AX2015-1152, AX2015-1277A, AX1270, and BA781. Other insertion sequences, such as IS2321, ISBwe, and ISBt variants, were less frequent than ISBce and ISBth. The four South African *B. anthracis* isolates contained more IS than the *B. anthracis* isolates in the same cluster (Figure 6). The anomalous *B. cereus* strain AX2016-1771Ac strain sequenced in this study included the insertion sequences IS231L, ISBce (17 and 19), ISBth (4 and 7) and the insertion cassette MICBan1. None of the insertion genes can discriminate typical *B. anthracis* from the anomalous *B. cereus* group strains compared in this study. The *B. anthracis* Ames TYPE-STRAIN and *B. thuringiensis* BGSC4AJ1 and BGSC4AC1 contained similar insertion sequence profiles, which included IS231 (C, E, H, L), ISBce (3 and 15), ISbth (16 and 17), and MICBan1.

4 Discussion

In this study, a comparative genomics approach was employed to analyze the nine *B. cereus* group isolates (Table 1), from animal blood smears that also included previously reported four *B. anthracis* strains from anthrax surveillance in KNP (Magome et al., 2024). Pan-genomic and wgSNP analysis was conducted to investigate previously reported "anomalous" *B. cereus* group strains (Liu et al., 2015; Carroll et al., 2020a,b), as well as newly sequenced isolates displaying typical *B. anthracis* and *B. cereus* phenotypes. The study highlights and supports some of the approaches used in the proposed framework by Carroll et al. (2020a,b). We further indicate through a restrictive ANI value of 99% and wgSNP analysis traditional *B. anthracis* can be distinguished from other closely related species which form part of the mosaicus group. Additionally, mobile genetic elements, including

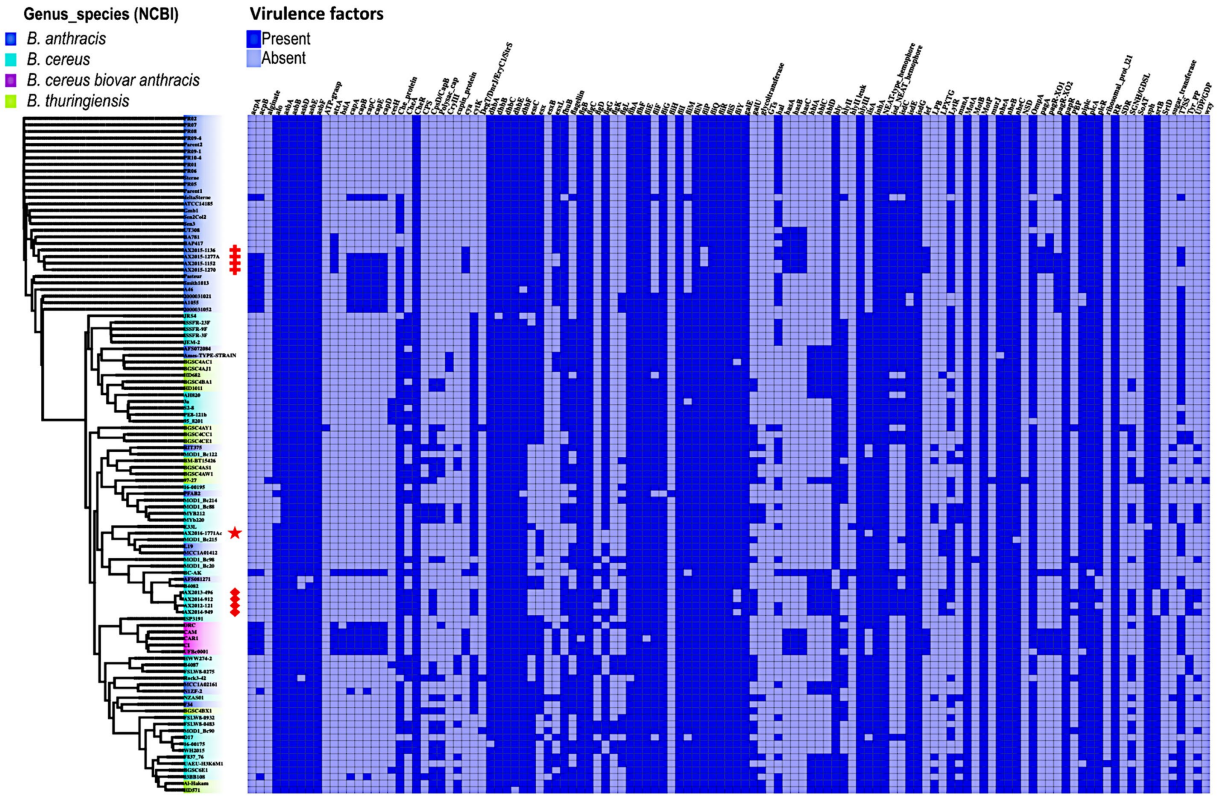


FIGURE 5
Presence and absence of virulence genes/factors identified in *Bacillus* genomes (n = 103) used in this study. The phylogenetic tree is based on the wgSNP of the *Bacillus* genomes. The South African isolates classified as *B. anthracis* (n = 4) are marked with red crosses; *B. cereus s. s* (n = 4) red diamonds shape, anomalous isolate AX2016-1771Ac marked with the red star.

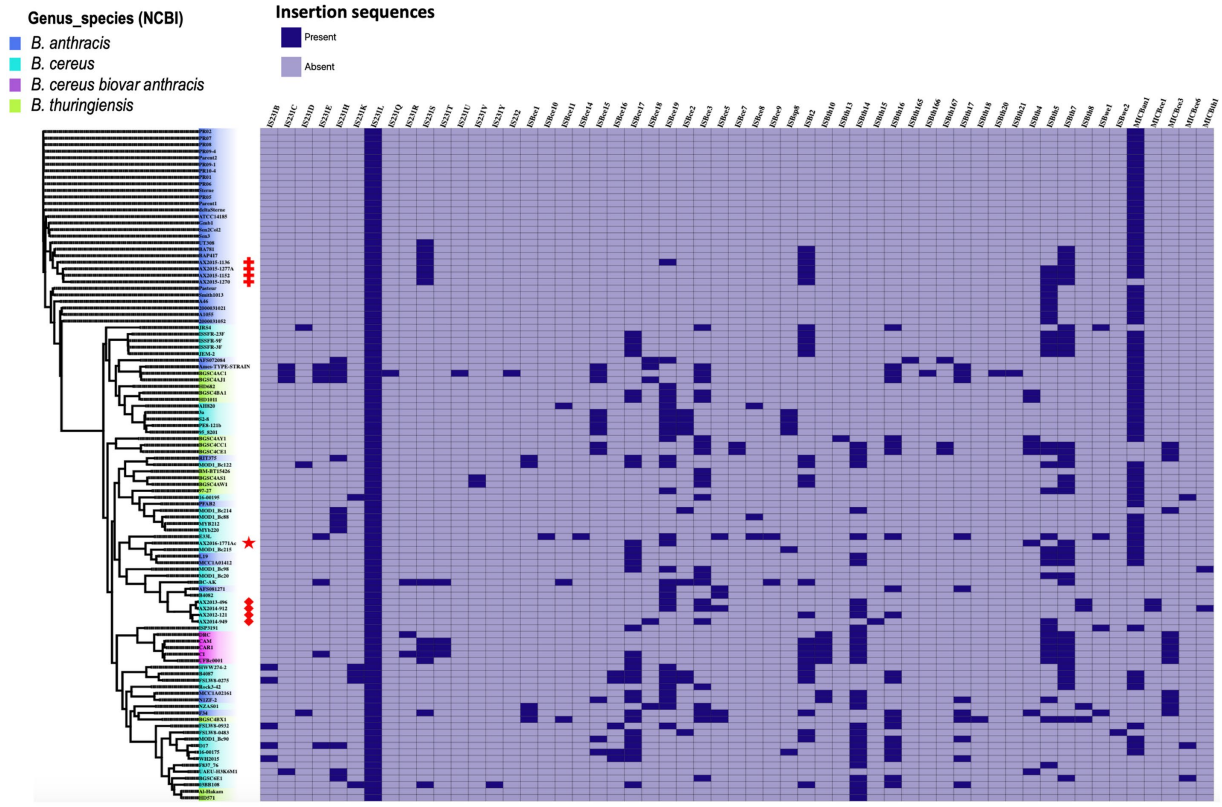


FIGURE 6
Heat map showing the presence and absence of insertion sequences and mobile insertion cassettes detected on 103 *B. cereus* group genomes analyzed in this study. The phylogenetic tree is based on the wgSNP analysis. The South African isolates classified as *B. anthracis* (n = 4) are marked with red crosses; *B. cereus s. s* (n = 4) red diamonds shape, anomalous isolate AX2016-1771Ac marked with a red star.

insertion sequences, virulence and antibiotic resistance genes across the genomes were investigated to contribute to the growing body isolates that form part of the *B. cereus* group.

The isolates in this study were initially screened for anthrax virulence markers (*pagA*, *lef* and *capB*) and the chromosomal marker Ba-1, with phenotypic characteristics of the isolated recorded as detailed in Ochai et al. (2024). An anomalous strain (AX2016-1771Ac) isolated from *Equus quagga* (Zebra) was discovered when it grouped phylogenetically with *B. anthracis* isolates Ochai et al. (2024). Notably, the strain was initially classified as *B. cereus* using classical microbiological tests, presenting as hemolytic, motile, and characterized as Gram-positive rods with short chains. Furthermore, together with other *Bacillus* isolates (Ochai et al., 2024), five *B. cereus* isolates sequenced in this study presented with positive qPCR detection for the anthrax virulence markers *lef* and/or *pagA* and/or chromosomal marker Ba-1 (Supplementary Table S1). The findings warranted further genomic investigations alongside reference atypical strains *B. cereus* and *B. cereus* biovar *anthracis*, which have reportedly caused anthrax-like diseases in humans and animals (Antonation et al., 2016; Baldwin, 2020).

While classical microbiology techniques, such as culturing and microscopy are still essential for bacterial identification, particularly in anthrax endemic regions (WHO, 2008; Ochai et al., 2024), molecular diagnostics often reveal discrepancies. The work of Ochai et al. (2024) showed that *Peribacillus*, *Lactobacillus*, and *Priestia* species were positive for anthrax virulence markers (*pagA*, *lef*, and *cap*). As an example, the PCR amplicons of *pagA* from *Peribacillus*, *Lactobacillus*, and *Priestia* isolates were sequenced and the BLASTn identification did not match with the *pagA* of *B. anthracis* (Ochai et al., 2024). Further analysis that involved genomic analysis and BLASTn comparison of the anthrax virulence markers in the *Priestia* and *B. cereus* group genomes revealed that these genes (*pagA*, *lef*, and *capB*) were only present in the four confirmed *B. anthracis* genomes with 99.9–100% alignment (Magome et al., 2024). In this study, several isolates with *B. cereus*-like phenotypes were positive for the anthrax-specific *lef* gene in qPCR assays but did not indicate *lef* gene in the genome after sequencing. The *lef* gene was found in the following isolates, with 99.9–100% in the four *B. anthracis* isolates, *B. cereus* biovar *anthracis* isolates and the atypical *B. cereus* strain. This is to be expected as anthrax virulence factors have been detected in these strains and were implicated in causing the anthrax disease to their host animal (Antonation et al., 2016; Baldwin, 2020). Previous studies have reported on false positives occurrence for anthrax virulence markers in non-*B. anthracis* strains from blood smears in anthrax endemic region in Kruger National Park (Lekota et al., 2016; Lekota et al., 2018). As thus the risk of misidentification in routine diagnostics, which could potentially compromise the reliability of anthrax detection assays, particularly in *B. cereus* group is increased if diagnosis is depended on molecular marker detection using PCR (Carroll et al., 2022a,b,c). In order to overcome the misidentification of anthrax using anthrax virulence genes on qPCR assay, Ochai et al. (2024) proposed using a combination of the PCR anthrax virulence

markers including the chromosomal Ba-1 marker for increased sensitivity in detecting *B. anthracis* in environmental samples. Other studies have proposed species-specific chromosomal markers found in only in *B. anthracis* that could assist in rapid and accurate detection of *B. anthracis* especially during anthrax outbreaks (Braun et al., 2021; Zorigt et al., 2024).

Genome sequencing remains a gold standard for accurate taxonomic classification, offering high resolution for species delineation (Carroll et al., 2022a,b,c). However, in the *B. cereus* group, genomic ambiguity persists due to the lack of standardized genomospecies thresholds (Carroll et al., 2022a,b,c). The nomenclatural discrepancies observed within the *B. cereus* group will not likely be resolved within the next few years especially when more of these anomalous strains are identified and could be either be misclassified and/or proposed as a new species in the *B. cereus* group based on the database and methodology applied (Gillis et al., 2024). Among the 36 isolates from GenBank classified as *B. anthracis*, 26 clustered with the four confirmed *B. anthracis*, all sharing an ANI threshold of 99%. This suggests that applying a 99% ANI threshold for classifying *B. anthracis* as initially proposed by Jain et al. (2018), could help distinguish typical *B. anthracis* from other group members. Moreover, this was augmented by wgSNP phylogenetic analysis, placing the 26 *B. anthracis* isolates together with the four confirmed South African *B. anthracis* as a cluster.

The remaining 10 anomalous GenBank genomes (Ames TYPE-STRAIN, AFS072084, PFAB2, L19, MCC1A01412, AFS081271, MCC1A02161, N1ZF-2, RIT375, and F34), classified as *B. anthracis*, diverged from typical *B. anthracis* clusters, raising caution in their usage for studies focused solely on typical *B. anthracis*. For instance, the Ames TYPE-STRAIN (GCA_000007845.1) was classified as *B. anthracis* across the genome identification tools (PubMLST and GTDB-TK) but exhibited high ANI with a *B. thuringiensis* isolate and contained insecticidal genes (*cryIII*), a marker typically associated with *B. thuringiensis*. Such findings underscore the challenges of relying solely on ANI or whole-genome SNP analysis for classification within the *B. cereus* group, as even wgSNP analysis did not resolve this discrepancy for the Ames TYPE-STRAIN including the other anomalous *B. cereus* and *B. thuringiensis* genomes.

The use of GTDB and PubMLST classified the sequenced isolates AX2012-121, AX2013-496, AX2014-912, and AX2014-949 from this study as *B. cereus*, belonging to *panC* group IV of the *B. cereus sensu stricto.* These isolates form part of the traditional *B. cereus* based on phenotypic and genomic traits, presented with an ANI of 98%, matching with *B. cereus* ATCC 14579 (GCA_000008725.1). The anomalous isolate AX2016-1771Ac sequenced in this study grouped together with four anomalous *B. anthracis* strains (L19 and MCC1A01412) and *B. cereus* strains (E33L and MOD1_Bc215). Pan-genome analysis identified 14 unique genes, which were solely found on these five isolates. These included genes involved in the physiological functions and processes such as metabolism, cellular homeostasis and transporter cassettes. The anomalous AX2016-1771A isolate contained 146 unique genes, which included ABC transporter permeases, integrase proteins and spore germination proteins (KA, KC, and KB) including aspartyl-phosphate phosphatase Spo0E family proteins known to act as regulator of sporulation which influences development pathways in bacteria such as *B. subtilis* (Perego, 2001).

According to the proposed nomenclature of Carroll et al. (2020a,b), strains that group closely with *B. anthracis* can be classified under the genomospecies *B. mosaicus* (*panC* group III) or simply use the strain number and report the isolate as a member of the *B. cereus* group (Carroll et al., 2022a,b,c). We endorse this classification, although we proposed that the traditional *B. anthracis* species name must be retained by those species that exhibit phenotypic characteristic of *B. anthracis* supported by 99% ANI reference *B. anthracis* with SNP shared by traditional *B. anthracis*. Thus, the anomalous isolate detected in this study based on phenotypic characteristics and genome analysis taking into consideration the impact the classification of species may have on clinical and environmental settings, the isolate AX2016-1771Ac was submitted and classified as *B. cereus* on GenBank. Further description is provided elucidating that the strain belongs to the *panC* group III of the mosaicus group. Recent studies have also reported on the difficulty of classifying novel isolates, they have however relied on placing the isolates in the eight genomospecies groups (I-VIII) as suggested by Carroll et al. (2020a,b). However, species names for novel isolates remain unresolved (Abdelli et al., 2023).

In this study, we also investigated the antimicrobial resistance (AMR) potential of bacterial isolates within the *B. cereus* groups, focusing on the prevalence of specific antibiotic resistance genes (ARGs) that could potentially affect therapeutic efficacy and inform public health monitoring. Monitoring the presence of these ARGs is critical for understanding the distribution of acquired resistance genes and assessing possible clinical interventions for antibiotic-resistant infections (Kompes et al., 2024). Our findings revealed that beta-lactamase genes, specifically *bla1*, *bla2*, and *BcII*, were the most prevalent ARGs among the *B. cereus* group isolates (Figure 4). These genes have been reported to present with a generally low expression in *B. anthracis* attributed to the mutation in regulatory genes such as the *plcR* (Chen et al., 2003; Materon et al., 2003). Among the 30 typical *B. anthracis* isolates studied, a recurring resistance gene profile was observed, characterized by the presence of *BcII*, *Bla1*, *Bla2*, *FosB2*, and *satA* (Figure 4). Notably, *FosB2*, a gene conferring Fosfomycin resistance, was exclusive to the typical *B. anthracis* isolates, distinguishing them from other *B. cereus* group members. The *satA* gene, a streptothricin N-acetyltransferase gene family, associated with nucleoside antibiotic resistance and typically found in actinomycetes in soil-dwelling actinomycetes, suggests a potential environmental acquisition of resistance (Burckhardt and Escalante-Semerena, 2019). Although these resistance determinants are consistent with previous characterizations of *B. anthracis*, further phenotypic testing would provide additional insights into the practical resistance potential of these isolates (Heine et al., 2024). The *B. cereus* showed presence of resistance genes such as beta-lactamases, rifamycin, fosfomycin, and vancomycin in the South African *B. cereus* strains. *Bacillus cereus s.s* are commonly reported in nosocomial infections in clinical settings, proliferation of ARGs among the isolates in the environment could potentially lead to the formation of superbugs which may pose public health risk if ignored (Kowalska et al., 2024).

The presence of virulence factors further emphasizes the pathogenic potential of these isolates. While typical anthrax-toxin genes (*cya*, *lef*, *pagA*) on the pXO1 plasmid (Turnbull, 2002; Liu et al., 2015; Carroll et al., 2020a,b) are absent in the anomalous strains, the anthrax-toxin genes and associated regulatory elements, i.e., *atxA*, were detected in the typical *B. anthracis* (AX2015-1152, AX2015-1270, and AX2015-1277A), as well as in atypical *B. cereus* BC-AK isolate and *B. cereus* biovar *anthracis* strains (UFBc0001, CAR, CAM,

and CI), which are implicated in anthrax related diseases in humans and animals (Ochai et al., 2024; Baldwin, 2020). The presence of the pXO1/pBCXO1 and pXO2/pBCXO1 plasmids suggests a shared virulence mechanism across these isolates. Interestingly, one sequenced *B. anthracis* strain, AX2015-1136, lacked the capsular genes (*cap-ABCDE*, *acpA*, *acpB* genes) despite carrying the tripartite anthrax toxin genes, indicative of the pXO1 plasmid's retention but pXO2 plasmid loss, conversely anomalous isolates deltaSterne, A1055, 2000031021, 2000013052, A46, Smith1013, and Pasteur contained capsular genes but lacked the tripartite anthrax toxin genes, indicative of the pXO2 plasmid retention and loss of pXO1 plasmid. This phenomenon is reportedly likely due to genetic instability from environmental pressures (Liang et al., 2016).

In atypical *B. cereus* strains, the functional expression of both PGA and hyaluronic acid capsules encoded by *has-ABC* located in the pBCXO1 plasmid, facilitates pathogenicity, with the hyaluronic capsule linked to cause anthrax-like disease (Antonation et al., 2016; Baldwin, 2020). This dual-capacity for capsule expression is unique to the atypical *B. cereus* biovar anthracis and is absent in *B. anthracis*, where capsule synthesis is disrupted by mutations within *hasA* (Okinaka et al., 1999). Therefore, the *B. anthracis* isolates only express the PGA capsule, whereas atypical *B. cereus* isolates and *B. cereus* biovar *anthracis* may express both a hyaluronic acid capsule and PGA capsule required for pathogenicity (Oh et al., 2011).

The diversity in the *Bacillus cereus* group's virulence and resistance factors is further demonstrated by the detection of the hydrolase *cesH* gene, linked to the emetic cereulide operon (*ces-ABCDPTH*), detected in six anomalous *B. cereus* strains (3a, 95_8201, BA087, BGSC6E1, PE8-121b, and S2-8). Although it is associated with a specific group of emetic *B. cereus* strains, non-emetic isolates have been found to carry *cesH* flanking genes which reportedly are closely related to the anthrax-toxin encoding pXO1 plasmid (Ehling-Schulz et al., 2006). The presence of toxin-related gene clusters (*nhe-ABC*, *hbl-ACD*, and *cytK*), which are the hemolysin BL, non-hemolytic enterotoxin, and cytotoxin K, respectively, found in sequenced *B. cereus* isolates AX2012-121, AX2013-496, AX2014-912, and AX2014-949, signals the potential for diarrheal illness, consistent with prior findings associating these clusters with enteric disease (Schoeni and Wong, 2005; Owusu-Kwarteng et al., 2017). The genomic plasticity within these isolates underscored by the presence of insertion sequences and mobile elements, may facilitate AMR and virulence gene transfer, impacting the adaptability and pathogenic potential of *Bacillus* species (Fayad et al., 2019; Bianco et al., 2023). Recent studies have also identified insertions contributing to clarithromycin resistance, underlining the evolutionary adaptability of *B. anthracis* and related species (Maxson et al., 2024).

5 Conclusion

In this study, *B. cereus* group isolates that were isolated from blood smear samples in the Kruger National Park presenting with amplification of anthrax-toxin markers during molecular analysis in Ochai et al. (2024) were investigated and resolved using genome-sequence analysis. Following identification of other anomalous strains, 26 out of 36 *B. anthracis* strains may be accepted as part of traditional *B. anthracis* strains. The above suggestion is based on the comparative genomics analysis applied which involved pan-genome, ANI (≥99%) and wgSNP analysis. The anomalous AX2016-17771Ac isolate from this study showed no SNP related to traditional *B. anthracis* and anthrax virulence genes, therefore the isolate was classified as *B. cereus panC* group III of the mosaicus group. The four newly sequenced isolates AX2012-121, AX2013-496, AX2014-912, and AX2014-949 are classified as *B. cereus sensu stricto* of *panC* group IV. Although we were able to distinguish traditional *B. anthracis* from other closely related strains, the relation between *B. thuringiensis* and *B. cereus* remains to be resolved especially those that group in the mosaicus clade. This study also calls into the reassessment of the *B. cereus* taxonomy committee which oversees the guidelines and monitoring of the submission or classification of species in the *B. cereus* group. Most of the anomalous isolates may have been assigned their species names due to the lack of advanced genomic tools, and/or the database available during analysis which contributed to the naming of the isolates within the *B. cereus* group. Our study further highlights the complex interplay of ARGs, virulence factors, and plasmid mobility within *B. anthracis* and other *B. cereus* group species. This genomic diversity necessitates ongoing surveillance and phenotypic assessment to better understand and mitigate the clinical risks posed by potential pathogens in this *B. cereus* group. The insights provided herein contribute to a growing body of evidence on *Bacillus*-related antimicrobial resistance, with potential implications for both therapeutic approaches and public health policies.

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 animal study was approved by Faculty of Natural and Agricultural Sciences Ethics Committee. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

TM: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. MS: Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – review & editing. AH: Conceptualization, Data curation, Investigation, Methodology, Resources, Supervision, Writing – review & editing. CB: Project administration, Resources, Supervision, Writing – review & editing, Validation. HH: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing. KL: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration,

Resources, Software, Supervision, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work is based on research supported by the National Research Foundation (NRF) of South Africa, Grant UID129525 awarded to K. E. Lekota. Views expressed are those of the authors and not of the funders.

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.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2025.1527049/full#supplementary-material

References

Abdelli, M., Falaise, C., Morineaux-Hilaire, V., Cumont, A., Taysse, L., Raynaud, F., et al. (2023). Get to know your neighbors: characterization of close *Bacillus anthracis* isolates and toxin profile diversity in the *Bacillus cereus* group. *Microorganisms* 11:2721. doi: 10.3390/microorganisms11112721

Afshinnekoo, E., Meydan, C., Chowdhury, S., Jaroudi, D., Boyer, C., Bernstein, N., et al. (2015). Geospatial resolution of human and bacterial diversity with city-scale metagenomics. *Cell Syst.* 1, 72–87. doi: 10.1016/j.cels.2015.01.001

Alcock, B. P., Huynh, W., Chalil, R., Smith, K. W., Raphenya, A. R., Wlodarski, M. A., et al. (2023). CARD 2023: expanded curation, support for machine learning, and resistome prediction at the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 51, D690–D699. doi: 10.1093/nar/gkac920

Antonation, K. S., Grützmacher, K., Dupke, S., Mabon, P., Zimmermann, F., Lankester, F., et al. (2016). *Bacillus cereus* biovar *anthracis* causing anthrax in sub-Saharan Africa—chromosomal monophyly and broad geographic distribution. *PLoS Negl. Trop. Dis.* 10:e0004923. doi: 10.1371/journal.pntd.0004923

Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., et al. (2018). KBase: the United States department of energy systems biology knowledgebase. *Nature biotechnology*, 36, 566–569. doi: 10.1038/nbt.4163

Baldwin, V. M. (2020). You can't *B. cereus*–a review of *Bacillus cereus* strains that cause anthrax-like disease. *Front. Microbiol.* 11:1731. doi: 10.3389/fmicb.2020.01731

Bazinet, A. L. (2017). Pan-genome and phylogeny of *Bacillus cereus sensu lato*. *BMC Evol. Biol.* 17:176. doi: 10.1186/s12862-017-1020-1

Bianco, A., Normanno, G., Capozzi, L., Del Sambro, L., Di Fato, L., Miccolupo, A., et al. (2023). High genetic diversity and virulence potential in *Bacillus cereus sensu lato* isolated from Milk and cheeses in Apulia region, Southern Italy. *Foods* 12:1548. doi: 10.3390/foods12071548

Bogaerts, B., Fraiture, M. A., Huwaert, A., Van Nieuwenhuysen, T., Jacobs, B., Van Hoorde, K., et al. (2023). Retrospective surveillance of viable *Bacillus cereus* group contaminations in commercial food and feed vitamin B2 products sold on the Belgian market using whole-genome sequencing. *Front. Microbiol.* 14:1173594. doi: 10.3389/fmicb.2023.1173594

Bottone, E. J. (2010). *Bacillus cereus*, a volatile human pathogen. *Clin. Microbiol. Rev.* 23, 382–398. doi: 10.1128/CMR.00073-09

Braun, P., Nguyen, M. D. T., Walter, M. C., and Grass, G. (2021). Ultrasensitive detection of *Bacillus anthracis* by real-time PCR targeting a polymorphism in multi-copy 16S rRNA genes and their transcripts. *Int. J. Mol. Sci.* 22:12224. doi: 10.3390/ijms222212224

Brézillon, C., Haustant, M., Dupke, S., Corre, J. P., Lander, A., Franz, T., et al. (2015). Capsules, toxins and *AtxA* as virulence factors of emerging *Bacillus cereus* biovar *anthracis*. *PLoS Negl. Trop. Dis.* 9:e0003455. doi: 10.1371/journal.pntd.0003455

Buchfink, B., Ashkenazy, H., Reuter, K., Kennedy, J. A., and Drost, H. G. (2023). Sensitive clustering of protein sequences at tree-of-life scale using DIAMOND DeepClust. *bioRxiv*. doi: 10.1101/2023.01.24.525373

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

Burckhardt, R. M., and Escalante-Semerena, J. C. (2019). Insights into the function of the N-acetyltransferase *SatA* that detoxifies streptothricin in *Bacillus subtilis* and *Bacillus anthracis*. *Appl. Environ. Microbiol.* 85, e03029–e03018. doi: 10.1128/AEM.03029-18

Carroll, L. M., Cheng, R. A., and Kovac, J. (2020b). No assembly required: using BTyper3 to assess the congruency of a proposed taxonomic framework for the *Bacillus cereus* group with historical typing methods. *Front. Microbiol.* 11:580691. doi: 10.3389/fmicb.2020.580691

Carroll, L. M., Cheng, R. A., Wiedmann, M., and Kovac, J. (2022a). Keeping up with the *Bacillus cereus* group: taxonomy through the genomics era and beyond. *Crit. Rev. Food Sci. Nutr.* 62, 7677–7702. doi: 10.1080/10408398.2021.1916735

Carroll, L. M., Matle, I., Kovac, J., Cheng, R. A., and Wiedmann, M. (2022b). Laboratory misidentifications resulting from taxonomic changes to *Bacillus cereus* group species, 2018–2022. *Emerg. Infect. Dis.* 28, 1877–1881. doi: 10.3201/eid2809.220293

Carroll, L. M., Pierneef, R., Mathole, A., Atanda, A., and Matle, I. (2022c). Genomic sequencing of *Bacillus cereus Sensu Lato* strains isolated from meat and poultry products in South Africa enables inter-and intranational surveillance and source tracking. *Microbiol. Spectr.* 10, e00700–e00722. doi: 10.1128/spectrum.00700-22

Carroll, L. M., Wiedmann, M., and Kovac, J. (2020a). Proposal of a taxonomic nomenclature for the *Bacillus cereus* group which reconciles genomic definitions of bacterial species with clinical and industrial phenotypes. *MBio* 11, 10–1128. doi: 10.1128/mBio.00034-20

Chen, Y., Succi, J., Tenover, F. C., and Koehler, T. M. (2003). β-Lactamase genes of the penicillin-susceptible *Bacillus anthracis* Sterne strain. *J. Bacteriol.* 185, 823–830. doi: 10.1128/JB.185.3.823-830.2003

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

Ehling-Schulz, M., Fricker, M., Grallert, H., Rieck, P., Wagner, M., and Scherer, S. (2006). Cereulide synthetase gene cluster from emetic *Bacillus cereus*: structure and location on a mega virulence plasmid related to *Bacillus anthracis* toxin plasmid pXO1. *BMC Microbiol.* 6, 1–11. doi: 10.1186/1471-2180-6-20

Ehling-Schulz, M., Lereclus, D., and Koehler, T. M. (2019). The *Bacillus cereus* group: Bacillus species with pathogenic potential. *Microbiol. Spectr.* 7, 10–1128. doi: 10.1128/microbiolspec.GPP3-0032-2018

Fagerlund, A., Ween, O., Lund, T., Hardy, S. P., and Granum, P. E. (2004). Genetic and functional analysis of the *cytK* family of genes in *Bacillus cereus*. *Microbiology* 150, 2689–2697. doi: 10.1099/mic.0.26975-0

Fayad, N., Kallassy Awad, M., and Mahillon, J. (2019). Diversity of *Bacillus cereus sensu lato* mobilome. *BMC Genomics* 20, 1–11. doi: 10.1186/s12864-019-5764-4

Feldbauer, R., Gosch, L., Lüftinger, L., Hyden, P., Flexer, A., and Rattei, T. (2020). DeepNOG: fast and accurate protein orthologous group assignment. *Bioinformatics* 36, 5304–5312. doi: 10.1093/bioinformatics/btaa1051

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, 10–1128. doi: 10.1128/AAC.00483-19

Gillis, A., Hendriksen, N. B., and Økstad, O. A. (2024). "Towards a consensus taxonomy of the *Bacillus cereus* group of bacteria in the sequencing era" in 2024 international congress on invertebrate pathology and microbial control (SIP 2024).

Girault, G., Blouin, Y., Vergnaud, G., and Derzelle, S. (2014). High-throughput sequencing of *Bacillus anthracis* in France: investigating genome diversity and population structure using whole-genome SNP discovery. *BMC Genomics* 15, 1–10. doi: 10.1186/1471-2164-15-288

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

Guardabassi, L., Perichon, B., van Heijenoort, J., Blanot, D., and Courvalin, P. (2005). Glycopeptide resistance *vanA* operons in *Paenibacillus* strains isolated from soil. *Antimicrob. Agents Chemother.* 49, 4227–4233. doi: 10.1128/AAC.49.10.4227-4233.2005

Guinebretière, M. H., Thompson, F. L., Sorokin, A., Normand, P., Dawyndt, P., Ehling-Schulz, M., et al. (2008). Ecological diversification in the *Bacillus cereus* group. *Environ. Microbiol.* 10, 851–865. doi: 10.1111/j.1462-2920.2007.01495.x

Heine, H. S., Drusano, G., Purcell, B. K., Anastasiou, D., Tanaka, S. K., and Serio, A. W. (2024). Omadacycline is active *in vitro* and *in vivo* against ciprofloxacin-resistant *Bacillus anthracis*. *Antimicrob. Agents Chemother.* 68:e0059524. doi: 10.1128/aac.00595-24

Hua, X., Liang, Q., Deng, M., He, J., Wang, M., Hong, W., et al. (2021). BacAnt: a combination annotation server for bacterial DNA sequences to identify antibiotic resistance genes, integrons, and transposable elements. *Front. Microbiol.* 12:649969. doi: 10.3389/fmicb.2021.649969

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

Johansson, M. H., 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

Jolley, K. A., Bray, J. E., and Maiden, M. C. (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

Kamar, R., Gohar, M., Jéhanno, I., Réjasse, A., Kallassy, M., Lereclus, D., et al. (2013). Pathogenic potential of *Bacillus cereus* strains as revealed by phenotypic analysis. *J. Clin. Microbiol.* 51, 320–323. doi: 10.1128/JCM.02848-12

Kim, Y., Koh, I., Young Lim, M., Chung, W. H., and Rho, M. (2017). Pan-genome analysis of Bacillus for microbiome profiling. *Sci. Rep.* 7:10984. doi: 10.1038/s41598-017-11385-9

Kolstø, A. B., Tourasse, N. J., and Økstad, O. A. (2009). What sets *Bacillus anthracis* apart from other Bacillus species? *Ann. Rev. Microbiol.* 63, 451–476. doi: 10.1146/annurev.micro.091208.073255

Kompes, G., Duvnjak, S., Reil, I., Mihaljević, Ž., Habrun, B., Benić, M., et al. (2024). Antimicrobial resistance profile, whole-genome sequencing and Core genome multilocus sequence typing of *B. anthracis* isolates in Croatia from 2001 to 2022. *Antibiotics* 13:639. doi: 10.3390/antibiotics13070639

Kotiranta, A., Lounatmaa, K., and Haapasalo, M. (2000). Epidemiology and pathogenesis of *Bacillus cereus* infections. *Microbes Infect.* 2, 189–198. doi: 10.1016/S1286-4579(00)00269-0

Kowalska, J., Maćkiw, E., Korsak, D., and Postupolski, J. (2024). Characterization of the *Bacillus cereus* group isolated from ready-to-eat foods in Poland by whole-genome sequencing. *Food Secur.* 13:3266. doi: 10.3390/foods13203266

Lekota, K. E., Bezuidt, O. K. I., Mafofo, J., Rees, J., Muchadeyi, F. C., Madoroba, E., et al. (2018). Whole genome sequencing and identification of *Bacillus endophyticus* and *B. anthracis* isolated from anthrax outbreaks in South Africa. *BMC Microbiol.* 18, 1–15. doi: 10.1186/s12866-018-1205-9

Lekota, K. E., Hassim, A., Ledwaba, M. B., Glover, B. A., Dekker, E. H., van Schalkwyk, L. O., et al. (2024). *Bacillus anthracis* in South Africa, 1975–2013: are some lineages vanishing? *BMC Genomics* 25:742. doi: 10.1186/s12864-024-10631-5

Lekota, K. E., Hassim, A., Mafofo, J., Rees, J., Muchadeyi, F. C., Van Heerden, H., et al. (2016). Polyphasic characterization of *Bacillus* species from anthrax outbreaks in animals from South Africa and Lesotho. *J. Infect. Dev. Ctries.* 10, 814–823. doi: 10.3855/jidc.7798

Liang, X., Zhang, H., Zhang, E., Wei, J., Li, W., Wang, B., et al. (2016). Identification of the pXO1 plasmid in attenuated *Bacillus anthracis* vaccine strains. *Virulence* 7, 578–586. doi: 10.1080/21505594.2016.1164366

Liu, Y., Lai, Q., Göker, M., Meier-Kolthoff, J. P., Wang, M., Sun, Y., et al. (2015). Genomic insights into the taxonomic status of the *Bacillus cereus* group. *Sci. Rep.* 5:14082. doi: 10.1038/srep14082

Liu, Y., Lai, Q., and Shao, Z. (2018). Genome analysis-based reclassification of *Bacillus weihenstephanensis* as a later heterotypic synonym of *Bacillus mycoides*. *Int. J. Syst. Evol. Microbiol.* 68, 106–112. doi: 10.1099/ijsem.0.002466

Liu, M., Li, X., Xie, Y., Bi, D., Sun, J., Li, J., et al. (2019). ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res.* 47, D660–D665. doi: 10.1093/nar/gky1123

Liu, D., Zhang, Y., Fan, G., Sun, D., Zhang, X., Yu, Z., et al. (2022). IPGA: a handy integrated prokaryotes genome and pan-genome analysis web service. *iMeta* 1:e55. doi: 10.1002/imt2.55

Liu, B., Zheng, D., Jin, Q., Chen, L., and Yang, J. (2019). VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res.* 47, D687–D692. doi: 10.1093/nar/gky1080

Magome, T. G., Ochai, S. O., Hassim, A., Bezuidenhout, C. C., van Heerden, H., and Lekota, K. E. (2024). A genome-based investigation of the *Priestia* species isolated from anthrax endemic regions in Kruger national Park. *Infect. Genet. Evol.* 123:105649. doi: 10.1016/j.meegid.2024.105649

Marston, C. K., Gee, J. E., Popovic, T., and Hoffmaster, A. R. (2006). Molecular approaches to identify and differentiate *Bacillus anthracis* from phenotypically similar Bacillus species isolates. *BMC Microbiol.* 6:22. doi: 10.1186/1471-2180-6-22

Marston, C. K., Hoffmaster, A. R., Wilson, K. E., Bragg, S. L., Plikaytis, B., Brachman, P., et al. (2005). Effects of long-term storage on plasmid stability in *Bacillus anthracis*. *Appl. Environ. Microbiol.* 71, 7778–7780. doi: 10.1128/AEM.71.12.7778-7780.2005

Materon, I. C., Queenan, A. M., Koehler, T. M., Bush, K., and Palzkill, T. (2003). Biochemical characterization of β-lactamases *Bla1* and *Bla2* from *Bacillus anthracis*. *Antimicrob. Agents Chemother.* 47, 2040–2042. doi: 10.1128/AAC.47.6.2040-2042.2003

Maxson, T., Overholt, W. A., Chivukula, V., Caban-Figueroa, V., Kongphet-Tran, T., Medina Cordoba, L. K., et al. (2024). Genetic basis of clarithromycin resistance in *Bacillus anthracis*. *Microbiol. Spectr.* 12, e04180–e04123. doi: 10.1128/spectrum.04180-23

Mogaji, H. O., Adewale, B., Smith, S. I., Igumbor, E. U., Idemili, C. J., and Taylor-Robinson, A. W. (2024). Combatting anthrax outbreaks across Nigeria's national land borders: need to optimize surveillance with epidemiological surveys. *Infect. Dis. Poverty* 13:10. doi: 10.1186/s40249-024-01179-3

Ochai, S. O., Hassim, A., Dekker, E. H., Magome, T., Lekota, K. E., Makgabo, M. S., et al. (2024). Comparing microbiological and molecular diagnostic tools for the surveillance of anthrax. *PLoS Negl. Trop. Dis.* 18:e0012122. doi: 10.1371/journal.pntd.0012122

Oh, S. Y., Budzik, J. M., Garufi, G., and Schneewind, O. (2011). Two capsular polysaccharides enable *Bacillus cereus* G9241 to cause anthrax-like disease. *Mol. Microbiol.* 80, 455–470. doi: 10.1111/j.1365-2958.2011.07582.x

Okinaka, R., Cloud, K., Hampton, O., Hoffmaster, A., Hill, K., Keim, P., et al. (1999). Sequence, assembly and analysis of pX01 and pX02. *J. Appl. Microbiol.* 87, 261–262. doi: 10.1046/j.1365-2672.1999.00883.x

Okinaka, R. T., and Keim, P. (2016). "The phylogeny of *Bacillus cereus sensu lato*" in The bacterial spore: from molecules to systems, 237–251.

Owusu-Kwarteng, J., Wuni, A., Akabanda, F., Tano-Debrah, K., and Jespersen, L. (2017). Prevalence, virulence factor genes and antibiotic resistance of *Bacillus cereus sensu lato* isolated from dairy farms and traditional dairy products. *BMC Microbiol.* 17, 1–8. doi: 10.1186/s12866-017-0975-9

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., et al. (2015). Roary: rapid large-scale prokaryote pan-genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421

Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E., and Larsson, D. J. (2014). BacMet: antibacterial biocide and metal resistance genes database. *Nucleic Acids Res.* 42, D737–D743. doi: 10.1093/nar/gkt1252

Pearson, T., Busch, J. D., Ravel, J., Read, T. D., Rhoton, S. D., U'ren, J. M., et al. (2004). Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. *Proc. Natl. Acad. Sci. USA* 101, 13536–13541. doi: 10.1073/pnas.0403844101

Pena-Gonzalez, A., Rodriguez-R, L. M., Marston, C. K., Gee, J. E., Gulvik, C. A., Kolton, C. B., et al. (2018). Genomic characterization and copy number variation of *Bacillus anthracis* plasmids pXO1 and pXO2 in a historical collection of 412 strains. *Msystems* 3, 10–1128. doi: 10.1128/msystems.00065-18

Perego, M. (2001). A new family of aspartyl phosphate phosphatases targeting the sporulation transcription factor Spo0A of *Bacillus subtilis*. *Mol. Microbiol.* 42, 133–143. doi: 10.1046/j.1365-2958.2001.02611.x

Rasko, D. A., Altherr, M. R., Han, C. S., and Ravel, J. (2005). Genomics of the *Bacillus cereus* group of organisms. *FEMS Microbiol. Rev.* 29, 303–329. doi: 10.1016/j.fmrre.2004.12.005

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277. doi: 10.1016/S0168-9525(00)02024-2

Sacchi, C. T., Whitney, A. M., Mayer, L. W., Morey, R., Steigerwalt, A., Boras, A., et al. (2002). Sequencing of 16S rRNA gene: a rapid tool for identification of *Bacillus anthracis*. *Emerg. Infect. Dis.* 8, 1117–1123. doi: 10.3201/eid0810.020391

Schoeni, J. L., and Wong, A. C. L. (2005). *Bacillus cereus* food poisoning and its toxins. *J. Food Prot.* 68, 636–648. doi: 10.4315/0362-028X-68.3.636

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

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

Seemann, T. (2015), Snippy: fast bacterial variant calling from NGS reads. Available at: https://github.com/tseemann/snippy

Seemann, T., Edwards, R., Goncalves da Silva, A., and Kiil, K. (2020). Shovill v1.1.0. Available at: https://github.com/tseemann/shovill

Shimoyama, Y. (2022). COG classifier: a tool for classifying prokaryote protein sequences into COG functional category. Computer Software. Available at: https://github.com/moshi4/COGclassifier.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312–1313.

Turnbull, P. C. B. (2002). Introduction: anthrax history, disease and ecology. *Curr Top Microbiol Immunol* 271, 1–19. doi: 10.1007/978-3-662-05767-4_1

Vilas-Bôas, G. T., Peruca, A. P. S., and Arantes, O. M. N. (2007). Biology and taxonomy of *Bacillus cereus*, *Bacillus anthracis*, and *Bacillus thuringiensis*. *Can. J. Microbiol.* 53, 673–687. doi: 10.1139/W07-029

WHO (2008). Anthrax in humans and animals. Geneva: World Health Organization.

Xie, J., Chen, Y., Cai, G., Cai, R., Hu, Z., and Wang, H. (2023). Tree visualization by one Table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees. *Nucleic Acids Res.* 51, W587–W592. doi: 10.1093/nar/gkad359

Yoon, S. H., Park, Y. K., and Kim, J. F. (2015). PAIDB v2. 0: exploration and analysis of pathogenicity and resistance islands. *Nucleic Acids Res.* 43, D624–D630. doi: 10.1093/nar/gku985

Zincke, D., Norris, M. H., Cruz, O., Kurmanov, B., McGraw, W. S., Daegling, D. J., et al. (2020). TaqMan assays for simultaneous detection of *Bacillus anthracis* and *Bacillus cereus* biovar *anthracis*. *Pathogens* 9:1074. doi: 10.3390/pathogens9121074

Zorigt, T., Furuta, Y., Paudel, A., Kamboyi, H. K., Shawa, M., Chuluun, M., et al. (2024). Pan-genome analysis reveals novel chromosomal markers for multiplex PCR-based specific detection of *Bacillus anthracis*. *BMC Infect. Dis.* 24:942. doi: 10.1186/s12879-024-09817-9

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 10 February 2025
DOI 10.3389/fmicb.2025.1548920

OPEN ACCESS

EDITED BY
Xiangyu Fan,
University of Jinan, China

REVIEWED BY
Agnieszka Kwiatek,
University of Warsaw, Poland
Yuhang Wang,
The University of Iowa, United States

*CORRESPONDENCE
Tonglei Wu
✉ 532966952@163.com

RECEIVED 20 December 2024
ACCEPTED 27 January 2025
PUBLISHED 10 February 2025

CITATION
Zhang L, Chen L, Zhang X, Li Y, Zheng Q, Li Y, Li N, Shi Q, Zhang Y and Wu T (2025) The *mcpC* mutant of *Salmonella enteritidis* exhibits attenuation and confers both immunogenicity and protective efficacy in mice.
Front. Microbiol. 16:1548920.
doi: 10.3389/fmicb.2025.1548920

The *mcpC* mutant of *Salmonella enteritidis* exhibits attenuation and confers both immunogenicity and protective efficacy in mice

Lu Zhang[1], Li Chen[1], Xuqiang Zhang[1], Yang Li[1], Qingfeng Zheng[2], Yun Li[3], Ning Li[4], Qiumei Shi[1], Yanying Zhang[1] and Tonglei Wu[1]*

[1]Hebei Key Laboratory of Preventive Veterinary Medicine, College of Animal Science and Technology, Hebei Normal University of Science and Technology, Qinghuangdao, China, [2]Qinhuangdao Animal Husbandry Station, Qinghuangdao, China, [3]Shijiazhuang Animal Products and Veterinary Drug Feed Quality Testing Center, Shijiazhuang, China, [4]Tangshan Academy of Agricultural Sciences, Tangshan, China

Background: *Salmonella enteritidis* (*SE*) is a Gram-negative, facultative anaerobic intracellular pathogen that not only causes disease and mortality in livestock and poultry but also contaminates animal-derived products, leading to foodborne illnesses in humans. This presents a significant threat to public health. To eliminate this pathogen, the development of novel vaccines targeting *SE* is imperative. Attenuated live vaccines are capable of eliciting robust immune protection against *SE*.

Methods: In this study, an *mcpC* gene deletion strain (Δ*mcpC*) was constructed by the wild strain C50336, to evaluate its potential as a genetically engineered attenuated live vaccine. The virulence of Δ*mcpC* was assessed by examining its resistance to environmental stresses, biofilm formation capacity, motility, adhesion, invasion ability, intracellular survival, LD_{50}, expression levels of virulence genes, and *in vivo* colonization ability. Furthermore, the immunogenicity of Δ*mcpC* was analyzed in mice by measuring specific IgG and SIgA antibody levels, lymphocyte proliferation, cytokine expression, and the protective efficacy of Δ*mcpC* vaccination.

Results: Compared to the wild-type strain, Δ*mcpC* exhibited no significant changes in biofilm formation or adhesion to Caco-2 cells. However, Δ*mcpC* showed significantly reduced survival under acidic, alkaline, thermal, and oxidative stress conditions; markedly diminished motility; weakened invasion of Caco-2 cells; and reduced intracellular survival in RAW264.7 macrophages. The LD_{50} of Δ*mcpC* increased by 30-fold, and the expression levels of certain virulence genes were significantly downregulated. Additionally, Δ*mcpC* demonstrated significantly decreased colonization in the liver, spleen, and cecum of mice, indicating attenuated virulence. Immunization with Δ*mcpC* induced the production of specific IgG and SIgA antibodies, enhanced lymphocyte proliferation, upregulated cytokine expression, and achieved a 100% survival rate in immunized mice. These findings indicate that Δ*mcpC* provides effective immune protection in mice.

Conclusion: This study demonstrates that deletion of the *mcpC* gene attenuates the virulence of *SE*. The Δ*mcpC* offers strong immune protection in mice, providing a solid foundation for the development of genetically engineered attenuated live vaccines against *SE*.

KEYWORDS

Salmonella enteritidis, *mcpC*, gene deletion, virulence, vaccine

Background

Salmonella enteritidis (*SE*) is a significant zoonotic pathogen that not only inflicts substantial economic losses on the livestock industry but has also emerged as one of the predominant serotypes causing salmonellosis in humans in recent years (Zhou et al., 2023; Chousalkar et al., 2018). While strict biosecurity measures can create *Salmonella*-free poultry farming environments, such measures are often economically unfeasible in many countries. Moreover, the widespread emergence of multidrug-resistant strains has raised concerns about the use of antibiotics. Vaccination remains a primary strategy for controlling *Salmonella* infections. Currently, inactivated vaccines, subunit vaccines, and live attenuated vaccines represent the major categories of *Salmonella* vaccines under development (Huberman et al., 2019).

Studies indicate that inactivated vaccines primarily induce humoral immunity and fail to elicit robust cell-mediated immune responses. As a result, booster immunizations are often required to achieve long-term protection (Deguchi et al., 2009). Subunit vaccines, while safe, typically require multiple doses to extend the duration of immunity and necessitate suitable adjuvants to trigger effective cellular immune responses. In contrast, live attenuated vaccines, when administered orally, adhere to the intestinal mucosa and mimic natural infection, thereby eliciting strong humoral and cell-mediated immune responses (Van Immerseel et al., 2005). The identification of virulence genes is crucial for developing gene-deletion-based live vaccines. This approach has been successfully employed in the development of various vaccines, including those for *Brucella* spp., *Salmonella* Typhimurium, *Yersinia pestis*, *Vibrio anguillarum*, and *Edwardsiella tarda* (Yang et al., 2015; Liu et al., 2018; Park et al., 2020; Cote et al., 2021; Zabalza-Baranguá et al., 2023).

SE must overcome multiple defense mechanisms inside and outside the gastrointestinal tract to establish infection in the host. Upon entering the digestive system, the first challenge it encounters is the bactericidal effect of gastric acid. Gastric acid, being a strongly acidic environment, inhibits or kills most microorganisms that reach the stomach. To survive, *SE* expresses high levels of acid shock proteins, such as ATPase, which help maintain intracellular pH balance and enhance its resistance to gastric acid. In the intestine, *SE* defends itself against bile and digestive enzymes by regulating its lipopolysaccharide (LPS) structure and producing anti-bile proteins, such as TolC. In the small intestine, it utilizes a type III secretion system (T3SS) to secrete effector molecules, including SipA, SipB, and SipC, which facilitate its invasion of intestinal epithelial cells. Within host cells, *Salmonella* forms a specialized membrane-bound compartment called the *Salmonella*-containing vacuole (SCV), which protects it from lysosomal degradation. After being phagocytosed by macrophages, *SE* injects T3SS effectors like SifA and SipC to inhibit the fusion of SCVs with lysosomes. Additionally, *Salmonella* can evade autophagy-mediated clearance by secreting effectors such as SseJ, allowing it to persist and proliferate within the SCV. In summary, the infection process of *SE* involves adapting to gastric acid, resisting bile, invading intestinal epithelium, evading the immune system, surviving intracellularly, and inducing inflammation. Each step relies on specific virulence factors and adaptive mechanisms, enabling *Salmonella* to establish infection in the gastrointestinal tract and beyond, leading to host disease. The loss of these virulence factors significantly attenuates the pathogen's virulence. Consequently, researchers have utilized these virulence factors to develop various gene-deletion vaccines, achieving promising immune protection results.

Methyl-accepting chemotaxis protein C (McpC) is a methylation-based chemotaxis protein involved in bacterial chemotaxis, enabling directional movement in response to chemical gradients. By sensing and transmitting environmental chemical signals, McpC helps bacteria regulate their motility. This chemotactic ability is crucial for pathogenic bacteria to colonize and spread within the host, thereby playing a significant role in virulence. Additionally, McpC is involved in regulating biofilm formation, flagellar biosynthesis, and toxin production (Hickman et al., 2005; Berleman and Bauer, 2005; Harkey et al., 1994). As such, modulating McpC expression or function in pathogens may impact their infection efficiency and pathogenicity, making McpC a key target for studying bacterial virulence and infection mechanisms. Previous studies have demonstrated the role of methyl-accepting chemotaxis proteins (MCPs) in the pathogenicity of various bacteria, including *Pseudomonas aeruginosa*, *Campylobacter jejuni*, and *Vibrio cholerae* (Li et al., 2014; Sampedro et al., 2014; Nishiyama et al., 2016). Based on these findings, it is hypothesized that McpC contributes to the virulence of *SE*, a hypothesis that requires further experimental validation. By precisely deleting the *mcpC* gene, the pathogen's virulence is significantly reduced while retaining its immunogenicity, thereby greatly enhancing safety and efficacy. Moreover, the vaccine strain produced by deleting the presumed virulence-associated gene *mcpC* can induce robust mucosal, cellular, and humoral immune responses, making it suitable for non-injection administration methods such as oral or spray vaccination, which facilitates large-scale application.

In this study, we constructed an *mcpC* gene deletion strain (Δ*mcpC*) using homologous recombination techniques and investigated its impact on *SE* virulence. The evaluation included analyses of stress resistance, biofilm formation, motility, adhesion, invasion, intracellular survival in macrophages, LD_{50}, expression levels of virulence genes, and bacterial load in host organs. Additionally, we assessed the ability of Δ*mcpC* to induce immune responses and its protective efficacy in mice using an infection model. The findings provide a foundational basis for the development of genetically engineered live-attenuated vaccines targeting *SE*.

Materials and methods

Bacterial strains, cells and plasmids

Bacterial strains and plasmids used in this study are listed in Table 1. *Salmonella enteritidis* strain C50336, a wild-type strain, was preserved in the Key Laboratory of Preventive Veterinary Medicine, Hebei Province, and used for constructing the Δ*mcpC* strain. The *mcpC* gene deletion strain in this study was constructed using the

TABLE 1 Bacterial strains and plasmids used in this study.

Strains	Relevant characteristics	Sources
C50336	*Salmonella enterica* serovar *Enteritidis*, wild-type	This study
Δ*mcpC*::*cat*	A first recombination strain	This study
Δ*mcpC*	A second recombination strain	This study
Δ*mcpC* + *mcpC*	Δ*mcpC*-complemented strain	This study

Plasmids	Characteristics	Sources
pKD3	Template plasmid; FRT-*aphT*-FRT (containing chloramphenicol resistance gene)	Datsenko and Wanner (2000)
pKD46	Red recombinase expression plasmid blapBAD gam bet exopSC101 oriTS (containing ampicillin resistance gene)	Datsenko and Wanner (2000)
pCP20	FLP recombinase expression plasmid	Datsenko and Wanner (2000)

λ-Red recombinase-mediated gene replacement method (Datsenko and Wanner, 2000). The bacteria were cultured in Luria-Bertani (LB) broth (Haibo Biotechnology Co., Ltd.) at 37°C unless otherwise specified. Human epithelial Caco-2 BBE cells and mouse macrophage RAW264.7 cells used in this study were provided by BeNa Culture Collection (Shanghai, China). Both cell types were cultured in DMEM (Thermo Fisher Scientific Co., Ltd.) supplemented with 10% fetal bovine serum (Thermo Fisher Scientific Co., Ltd.). Antibiotics were added as necessary, such as 50 μg/mL streptomycin and 50 U/mL penicillin, or 50 μg/mL gentamicin, in an incubator with 5% CO_2.

Experimental animals and ethical statement

Kunming (KM) mice were obtained from Beijing Speifu Biotechnology Co., Ltd. All animal experiments were conducted in full compliance with international ethical standards and the Experimental Animal Regulation Ordinances (HPDST 2020-17) as stipulated by the Hebei Provincial Department of Science and Technology. The study protocol was reviewed and approved by the Animal Care and Use Committee of Hebei Normal University of Science and Technology.

Construction of the *mcpC* deletion strain and complemented strain

C50336 (pKD46) was cultured in Luria-Bertani agar (LBA) containing 225 mg/mL L-arabinose at 30°C until the optical density (OD) at 600 nm reached 0.6–0.8. The cells were then washed three times with pre-chilled autoclaved ultrapure water to prepare electrocompetent cells. Using DH5α (pKD3) as a template, a homologous targeting fragment was amplified with specific primers P1 and P2 (Table 2). The purified fragment was electroporated into electrocompetent C50336 (pKD46) cells. Subsequently, 1 mL of LB broth was added, and the cells were incubated at 30°C for 2 h. The culture was then plated onto LB agar containing 50 μg/mL chloramphenicol (Cm). Colonies were identified using primers P3 and P4 (Table 2). Positive strains were cured of the pKD46 plasmid by incubation at 42°C and were designated as Δ*mcpC*::*cat*.

To remove the Cm cassette, the pCP20 plasmid was introduced into Δ*mcpC*::*cat* via electroporation. Mutants were identified using primers P3 and P4, and positive strains were cured of the pCP20 plasmid by incubation at 42°C, resulting in the final Δ*mcpC* strain.

For the construction of the complemented strain, the open reading frame of the *mcpC* gene was amplified using primers P5 and P6 (Table 2). The purified PCR product was cloned into the pMD-19 T vector [Takara Biomedical Technology (Beijing) Co., Ltd.], resulting in the recombinant plasmid pMD-19 T-*mcpC*. This plasmid was then introduced into Δ*mcpC* by electroporation, and transformants were confirmed using primers P5 and P6. The complemented strain was designated as Δ*mcpC* + *mcpC*.

In vitro stress simulation experiments

Overnight cultures of C50336, Δ*mcpC* and Δ*mcpC* + Δ*mcpC* were washed three times with physiological saline and enumerated using the traditional plate counting method to determine the initial bacterial count. The bacterial suspensions were then subjected to static incubation under various stress conditions: physiological saline at pH 3.5, physiological saline at pH 10, and at 42°C for 1 h. Additionally, the bacterial suspensions were exposed to physiological saline containing 10 mmol/L H_2O_2 for 10 min. After stress treatment, bacterial counts were determined, representing the post-stress bacterial count. The survival rate of each strain under different conditions was calculated as follows: survival rate = (post-stress bacterial count)/(initial bacterial count).

Biofilm formation assay

The crystal violet (CV) staining method was used to assess biofilm formation (Zhang et al., 2020). Bacterial suspensions of C50336, Δ*mcpC*, and Δ*mcpC* + *mcpC* were inoculated at a 1:100 ratio into glass tubes containing 6 mL of LB broth and incubated statically at 28°C for 3 days. The tubes were washed 2–3 times with PBS, fixed with anhydrous methanol for 15 min, and stained with 2% CV for 15 min. The presence and thickness of stained bacterial rings on the tube walls were observed.

For quantitative analysis, a 96-well plate assay was performed. Each well was inoculated with 150 μL of bacterial suspension and stained as described above. After staining, 200 μL of anhydrous ethanol was added to each well to dissolve the CV. Absorbance at 570 nm was measured, and the experiment was repeated three times.

To detect two major components of the biofilm, curli fimbriae and cellulose, 5 μL of bacterial suspension from C50336, Δ*mcpC*, and Δ*mcpC* + *mcpC* was spotted onto salt-free LB agar containing 160 mg/L Congo red and 10 mg/L Coomassie brilliant blue (for curli detection) or onto agar containing 200 mg/L Calcofluor White Stain (CWS) for cellulose detection. Plates were incubated at 28°C for 2 days. Colony color, morphology, and fluorescence intensity under UV light (366 nm) were observed (Dong et al., 2011).

TABLE 2 Primers used for constructing the mutant and the complemented strain.

Primer	Sequence (5′–3′)
P1:	AGCAGCTCATGTTACTGGATGAAGAGGGGCGCTGGAGCCAGAGTTCGCAGAAAGAGCTGTGTGTAGGCTGGAGCTGCTTCG
P2:	CGCACGCGCCGCTTCAACCGCCGCGTTCAGCGCCAGAATATTGGTCTGAAAGGCAATGGCATATGAATATCCTCCTTAG
P3:	CGCTCTGTCTTTGTTTAGCCTTGA
P4:	ATCCCTTCCTGAGTCTGACTGGTT
P5:	GAAAATATGTTTTTGCATAACATTAAAA
P6:	TTAAGCGGGCTGCGTGTCCTCTTCGCGGA

Motility assay

Bacterial motility was assessed using semi-solid agar plates (0.3% agar). Briefly, 5 μL of overnight bacterial culture was carefully inoculated into the center of the semi-solid agar plate by gently stabbing it with a pipette tip to avoid agitation. Plates were incubated upright at 37°C for 5–6 h. Motility was determined by observing the migration of bacteria from the inoculation site toward the periphery of the plate. The experiment was repeated three times (Eakley et al., 2011).

Adhesion, invasion, and intracellular survival assays

Bacterial suspensions of C50336, Δ*mcpC*, and Δ*mcpC* + *mcpC* were washed three times with PBS and enumerated using CFU calculations from serial dilutions on agar plates. Caco-2 BBE cells were seeded into 12-well plates at a density of 10^5 CFU/well and cultured overnight in antibiotic-free DMEM supplemented with 10% FBS to achieve 80% confluence. The cells were washed three times with sterile PBS, and 1 mL of bacterial suspension was added to each well at a multiplicity of infection (MOI) of 100. Plates were centrifuged at 1,000 rpm for 5 min and incubated in a 37°C, 5% CO_2 incubator for 1 h (Xiong et al., 2023).

Adhesion assay

After a 1-h incubation, the cells were washed three times with PBS and lysed with 1% Triton X-100 for 8 min. The cell lysates were serially diluted and plated on SS agar for bacterial enumeration. Adhesion rate = (number of adhered bacteria/number of bacteria in the inoculum per well) × 100%.

Invasion assay

After a 1-h incubation, the cells were washed three times with PBS and further incubated for 1 h in DMEM containing gentamicin (100 μg/mL) to kill extracellular bacteria. The cells were then lysed with 1% Triton X-100 for 8 min, and the lysates were plated for bacterial enumeration. Invasion rate = (number of intracellular bacteria/number of bacteria in the inoculum per well) × 100%.

For the intracellular survival assay, RAW264.7 cells were seeded at a density of 10^5 cells per well into two 12-well plates. Cells were infected with bacteria at an MOI of 100 and incubated at 37°C with 5% CO_2 for 2 h. Non-adherent and non-invaded bacteria were removed by washing the cells twice with PBS. The cells were then incubated for 1 h in DMEM containing 100 μg/mL gentamicin to eliminate remaining extracellular bacteria. The cells were lysed with 1% Triton X-100, and intracellular bacteria were enumerated, representing the bacterial count at 3 h post-infection. For the second plate, the cells were further incubated in DMEM containing 10 μg/mL gentamicin for 20 h. Cells were then lysed with 1% Triton X-100, and intracellular bacteria were enumerated to represent the bacterial count at 23 h post-infection. Intracellular survival rate = (intracellular bacteria at 23 h/intracellular bacteria at 3 h) × 100%.

Determination of LD_{50} in mice

Eighty KM mice aged 4 to 6 weeks were randomly divided into 16 groups (n = 5). The first five groups were intraperitoneally (i.p.) injected with C50336 at doses of 2×10^7, 2×10^6, 2×10^5, 2×10^4, and 2×10^3 CFU/mouse, respectively. The second five groups were i.p. injected with Δ*mcpC* at doses of 3.8×10^9, 3.8×10^8, 3.8×10^7, 3.8×10^6, and 3.8×10^5 CFU/mouse, respectively. The third five groups were i.p. injected with Δ*mcpC* + *mcpC* at doses of 2×10^7, 2×10^6, 2×10^5, 2×10^4, and 2×10^3 CFU/mouse, respectively. The remaining group was i.p. injected with an equal volume of PBS. The mortality of the mice was observed and recorded over a 14-day period following the injection (Yin et al., 2019).

The LD_{50} value was calculated using the formula of $\log_{10}$ (50% endpoint) = A + (B × C), where A = $\log_{10}$ (infectious dose showing a mortality next below 50%), B = difference of logarithms = [50% − (mortality at infectious dose next below 50%)]/[(mortality next above 50%) − (mortality next below 50%)], and C = $\log_{10}$ (difference between serial infectious doses used in challenge studies) (Park et al., 2022).

RNA extraction and quantitative real-time PCR

To further investigate the effect of the *mcpC* gene on the virulence of *SE*, quantitative real-time PCR (qPCR) was employed to analyze the expression levels of *SE* virulence-related genes after the deletion of the *mcpC* gene. The procedure was as follows: RNA was extracted using a bacterial RNA extraction kit (Beijing Aidlab Biotechnologies Co., Ltd.), and genomic DNA was removed via DNase I treatment. The RNA was then reverse-transcribed into cDNA using a reverse transcription kit (Bohang Biotechnology Co., Ltd.). This cDNA was used as the template for qPCR analysis with the SYBR Green dye method. The UltraSYBR Mixture used in this method is from Jiangsu Cowin Biotech Co, Ltd. Virulence factor-related genes and their

primers were selected based on previously reported literature (Table 3), with the *16S rRNA* gene serving as the internal control (Frye et al., 2006; Upadhyaya et al., 2013). The qPCR thermal cycling conditions were as follows: initial denaturation at 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min.

TABLE 3 Primers used for qPCR.

Gene	Sequence (5′–3′)
fimD-F	CGCGGCGAAAGTTATTTCAA
fimD-R	CCACGGACGCGGTATCC
flgG-F	GCGCCGGACGATTGC
flgG-R	CCGGGCTGGAAAGCATT
prot6E-F	GAACGTTTGGCTGCCTATGG
prot6E-R	CGCAGTGACTGGCATCAAGA
csgA-F	AATGCCACCATCGACCAGTG
csgA-R	CAAAACCAACCTGACGCACC
csgD-F	GCCTCATATTAACGGCGTG
csgD-R	AGCGGTAATTTCCTGAGTGC
bcsA-F	GCCCAGCTTCAGAATATCCA
bcsA-R	TGGAAGGGCAGAAAGTGAAT
ompR-F	TGTGCCGGATCTTCTTCCA
ompR-R	CTCCATCGACGTCCAGATCTC
hflK-F	AGCGCGGCGTTGTGA
hflK-R	TCAGACCTGGCTCTACCAGATG
tatA-F	AGTATTTGGCAGTTGTTGATTGTTG
tatA-R	ACCGATGGAACCGAGTTTTTT
lrp-F	TTAATGCCGCCGTGCAA
lrp-R	GCCGGAAACCAAATGACACT
sipA-F	CAGGGAACGGTGTGGAGGTA
sipA-R	AGACGTTTTTGGGTGTGATACGT
sipB-F	GCCACTGCTGAATCTGATCCA
sipB-R	CGAGGCGCTTGCTGATTT
pipB-F	GCTCCTGTTAATGATTTCGCTAAAG
pipB-R	GCTCAGACTTAACTGACACCAAACTAA
invH-F	CCCTTCCTCCGTGAGCAAA
invH-R	TGGCCAGTTGCTCTTTCTGA
mgtC-F	CGAACCTCGCTTTCATCTTCTT
mgtC-R	CCGCCGAGGGAGAAAAAC
sodC-F	CACATGGATCATGAGCGCTTT
sodC-R	CTGCGCCGCGTCTGA
orf245-F	CAGGGTAATATCGATGTGGACTACA
orf245-R	GCGGTATGTGGAAAACGAGTTT
rfbH-F	ACGGTCGGTATTTGTCAACTCA
rfbH-R	TCGCCAACCGTATTTTGCTAA
xthA-F	CGCCCGTCCCCATCA
xthA-R	CACATCGGGCTGGTGTTTT
mrr1-F	CCATCGCTTCCAGCAACTG
mrr1-R	TCTCTACCATGAACCCGTACAAATT
16S rRNA-F	CCAGGGCTACACACGTGCTA
16S rRNA-R	TCTCGCGAGGTCGCTTCT

Bacterial colonization and persistence in organs

Thirty-five 4–6-week-old KM mice were randomly divided into two groups: the C50336 infection group (Group A, $n = 20$) and the Δ*mcpC* infection group (Group B, $n = 15$). Mice in Group A were i.p. injected with the wild-type strain C50336, while mice in Group B were i.p. injected with the gene-deleted strain Δ*mcpC*. The injection dose for both groups was 2×10^5 CFU/mouse. On days 3, 7, and 14 post-infection, five mice were randomly selected from each group and euthanized. Under sterile conditions, the liver, spleen, and cecum were collected, weighed, and homogenized in 1 mL of sterile PBS. Serial dilutions of the homogenate were plated on SS agar for bacterial enumeration. The bacterial load in each organ was calculated using the formula $\log_{10}$ CFU/g (Negi et al., 2007; Milanez et al., 2018).

Immunization with Δ*mcpC* in mice

Twenty-four 6 to 8-week-old female KM mice were randomly divided into two groups: the immunized group ($n = 12$) and the control group ($n = 12$). The immunized group received an oral dose (p.o.) of 2×10^7 CFU/mouse of the Δ*mcpC* strain. The procedure was as follows: after fasting and water deprivation for 6 h, each mouse was orally administered 100 μL of 5% $NaHCO_3$ to neutralize gastric acid. Two hours later, oral inoculation was performed using a gavage needle. On day 14 post-immunization (dpi), a booster immunization with the same dose was administered. The control group received an equivalent volume of PBS via oral administration.

Detection of IgG and IgA

The indirect ELISA method was employed to detect *SE*-specific IgG and IgA in the serum and feces of immunized mice. The strain C50336 was cultured to the logarithmic growth phase, washed twice with PBS, and resuspended. The suspension was ultrasonically lysed for 30 min at 4°C, followed by centrifugation at 8,000 rpm for 10 min. The supernatant was collected and stored at −80°C for use as the coating antigen (Ji et al., 2022). At 14 dpi and 28 dpi, blood samples were randomly collected from three mice in the immunized group and three in the control group. The blood was centrifuged to separate the serum, which was stored at −80°C for IgG detection. Fecal samples were also collected from the same three mice. Five fecal pellets from each mouse were suspended in 0.5 mL of extraction buffer containing 0.1 mg/mL soybean trypsin inhibitor (Shanghai Yuanye Bio-Technology Co., Ltd.), 10 mg/mL bovine serum albumin (Biofroxx), and 30 mM EDTA disodium (pH 7.6). After homogenization and centrifugation, the supernatant was collected and stored at −80°C for IgA detection (Nandre et al., 2011; Pasquevich et al., 2011).

Using a checkerboard titration method, the optimal antigen coating concentration and sample dilution were determined. The coating antigen (500 ng/100 μL) was added to a 96-well microtiter plate and incubated overnight at 4°C. Blocking was performed using

5% skim milk (prepared in PBS containing 0.05% Tween 20, PBST) at 37°C for 4 h (200 μL/well). After removing the blocking solution, the plate was washed three times with PBST. Serum samples diluted 1:400 or undiluted fecal supernatant (100 μL/well) were added and incubated at 37°C for 1 h. The plate was then washed three times with PBST. HRP-conjugated goat anti-mouse IgG (diluted 1:10,000, Beijing Solarbio Science & Technology Co., Ltd.) or HRP-conjugated goat anti-mouse IgA (diluted 1:10,000, Beijing Pulilai Gene Technology Co., Ltd.) was added at 100 μL/well and incubated at 37°C for 1 h. Following another washing step, 100 μL of TMB substrate solution (Beijing Solarbio Science & Technology Co., Ltd.) was added to each well and incubated at 37°C for 10 min. The reaction was terminated by adding 50 μL of 2 M H_2SO_4 to each well, and the OD_{450nm} was measured (Park et al., 2022).

Lymphocyte proliferation assay

The C50336 bacterial antigen was prepared following the previously described method and used as a stimulant for the lymphocyte proliferation assay. At 14 dpi and 28 dpi, three mice from each group were euthanized. The spleens of the immunized mice were aseptically isolated, homogenized, and filtered through a 70 μm cell strainer (Beijing Labgic Technology Co., Ltd.) to obtain spleen cells. Red blood cells were lysed using a red blood cell lysis buffer (Beijing Solarbio Science & Technology Co., Ltd.). The spleen lymphocytes were suspended in RPMI Medium 1,640 (Thermo Fisher Scientific Co., Ltd.) supplemented with 10% fetal bovine serum (FBS), 50 U/mL penicillin, and 50 μg/mL streptomycin. Cell viability was assessed using the trypan blue exclusion test, and cells were counted using a hemocytometer.

In a 96-well tissue culture plate, 10^5 cells/100 μL of cell suspension were added to each well. For the stimulation group, bacterial antigen was added at a final concentration of 7.5 μg/mL (11 μL/well). For the non-stimulation group, 11 μL of RPMI Medium 1,640 containing 10% FBS, 50 U/mL penicillin, and 50 μg/mL streptomycin was added to each well. A medium-only control group was also included (111 μL/well). The 96-well tissue culture plates were incubated at 37°C in a humidified environment with 5% CO_2 for 72 h. Lymphocyte proliferation was measured using the MTT Cell Proliferation and Cytotoxicity Assay Kit (Shanghai Beyotime Biotechnology Co., Ltd.). Absorbance was measured at 450 nm, and the stimulation index (SI) was calculated as follows (Lin et al., 2017):

$$SI = \left(\begin{array}{l}\text{Mean OD value of stimulation group} - \\ \text{Mean OD value of medium} - \text{only group}\end{array}\right) / \left(\begin{array}{l}\text{Mean OD value of non} - \\ \text{stimulation group} - \\ \text{Mean OD value of} \\ \text{medium} - \text{only group}\end{array}\right)$$

The expression of cytokines in the spleen

qPCR was used to evaluate the expression levels of cytokines *IL-1β*, *IL-2*, *IL-4*, *IL-6*, *IL-10*, *IFN-γ*, and *TNF-α* in the spleen of immunized mice. At 14 dpi and 28 dpi, three mice from each group were randomly selected. Spleens were aseptically isolated, and total RNA was extracted using the TriQuick Total RNA Extraction Kit (Beijing Solarbio Science & Technology Co., Ltd.). DNA was removed by DNase I treatment, and RNA was reverse transcribed into cDNA using a reverse transcription kit (Bohang Biotechnology Co., Ltd.). Samples were stored at −80°C until use. RT-PCR for gene expression studies was performed using UltraSYBR Mixture (Jiangsu Cowin Biotech Co, Ltd., China).

The primers used for the qPCR are shown in Table 4. Cytokine expression levels were normalized to the internal control gene *gapdh* and *β-actin* and calculated using the $2^{-\Delta\Delta C_T}$ method. Thermal cycling conditions: initial denaturation at 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min (Wang et al., 2021; Kang et al., 2022).

Immune protection assessment in mice

Forty 6 to 8-week-old KM female mice were randomly divided into four groups (n = 10): the immunization groups (Group A and Group B), the challenge group (Group C), and the control group (Group D). Mice in Group A were orally administered 2×10^7 CFU/mouse of Δ*mcpC*, with a booster immunization at the same dose at 14 dpi. Mice in Group B were orally administered 2×10^6 CFU/mouse of Δ*mcpC*, also with a booster immunization at the same dose at 14 dpi. Mice in Groups C and D were orally administered an equivalent volume of PBS. At 28 dpi, mice in the immunization groups (Groups A and B) and the challenge group (Group C) were i.p. injected with 2×10^7 CFU/mouse of C50336, while mice in the control group (Group D) were i.p. injected with an equivalent volume of PBS. For 14 days post-challenge, mouse mortality was recorded daily, and the relative survival rate was calculated as follows:

TABLE 4 Primers used for the qPCR amplification of cytokines.

Primers	Sequence (5′–3′)
IL-1β-F	GACTGTTTCTAATGCCTTCCC
IL-1β-R	ATGGTTTCTTGTGACCCTGA
IL-2-F	TGAGCAGGATGGAGAATTACAGG
IL-2-R	GTCCAAGTTCATCTTCTAGGCAC
IL-4-F	GGTCTCAACCCCCAGCTAGT
IL-4-R	GCCGATGATCTCTCTCAAGTGAT
IL-6-F	TAGTCCTTCCTACCCCAATTTCC
IL-6-R	TTGGTCCTTAGCCACTCCTTC
IL-10-F	CTTACTGACTGGCATGAGGATCA
IL-10-R	GCAGCTCTAGGAGCATGTGG
IFN-γ-F	ATGAACGCTACACACTGCATC
IFN-γ-R	CCATCCTTTTGCCAGTTCCTC
TNF-α-F	CCCTCACACTCAGATCATCTTCT
TNF-α-R	GCTACGACGTGGGCTACAG
GAPDH-F	AGGTCGGTGTGAACGGATTTG
GAPDH-R	TGTAGACCATGTAGTTGAGGTCA
β-Actin-F	TTCAACACCCCAGCCATG
β-Actin-F	CCTCGTAGATGGGCACAGT

$$\text{Relative survival rate} = \left(\begin{array}{l} \text{Mortality rate of the} \\ \text{control group} - \\ \text{Mortality rate of the} \\ \text{experimental group} \end{array} \right) / \text{Mortality rate of the control group} \times 100\%$$

Statistical analysis

Data are expressed as the mean ± standard error of the mean (SEM). All statistical analysis were two-way ANOVA and post-test. Differences between two samples were evaluated using Student's *t*-test. Significant differences are indicated with an asterisk (*), where $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$ are considered to represent statistically significant differences in mean values. Statistical analyses were conducted using IBM SPSS Statistics 26.

Results

The Δ*mcpC* mutation results in a reduced stress defense capacity

Genomic DNA from the strains C50336, Δ*mcpC*::*cat*, and Δ*mcpC* was used as templates for PCR identification with primers P3 and P4. The results (Figure 1A) showed bands at 1203 bp, 1,506 bp, and 489 bp, respectively, consistent with the expected sizes, indicating the successful knockout of the *mcpC* gene. Genomic DNA from the suspected complemented strain was used as a template, and primers P5 and P6 were employed for PCR identification. The results (Figure 1B) revealed a target band at approximately 1,572 bp, matching the expected size, confirming the successful construction of the *mcpC* complemented strain, designated as Δ*mcpC* + *mcpC*.

To investigate whether the *mcpC* gene affects the resistance of *SE* to various environmental stresses, the survival rates of C50336 and Δ*mcpC* were compared under conditions of pH 3.5, pH 10, 42°C, and 10 mmol/L H_2O_2. The results showed that (Figure 1C), compared to C50336, the survival rate of Δ*mcpC* was significantly reduced under all tested conditions, indicating that the deletion of the *mcpC* gene weakens *SE*'s resistance to acid, alkaline, thermal, and oxidative stresses.

The *mcpC* gene does not affect biofilm formationand drug resistance

The biofilm formation ability of strains C50336, Δ*mcpC*, and Δ*mcpC* + *mcpC* was evaluated. Tube assay results showed no significant differences in biofilm formation ability among the three strains (Figure 2A). Quantitative analysis using a 96-well plate assay also indicated no significant differences in biofilm formation among the strains (Figure 2B).

Detection of curli fimbriae, a major component of biofilms, revealed that all three strains formed red, rough colonies (Figure 2C), indicating their ability to produce curli fimbriae. Cellulose production analysis showed that colonies of all three strains exhibited the same fluorescence intensity under UV light (Figure 2D), confirming their ability to produce cellulose. These findings demonstrate that the *mcpC* gene does not affect the biofilm formation ability of *SE*.

The Δ*mcpC* mutation results in motility reduction

To assess whether the deletion of the *mcpC* gene affects the motility of *SE*, motility assays were conducted. The results showed that the average migration diameter of the Δ*mcpC* strain (17.5 mm) was significantly smaller than that of the wild-type C50336 strain (25 mm) and the complemented strain Δ*mcpC* + *mcpC* (21 mm) (Figure 3). These findings indicate that the deletion of the *mcpC* gene impairs the motility of *SE*.

Removal of *mcpC* weakens the invasion and intracellular survival of *SE*

The adhesion and invasion abilities of the Δ*mcpC* strain toward intestinal epithelial cells were assessed using human colorectal adenocarcinoma cells Caco-2. As shown in Figure 4A, the adhesion rate of Δ*mcpC* was comparable to that of the wild-type C50336 strain, but its invasion rate was significantly reduced.

The intracellular survival of Δ*mcpC* was evaluated using mouse-derived macrophages RAW264.7. The results revealed a significantly lower survival rate of Δ*mcpC* within macrophages compared to C50336 (Figure 4B). These findings indicate that the *mcpC* gene is critical for the invasion and intracellular survival capabilities of *SE*.

Δ*mcpC* exhibits increased LD_{50} in mice

The virulence of C50336 and Δ*mcpC* strains was evaluated in mice via i.p. injection. Mice infected with Δ*mcpC* began to die on day 2, whereas those infected with C50336 showed mortality starting on day 3. In contrast, the control group exhibited no mortality and maintained normal behavior, including smooth fur, good mental state, and an absence of symptoms such as shivering, hunching, eye crusting, or disheveled fur.

The calculated LD_{50} values (Table 5) were 6.3×10^5 CFU/mouse for C50336 and 1.9×10^7 CFU/mouse for Δ*mcpC*, with the LD_{50} of Δ*mcpC* being approximately 30 times higher than that of C50336 ($1.9 \times 10^7 / 6.3 \times 10^5 \approx 30$). The LD_{50} of the gene-complemented strain Δ*mcpC* + *mcpC* is 1.6×10^6 CFU/mouse. The results suggest that the deletion of the *mcpC* gene significantly attenuates the virulence of *SE*.

Removal of mcpC results in a down-regulation of the multiple virulence gene expressionof in *SE*

To investigate the mechanisms underlying the attenuated virulence of *SE* caused by the deletion of the *mcpC* gene, this study employed qPCR to examine the expression levels of various virulence genes. The results revealed that the deletion of *mcpC* significantly downregulated the expression of genes associated with bacterial motility (*fimD*), biofilm formation (*csgA*, *csgD*), cell membrane and cell wall integrity

10.3389/fmicb.2025.1548920

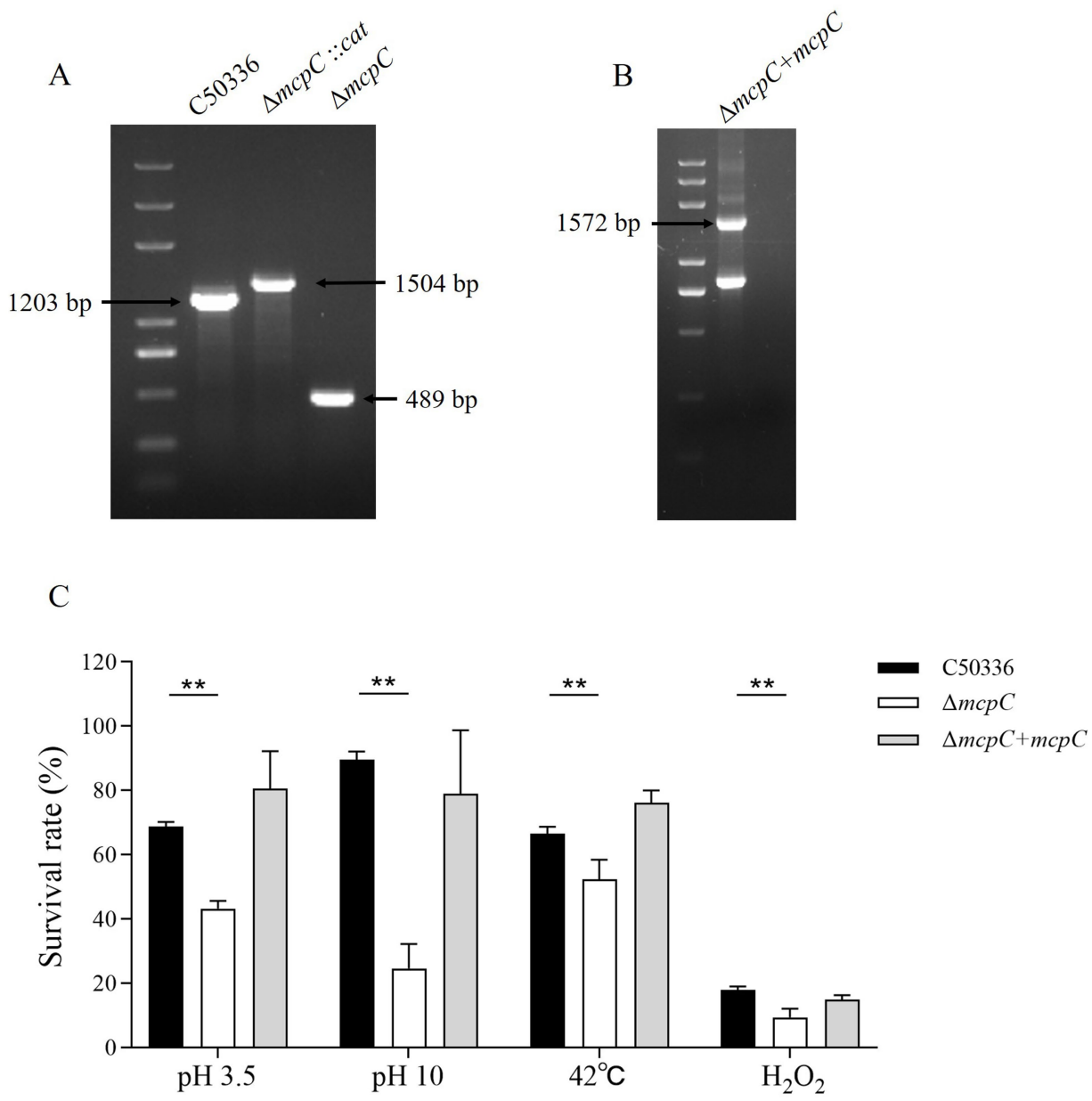


FIGURE 1
(A) PCR verification of the *mcpC* gene deletion strain. The PCR product of C50336 has a length of 1,203 bp, the product of Δ*mcpC*::*cat* has a length of 1,504 bp, the product of Δ*mcpC* has a length of 489 bp. **(B)** PCR verification of Δ*mcpC* + *mcpC*. The PCR product has a length of 1,572 bp. **(C)** The survival rate of C50336, Δ*mcpC* and Δ*mcpC* + *mcpC* under various environmental stresses. The data represents the average of 3 replicates.

(*hflK*, *lrp*), the type III secretion system (T3SS) (*sipA*, *sipB*, *pipB*), adhesion and invasion (*invH*), intracellular survival (*mgtC*, *sodC*), and nucleic acid exonuclease/endonuclease activity (*mrr1*) (Figure 5). These findings suggest that *mcpC* may regulate the expression of multiple virulence genes, thereby contributing to the overall virulence of *SE*.

Removal of *mcpC* results in reduced colonization and persistence of bacteria in the organ

Organ bacterial load is another critical indicator of *SE* virulence. In this study, the bacterial loads in the liver, spleen, and cecum of mice infected with C50336 and Δ*mcpC* were assessed. After homogenization of the organs, bacterial counts were measured.

As shown in Figure 6, bacteria were isolated from the liver, spleen, and cecum on days 3, 7, and 14 post-infection. The colonization levels of both C50336 and Δ*mcpC* peaked on day 7 and subsequently declined. Notably, compared to C50336, Δ*mcpC* exhibited significantly lower bacterial loads in the liver, spleen, and cecum on day 3 and day 7; in the liver and cecum on day 14.

These findings indicate that while Δ*mcpC* can still colonize organs, its colonization ability in certain organs is significantly reduced, suggesting that the deletion of the *mcpC* gene diminishes the organ bacterial load and virulence of *SE*.

10.3389/fmicb.2025.1548920

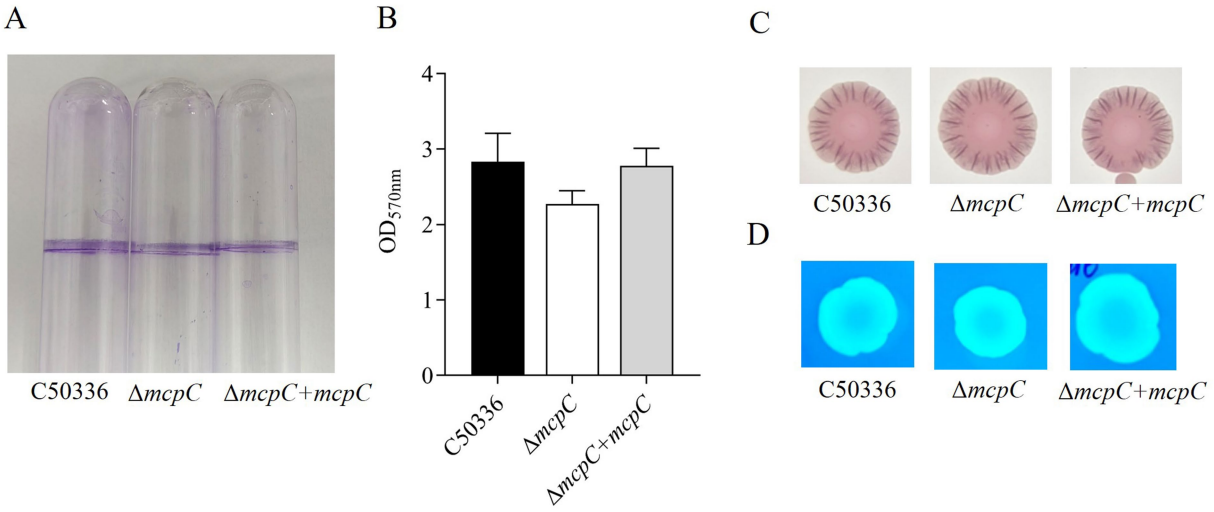


FIGURE 2
(A) Detection of biofilm formation in glass test tubes. **(B)** Qualitative detection of biofilm formation in microtiter plates, with absorbance measured at 570 nm. **(C)** Curli formation detection. **(D)** Cellulose formation detection.

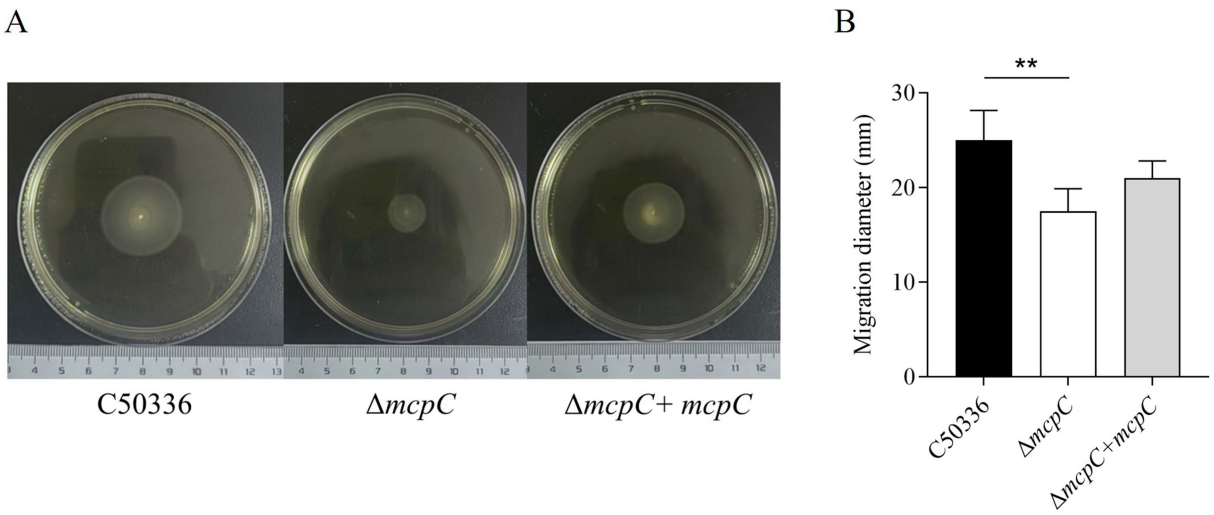


FIGURE 3
(A) The images shown are representatives of several independent assays. **(B)** The motility of the strains was evaluated on 0.3% agar plates, measured after 5 h of incubation. The data represents the average of 3 replicates.

Δ*mcpC* can induce mucosal, humoral, and cellular immune responses

To evaluate the ability of Δ*mcpC* to induce specific humoral and mucosal immune responses, serum IgG and fecal SIgA levels were measured in immunized mice using indirect ELISA. The results showed that at 14 dpi and 28 dpi, mice immunized with Δ*mcpC* exhibited significantly higher serum IgG levels compared to the control group (Figure 7A). Additionally, elevated SIgA levels were detected in fecal samples (Figure 7B). Both serum IgG and fecal SIgA levels increased significantly following the second immunization compared to the first. These findings suggest that Δ*mcpC* effectively induces robust specific humoral and mucosal immune responses, which are enhanced with repeated immunizations.

Lymphocyte proliferation assays are a primary method for evaluating cellular immune responses. In this study, a lymphocyte proliferation assay was conducted using C50336 antigens as stimulants. The results revealed that lymphocyte proliferation, measured by the stimulation index (SI), was significantly higher in mice immunized with Δ*mcpC* compared to the control group (Figure 7C). Moreover, after the booster immunization, lymphocyte proliferation levels further increased. These findings indicate that Δ*mcpC* induces a strong specific immune response, which is enhanced with repeated immunizations.

The expression of different cytokines reflects the strength and polarization of the immune response. At 14 dpi and 28 dpi, the expression levels of *IL-1β*, *IL-2*, *IL-4*, *IL-6*, *IL-10*, *TNF-α*, and *IFN-γ* were analyzed using qPCR. The results showed that (Figures 7D,E), compared

10.3389/fmicb.2025.1548920

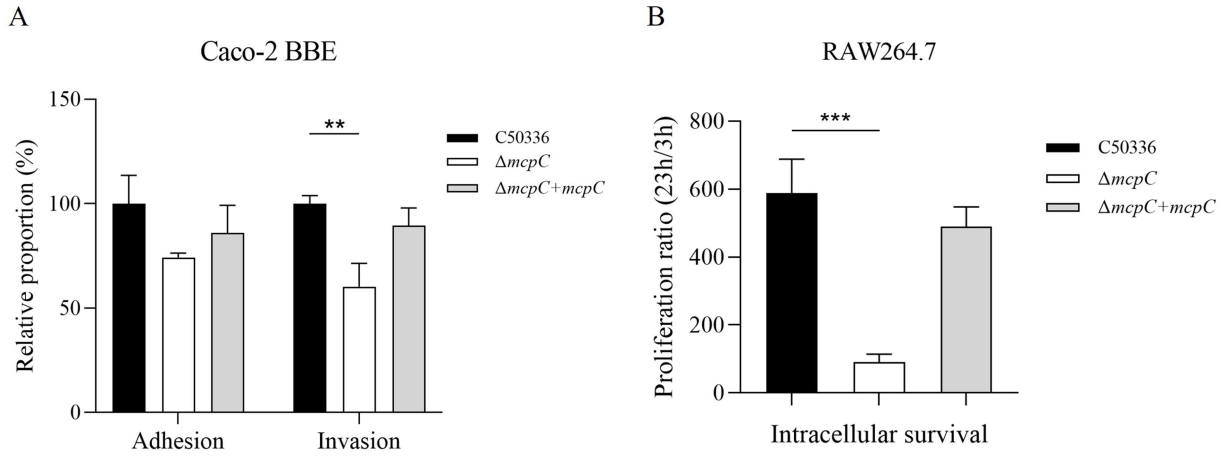


FIGURE 4
(A) Adhesion and invasion of bacteria in Caco-2 cells. **(B)** Intracellular survival in RAW264.7 cells. The results of the adhesion and invasion assays are presented as a ratio to the C50336 (the results of C50336 were considered as 100%). The data represents the average of 3 replicates.

TABLE 5 LD_{50} of C50336, Δ*mcpC* and Δ*mcpC* + *mcpC* in KM mice.

Strain	Inoculation dose (CFU/mouse)	No. of deaths/total No. of mice	LD_{50}
C50336	2×10^7	5/5	6.3×10^5
	2×10^6	3/5	
	2×10^5	2/5	
	2×10^4	0/5	
	2×10^3	0/5	
Δ*mcpC*	3.8×10^9	5/5	1.9×10^7
	3.8×10^8	5/5	
	3.8×10^7	3/5	
	3.8×10^6	1/5	
	3.8×10^5	0/5	
Δ*mcpC* + *mcpC*	2×10^7	5/5	1.6×10^6
	2×10^6	2/5	
	2×10^5	1/5	
	2×10^4	0/5	
	2×10^3	0/5	

to the control group, the expression levels of *IL-6*, *IL-10*, and *IFN-γ* significantly increased after the first immunization. Following the second immunization, *IL-6* and *IFN-γ* expression levels further increased, although *IL-6* remained similar to the levels observed after the first immunization, while *IFN-γ* continued to rise. These findings suggest that Δ*mcpC* effectively induces a strong immune response in mice.

Δ*mcpC* immunization provides powerful protective immune protection for mice

To further evaluate the protective immunity conferred by Δ*mcpC*, this study challenged immunized mice with a virulent strain via i.p. injection, recorded survival rates, and plotted survival curves. The results showed that mice in Group A (oral immunization with a dose of 2×10^7 CFU/mouse) exhibited 100% survival with no deaths recorded. In Group B (oral immunization with a dose of 2×10^6 CFU/mouse), partial mortality was observed. In contrast, all mice in Group C (non-immunized, challenged group) succumbed within 5 days, while mice in Group D (non-immunized, non-challenged control group) showed no mortality (Figure 8).

The percentage survival of mice at 14 days post-challenge is summarized in Table 6. The results showed that Groups A and D achieved a relative survival rate of 100%, Group B exhibited a survival rate of 60%, while Group C had no survivors. These findings indicate that oral immunization with Δ*mcpC* offers robust protection against *SE* infection in mice.

Discussion

SE can cause subclinical infections in adult livestock and poultry, leading to bacterial shedding through feces into the external environment. This creates challenges in pathogen eradication, resulting in severe systemic infections and high mortality rates in animals. The use of antibiotics and vaccination are key strategies for controlling *Salmonella* infections. However, widespread antibiotic use has led to the emergence of multidrug-resistant strains and severe drug residue issues, posing a significant threat to public health. Therefore, there is an urgent need for an effective vaccine to control this important zoonotic pathogen. As an intracellular pathogen, the host's cellular immune response plays a crucial role in limiting *SE* infections. Live attenuated vaccines are considered more effective than inactivated vaccines in combating both intestinal and systemic infections, as they can simultaneously stimulate robust cellular and humoral immune responses (Wang et al., 2022). In this study, a *mcpC* mutant strain of *SE* Δ*mcpC* exhibited significantly attenuated virulence and provided strong immune protection in mice.

During its pathogenesis, SE must overcome numerous adverse environmental conditions, including antimicrobial peptides,

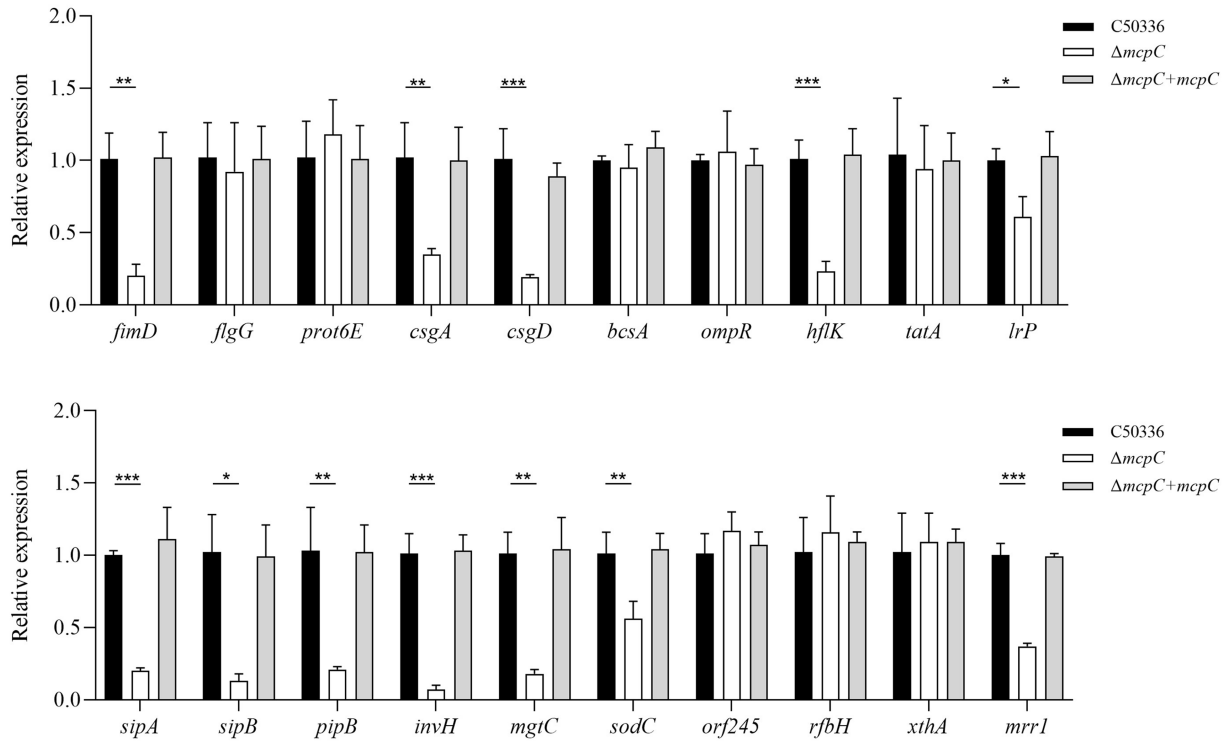


FIGURE 5
The expression levels of virulence genes in C50336, Δ*mcpC* and Δ*mcpC* + *mcpC* were detected by using qPCR, with *16S rRNA* as the housekeeping gene. The data represents the average of 3 replicates.

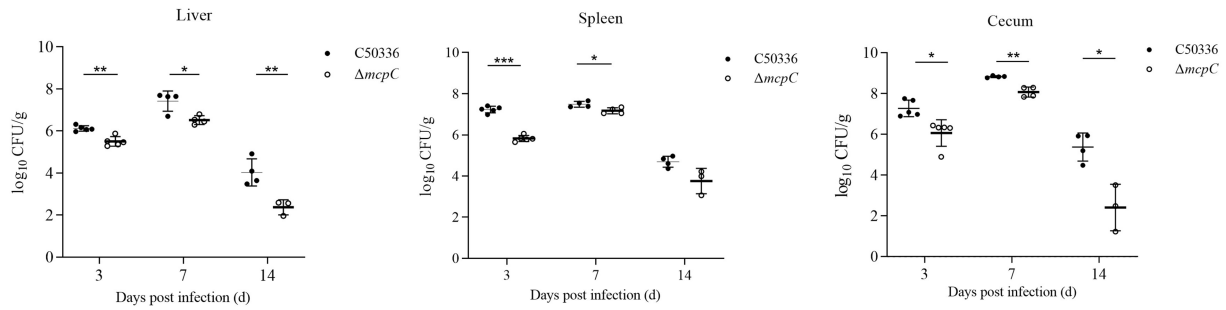


FIGURE 6
The colonization and persistence of C50336 and Δ*mcpC* colonization in liver, spleen and cecum after challenge. Values are represented as $\log_{10}$ CFU/g sample. Negative samples are shown as 0 CFU/g; all samples from the control group were negative.

temperature and pH fluctuations, and nutrient limitations (Rana et al., 2021). The ability to survive under these diverse environmental conditions is a fundamental characteristic of SE virulence (Arunima et al., 2020). In this study, the deletion of the mcpC gene impairs SE's ability to sense and respond to environmental changes, such as acid, alkali, heat, and oxidative stress. A large number of studies showed that Lon, CpxR, RfaL, and RpoS significantly affect the strain's stress defense capacity, thereby influencing the bacteria's pathogenicity (Kirthika et al., 2024; Badie et al., 2021). Thus, it is speculated that McpC helps *Salmonella* cope with challenges such as oxidative stress, high temperature, and hypoxia, and may influence the bacteria's virulence.

Numerous studies have demonstrated that *Salmonella* can form biofilms on a variety of contact surfaces (Römling et al., 2003). These biofilms exhibit high resistance to antimicrobial agents and contribute to the increased virulence of *Salmonella*, facilitating the establishment of chronic infections (Merino et al., 2019). Biofilm formation allows *Salmonella* to persist in poultry farming environments and contaminate poultry meat and eggs, which remain major vehicles for foodborne *Salmonella* outbreaks. Curli fimbriae and cellulose are the primary components of *Salmonella* biofilms. The Δ*mcpC* was no significant reduction in the formation of biofilm, curli fimbriae, and cellulose. But the expression levels of biofilm-related genes, *csgA* and *csgD*, were significantly reduced. One possible explanation is that

10.3389/fmicb.2025.1548920

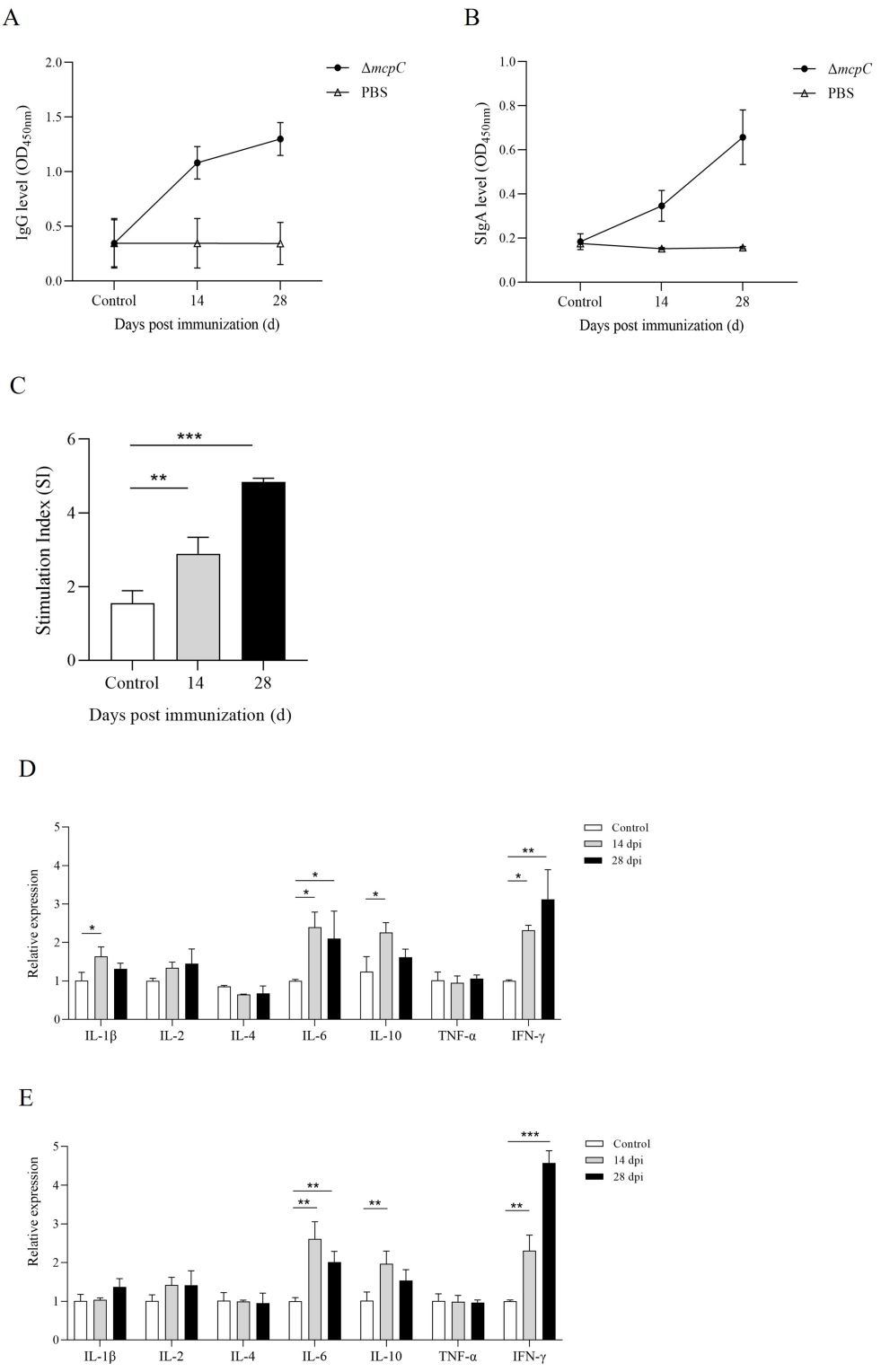


FIGURE 7
Production of *Salmonella*-specific antibodies in mouse sera and fecal. Levels of **(A)** IgG, **(B)** SIgA. **(C)** The lymphocyte proliferation was assessed at 14 dpi and 28 dpi using soluble antigen. At each time point, the mean stimulation index (SI) of the vaccinated group was significantly higher than that of the control group. **(D,E)** The expression level of cytokines *IL-1β*, *IL-2*, *IL-4*, *IL-6*, *IL-10*, *TNF-α* and *IFN-γ*. The internal control gene were *gapdh* and *β-actin*, respectively.

while the deletion of the *mcpC* gene downregulates certain biofilm-related genes, biofilm formation is regulated by the collective interaction of multiple genes.

Bacterial motility is closely associated with chemotaxis. The combination of motility and chemotaxis enables bacteria to detect and pursue nutrients, allowing them to reach and maintain their preferred

colonization niches (Josenhans and Suerbaum, 2002). The level of motility plays a critical role in determining whether *Salmonella* can reach specific sites within the host, thereby influencing its virulence. We observed that the motility of the Δ*mcpC* was significantly reduced. The expression of the motility-associated gene *fimD was* down-regulated in Δ*mcpC*. These findings suggest that the deletion of the *mcpC* gene impairs the motility of *SE*, which in turn affects its pathogenic process. This hypothesis is supported by many studies, such as the finding that the c-di-GMP binding effector STM0435 regulates flagella synthesis, controls biofilm formation, and affects Salmonella virulence (Dai et al., 2024). In addition, the knockout of YeiE reduces the expression of flagella-related genes such as fliA, flgM, and fliD in *Salmonella enterica*, leading to a decrease in flagella formation, which affects the bacteria's colonization of the intestines and its virulence (Westerman et al., 2021).

The pathogenic process of *Salmonella* involves adhesion and invasion of intestinal epithelial cells, survival and replication within host cells, and dissemination beyond the intestine. Invasion of host cells is a critical characteristic of *Salmonella* pathogenicity, while survival within macrophages is considered a more robust indicator of bacterial virulence (Fields et al., 1986; Eakley et al., 2011). We found that that the deletion of the *mcpC* gene impairs the invasion and intracellular survival capabilities of *SE*, both of which are closely linked to its virulence.

An ideal attenuated live vaccine strain should effectively withstand host-induced stress, provide robust protection against the target pathogen, and successfully colonize host lymphoid tissues, all while maintaining an avirulent profile (Pati et al., 2013). This study showed that the LD_{50} of Δ*mcpC* was significantly increased, and the expression of multiple virulence genes is down-regulated, indicating a reduction in its virulence. Following ingestion, *Salmonella* replicates in mucosa-associated lymphoid tissues, such as Peyer's patches, and disseminates via mesenteric lymph nodes to systemic organs, including the spleen and liver. The cecum is also a primary colonization site for *SE* in poultry. Organ bacterial load is a critical indicator of *Salmonella* virulence. Our results showed that Δ*mcpC* exhibited significantly reduced bacterial loads in these organs compared to the wild-type C50336 strain, indicating that *mcpC* deletion diminishes the pathogenicity of *SE*. Collectively, these findings suggest that the Δ*mcpC* mutant strain has significant potential as a candidate for an attenuated live vaccine.

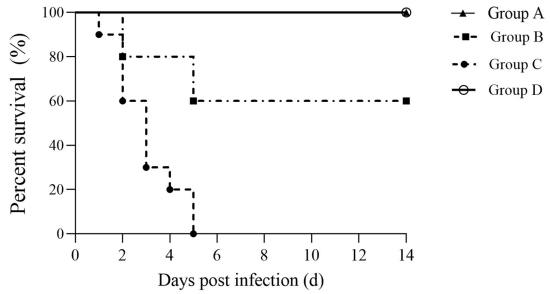


FIGURE 8
Survival curve. Female KM mice aged 6–8 weeks were p.o. inoculated with the Δ*mcpC* and i.p. injected with a lethal dose of C50336 at 28 dpi. The percent survival of mice was monitored daily.

Another critical requirement for developing an attenuated live vaccine is its ability to elicit both humoral and cellular immune responses in the host (Lin et al., 2017). A robust humoral immune response, particularly the secretion of secretory IgA (SIgA) in the intestinal mucosa, is essential for combating *Salmonella* infections (Roesler et al., 2006). Furthermore, as *Salmonella* is a facultative intracellular pathogen, cellular immune responses are crucial for effective host defense (Nandre et al., 2011). In this study, indirect ELISA revealed that Δ*mcpC* induced strong specific humoral and mucosal immune responses. Given that this pathogen primarily resides in the intestinal tract, IgA secreted into the intestinal lumen is likely to play a key role in protective immunity (Nandre et al., 2011). Our results also showed that Δ*mcpC* could induce a specific cellular immune response, and up-regulate the expression of IL-6, *IL-10*, and *IFN-γ*. IL-6 is pro-inflammatory cytokines that coordinate inflammatory and host defense responses (Huang and Sheng, 2010). *IL-10*, predominantly produced by Th2 cells, is a multifunctional anti-inflammatory cytokine associated with immune regulation, defense, and infection (Yuan et al., 2022). *IFN-γ*, produced by activated T cells and natural killer (NK) cells, plays a vital role in host defense against intracellular pathogens such as *Salmonella* Typhimurium (Benbernou and Nauciel, 1994). Immune protection rate is a key indicator of vaccine efficacy, reflecting the vaccine's ability to protect against the target pathogen post-immunization. Oral immunization with Δ*mcpC* achieved a relative protection rate of 100%, demonstrating its strong protective efficacy in mice.

In summary, the *mcpC* gene is involved in multiple biological processes in *SE*, and its deletion significantly attenuates bacterial virulence in mice. The Δ*mcpC* strain was shown to induce robust immune responses and provide excellent immune protection in mice. These findings suggest that the Δ*mcpC* strain is a promising candidate for an attenuated live vaccine against *SE*, laying a preliminary foundation for the development of genetically engineered vaccines. Considering the broad host range of *SE*, future studies will validate its protective efficacy and safety in poultry and other animal models.

TABLE 6 Protective effects of Δ*mcpC* in mice.

Group	Vaccination			Number	Challenge			Survivors/ total	Percent survival (%)
	Strain	Route	Dose (CFU)		Strain	Route	Dose (CFU)		
A	Δ*mcpC*	p.o.	2×10^7	10	C50336	i.p.	2×10^7	10/10	100
B	Δ*mcpC*	p.o.	2×10^6	10	C50336	i.p.	2×10^7	6/10	60
C	PBS	p.o.	—	10	C50336	i.p.	2×10^7	0/10	0
D	PBS	p.o.	—	10	PBS	i.p.	—	10/10	100

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 animal study was approved by Animal Care and Use Committee of Hebei Normal University of Science and Technology. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

LZ: Writing – original draft, Data curation, Formal analysis, Methodology, Supervision. LC: Data curation, Formal analysis, Methodology, Writing – original draft. XZ: Methodology, Writing – original draft. YaL: Methodology, Writing – original draft. QZ: Software, Writing – original draft. YuL: Software, Writing – original draft. NL: Software, Writing – original draft. QS: Supervision, Validation, Writing – review & editing. YZ: Funding acquisition, Writing – review & editing. TW: Conceptualization, Funding acquisition, Investigation, Methodology, Writing – original draft, 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 the Hebei Agriculture Research System (HBCT2024280205, HBCT2024280401, and HBCT2024280404); Central Government Guides Local Projects, Hebei Province, Department of Science and Technology (23626604G).

Acknowledgments

The authors thank Prof. Zheng Wang from Beijing University of Agriculture for their valuable help in our experiment.

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.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

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

Arunima, A., Swain, S. K., Patra, S. D., Das, S., Mohakud, N. K., Misra, N., et al. (2020). Role of OB-fold protein YdeI in stress response and virulence of *Salmonella enterica* serovar Enteritidis. *J. Bacteriol.* 203:e00237. doi: 10.1128/JB.00237-20

Badie, F., Saffari, M., Moniri, R., Alani, B., Atoof, F., Khorshidi, A., et al. (2021). The combined effect of stressful factors (temperature and pH) on the expression of biofilm, stress, and virulence genes in *Salmonella enterica* ser. Enteritidis and Typhimurium. *Arch. Microbiol.* 203, 4475–4484. doi: 10.1007/s00203-021-02435-y

Benbernou, N., and Nauciel, C. (1994). Influence of mouse genotype and bacterial virulence in the generation of interferon-gamma-producing cells during the early phase of *Salmonella* Typhimurium infection. *Immunology* 83, 245–249

Berleman, J. E., and Bauer, C. E. (2005). Involvement of a Che-like signal transduction cascade in regulating cyst cell development in *Rhodospirillum centenum*. *Mol. Microbiol.* 56, 1457–1466. doi: 10.1111/j.1365-2958.2005.04646.x

Chousalkar, K., Gast, R., Martelli, F., and Pande, V. (2018). Review of egg-related salmonellosis and reduction strategies in United States, Australia, United Kingdom and New Zealand. *Crit. Rev. Microbiol.* 44, 290–303. doi: 10.1080/1040841X.2017.1368998

Cote, C. K., Biryukov, S. S., Klimko, C. P., Shoe, J. L., Hunter, M., Rosario-Acevedo, R., et al. (2021). Protection elicited by attenuated live *Yersinia pestis* vaccine strains against lethal infection with virulent *Y. pestis*. *Vaccines* 9:161. doi: 10.3390/vaccines9020161

Dai, Y., Liu, R., Yue, Y., Song, N., Jia, H., Ma, Z., et al. (2024). A c-di-GMP binding effector STM0435 modulates flagellar motility and pathogenicity in *Salmonella*. *Virulence* 15:2331265. doi: 10.1080/21505594.2024.2331265

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. U.S.A.* 97, 6640–6645. doi: 10.1073/pnas.120163297

Deguchi, K., Yokoyama, E., Honda, T., and Mizuno, K. (2009). Efficacy of a novel trivalent inactivated vaccine against the shedding of *Salmonella* in a chicken challenge model. *Avian Dis.* 53, 281–286. doi: 10.1637/8516-110908-Reg.1

Dong, H., Peng, D., Jiao, X., Zhang, X., Geng, S., and Liu, X. (2011). Roles of the spiA gene from *Salmonella enteritidis* in biofilm formation and virulence. *Microbiology* 157, 1798–1805. doi: 10.1099/mic.0.046185-0

Eakley, N. M., Bochsler, P. N., Gopal Reddy, P., and Fadl, A. A. (2011). Biological and virulence characteristics of the YqhC mutant of *Salmonella*. *Microbiol. Immunol.* 55, 830–840. doi: 10.1111/j.1348-0421.2011.00387.x

Fields, P. I., Swanson, R. V., Haidaris, C. G., and Heffron, F. (1986). Mutants of *Salmonella* Typhimurium that cannot survive within the macrophage are avirulent. *Proc. Natl. Acad. Sci. U.S.A.* 83, 5189–5193. doi: 10.1073/pnas.83.14.5189

Frye, J., Karlinsey, J. E., Felise, H. R., Marzolf, B., Dowidar, N., McClelland, M., et al. (2006). Identification of new flagellar genes of *Salmonella enterica* Serovar Typhimurium. *J. Bacteriol.* 188, 2233–2243. doi: 10.1128/JB.188.6.2233-2243.2006

Harkey, C. W., Everiss, K. D., and Peterson, K. M. (1994). The *Vibrio cholerae* toxin-coregulated-pilus gene tcpI encodes a homolog of methyl-accepting chemotaxis proteins. *Infect. Immun.* 62, 2669–2678. doi: 10.1128/iai.62.7.2669-2678.1994

Hickman, J. W., Tifrea, D. F., and Harwood, C. S. (2005). A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14422–14427. doi: 10.1073/pnas.0507170102

Huang, Z.-B., and Sheng, G.-Q. (2010). Interleukin-1β with learning and memory. *Neurosci. Bull.* 26, 455–468. doi: 10.1007/s12264-010-6023-5

Huberman, Y. D., Velilla, A. V., and Terzolo, H. R. (2019). Evaluation of different live *Salmonella enteritidis* vaccine schedules administered during layer hen rearing to reduce excretion, organ colonization, and egg contamination. *Poult. Sci.* 98, 2422–2431. doi: 10.3382/ps/pez003

Ji, H. J., Jang, A.-Y., Song, J. Y., Ahn, K. B., Han, S. H., Bang, S. J., et al. (2022). Development of live attenuated *Salmonella* Typhimurium vaccine strain using radiation mutation enhancement technology (R-MET). *Front. Immunol.* 13:931052. doi: 10.3389/fimmu.2022.931052

Josenhans, C., and Suerbaum, S. (2002). The role of motility as a virulence factor in bacteria. *Int. J. Med. Microbiol.* 291, 605–614. doi: 10.1078/1438-4221-00173

Kang, X., Yang, Y., Meng, C., Wang, X., Liu, B., Geng, S., et al. (2022). Safety and protective efficacy of *Salmonella pullorum* spiC and rfaH deletion rough mutant as a live attenuated DIVA vaccine candidate. *Poult. Sci.* 101:101655. doi: 10.1016/j.psj.2021.101655

Kirthika, P., Senevirathne, A., Park, S., Aganja, R. P., Kim, I.-S., Tae, H.-J., et al. (2024). Intracellular survival and pathogenicity modulation of *Salmonella* Lon, CpxR, and RfaL mutants used as live bacterial vectors under abiotic stress, unveiling the link between stress response and virulence in epithelial cells. *Int. J. Mol. Sci.* 25:9056. doi: 10.3390/ijms25169056

Li, Z., Lou, H., Ojcius, D. M., Sun, A., Sun, D., Zhao, J., et al. (2014). Methyl-accepting chemotaxis proteins 3 and 4 are responsible for *Campylobacter jejuni* chemotaxis and jejuna colonization in mice in response to sodium deoxycholate. *J. Med. Microbiol.* 63, 343–354. doi: 10.1099/jmm.0.068023-0

Lin, Z., Tang, P., Jiao, Y., Kang, X., Li, Q., Xu, X., et al. (2017). Immunogenicity and protective efficacy of a *Salmonella enteritidis* sptP mutant as a live attenuated vaccine candidate. *BMC Vet. Res.* 13:194. doi: 10.1186/s12917-017-1115-3

Liu, X., Jiao, C., Ma, Y., Wang, Q., and Zhang, Y. (2018). A live attenuated *Vibrio anguillarum* vaccine induces efficient immunoprotection in Tiger puffer (*Takifugu rubripes*). *Vaccine* 36, 1460–1466. doi: 10.1016/j.vaccine.2018.01.067

Merino, L., Trejo, F. M., De Antoni, G., and Golowczyc, M. A. (2019). *Lactobacillus* strains inhibit biofilm formation of *Salmonella* sp. isolates from poultry. *Food Res. Int.* 123, 258–265. doi: 10.1016/j.foodres.2019.04.067

Milanez, G. P., Werle, C. H., Amorim, M. R., Ribeiro, R. A., Tibo, L. H. S., Roque-Barreira, M. C., et al. (2018). HU-lacking mutants of *Salmonella enterica* Enteritidis are highly attenuated and can induce protection in murine model of infection. *Front. Microbiol.* 9:1780. doi: 10.3389/fmicb.2018.01780

Nandre, R. M., Chaudhari, A. A., Matsuda, K., and Lee, J. H. (2011). Immunogenicity of a *Salmonella enteritidis* mutant as vaccine candidate and its protective efficacy against salmonellosis in chickens. *Vet. Immunol. Immunopathol.* 144, 299–311. doi: 10.1016/j.vetimm.2011.08.015

Negi, V. D., Singhamahapatra, S., and Chakravortty, D. (2007). *Salmonella enterica* serovar Typhimurium strain lacking pmrG-HM-D provides excellent protection against salmonellosis in murine typhoid model. *Vaccine* 25, 5315–5323. doi: 10.1016/j.vaccine.2007.05.015

Nishiyama, S., Takahashi, Y., Yamamoto, K., Suzuki, D., Itoh, Y., Sumita, K., et al. (2016). Identification of a *Vibrio cholerae* chemoreceptor that senses taurine and amino acids as attractants. *Sci. Rep.* 6:20866. doi: 10.1038/srep20866

Park, S., Jung, B., Kim, E., Hong, S.-T., Yoon, H., and Hahn, T.-W. (2020). *Salmonella* Typhimurium lacking YjeK as a candidate live attenuated vaccine against invasive *Salmonella* infection. *Front. Immunol.* 11:1277. doi: 10.3389/fimmu.2020.01277

Park, S., Jung, B., Kim, E., Yoon, H., and Hahn, T.-W. (2022). Evaluation of *Salmonella* Typhimurium lacking fruR, ssrAB, or hfq as a prophylactic vaccine against *Salmonella* lethal infection. *Vaccine* 10:1413. doi: 10.3390/vaccines10091413

Pasquevich, K. A., Ibañez, A. E., Coria, L. M., García Samartino, C., Estein, S. M., Zwerdling, A., et al. (2011). An oral vaccine based on U-Omp19 induces protection against *B. abortus* mucosal challenge by inducing an adaptive IL-17 immune response in mice. *PLoS One* 6:e16203. doi: 10.1371/journal.pone.0016203

Pati, N. B., Vishwakarma, V., Selvaraj, S. K., Dash, S., Saha, B., Singh, N., et al. (2013). *Salmonella* Typhimurium TTSS-2 deficient mig-14 mutant shows attenuation in immunocompromised mice and offers protection against wild-type *Salmonella* Typhimurium infection. *BMC Microbiol.* 13:236. doi: 10.1186/1471-2180-13-236

Rana, K., Nayak, S. R., Bihary, A., Sahoo, A. K., Mohanty, K. C., Palo, S. K., et al. (2021). Association of quorum sensing and biofilm formation with Salmonella virulence: story beyond gathering and cross-talk. *Arch. Microbiol.* 203, 5887–5897. doi: 10.1007/s00203-021-02594-y

Roesler, U., Heller, P., Waldmann, K.-H., Truyen, U., and Hensel, A. (2006). Immunization of sows in an integrated pig-breeding herd using a homologous inactivated *Salmonella* vaccine decreases the prevalence of *Salmonella* Typhimurium infection in the offspring. *J. Veterinary Med. B* 53, 224–228. doi: 10.1111/j.1439-0450.2006.00951.x

Römling, U., Bokranz, W., Rabsch, W., Zogaj, X., Nimtz, M., and Tschäpe, H. (2003). Occurrence and regulation of the multicellular morphotype in *Salmonella* serovars important in human disease. *Int. J. Med. Microbiol.* 293, 273–285. doi: 10.1078/1438-4221-00268

Sampedro, I., Parales, R. E., Krell, T., and Hill, J. E. (2014). *Pseudomonas* chemotaxis. *FEMS Microbiol Rev.* 39, 17–46. doi: 10.1111/1574-6976.12081

Upadhyaya, I., Upadhyay, A., Kollanoor-Johny, A., Darre, M., and Venkitanarayanan, K. (2013). Effect of plant derived antimicrobials on *Salmonella enteritidis* adhesion to and invasion of primary chicken oviduct epithelial cells *in vitro* and virulence gene expression. *Int. J. Mol. Sci.* 14, 10608–10625. doi: 10.3390/ijms140510608

Van Immerseel, F., Methner, U., Rychlik, I., Nagy, B., Velge, P., Martin, G., et al. (2005). Vaccination and early protection against non-host-specific *Salmonella* serotypes in poultry: exploitation of innate immunity and microbial activity. *Epidemiol. Infect.* 133, 959–978. doi: 10.1017/S0950268805004711

Wang, Y., Huang, C., Tang, J., Liu, G., Hu, M., Kang, X., et al. (2021). *Salmonella pullorum* spiC mutant is a desirable LASV candidate with proper virulence, high immune protection and easy-to-use oral administration. *Vaccine* 39, 1383–1391. doi: 10.1016/j.vaccine.2021.01.059

Wang, X., Kang, X., Pan, M., Wang, M., Zhang, J., and Song, H. (2022). Evaluation of the protective immune response induced by an rfbG-deficient *Salmonella enterica* serovar Enteritidis strain as a live attenuated DIVA (differentiation of infected and vaccinated animals) vaccine in chickens. *Microbiol. Spectr.* 10, e01574–e01522. doi: 10.1128/spectrum.01574-22

Westerman, T. L., McClelland, M., and Elfenbein, J. R. (2021). YeiE regulates motility and gut colonization in *Salmonella enterica* serotype Typhimurium. *mBio* 12:e0368020. doi: 10.1128/mBio.03680-20

Xiong, D., Song, L., Chen, Y., Jiao, X., and Pan, Z. (2023). *Salmonella enteritidis* activates inflammatory storm via SPI-1 and SPI-2 to promote intracellular proliferation and bacterial virulence. *Front. Cell. Infect. Microbiol.* 13:1158888. doi: 10.3389/fcimb.2023.1158888

Yang, W., Wang, L., Zhang, L., Qu, J., Wang, Q., and Zhang, Y. (2015). An invasive and low virulent *Edwardsiella tarda* esrB mutant promising as live attenuated vaccine in aquaculture. *Appl. Microbiol. Biotechnol.* 99, 1765–1777. doi: 10.1007/s00253-014-6214-5

Yin, J., Cheng, Z., Wu, Y., He, Q., Zhang, J., Yang, Z., et al. (2019). Characterization and protective efficacy of a *Salmonella* pathogenicity island 2 (SPI2) mutant of *Salmonella paratyphi* A. *Microb. Pathog.* 137:103795. doi: 10.1016/j.micpath.2019.103795

Yuan, Z., Song, H., Huang, Q., Liu, J., Sun, H., Meng, X., et al. (2022). Immune enhancement effects of inactivated vaccine against extracellular products of *Aeromonas caviae* AC-CY on crucian carp. *Fish Shellfish Immunol.* 127, 1001–1011. doi: 10.1016/j.fsi.2022.07.046

Zabalza-Baranguá, A., Poveda-Urkixo, I., Mena-Bueno, S., Ramírez, G. A., De Bolle, X., and Grilló, M. J. (2023). Vaccine properties of *Brucella melitensis* 16MΔ*wzm* and reactivation of placental infection in pregnant sheep. *Vaccine* 41, 1554–1566. doi: 10.1016/j.vaccine.2023.01.017

Zhang, Z., Du, W., Wang, M., Li, Y., Su, S., Wu, T., et al. (2020). Contribution of the colicin receptor CirA to biofilm formation, antibotic resistance, and pathogenicity of *Salmonella enteritidis*. *J. Basic Microbiol.* 60, 72–81. doi: 10.1002/jobm.201900418

Zhou, Y., Xiong, D., Guo, Y., Liu, Y., Kang, X., Song, H., et al. (2023). *Salmonella enteritidis* RfbD enhances bacterial colonization and virulence through inhibiting autophagy. *Microbiol. Res.* 270:127338. doi: 10.1016/j.micres.2023.127338

frontiers | Frontiers in Microbiology

TYPE Brief Research Report
PUBLISHED 17 February 2025
DOI 10.3389/fmicb.2025.1535893

OPEN ACCESS

EDITED BY
Xuanyu Tao,
University of Oklahoma, United States

REVIEWED BY
Dongyu Wang,
University of Oklahoma, United States
Mengyuan Ji,
University of Oklahoma, United States

*CORRESPONDENCE
Xiao Yu
✉ Yuxiaowz@163.com
Yiwei Shi
✉ shiyw@sxmu.edu.cn

RECEIVED 28 November 2024
ACCEPTED 27 January 2025
PUBLISHED 17 February 2025

CITATION
Zhang J, Li Q, Liu J, Fan F, Shi Y and Yu X (2025) Prevalence of hypervirulent *Klebsiella pneumoniae* strains in COVID-19 patients with bacterial co-infections.
Front. Microbiol. 16:1535893.
doi: 10.3389/fmicb.2025.1535893

Prevalence of hypervirulent *Klebsiella pneumoniae* strains in COVID-19 patients with bacterial co-infections

Jingfen Zhang[1], Qiaoyu Li[2], Jingjing Liu[1], Fangfang Fan[3], Yiwei Shi[3]* and Xiao Yu[3]*

[1]The First Clinical Medical College of Shanxi Medical University, Taiyuan, China, [2]Academy of Medical Sciences, Department of Occupational Health, Shanxi Medical University, Taiyuan, Shanxi, China, [3]National Health Commission of the People's Republic of China (NHC) Key Laboratory of Pneumoconiosis, Shanxi Key Laboratory of Respiratory Diseases, Department of Pulmonary and Critical Care Medicine, The First Hospital of Shanxi Medical University, Taiyuan, Shanxi, China

Objective: A recent alarming report from the World Health Organization highlighted the rapid global spread of a hypervirulent, carbapenem-resistant strain of *Klebsiella pneumoniae*. The COVID-19 pandemic frequently led to bacterial co-infections, with *K. pneumoniae* being a common and highly pathogenic agent. This study aimed to assess KP characteristics via whole-genome sequencing and clarify its molecular epidemiology to guide standardized clinical treatment.

Methods: Our retrospective analysis of clinical data from COVID-19 patients admitted to our hospital between 7 December 2022, and 2 January 2023-following China's policies changes, which led to a significant influx of patients-identified 17 *K. pneumoniae* isolates from sputum samples with bacterial co-infections. These isolates underwent whole-genome sequencing for ST typing, virulence gene annotation, plasmid profiling, and antimicrobial susceptibility testing.

Results: Of the 17 *K. pneumoniae* isolates, 52.9% were hypermucoviscous. Whole genome sequencing identified eight sequence types (STs), with ST23/KL1 being the most prevalent at 35.3%. Virulence genes were present in 94.1% of strains, including Yersiniabactin (70.6%), Aerobactin (82.3%), and Salmochelin (88.2%). Plasmid analysis revealed common IncHI1B/FIBk or IncFIBk types. All isolates were highly sensitive to antibiotics, except for blaSHV resistance. The 17 patients had a median age of 71 years and significant comorbidities, such as hypertension (64.7%) and diabetes (41.2%).

Conclusion: The ST types and virulence gene profiles indicate that most *K. pneumoniae* strains co-infecting COVID-19 patients are common, high-virulence strains prevalent in the Asia-Pacific region. Our findings suggest that COVID-19 may contribute to the spread of hypervirulent *K. pneumoniae* strains, potentially informing the ongoing WHO epidemic alert.

KEYWORDS

hypervirulent *Klebsiella pneumoniae*, COVID-19, bacterial co-infections, sequence type, virulence factor

Introduction

The COVID-19 pandemic has posed unprecedented challenges to global public health systems, significantly exacerbating the spread of various bacterial pathogens (Fan et al., 2023; Lansbury et al., 2020). A recent alarming report by the World Health Organization (WHO) highlighted the rapid global dissemination of a hypervirulent, carbapenem-resistant strain of *K. pneumoniae*, predominantly the ST23 strain, identified in at least 16 countries and regions since early 2024 (World Health Organization, 2024). Traditionally reported as sporadic or localized clonal outbreaks, the widespread prevalence of this strain under current circumstances remains unexplained. Bacterial co-infections in COVID-19 can lead to severe illness and warrant close attention.

Existing studies have shown that the prevalence of *K. pneumoniae* co-infection with COVID-19 is highest in Asia at 23% (95% CI: 14–35%), followed by Europe at 15% (95% CI: 6–32%), and the Americas at 4% (95% CI: 4–5%). Globally, approximately 17–40% of COVID-19 patients develop bacterial co-infections, with *K. pneumoniae* infections accounting for 19% (ranging from 13 to 28%) of these cases (das Chagas et al., 2024).

Recent studies have shown that *K. pneumoniae* is frequently detected in critically ill COVID-19 patients admitted to the Intensive Care Unit (ICU), as evidenced by respiratory tract and blood samples from these patients (Arcari et al., 2021; Pourajam et al., 2022). The prevalence of multi-drug resistant bacterial infections, particularly those caused by *K. pneumoniae*, is notably associated with the deterioration of patient outcomes, characterized by increased complications, higher morbidity and mortality (Ficik et al., 2023; Navon-Venezia et al., 2017). These findings highlight the crucial role that co-infection with *K. pneumoniae* plays in the progression of COVID-19 to severe conditions.

However, these reports lacked in-depth genomic analysis, focusing primarily on resistance profiles and clinical outcomes. This study aims to use genomic analysis to further investigate the impact of SARS-CoV-2 on the epidemiology of *K. pneumoniae* in our hospital.

Materials and methods

Study design and sample collection

This study retrospectively analyzed clinical and microbiological data from COVID-19 patients admitted to our hospital between 7 December 2022, and 2 January 2023, during a period of increased patient influx following a policy shift in COVID-19 prevention. We confirmed KP co-infection in all 17 patients using a comprehensive diagnostic approach that integrated clinical signs, imaging studies (CT or X-ray), blood tests (WBC, NEU%, LYM%, CRP, PCT, etc.), and sputum cultures, ensuring accurate identification. A total of 17 *K. pneumoniae* isolates were obtained from the sputum samples of these patients. Additionally, we extracted clinical data, including demographic characteristics, comorbidities, and laboratory findings, from electronic medical records for further analysis.

Phenotypic analysis

Phenotypic analysis of the isolates was performed using the string test to identify hypermucoviscosity. A positive result, indicated by a mucoviscous string greater than 5 mm, classified the strain as hypermucoviscous.

Whole genome sequencing and genomic analysis

Whole genome sequencing (WGS) was conducted to analyze the genomic features of the isolates. Genomic DNA was extracted using a Qiagen DNA Mini Kit, and libraries were prepared using the DNA Library Prep Kit before sequencing on the Illumina HiSeq platform. Sequence assembly was performed with SPAdes, and genome annotation was conducted using Prokka. Sequence types (STs) were determined using the MLST tool, while virulence-associated genes were identified using the BIGSdb-Kp database. Plasmid replicon types were determined with PlasmidFinder, and phylogenetic relationships were analyzed using CSI Phylogeny. Virulence scores were calculated based on the presence of key genes: yersiniabactin, colibactin, and aerobactin. Scores ranged from 0 (no virulence genes detected) to 5 (all three virulence genes present). This scoring system was used to quantify the virulence potential of the isolates. Resistance genes were screened using the ResFinder database.

Antimicrobial susceptibility testing

Antimicrobial susceptibility testing was performed using the disk diffusion method, following CLSI guidelines. The tested antibiotics included ceftriaxone, meropenem, ciprofloxacin, amikacin, and tigecycline.

Clinical data analysis

Clinical data were analyzed to identify patient characteristics, including age, sex, comorbidities, and hospitalization duration. Patients were categorized as having either hospital-acquired or community-acquired infections. Statistical analyses were conducted using SPSS, with results presented as medians with interquartile ranges (IQR) or means with standard deviations (SD).

Ethical approval

Ethical approval for the study was obtained from the Ethics Committee Of Our Hospital (NO.KYLL-2023-091). Informed consent was waived due to the retrospective nature of the research. To protect patient privacy, all genomic data were anonymized and handled in compliance with institutional guidelines and applicable privacy regulations.

Results and discussion

Distribution and virulence of *K. pneumoniae* isolates based on phenotypic and genotypic analysis

Phenotypic analysis indicated that 9 of these isolates were hypermucoviscous. Whole genome sequencing identified 8 different sequence types (ST), with 7 of these belonging to highly virulent clonal groups, ST23/KL1 was the most prevalent at 35.3%, followed by ST412/KL57 at 17.6%, and ST86/KL2 at 11.8%. The other isolates corresponded to five different STs: ST35, ST380, ST65, ST2906, and ST1550-2LV, with ST1550-2LV not classified as a highly virulent type (Figure 1). The clustering patterns did not strongly support hospital-based clonal transmission, highlighting the likelihood of community acquisition in most cases. Additionally, ST23/KL1 has been reported to be more prevalent in East Asian regions, which could also explain its dominance in our study.

Upon further analysis of Figure 1, we observed that there was no significant correlation between virulence factor scores and hospitalization duration, clinical signs, and infection modes. This suggests that, while virulence factors such as *rmpA*, aerobactin, and colibactin contribute to the pathogenicity of *K. pneumoniae*, they do not appear to directly impact the length of hospital stay. Instead, factors such as patient comorbidities, immune status, and the severity of underlying COVID-19 infections likely play a more substantial role in determining hospitalization duration. This finding implies that the patients in this cohort may have been carriers of hypermucoviscous *K. pneumoniae*, rather than experiencing severe infections primarily driven by these high-virulence strains. The absence of a strong correlation further supports the notion that co-infection dynamics are multifaceted, influenced by a complex interplay of both bacterial virulence and host-related factors.

Virulence gene profiles, plasmid types, and antimicrobial susceptibility of *K. pneumoniae* strains

Strains with the same ST carry similar types of high-virulence-associated plasmids (plasmid type: IncHI1B/FIBκ or IncFIBκ) and exhibit comparable virulence gene profiles. The prevalence of virulence genes among these strains is detailed below (Figure 1): yersiniabactin (facilitates iron acquisition, enhancing the survival and virulence of bacteria in iron-limited environments) was detected in 70.6% of the strains (12 strains), with the highest incidence found in *ybt1* (ICE*Kp10*) at 35.2% (6 strains); *ybt4* (plasmid) and *ybt9* (ICE*Kp3*) were each present in 11.7% (2 strains). Salmochelin (glycosylated derivative of enterobactin, another siderophore) had the highest detection rate at 88.2%; aerobactin (siderophore) and *RmpA/RmpA2* (synthesis of the bacterial capsule) were found in 82.3% of strains, and the *wzi* (capsule assembly protein) gene was present in 70.6% of strains. Colibactin (a genotoxin, induces DNA double-strand breaks) displayed the lowest detection rate at 35.3%. The mucoviscosity phenotype-regulating gene, *rmpA*, which is located on the large plasmids of these virulent strains, primarily controls capsule synthesis, resulting in a highly mucoviscous phenotype, observed in 52.9% of the strains. The string test has traditionally identified hvKp strains, but recent studies show that several genetic markers offer high diagnostic accuracy (> 0.95) for detecting hypervirulent strains. These markers include the salmochelin gene (*iroB*), aerobactin gene (*iucA*), and mucoid phenotype regulators (*rmpA* and *rmpA2*). These genes serve as biomarkers to distinguish hypervirulent *K. pneumoniae* (hvKp) from classical strains (Russo et al., 2018). In this study, 16 out of the 17 strains carried one or more of these virulence genes, with only one hospital-acquired strain lacking these genes.

Antimicrobial susceptibility tests revealed that the 17 isolated strains of *K. pneumoniae* were highly sensitive to most antibiotics. Apart from the inherent bla_{SHV} gene, no other resistance genes were detected in any of the strains. Our whole-genome SNP analysis revealed significant genetic diversity among the *K. pneumoniae* isolates. Combined with clustering analysis and consideration of isolation sites and collection times, the findings suggest these strains were primarily community-acquired infections, not hospital-based clonal transmissions. Strains with the same ST type also showed no evidence of clonal spread (Supplementary Table 1). The analysis of ST types and virulence gene profiles reveals that the majority of the *K. pneumoniae* strains co-infected with COVID-19 represent common, widespread high-virulence strains prevalent in the Asia-Pacific region.

Traditionally, hypervirulent *K. pneumoniae* strains are mostly sensitive to conventional antibiotics, while clinically prevalent multidrug-resistant strains are generally not hypervirulent, while the strains exhibit high virulence, the limited resistance profiles may reflect an evolutionary trade-off, where acquiring extensive resistance genes could compromise fitness or virulence. Additionally, hypervirulent *K. pneumoniae* strains are traditionally associated with community-acquired infections, where antibiotic selection pressure is lower compared to hospital settings, potentially limiting the acquisition of resistance genes.

Clinical profiles and emergence of hypervirulent *K. pneumoniae* in COVID-19 patients post policy shift

Clinical data indicate that these 17 patients are elderly and suffer predominantly from underlying health conditions. The median age is 71 years (IQR: 64–78), comprising 9 males and 8 females. The average duration of hospitalization was 11.2 ± 6.7 days. A significant proportion of the patients have comorbidities; 64.7% are diagnosed with hypertension, 41.2% with diabetes, and 35.3% with cardiovascular diseases. These patients were hospitalized due to COVID-19 infections or related complications from their existing health conditions. The clinical data, including imaging and laboratory findings, suggest that the majority of these COVID-19 patients were carriers of high-virulence strains. Furthermore, due to China's COVID-19 control policies, most hospitals in the country did not admit COVID-19 patients, as these patients were restricted to treatment in designated hospitals. It was only after the policy shift in December 2022 that

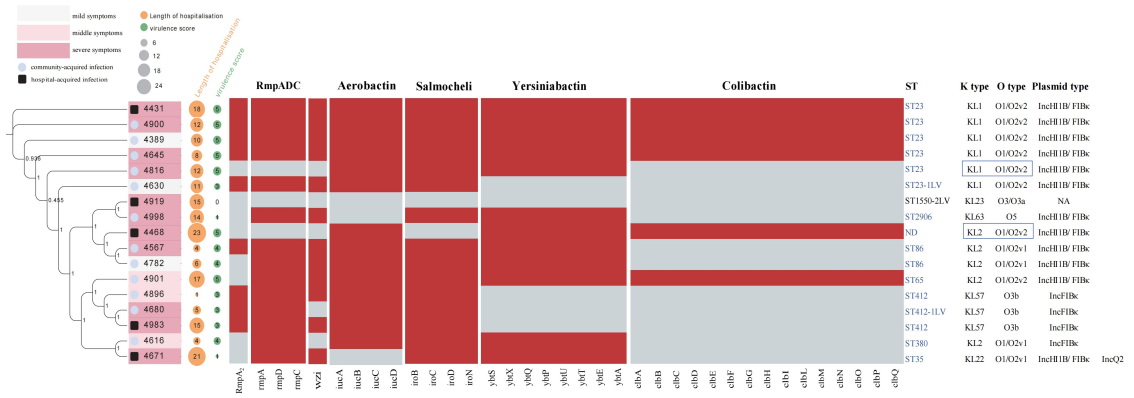


FIGURE 1
The phylogenetic tree of 17 Kp strains, along with their hospitalization durations, virulence genes, and MLST typing. A total of 17 *K. pneumoniae* (Kp) strains were analyzed, as shown in the left figure, illustrating the evolutionary relationship of Kp strains associated with COVID-19 pneumonia. The intensity of pink within the boxes represents varying degrees of clinical subtypes: deep pink indicates severe conditions, light pink indicates moderate conditions, and light gray represents mild conditions. A black frame around the sample number denotes hospital-acquired infections, while a blue circle indicates community-acquired infections. An orange circle represents the duration of hospitalization, and a green circle reflects virulence gene scores, with the size of each circle proportional to its respective value. Virulence scores are defined as follows: score 0 = no yersiniabactin, colibactin, or aerobactin; 1 = yersiniabactin only; 2 = yersiniabactin and colibactin (or colibactin only); 3 = aerobactin without yersiniabactin or colibactin; 4 = aerobactin with yersiniabactin (no colibactin); and 5 = yersiniabactin, colibactin, and aerobactin. The central heatmap indicates gene presence, with red representing positive genes and gray for negative ones. The right side displays MLST typing and serotypes, with hypervirulent strain ST types highlighted in blue font. Strains with K locus and O locus sequencing issues are shown in the blue boxes on the right.

COVID-19 patients were widely admitted to general hospitals, including ours. This sudden influx provided a unique opportunity to observe bacterial co-infections, including *K. pneumoniae*, under a high patient-load scenario. According to the experience of the clinical microbiology laboratory, there has been a noticeable increase in hypervirulent *K. pneumoniae* infections following the admission of COVID-19 patients. However, standardized testing and reporting for hypervirulent strains remain absent in current clinical microbiology practices, preventing the collection of comprehensive statistical data.

Also, During the COVID-19 pandemic, the World Health Organization observed a noticeable global increase in hypervirulent *K. pneumoniae* strains. This study identified a considerable number of hypervirulent *K. pneumoniae* strains despite the extremely high workload and staff shortages during the study period. Notably, none of these strains belonged to the ST11 hyper-resistant epidemic clone, which is encouraging as patients generally experienced favorable clinical outcomes.

Apart from the patient flow driven by policy changes or factors related to COVID-19 infection, several additional factors may have influenced the prevalence of these strains. Specifically, in healthcare settings, the increase in antimicrobial use was evident in both COVID-19 wards and non-COVID-19 wards. The overuse of antimicrobials, combined with changes in hospital microbiological ecosystems following the large-scale admission of COVID-19 patients, and the increased number of critically ill patients in overcrowded ICUs, likely created an environment more conducive to the survival and dissemination of multidrug-resistant (MDR) and hypervirulent strains. Additionally, altered immune responses in COVID-19 patients may have rendered them more susceptible to infections by hypervirulent bacteria. Moreover, the significant strain on healthcare systems during the pandemic—characterized by surging patient numbers and insufficient resource allocation—may have led to delays in infection control measures and antimicrobial treatments, further facilitating the spread of these hypervirulent strains. Collectively, these factors likely contributed to the post-COVID-19 emergence and prevalence of hypervirulent *K. pneumoniae*.

We acknowledge that the limited data collection timeframe coincided with this policy transition, focusing our observations on a unique and dynamic period. While this context underscores the relevance of our findings, it also limits the ability to observe trends across different COVID-19 waves or under varying healthcare system pressures. This limitation affects the generalizability of our conclusions to other timeframes or regions. To address this, we emphasize the need for future multi-center studies with extended observation periods to comprehensively assess bacterial co-infection trends over time. We plan to collaborate with more hospitals across different regions to investigate regional variations in the prevalence of hypervirulent *K. pneumoniae* strains. Additionally, we aim to analyze factors such as patient demographics, hospital settings, and infection control practices to better understand their impact on strain transmission and clinical outcomes. These studies would help elucidate the broader epidemiology of hypervirulent *K. pneumoniae* strains and their interaction with SARS-CoV-2, particularly under diverse healthcare and pandemic conditions.

Conclusion

In conclusion, the widespread prevalence of carbapenem-resistant hypervirulent *K. pneumoniae* strains, as highlighted by WHO, is likely predicated on the prior widespread circulation of hypervirulent strains. However, current global and national antimicrobial resistance surveillance systems primarily focus on detecting multidrug resistant strains or isolates from sterile body fluids, with limited attention given to strains isolated from

sputum, leaving the true prevalence of hypervirulent strains unclear. It is possible that these hypervirulent strains have acquired carbapenem-resistant plasmids, leading to the observed epidemic. Recently, a study has revealed that during the COVID-19 pandemic, multiple clusters of carbapenem-resistant and hypervirulent *K. pneumoniae* (CR-hvKp) emerged and exhibited significant clonal spread. Post-pandemic, both the prevalence and incidence of CR-hvKp have increased markedly, providing robust support for our hypothesis (Liu et al., 2024). Although our findings are based on a single-center report, they suggest that COVID-19 may have facilitated the widespread dissemination of hypervirulent *K. pneumoniae* strains, potentially explaining the current epidemic. More epidemiological data are needed to confirm this, as these strains have highly mutable genomes and can easily acquire carbapenem-resistant plasmids.

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 below: https://ngdc.cncb.ac.cn/, CRA019953.

Ethics statement

The studies involving humans were approved by the Ethics Committee of the First Hospital of Shanxi Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. 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

JZ: Conceptualization, Data curation, Formal analysis, Writing – original draft. QL: Data curation, Formal analysis, Investigation, Writing – original draft. JL: Data curation, Formal analysis, Investigation, Writing – original draft. FF: Validation, Visualization, Writing – original draft. YS: Funding acquisition, Project administration, Supervision, Writing – review and editing. XY: Funding acquisition, Project administration, Supervision, Writing – review and editing.

Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the National Natural Science Foundation of China (NSFC): 82202569; supported by the Central Guidance for Regional Science and Technology Development Projects: YDZJSX2024B010; and supported by The Basic Research Program of Shanxi Province: 202403021212264.

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.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

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.2025.1535893/full#supplementary-material

References

Arcari, G., Raponi, G., Sacco, F., Bibbolino, G., Di Lella, F., Alessandri, F., et al. (2021). *Klebsiella pneumoniae* infections in COVID-19 patients: A 2-month retrospective analysis in an Italian hospital. *Int. J. Antimicrob. Agents* 57:106245. doi: 10.1016/j.ijantimicag.2020.106245

das Chagas, L., Araujo, S., Pontes Serra, C., Araujo, K., Cunha, M., and Correie, A. (2024). Co-infection of SARS-CoV-2 and *Klebsiella pneumoniae*: A systematic review and meta-analysis. *Diagnostics* 14:1149.

Fan, H., Zhou, L., Lv, J., Yang, S., Chen, G., Liu, X., et al. (2023). Bacterial coinfections contribute to severe COVID-19 in winter. *Cell Res.* 33, 562–564. doi: 10.1038/s41422-023-00821-3

Ficik, J., Andrezál, M., Drahovská, H., Böhmer, M., Szemes, T., Liptáková, A., et al. (2023). Carbapenem-resistant *Klebsiella pneumoniae* in COVID-19 Era—challenges and solutions. *Antibiotics* 12:1285. doi: 10.3390/antibiotics12081285

Lansbury, L., Lim, B., Baskaran, V., and Lim, W. S. (2020). Co-infections in people with COVID-19: A systematic review and meta-analysis. *J. Infect.* 81, 266–275. doi: 10.1016/j.jinf.2020.05.046

Liu, C., Guo, J., Fan, S., Guo, W., Qi, H., Baker, S., et al. (2024). An increased prevalence of carbapenem-resistant hypervirulent *Klebsiella pneumoniae* associated with the COVID-19 pandemic. *Drug Resistance Updates* 77:101124. doi: 10.1016/j.drup.2024.101124

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

Pourajam, S., Kalantari, E., Talebzadeh, H., Mellali, H., Sami, R., Soltaninejad, F., et al. (2022). Secondary bacterial infection and clinical characteristics in patients with COVID-19 admitted to two intensive care units of an academic hospital in iran during the first wave of the pandemic. *Front. Cell. Infect. Microbiol.* 12:784130. doi: 10.3389/fcimb.2022.784130

Russo, T., Olson, R., Fang, C., Stoesser, N., Miller, M., MacDonald, U., et al. (2018). Identification of biomarkers for differentiation of hypervirulent *Klebsiella pneumoniae* from classical *K. pneumoniae*. *J. Clin. Microbiol.* 56:e00776-18. doi: 10.1128/JCM.00776-18

World Health Organization (2024). *Antimicrobial Resistance, Hypervirulent Klebsiella pneumoniae - Global Situation*. Available: https://www.who.int/emergencies/disease-outbreak-news/item/2024-DON527 (accessed September 13, 2024).

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 26 February 2025
DOI 10.3389/fmicb.2025.1540843

OPEN ACCESS

EDITED BY
Xiangyu Fan,
University of Jinan, China

REVIEWED BY
Yanan Wang,
Henan Agricultural University, China
Sudhir Kumar,
Iowa State University, United States

*CORRESPONDENCE
Bin Zhou
✉ zhoubin@wzhospital.cn

RECEIVED 17 December 2024
ACCEPTED 24 January 2025
PUBLISHED 26 February 2025

CITATION
Yuan Y, Li P, Shen W, Li M, He X and Zhou B (2025) Genomic identification of a pair of multidrug-resistant but non-pathogenic *Salmonella enterica* serovar Goldcoast isolates in southeast China. *Front. Microbiol.* 16:1540843. doi: 10.3389/fmicb.2025.1540843

Genomic identification of a pair of multidrug-resistant but non-pathogenic *Salmonella enterica* serovar Goldcoast isolates in southeast China

Yongjuan Yuan[1], Ping Li[2], Wei Shen[1], Min Li[1], Xiaofei He[3] and Bin Zhou[3]*

[1]Jiashan County Center for Disease Control and Prevention, Jiaxing, China, [2]Jiaxing Center for Disease Control and Prevention, Jiaxing, China, [3]Medical Research Center, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang Province, China

Introduction: *Salmonella* is an important foodborne pathogen that can induce severe diseases such as gastrointestinal disease and typhoid fever. Accumulating evidence revealed that *Salmonella*'s resistance to antibiotics also seriously affects human health. Pathogenic *Salmonella enterica* serovar Goldcoast (*S.* Goldcoast) was first detected in 2010 in China and was predicted to have an increasing tendency.

Methods: The MacConkey agar, Salmonella Shigella agar, three-sugar iron agar slant, and Gram-stained microscopic examination were used for strain identification. Gram-negative bacteria identification cards explored more properties of the isolates, while antimicrobial susceptibility testing was used to examine the multidrug resistance. The 2nd and 3rd generation sequencing revealed the genetic information of the isolates.

Results: Two non-pathogenic isolates with multidrug resistance, JS33 and JS34, harbored 42 antibiotic-resistant genes (ARGs) in contig1 and 13 ARGs in contig2, were isolated from a healthy donor living in southeast China and identified as *S.* Goldcoast (6,8:r:l,w). Interestingly, JS33 and JS34 showed identical responses to more than 20 antimicrobial agents and were resistant to ampicillin, selectrin, chloramphenicol, tetracycline, and streptomycin. However, JS33 differed from JS34 in hydrogen sulfide (H_2S) generation. The genomic sequencing identified a deletion in thiosulfate reductase (K08352) in JS34.

Discussion: H_2S is an essential physiological regulator linked to inflammation and cancer. Therefore, genomic identification of JS33 and JS34 provided us with a better understanding of drug resistance and could be used as model strains to study the effects of microbial H_2S production on the host. Since JS33 and JS34 did not induce gastrointestinal infection or other clinical symptoms as previously reported, the appearance of non-pathogenic *S.* Goldcoast in southeast China warned us to prepare for the prevalence of antimicrobial-resistant *S.* Goldcoast in China.

KEYWORDS

***Salmonella enterica* serovar Goldcoast, multidrug resistance, β-Lactamase, H_2S generation, thiosulfate reductase**

1 Introduction

Salmonella is a common foodborne pathogen, and *S. enterica* and *S. bongori* can induce severe diseases, such as gastrointestinal disease and typhoid fever (Lamichhane et al., 2024). The Disease Control and Prevention (CDC) center estimates 1.35 million infections, 26,500 hospitalizations, and 420 deaths annually, with about $3.3 billion in costs in the United States due to *Salmonella* infection (Hoffmann et al., 2012). *S. enterica* was dominant in causing human infection in China; that serovars *S. typhi* and *S. paratyphi*-A, B, or C can cause typhoid and paratyphoid fevers in humans, whereas other serovars are loosely described as non-typhoidal *Salmonella* (NTS), accounting for more than 98% of *S. enterica* isolates (Manesh et al., 2021).

S. Goldcoast is a NTS with high plasmid carrier rates and the cytolethal distending toxin subunit B (cdtB toxin) commonly, accounting for 2.14% frequency of serovars in *S. enterica* isolates and 2.59% of human origin in China. The trend was expected to increase according to the analysis of temporal and spatial dynamics of antimicrobial-resistant *S. enterica* from 2006 to 2019 in China (Wang et al., 2023a). Based on the extensive study, we derived the first appearance of *S.* Goldcoast in 2010 in Fujian province and its prevalence in Shanghai (Wang et al., 2023a). Few cases were reported in Zhejiang province, and all reported cases exhibited gastrointestinal infection or other clinical symptoms, independent of age and gender (Wang et al., 2023a).

Broad-spectrum antibiotics are used for bacteremia, invasive NTS infections, and disseminated typhoidal *Salmonella* infections (Smith et al., 2016; Gal-Mor et al., 2014). However, the response toward antibiotics varied depending on the bacteria's serotype and the host's immune response. *Salmonella*'s multidrug resistance (MDR), a global issue affecting countries at all income levels, leads to economic problems worldwide (Aleksandrowicz et al., 2023). *S.* Goldcoast showed a higher proportion of MDR rate of human origin (66.67%) than that of non-human origin (41.67%) (Wang et al., 2023a). Although the current understanding of MDR, including gene mutation, efflux pumps, passivating and inactivating enzymes encoded by drug resistance genes, and the transfer of genetic resistance gene elements in bacteria, has shed some light on the issue, the global community is still grappling with antimicrobial resistance, and further research, particularly into the underlying mechanisms of MDR in *Salmonella,* is crucial (Gaurav et al., 2023; Darby et al., 2023).

Moreover, most *Salmonella* produce hydrogen sulfide (H_2S), a beneficial gas regulating cardiovascular activity, nerve conduction, anti-inflammation, and metabolism if properly activated (Han et al., 2022). However, the mechanisms by which H_2S regulates various physiological functions remain unclear. In the present study, a pair of non-pathogenic *Salmonella* isolates were isolated from a healthy female and identified as *Salmonella enterica* serovar Goldcoast according to the White-Kauffmann-Le Minor antigenic table, indicating a growth of microbial diversity of *S.* Goldcoast in southeast China. Antibacterial drug sensitivity tests showed that both isolates had MDR to ampicillin, chloramphenicol, tetracycline, and streptomycin. However, their ability to generate H_2S was quite different. Therefore, deep sequencing of these two isolates was adapted to help us understand the mechanism underlying *Salmonella*'s multidrug resistance and H_2S generation.

2 Material and method

2.1 Reagents

Selenite Brilliant Green (SBG) enrichment solution (#HB8606, Qingdao Haibo), Blood plate (#CP10002, Shanghai Kemagar), *Salmonella Shigella* (SS) agar medium (#HB4089, Qingdao Haibo), Xylose lysine deoxycholate (XLD) agar medium (#HB4105, Qingdao Haibo), MacConkey agar medium (#HB6238-9, Qingdao Haibo), three-sugar iron agar slant (#HB4088, Qingdao Haibo), Gram staining solution (#HB8278, Qingdao Haibo), Gram-negative bacteria identification card (#21341, Merieux, France), *Salmonella* typing diagnostic serum (#882116,#152,116,#332,106, Senyan, Japan), Antimicrobial Susceptibility Testing (AST) panel for aerobic Gram-negative bacilli (#B3226B, Thermo Fisher, America), DNA extraction reagents (#51304, QIAGEN); all reagents were used within their expiry dates.

2.2 Equipment

Constant temperature incubator (MIR-H263L-PC, PHCBI), Optical microscope (CX21FS1, Olympus), Automatic microbial identification and drug sensitivity analysis system (VITEK 2 COMPACT, Merieux, France), Turbidimeter (DensiCHEK plus, Merieux, France), Microbial susceptibility instrument (Vizion®, Thermo), High-throughput sequencer (model: 550, Illumina) were used for sequencing bacterial genomes (2nd generation), etc.

2.3 Materials

The *S. enterica* isolates 2023JS33 and 2023JS34 were extracted from the stool sample of a healthy female, 52 years old, located in southeast China, without typhoid fever or any gastrointestinal complaints. The studies involving humans were approved by Committee of Zhejiang Provincial Center for Disease Control and Prevention. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided.

2.4 Isolation

An appropriate amount of feces was inoculated in SBG enrichment broth and incubated at 36 °C for 24 h. Then, broth containing bacteria was inoculated on SS agar medium and MacConkey agar medium by drawing lines in sections and incubated at 36 °C for 24 h. After that, a single colony was selected and inoculated with Triple Sugar Iron (TSI) and incubated at 36 °C for 24 h. Pick the interested bacterial species for microscopic examination with Gram stain and subsequently inoculate into blood plates and incubate at 36°C for 24 h. The purified bacterial species were identified by automatic biochemical identification and examined with the serum agglutination test.

2.5 Automatic biochemical identification

One to two single colonies were picked by inoculation rings and emulsified in sterile water. Adjust solution to 0.5 McFarland turbidity for biochemical identification with Gram-negative bacteria identification cards that had been rewarmed in advance. The sterilized saline was used as agglutination control and the agglutination phenomenon was observed within 2 min.

2.6 Serum agglutination test

An appropriate amount of *Salmonella* serum was dropped on a clean slide and mixed with the bacterial moss, picked out by an inoculation ring, and sterilized saline thoroughly. If no agglutination was observed, other commercial serums were used to conduct serum agglutination tests one by one, according to the instructions of reagents.

2.7 Antibacterial drug sensitivity test

The antimicrobial susceptibility of the isolates was determined by microdilution broth assay. In detail, tested isolates (2023JS33,2023JS34) and quality control strain (ATCC25922) were streaked and inoculated on blood agar plates and incubated at 36°C for 24 h. Individual colonies were picked with an inoculation ring seeded again on blood agar plates and incubated for 24 h at 36 °C. One or two colonies were picked from freshly prepared blood agar plates and emulsified in sterile water. Adjusted the solution to 0.5 McFarland turbidity and mixed thoroughly. Then, the bacterial suspension prepared above 10 μL was added to a test tube containing 11 mL cation-adjusted Mueller-Hinton broth (CAMHB) and mixed well. The mixture should be used within 15 min. Replace the test tube cover with a Sensititre® disposable sampling head and add the sample to the CHNENF drug sensitivity test plate according to AIM® instructions. Remove the test tube/sampling head combination from AIM® within 30 s after completion of sample loading in the drug susceptibility plate.

After the inoculation of the drug sensitivity plate, the purity of the final culture solution was checked, and all micropores were covered with a sealing film. After the incubation at 36°C for 24 h, all samples were read with a microbial susceptibility instrument, Vizion®. The minimum inhibitory concentration (MIC) of the drugs that naked eye could see was recorded and defined as sensitive (S), moderately sensitive (I) and resistant (R) according to the standard of Clinical and Laboratory Standards Institute (CLSI) (2023). The quality control strain was *Escherichia coli* ATCC25922. As CLSI does not provide streptomycin resistance breakpoint, it was determined according to the National Antimicrobial Resistance Monitoring System (NARMS) MIC criteria [Centers for Disease Control and Prevention (CDC), 2018].

2.8 Sequencing

Two isolates (2023JS33 and 2023JS34) were sent to the genetic testing laboratory of Zhejiang Tianke High-tech Development Co., Ltd. for deep sequencing (3rd generation sequencing). Whole genomic DNA was extracted by Gentra Puregene Yeast/Bact Kit (Qiagen, Valencia, CA) and sequenced using the GridION X5 platform (Oxford Nanopore Technology).

3 Results

3.1 Identifying *Salmonella enterica* serovar Goldcoast strains with different H_2S generation capacities

Colorless, translucent, and smooth round colonies were observed on MacConkey agar medium (Figure 1A), supporting that JS33 and JS34 belonged to *Salmonella* (Farhoudi Moghaddam et al., 1988). In addition, JS33 and JS34 formed round, moist, smooth, translucent colonies that became lighter in color on the SS agar medium (Figure 1B). However, JS33 was a colony with a black center, which differed from JS34 (Figure 1B). These two bacterial isolates with inconsistent morphology on the SS agar medium were selected and inoculated on the three-sugar iron agar slant (Figure 1C). It showed that both isolates fermented glucose and produced acid and gas but did not ferment lactose and sucrose, as these two gas-produced (+) isolates showed acid (K) on the slant and alkali (A) on the bottom (Figure 1C).

Interestingly, JS33 (K/A++) generated H_2S (+) (black), while JS34 (K/A + -) did not produce H_2S (−) (Figure 1C). Gram-stained microscopic examination revealed both isolates as Gram-negative bacilli due to the appearance of a loosely distributed red color (Figure 1D). Gram-negative bacterial identification cards also identified high similarity in JS33 and JS34 except for the production of H_2S (Table 1, bold text). The serotypes of the two suspected *Salmonella* isolates were 6, 8: r: l, w, which could be identified as *Salmonella enterica* serovar Goldcoast according to the White-Kauffmann-Le Minor antigenic table. The 2nd generation sequencing also identified JS33 and JS34 as *S.* Goldcoast and there were structural variations in the JS34 assembled genome and some of the original reads, compared with the reference genome (Figure 1E).

3.2 JS33 and JS34 isolates show multidrug resistance

Based on the latest version of CLSI breakpoints, JS33 and JS34 were evaluated with MDR to ampicillin (AMP), compound sulfamethoxazole (or selectrin, SXT), chloramphenicol (CHL), tetracycline (TET), and streptomycin (STR) (Table 2). They were both intermediate-resistant to ampicillin/sulbactam (AMS), colistin (CT), and polymyxin (BPOL). Besides, JS33 was sensitive to cefazolin (CFZ), while JS34 was intermediate-resistant (Table 2). Both JS33 and JS34 were sensitive to azithromycin (AZM), ciprofloxacin (CIP), nalidixic acid (NAL), and gentamicin (GEN), as well as cefotaxime (CTX), ceftazidime (CAZ), cefoxitin (CFX), imipenem (IPM), amoxicillin/clavulanic acid (AMC), cefuroxime (CXM), cefepime (CPM), ceftazidime-avibactam (CZA), meropenem (MEM), ertapenem (ETP), tigecycline (TGC), and amikacin (AMI).

3.3 Deep sequencing reveals ten protein-coding genes exclusively expressed in either JS33 or JS34

To elucidate the genetic background, we extracted DNA samples from purified JS33 and JS34 isolates and subjected them to deep sequencing using GridION (Oxford Nanopore Technology). This process yielded a comparable annotated sequence number in Non-Redundant (NR), Swiss-port, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Clusters of Orthologous Genes (COG) databases in JS33 and JS34, respectively (Figure 2A). Ten protein-coding genes were exclusively expressed in either JS33 or JS34 (Figures 2A,B). Among them, the thiosulfate reductase (K08352), nitrate reductase (K02567), β-lactamase

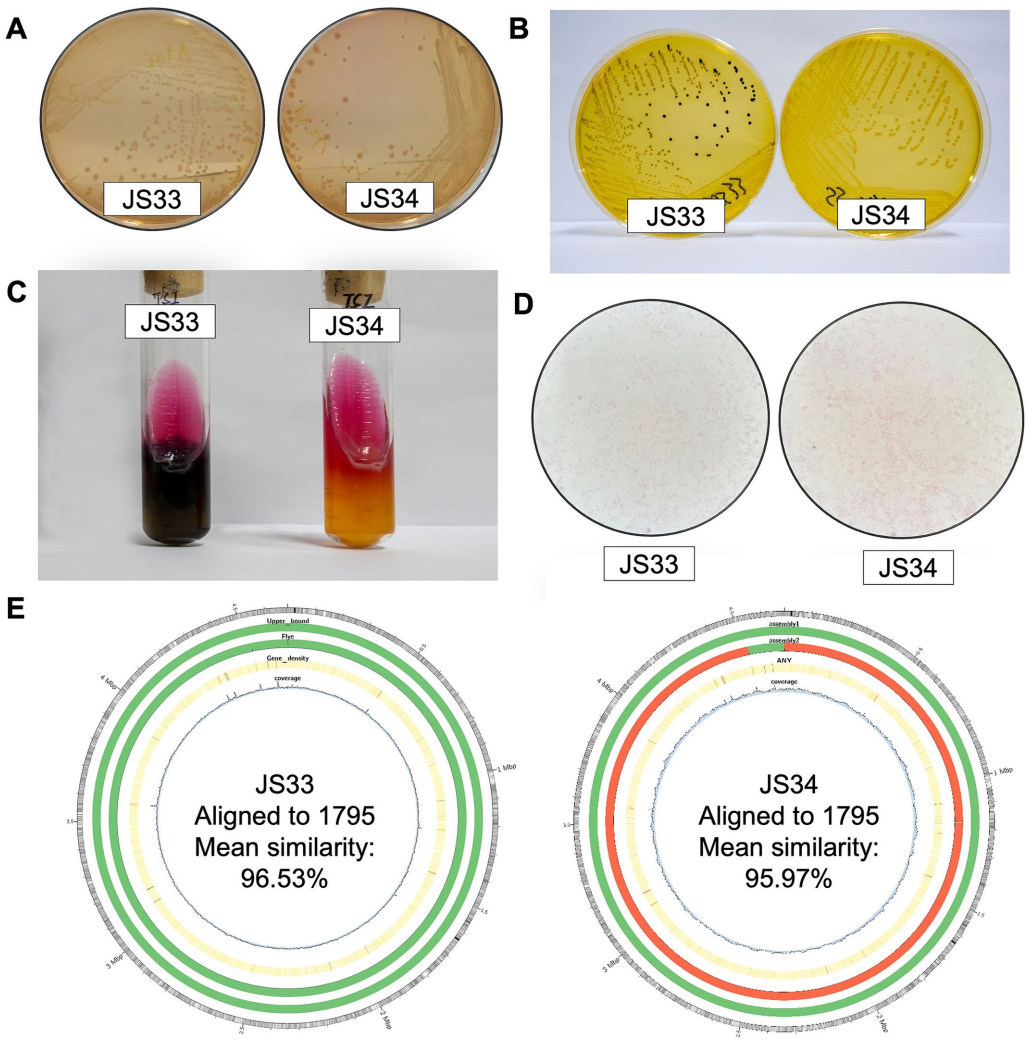


FIGURE 1
JS33 and JS34 are different in H_2S generation. **(A–C)** JS33 and JS34 were cultured in MacConkey agar **(A)**, SS agar medium **(B)**, and three-sugar iron agar slant **(C)**. **(D)** The Gram-stained microscopic examination of JS33 and JS34. **(E)** Aligned to CP091611.1 *Salmonella enterica* strain 1795 chromosome, complete genome (from outer to the inner: 1. GC ratio; 2. QUAST gives parametric assembly results; 3. Flye *De Novo* assembly results; 4. Gene density; 5. Depth of sequencing coverage).

(K18698), and clavulanate-9-aldehyde reductase (K12677) were unique to JS33. Other proteins, such as fibronectin-binding autotransporter adhesin (K19231), DNA (cytosine-5)-methyltransferase 1(K00558), and REP-associated tyrosine transposase (K07491), were expressed in both JS33 and JS34, but with different gene numbers (Figure 2B).

Despite the distinct genome, the function classification identified high genetic similarities between JS33 and JS34. The COG function classification was consistent in both isolates, with only minor variations in the number of genes (Figure 2C). Amino acid transport and metabolism, carbohydrate transport and metabolism, transcription, cell wall/membrane/envelope biogenesis, and energy production and conversion were the top five gene-enriched functions, highlighting the shared roles of JS33 and JS34. The KEGG analysis consistently emphasized the roles of JS33 and JS34 in regulating metabolism and participating in genetic information processing (Supplementary Figure S1).

3.4 β-Lactamase is related to MDR, while the deficiency of thiosulfate reductase inhibits H_2S production in JS34

Deep sequencing also revealed that JS33 and JS34 were closely related to infectious disease and drug resistance, while one more gene was identified in JS33, which encoded β-lactamase class A (Figure 2B; Supplementary Figure S1). β-lactamases are the most common reason resulting in resistance to β-lactam antibiotics in Gram-negative bacteria (Bush and Bradford, 2019). By combing the detailed KEGG classification with gene identification, we found that the exclusively expressed K12677 and K18698 participated in the biosynthesis of secondary metabolites, butanoate metabolism, β-lactam resistance, and clavulanic acid biosynthesis. They were responsible for mild differences in drug response between JS33 and JS34 (Figure 3A). K02567 and K08352 participated in energy metabolism by regulating nitrogen and

10.3389/fmicb.2025.1540843

TABLE 1 Gram-negative testing of JS33 and JS34.

JS33 *Salmonella* group (98% probability)											
APPA	–	ADO	–	PyrA	–	IARL	–	dCEL	–	BGAL	–
H2S	+	BNAG	–	AGLTp	–	dGLU	+	GGT	–	OFF	+
BGLU	–	dMAL	+	dMAN	+	dMNE	+	BXYL	–	BAlap	–
ProA	–	LIP	–	PLE	–	TyrA	+	URE	–	dSOR	+
SAC	–	dTAG	–	dTRE	+	CIT	+	MNT	–	5KG	+
lLATk	+	AGLU	–	SUCT	+	NAGA	–	AGAL	+	PHOS	+
GlyA	–	ODC	+	LDC	+	IHISa	–	CMT	+	BGUR	–
O129R	+	GGAA	–	lMLTa	–	ELLM	–	lLATa	–		
JS34 *Salmonella* group (97% probability)											
APPA	–	ADO	–	PyrA	–	IARL	–	dCEL	–	BGAL	–
H2S	–	BNAG	–	AGLTp	–	dGLU	+	GGT	–	OFF	+
BGLU	–	dMAL	+	dMAN	+	dMNE	+	BXYL	–	BAlap	–
ProA	–	LIP	–	PLE	–	TyrA	+	URE	–	dSOR	+
SAC	–	dTAG	–	dTRE	+	CIT	+	MNT	–	5KG	+
lLATk	+	AGLU	–	SUCT	+	NAGA	–	AGAL	+	PHOS	+
GlyA	–	ODC	+	LDC	+	IHISa	–	CMT	+	BGUR	–
O129R	+	GGAA	–	lMLTa	–	ELLM	–	lLATa	–		

"+" represents a positive result, "–" represents a negative result. Bold text highlights the difference in the production of H_2S.

sulfur metabolism, respectively. Particularly, K08352 played an essential role in H_2S production (Figure 3B).

Moreover, the resistance gene identifier (RGI) identified 55 resistance genes (42 in config1 and 13 in config2, >50% identities, E-value<0.00001), while 53 ARGs were common in JS33 and JS34, and 45 of 55 ARGs showed more than 90% identities (Figure 4). Based on the analysis of virulence factors in pathogenic bacteria, we found that the gene encoded K19231 in JS33 was linked to the upaH gene, which regulates the AIDA-I type autotransporter protein, a rarely glycosylated protein. The gene encoded K02567 was associated with the nuoG gene and functioned as an anti-apoptosis factor. The gene encoded K08352 was related to narG and was involved in anaerobic respiration. However, the differentially expressed genes did not correlate with bacterial virulence in JS34.

Furthermore, the subcellular localization of secretory proteins in JS33 and JS34 were similar based on PSORTb analysis[1] (Yu et al., 2010). Most secretory proteins were located at cytoplasmic and fewer were in cytoplasmic membrane, while a few were in periplasmic. The Prophage prediction based on PHAge Search Tool Enhanced Release software (PHASTER) also showed high similarity in contig_1 and relatively less similarity in contig_2 between JS33 and JS34[2] (Arndt et al., 2016). Genomic island prediction based on Island Viewer and Crispr-Cas prediction based on CRISPR finder[3] were the same as each other (Couvin et al., 2018). No difference was observed in JS33 and JS34 based on the carbohydrate-active enzymes database.

1 http://www.psort.org

2 https://phaster.ca

3 https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index

4 Discussion and conclusion

JS33 and JS34 were identified as *Salmonella enterica* serovar Goldcoast with almost identical biological properties. The only remarkable difference between JS33 and JS34 was H_2S generation based on the three-sugar iron agar slant and Gram-negative bacterial identification. Consistent with the phenotypic observation, deep sequencing also identified high similarity in genetic information. Only ten genes and four proteins were exclusively expressed in JS33 or JS34. There were two different genes relating to antibiotic resistance, i.e., TEM-185 and TEM-163, and three relating to virulence factors, i.e., narG, nuoG, and upaH.

The K18698 represented *β*-lactamase, and K12677 represented clavulanate-9-aldehyde reductase affected resistance to antibiotics in JS33. Specifically, β-lactams are the most widely used antibacterial agents worldwide, while β-lactamases are capable of deacylating β-lactam-derived covalent complexes, representing the most critical resistance mechanism in Gram-negative bacteria (Mora-Ochomogo and Lohans, 2021). However, the presence of β-lactamases did not generate a remarkable impact on the efficacy of antibiotics in JS33 compared with JS34. It was because of the coexistence of clavulanate-9-aldehyde reductase catalyzed the biogenesis of clavulanic acid, an important inhibitor of β-lactamases in JS33 (Docquier and Mangani, 2018). Clinically, β-lactam antibiotics are frequently administered with a β-lactamase inhibitor, such as clavulanic acid, that protects the antibiotic from β-lactamase catalyzed degradation (Huttner et al., 2020).

K02567 represented nitrate reductase, and K08352 represented thiosulfate reductase, which participated in nitrogen metabolism and sulfur metabolism, respectively. Particularly, thiosulfate reductase deficiency resulted in an ultimate inhibition of H_2S production in JS34. Thiosulfate reductase activity is found in

TABLE 2 Results of drug-susceptibility testing.

Antimicrobial agents	JS33	JS34	MIC (ug/ml)		
			S	I	R
AZM	4(S)	4(S)	≤16/8	/16	≥32
AMP	>64(R)	>64(R)	≤8	16^	≥32
AMS	16(I)	16(I)	≤8/4	16/8^	≥32/16
CIP	0.25(S)	0.25(S)	≤0.25	0.5^	≥1
SXT	>8(R)	>8(R)	≤2/38	-	≥4/76
CHL	>64(R)	>64(R)	≤8	16	≥32
NAL	8(S)	8(S)	≤16	-	≥32
GEN	≤1(S)	≤1(S)	≤2	4^	≥8
TET	>32(R)	>32(R)	≤4	8	≥16
CTX	≤0.25(S)	≤0.25(S)	≤1	2^	≥4
CAZ	≤0.5(S)	1(S)	≤4	8^	≥16
CFX	4(S)	4(S)	≤8	16^	≥32
CFZ	2(S)	4(I)	≤2	4	≥8
IPM	≤0.25(S)	≤0.25(S)	≤1	2^	≥4
CT	0.25(I)	1(I)	-	≤2	≥4
BPOL	0.25(I)	2(I)	-	≤2	≥4
AMC	8(S)	8(S)	≤8/4	16/8^	≥32/16
CXM	8(S)	8(S)	≤8	16^	≥32
CPM	≤1(S)	≤1(S)	≤2	-	≥16
CZA	0.5(S)	0.5(S)	≤8/4	-	≥16/4
MEM	≤0.12(S)	≤0.12(S)	≤1	2^	≥4
ETP	≤0.25(S)	≤0.25(S)	≤0.5	1^	≥2
TGC	0.5(S)	0.5(S)	≤4	8	≥16
AMI	≤4(S)	≤4(S)	≤4	8^	≥16
STR	32(R)	>32(R)	≤16	-	≥32

"^" Drugs with potential to concentrate in urine.

numerous microorganisms, whereas the molecular mechanism of reductive cleavage of thiosulfate is not yet known in bacteria (Le Faou et al., 1990). The comparison between JS33 and JS34 could help to illustrate the role of thiosulfate reductase in H_2S production in bacteria. No other significant difference was observed between JS33 and JS34 in the subcellular localization of secretory protein, the Prophage prediction, or genomic island prediction.

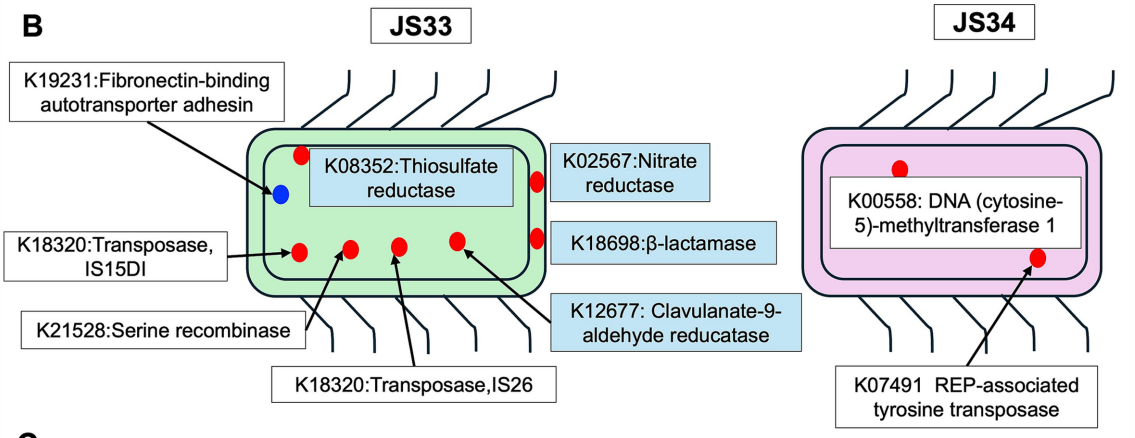
However, JS33 and JS34 differed from previously reported pathogenic *S.* Goldcoast in Zhejiang province in that they did not induce any clinical symptoms. There were three cases of *S.* Goldcoast that were collected in Zhejiang province, and two of them came from human hosts (Wang et al., 2023a, 2023b). The XXB830 (2015, 1-year-old, Female) was extracted from human feces which caused the gastrointestinal infection, while XXB1582 (2017, 67-year-old, Male) was extracted from the blood which caused the extra-intestinal infection. By downloading public files from the Chinese Local Salmonella Genome DataBase version 2, we compared previously reported cases with JS33 and JS34 (Wang et al., 2023b). Only 1 of 1,323 genes relating to virulence factors were different between XXB830 and XXB1582, but 169 of 1,323 genes relating to virulence factors disappeared in JS33 and JS34. Of course, the difference in accuracy between the second-generation sequencing and the third-generation sequencing may cause some errors in data analysis. On the contrary, JS33 showed the exact same antibiotic-resistant genes compared with XXB830. Moreover, the annotated sequence number ascribed in NR, Swiss-port, KEGG, and COG, as well as function classification, was similar in all *S.* Goldcoast samples.

Identifying new *S.* Goldcoast strains was consistent with previously estimated increased tendency and alerted a prevalence of *S.* Goldcoast with MDR in southeast China. Besides, due to the extensive similarities and specific differences between JS33 and JS34, they could perform as model strains to help us understand microbial antibiotic resistance and study microbial H_2S. However, the conclusions of this study are limited by the small sample size. As *S.* Goldcoast was the 16th serotype of *S. enterica* in China, accounting for 0.91% of geographical distribution, a comprehensive monitor of *S.* Goldcoast was needed but has not drawn much attention from the public (Wang et al., 2023a). The surveillance of non-pathogenic but multidrug-resistant isolates from healthy populations was especially unsatisfying and urgently needed.

10.3389/fmicb.2025.1540843

A

	JS33		JS34	
Database	Number	Percentage %	Number	Percentage %
NR	4499	99.60	4493	99.62
Swiss-Prot	3949	87.43	3941	87.38
KEGG	3733	82.64	3725	82.59
COG	3768	83.42	3760	83.37
ALL	4504	99.71	4498	99.73



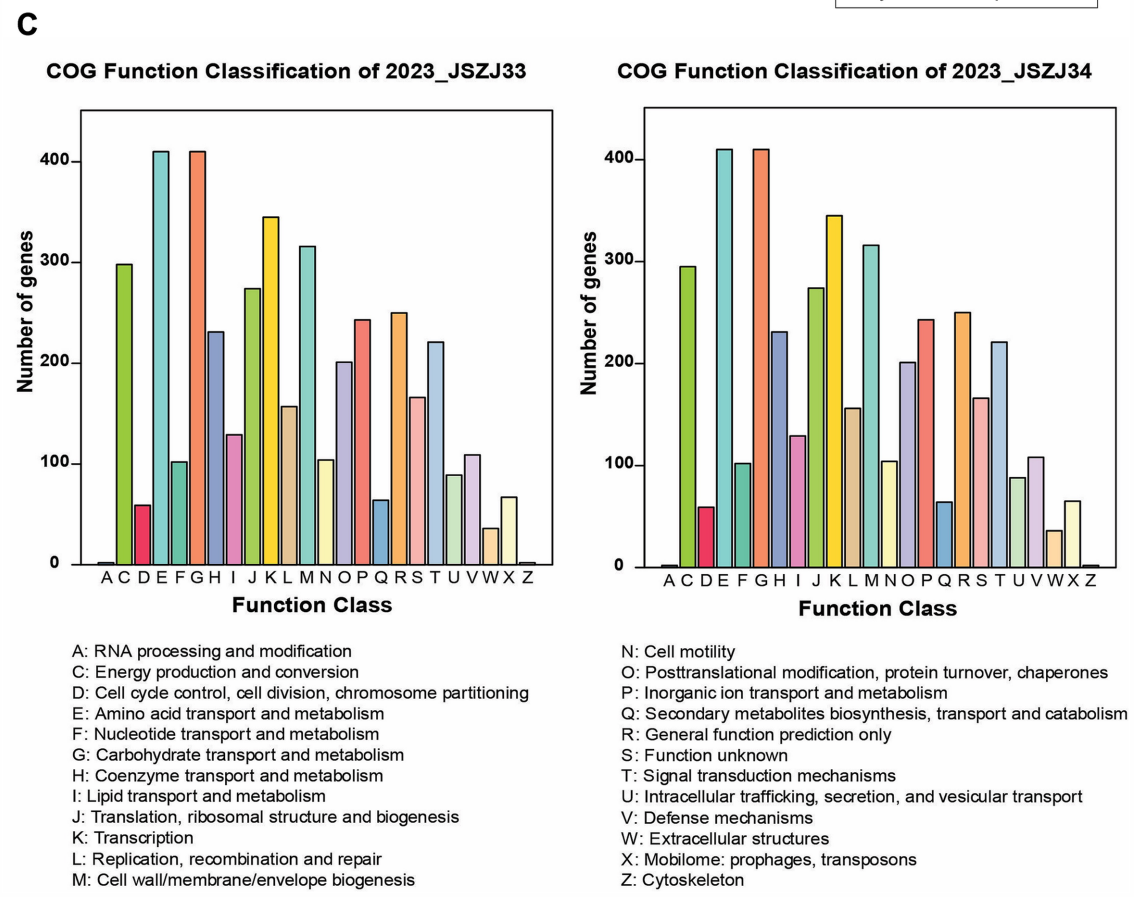


FIGURE 2
Ten protein-coding genes are exclusively expressed either in JS33 or JS34. **(A)** Annotated sequence number ascribed in NR, Swiss-port, KEGG, and COG in JS33 and JS34, respectively. **(B)** Exclusive protein-encoding genes in JS33 or JS34. The blue point represents proteins with unknown subcellular localization, while the red points represent proteins with specific localization. **(C)** The COG function classification of JS33 and JS34 samples.

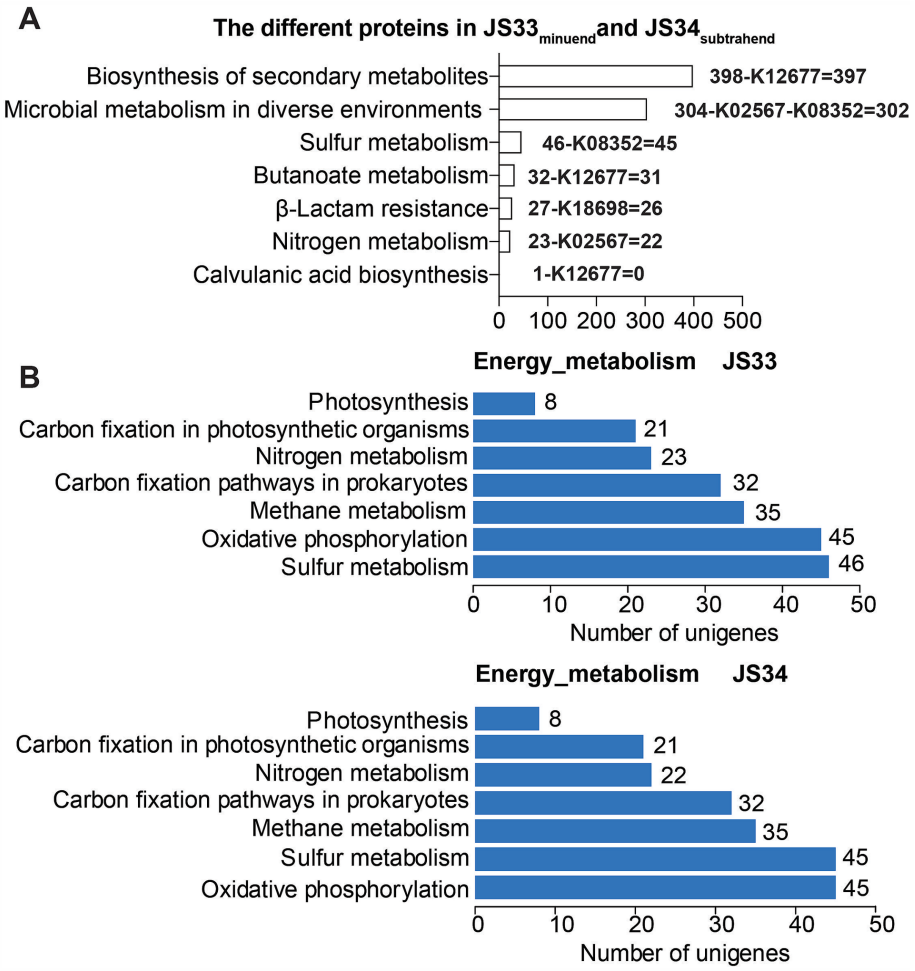


FIGURE 3
Distinct genome affects drug resistance and H_2S production. **(A)** The annotated sequence number of JS33 differed from JS34 in KEGG pathways. **(B)** The annotated sequence number of JS33 and JS34 in KEGG energy metabolism pathways.

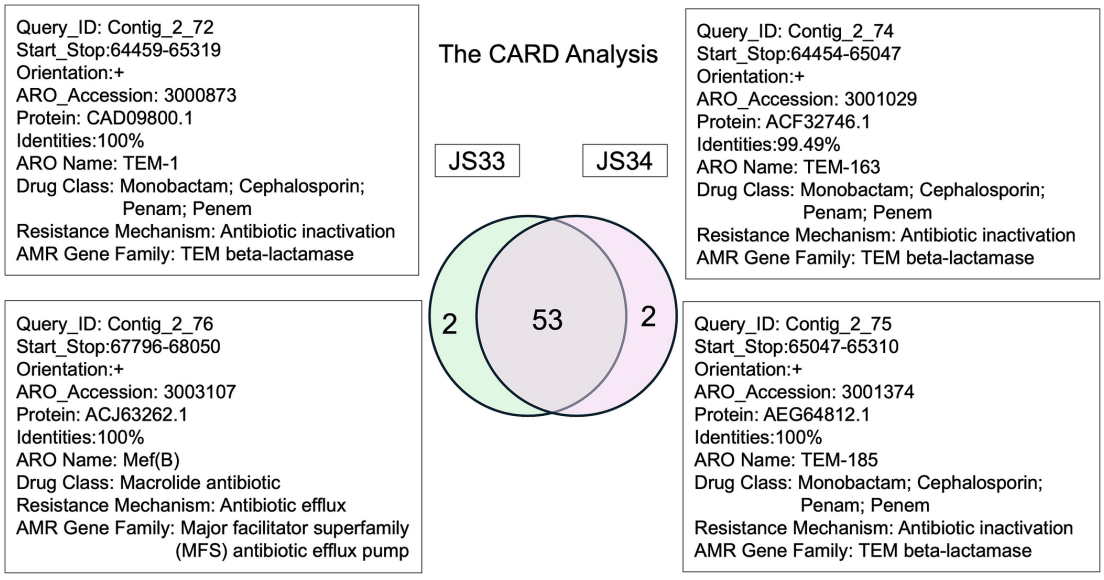


FIGURE 4
The comprehensive antibiotic resistance database in JS33 and JS34.

Data availability statement

The genome sequence of Salmonella enterica serovar Goldcoast strains JS33 and JS34 (Biosample ID: 2023_JSZJ_022 and 2023_JSZJ_023) can be accessed at DDBJ/ENA/GenBank under the accession number PQ613620. The R script and analysis report was deposited in GitHub (https://github.com) public repository, zhouli3-wz/JS33-and JS34.

Ethics statement

The studies involving humans were approved by Committee of Zhejiang Provincial Center for Disease Control and Prevention. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

YY: Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing. PL: Data curation, Formal analysis, Investigation, Software, Visualization, Writing – review & editing. WS: Investigation, Methodology, Supervision, Visualization, Writing – review & editing. ML: Investigation, Methodology, Resources, Writing – review & editing. XH: Formal analysis, Writing – original draft. BZ: Investigation, Project administration, Software, 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 supported by the Key Laboratory of Public Health Detection and Etiological Research of Zhejiang Province.

Acknowledgments

Authors are grateful to all staffs who were responsible for the specimen collection.

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.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

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.2025.1540843/full#supplementary-material

References

Aleksandrowicz, A., Carolak, E., Dutkiewicz, A., Błachut, A., Waszczuk, W., and Grzymajlo, K. (2023). Better together-Salmonella biofilm-associated antibiotic resistance. *Gut Microbes* 15:2229937. doi: 10.1080/19490976.2023.2229937

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

Bush, K., and Bradford, P. A. (2019). Interplay between β-lactamases and new β-lactamase inhibitors. *Nat. Rev. Microbiol.* 17, 295–306. doi: 10.1038/s41579-019-0159-8

Centers for Disease Control and Prevention (CDC). National antimicrobial resistance monitoring system NARMS 2015 human isolates surveillance report. Atlanta, GA: (2018). Available at: https://www.cdc.gov/narms/pdf/2015-NARMS-Annual-Report-cleared_508.pdf

Clinical and Laboratory Standards Institute (CLSI). (2023). M100 performance standards for antimicrobial susceptibility testing. 28 ed. Wayne, PA. Available at: https://community.clsi.org/media/1930/m100ed28_sample.pdf

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., et al. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 46, W246–w251. doi: 10.1093/nar/gky425

Darby, E. M., Trampari, E., Siasat, P., Gaya, M. S., Alav, I., Webber, M. A., et al. (2023). Molecular mechanisms of antibiotic resistance revisited. *Nat. Rev. Microbiol.* 21, 280–295. doi: 10.1038/s41579-022-00820-y

Docquier, J.-D., and Mangani, S. (2018). An update on β-lactamase inhibitor discovery and development. *Drug Resist. Updat.* 36, 13–29. doi: 10.1016/j.drup.2017.11.002

Farhoudi Moghaddam, A. A., Katouli, M., and Parsi, M. (1988). Comparison of Teknaf enteric agar and MacConkey/salmonella-shigella agar in evaluation of Shigella infection. *Lancet* 1, 1165–1166. doi: 10.1016/S0140-6736(88)91979-4

Gal-Mor, O., Boyle, E. C., and Grassl, G. A. (2014). Same species, different diseases: how and why typhoidal and non-typhoidal *Salmonella enterica* serovars differ. *Front. Microbiol.* 5:391. doi: 10.3389/fmicb.2014.00391

Gaurav, A., Bakht, P., Saini, M., Pandey, S., and Pathania, R. (2023). Role of bacterial efflux pumps in antibiotic resistance, virulence, and strategies to discover novel efflux pump inhibitors. *Microbiology (Reading)* 169:1333. doi: 10.1099/mic.0.001333

Han, S., Li, Y., and Gao, H. (2022). Generation and physiology of hydrogen sulfide and reactive sulfur species in Bacteria. *Antioxidants (Basel)* 11, 2487–2512. doi: 10.3390/antiox11122487

Hoffmann, S., Batz, M. B., and Morris, J. G. Jr. (2012). Annual cost of illness and quality-adjusted life year losses in the United States due to 14 foodborne pathogens. *J. Food Prot.* 75, 1292–1302. doi: 10.4315/0362-028X.JFP-11-417

Huttner, A., Bielicki, J., Clements, M. N., Frimodt-Møller, N., Muller, A. E., Paccaud, J. P., et al. (2020). Oral amoxicillin and amoxicillin-clavulanic acid: properties, indications and usage. *Clin. Microbiol. Infect.* 26, 871–879. doi: 10.1016/j.cmi.2019.11.028

Lamichhane, B., Mawad, A. M. M., Saleh, M., Kelley, W. G., Harrington, P. J., Lovestad, C. W., et al. (2024). Salmonellosis: an overview of epidemiology, pathogenesis, and innovative approaches to mitigate the antimicrobial resistant infections. *Antibiotics (Basel)* 13, 76–126. doi: 10.3390/antibiotics13010076

Le Faou, A., Rajagopal, B. S., Daniels, L., and Fauque, G. (1990). Thiosulfate, polythionates and elemental sulfur assimilation and reduction in the bacterial world. *FEMS Microbiol. Rev.* 6, 351–381. doi: 10.1016/0378-1097(90)90688-M

Manesh, A., Meltzer, E., Jin, C., Britto, C., Deodhar, D., Radha, S., et al. (2021). Typhoid and paratyphoid fever: a clinical seminar. *J. Travel Med.* 28, 1–13. doi: 10.1093/jtm/taab012

Mora-Ochomogo, M., and Lohans, C. T. (2021). β-Lactam antibiotic targets and resistance mechanisms: from covalent inhibitors to substrates. *RSC Med. Chem.* 12, 1623–1639. doi: 10.1039/D1MD00200G

Smith, S. I., Seriki, A., and Ajayi, A. (2016). Typhoidal and non-typhoidal Salmonella infections in Africa. *Eur. J. Clin. Microbiol. Infect. Dis.* 35, 1913–1922. doi: 10.1007/s10096-016-2760-3

Wang, Y., Liu, Y., Lyu, N., Li, Z., Ma, S., Cao, D., et al. (2023a). The temporal dynamics of antimicrobial-resistant Salmonella enterica and predominant serovars in China. *Natl. Sci. Rev.* 10, 1–17. doi: 10.1093/nsr/nwac269

Wang, Y., Xu, X., Zhu, B., Lyu, N., Liu, Y., Ma, S., et al. (2023b). Genomic analysis of almost 8,000 Salmonella genomes reveals drivers and landscape of antimicrobial resistance in China. *Microbiol. Spectr* 11, e02080–e02023. doi: 10.1128/spectrum.02080-23

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., et al. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26, 1608–1615. doi: 10.1093/bioinformatics/btq249

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 19 March 2025
DOI 10.3389/fmicb.2025.1549260

OPEN ACCESS

EDITED BY
Renmao Tim Tian,
Illinois Institute of Technology, United States

REVIEWED BY
Eva Sanjuan,
University of La Laguna, Spain
Iddya Karunasagar,
Nitte University, India

*CORRESPONDENCE
Abani K. Pradhan
✉ akp@umd.edu

RECEIVED 20 December 2024
ACCEPTED 03 February 2025
PUBLISHED 19 March 2025

CITATION
Feng S, Ramachandran P, Blaustein RA and Pradhan AK (2025) Bioinformatics combined with machine learning unravels differences among environmental, seafood, and clinical isolates of *Vibrio parahaemolyticus*. *Front. Microbiol.* 16:1549260. doi: 10.3389/fmicb.2025.1549260

Bioinformatics combined with machine learning unravels differences among environmental, seafood, and clinical isolates of *Vibrio parahaemolyticus*

Shuyi Feng[1], Padmini Ramachandran[1,2], Ryan A. Blaustein[1] and Abani K. Pradhan[1,3]*

[1]Department of Nutrition and Food Science, University of Maryland, College Park, MD, United States, [2]Human Foods Program U.S. Food and Drug Administration, College Park, MD, United States, [3]Center for Food Safety and Security Systems, University of Maryland, College Park, MD, United States

Vibrio parahaemolyticus is the leading cause of illnesses and outbreaks linked to seafood consumption across the globe. Understanding how this pathogen may be adapted to persist along the farm-to-table supply chain has applications for addressing food safety. This study utilized machine learning to develop robust models classifying genomic diversity of *V. parahaemolyticus* that was isolated from environmental (n = 176), seafood (n = 975), and clinical (n = 865) sample origins. We constructed a pangenome of the respective genome assemblies and employed random forest algorithm to develop predictive models to identify gene clusters encoding metabolism, virulence, and antibiotic resistance that were associated with isolate source type. Comparison of genomes of all seafood-clinical isolates showed high balanced accuracy (≥0.80) and Area Under the Receiver Operating Characteristics curve (≥0.87) for all of these functional features. Major virulence factors including *tdh*, *trh*, type III secretion system-related genes, and four alpha-hemolysin genes (*hlyA*, *hlyB*, *hlyC*, and *hlyD*) were identified as important differentiating factors in our seafood-clinical virulence model, underscoring the need for further investigation. Significant patterns for AMR genes differing among seafood and clinical samples were revealed from our model and genes conferring to tetracycline, elfamycin, and multidrug (phenicol antibiotic, diaminopyrimidine antibiotic, and fluoroquinolone antibiotic) resistance were identified as the top three key variables. These findings provide crucial insights into the development of effective surveillance and management strategies to address the public health threats associated with *V. parahaemolyticus*.

KEYWORDS

comparative genomics, machine learning, *Vibrio parahaemolyticus*, virulence, antibiotic resistance

1 Introduction

Vibrio parahaemolyticus is a Gram-negative, halophilic bacterium that is widely distributed in estuarine, marine, and coastal surroundings, and frequently detected in diverse seafood products such as clams, shrimps, crabs, and oysters (Su and Liu, 2007). *V. parahaemolyticus* is an important foodborne pathogen that is responsible for illnesses associated with seafood

throughout the world and is often linked to the consumption of raw or improperly handled seafood (DePaola et al., 2003). Typical signs and symptoms triggered by the infection of *V. parahaemolyticus* encompass watery diarrhea, abdominal cramps, nausea, vomiting, fever, headache, and bloody diarrhea (Centers for Disease Control and Prevention, 2013). Immunocompromised persons are at highest risk for morbidity and mortality (Centers for Disease Control and Prevention, 2013). Outbreaks/infections caused by *V. parahaemolyticus* usually happen in regions with high water temperatures. However, the ongoing climate changes are believed to expand the prevalence of *V. parahaemolyticus* geographically and increase human exposure to *V. parahaemolyticus* on a global scale (Zhang W. et al., 2023). Therefore, the development of efficient management strategies to control the spread of *V. parahaemolyticus* and minimize related food safety risks is needed.

Native to estuarine environments, *V. parahaemolyticus* can become a problematic contaminant among the microflora in shellfish as it takes on a broad niche range. In general, bacterial attachment and internationalization are described as the two critical processes mediating its transmission to and persistence in raw seafood (Brauge et al., 2024). Human consumption of contaminated seafood products may then result in the development of foodborne illness. However, the specific mechanisms involved in *V. parahaemolyticus* transmission and survival across diverse lifestyles, from the environment to seafood and consumers remain unclear. Thermostable direct haemolysin (TDH) and thermostable-related haemolysin (TRH) are the two major virulence factors in *V. parahaemolyticus* that may play important roles, as *tdh* and *trh* have been identified as reliable gene markers for the detection of pathogenic strains due to their prevalence in clinical isolates (Raghunath, 2015). Nevertheless, several studies have demonstrated that *tdh* and *trh* negative strains also cause infection, which indicates that additional virulence factors may be involved as well (Chao et al., 2010; Velazquez-Roman et al., 2012; Zha et al., 2023). Furthermore, while antibiotics have been widely adopted as the major treatment for *V. parahaemolyticus* infection, especially for severe cases (Loo et al., 2020), there is a growing concern for the emergence of antibiotic resistance among the species (Letchumanan et al., 2015; Letchumanan et al., 2016; Loo et al., 2020). Comparing the metabolism, virulence, and antibiotic resistance profiles of different *V. parahaemolyticus* isolates representing alternative lifestyles (i.e., waterborne, food-associated, and clinical) may provide a better understanding of its mechanisms for contamination, pathogenicity, and overall health risk.

Whole genome sequencing technologies have become increasingly utilized in the food industry for food safety monitoring assessment (Brown et al., 2019; Unrath et al., 2021). Given the complexity of sequencing data, machine learning (ML) can be applied to capture patterns in datasets with large quantities, and make robust predictions based on identified patterns (Tanui et al., 2022b; Karanth et al., 2022; Benefo et al., 2024a; Feng et al., 2024). Machine learning, particularly supervised ML, has demonstrated great applications in food safety such as predicting the disease outcome of *Salmonella*, the virulence potential and food source attribution of *Listeria monocytogenes*, as well as the abundance of *V. parahaemolyticus* (Tanui et al., 2022a; Ndraha et al., 2021; Karanth et al., 2022; Gmeiner et al., 2024). According to the models with good performance, the most influential predictors could also be retrieved, which shows great promise in managing and controlling food safety accurately. For example, Benefo et al. (2024a) adopted six different ML algorithms and identified the critical *Salmonella* stress response gene during poultry processing with high accuracy. Random forest (RF), as one of the most used ML algorithms in food safety, has been highlighted for its robust performance when the number of predictors is much larger than the number of observations, such as in WGS data (Biau and Scornet, 2016). Generally, the RF algorithm aggregates the prediction of several randomized decision trees through averaging, to obtain a final prediction/decision (Biau and Scornet, 2016). Thus, applying RF and alternative modeling efforts holds the potential to retrieve and reveal the information underlying bacterial behaviors from a genetic level via analyzing WGS data.

For this study, we aimed to perform a pangenomic analysis and apply RF to identify key genetic signatures of *V. parahaemolyticus* isolated from environmental, seafood, and clinical samples (i.e., potential differences in metabolism, virulence, and antibiotic resistance as a factor of source type). The findings from this study could help to (1) understand the adaptive response of *V. parahaemolyticus* as it transmits along the farm-to-table supply chain (environment-seafood-consumer) and (2) identify potential virulence factors and antibiotic resistance genes in *V. parahaemolyticus* that may have implications for consumer health and food safety.

2 Materials and methods

2.1 Sample collection

Genome assemblies of *V. parahaemolyticus* were collected from the National Center for Biotechnology Information (NCBI) Pathogen Detection database.[1] A total of 6,227 assemblies consisting of environmental (n = 633), seafood (n = 2,284), and clinical (n = 3,310) isolates were downloaded and used in this study after checking the isolation type and isolation source manually for each assembly. Assemblies were subset for further analysis based on specific inclusion criteria for having corresponding metadata that indicated specific sample sources (i.e., environmental, seafood, and clinical), as described in Supplementary Table S1.

2.2 Bioinformatics analysis

The selected genome assemblies were processed with CheckM (v1.2.2) (Parks et al., 2015) for quality control, and those predicted to have greater than 97% completeness and less than 3% contamination (n = 176, 975, and 865 for environmental, seafood, and clinical isolates, respectively) were further processed (Blaustein et al., 2019). Annotation and pangenome construction of these high-quality assemblies were performed with Prokka (v1.14.6) and Panaroo (v1.3.4), sequentially (Seemann, 2014; Tonkin-Hill et al., 2020). Genes identified in the pangenome were categorized into three different sets based on their prevalence across all strains analyzed: core genes were present in over 95% of isolates, shell genes were found between 15 to 95% isolates, while cloud genes were defined as those with a prevalence less than 15% isolates (Livingstone et al., 2018). In addition to the

1 https://www.ncbi.nlm.nih.gov/pathogens/, accessed on March 4, 2024.

comprehensive pangenome for all isolates, pangenomes for the subgroups of seafood and clinical isolates were constructed as well.

The nucleotide sequences of all gene clusters in the respective pangenomes were translated with Prodigal (v2.6.3) (Hyatt et al., 2010). Amino acid sequences were then screened for homology to proteins in the Database of Clusters of Orthologous Genes (COG), the Virulence Factor Database (VFDB) and the Comprehensive Antibiotic Resistance Database (CARD) using BLASTp (v2.14.1) (Camacho et al., 2009; Liu et al., 2022; Alcock et al., 2023) to identify the gene profiles with homology to features for metabolism, virulence and antibiotic resistance, respectively. During our preliminary analysis, different cutoff thresholds ranging from 99 to 50% (99, 98, 97, 96, 95, 90, 85, 80, 75, 70, 65, 60, 55, and 50%) were employed to query coverage and percent identity, as we aimed to get the threshold as high as possible while 50% is the common choice for BLASTp. The filtered genes with different thresholds were fed into RF models as the predictors. After comparing the performance of models (sensitivity, specificity, balanced accuracy, and Area Under the Receiver Operating Characteristics curve (AUROC)) using filtered genes with different cutoff values (Supplementary Tables S2–S7), the thresholds for both query coverage and percent identity were set as 90, 80, and 50% for metabolism, virulence, and antibiotic resistance models, respectively.

2.3 Machine learning

Random forest was adopted to develop predictive models for isolation sources of *V. parahaemolyticus* (environmental vs. seafood (ES) and seafood vs. clinical (SC)). The presence and absence of genes related to metabolism, virulence, and antibiotic resistance were separately used as the predictors. The overview of the prediction strategy used in this study is simplified as a workflow and displayed in Figure 1. Further details regarding this approach are described in the following sections.

2.3.1 Data preprocessing

The presence and absence of gene clusters (denoted by 1 and 0, respectively) with homology to each functional category (metabolism, virulence, and antibiotic resistance) were used as the input variables for the ML models. Predictors (gene clusters) possessing only one unique value (zero variance predictor) or a limited number of unique values (near-zero variance predictor) were removed as they could introduce unnecessary complexity to the model and lead to increased computational time without significantly increasing the accuracy of the model (Kuhn, 2019). Predictors with near-zero variance were detected by estimating frequency ratio (the frequency of the most prevalent value over the second most frequent value) and unique value percentage (the number of unique values to the total number of samples expressed as a percentage). For this study, a predictor with a frequency ratio greater than 19 and a unique value percentage less than 10% was considered as near-zero variance and, therefore, excluded from model building (Kuhn, 2019; Benefo et al., 2024a).

Class imbalance, which could result in potential bias in the model, was observed for ES (15.29% for the minority class (environmental isolates) and 84.71% for the majority class (seafood isolates)) while was not found in SC (47.01% for the

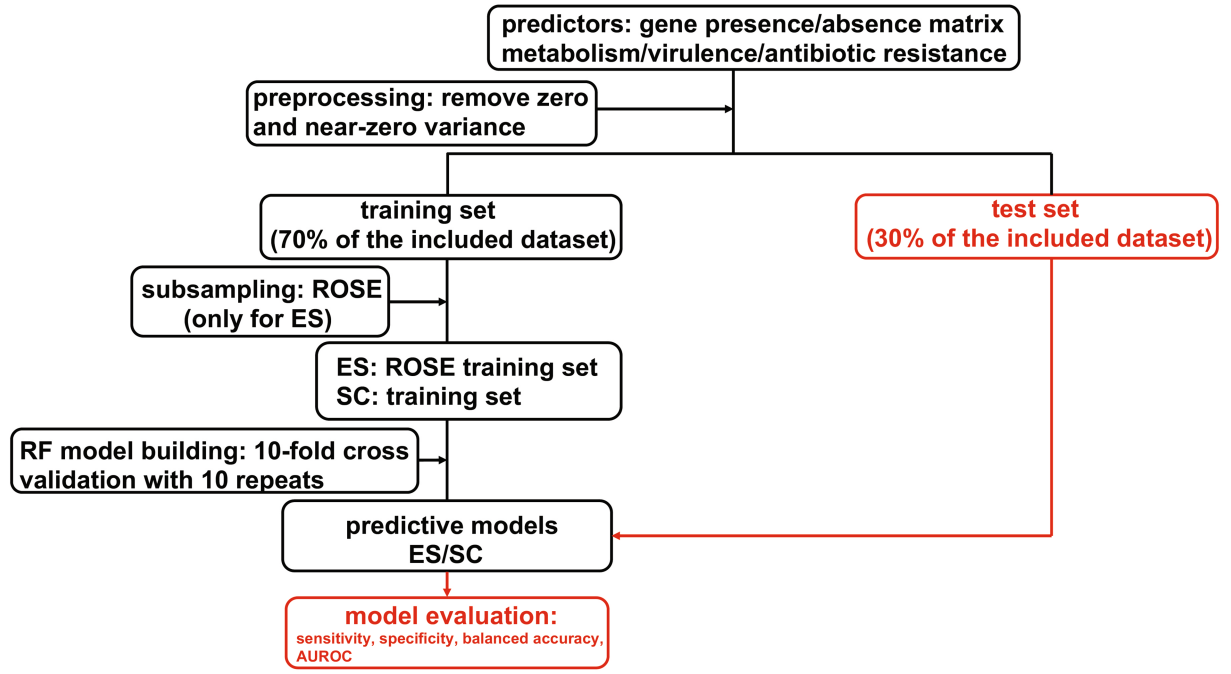


FIGURE 1
A simplified workflow for the approach used in this study.

minority class (clinical isolates) and 52.99% for the majority class (seafood isolates)). Upsampling, downsampling, random over-sampling examples (ROSE) technique, and Synthetic Minority Oversampling TEchnique were applied to attenuate the imbalance in the ES dataset during the preliminary analysis, and ROSE was selected for subsampling in the ES models due to having better predictive performance than the other methods. Through ROSE, the majority class is downsampled while new instances are generated via a smoothed-bootstrap approach for the minority class (Lunardon et al., 2014).

2.3.2 Model building

Six classification models were developed using RF for both ES and SC datasets across each functional category: metabolism, virulence, and antibiotic resistance. For each model, the dataset was randomly split into a training set (70% of included data) and a test set (30% of included data), which were used for model building and model test and validation, respectively (Benefo et al., 2024b). Ten-fold cross-validation with 10 repeats was adopted to train the model as it helps to reduce the potential bias (Kohavi, 1995). Specifically, the training set was randomly partitioned into 10 subsets, and 10 models were built out of these 10 subsets. For each model/iteration, nine subsets were employed to train the model while the remaining set was kept aside to test and evaluate the model performance. The aforementioned procedures were repeated 10 times, resulting in an average performance for all models generated throughout the process (Kuhn, 2019). Randomized search was adopted to tune the hyperparameters and identify the optimal ones for each model. The test of the developed models was conducted using the hold-out test set (30% of included data), and a confusion matrix was generated according to model performance on the test set.

2.3.3 Model evaluation

Sensitivity, specificity, balanced accuracy, and AUROC were used to evaluate the performance of the developed models. Sensitivity and specificity are commonly used metrics to evaluate the performance of classification models. Sensitivity is defined as the ratio of the correctly identified positives to all true positives, while specificity refers to the proportion of true negatives that are correctly predicted (Sidey-Gibbons and Sidey-Gibbons, 2019). Balanced accuracy, which is defined as the average of sensitivity and specificity, outperforms traditional accuracy when evaluating the performance of models with imbalanced data as it considers accuracies for both positive and negative classes (Thölke et al., 2023). Sensitivity, specificity, and balanced accuracy all range from 0 to 1; and the closer these values to 1, the better performance the model has. AUROC characterizes the classification (discrimination) ability of the model. Specifically, the value of AUROC varies from 0.5 to 1, with AUROC = 0.5 (baseline) linked to random classification while AUROC = 1 indicates a perfect classifier (D'Agostino et al., 2013). Moreover, the plots of AUROC were generated as well. In the AUROC graph, the false positive rate (1- specificity) of the model is the x-axis while the true positive rate (sensitivity) of the model is the y-axis. An AUROC curve which is close to the upper left corner of the graph is considered as the indicator of high AUROC value and therefore, good predictive ability of a model.

2.3.4 Significant genes identification

The twenty most significant genes for each reliable model were identified and ranked based on their importance (note: only 17 genes were listed for the SC-antibiotic resistance model since it only had 17 genes as predictors). It was estimated by computing the difference in the prediction accuracies of the model caused by permuting the values of each predictor variable. The calculated difference between the two accuracies was averaged over all trees and normalized by the standard error. The more significantly permuting the value of a predictor impacts the accuracy, the more important that predictor (Kuhn, 2019). All the ML analyses were performed using the *caret* and *MLeval* package (Kuhn, 2019) in R (v. 4.1.1). The prevalence rate (the ratio of positive genomes to the total genomes) of the identified genes was calculated. The Proportion test was performed using the *prop.test* package in R (v. 4.1.1) to evaluate the homogeneity of proportions in different isolate sources. In addition, relevant information about the specific genes in COG, VFDB, and CARD that were homologous to the most important pangenome gene cluster predictors (e.g., homologous gene COG category) were retrieved from the respective databases.

2.4 Data visualization

A pie chart was generated for the pangenome for all isolates. AUROC curves and heatmaps were generated for the prevalence of the identified important genes via R (v. 4.1.1) using the *autoplot* and *pheatmap* packages, respectively.

3 Results

3.1 Pangenome characteristics

A total of 42,324 gene clusters were identified in the *V. parahaemolyticus* pangenome, with 4,608 ± 160 genes per genome (mean ± SD). Specifically, our pangenome identified 3,880 core genes, 1,081 shell genes, and 37,363 cloud genes. The pie chart demonstrating the distribution of total genes and respective percentages is shown in Figure 2.

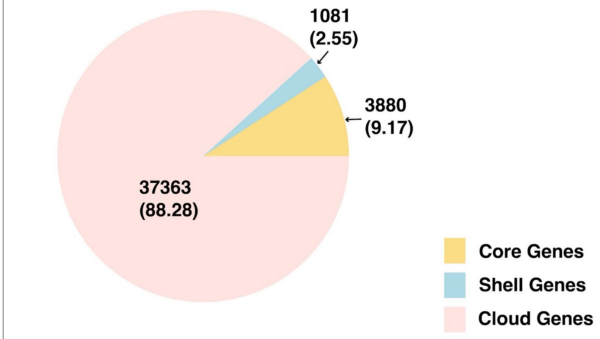


FIGURE 2
Pie chart of genes of the overall pangenome. Total genes (percent) are listed.

3.2 Predictive models

Six ML classification models were built based on the presence and absence of genes with homology to metabolism, virulence, and antibiotic resistance for the ES and SC datasets. Based on the thresholds for query coverage and percentage of identity, 4,132, 273, and 160 genes were selected as inputs for the metabolism, virulence, and antibiotic resistance models, respectively. After removing zero and near-zero variance, 390, 23, 19, 380, 48, and 17 genes were used as the predictors for ES-metabolism, ES-virulence, ES-antibiotic resistance, SC-metabolism, SC-virulence, and SC-antibiotic resistance models, respectively.

The performance of all six models was measured using sensitivity, specificity, balanced accuracy, and AUROC (shown in Table 1). Generally, SC models provided better predictions compared to ES models, and models for metabolism surpassed those for virulence and antibiotic resistance.

Specifically, sensitivity, specificity, and balanced accuracy varied from 0.44 to 0.52, 0.72 to 0.88, and 0.58 to 0.70, respectively for ES models; while for SC models, the range for sensitivity, specificity, and balanced accuracy were 0.73 to 0.88, 0.87 to 0.96, and 0.80 to 0.90, respectively. On the other hand, all models, except for ES-virulence and ES-antibiotic resistance, resulted in an AUROC value greater than 0.80 (ranging from 0.82 to 0.96), and a model with an AUROC value above 0.80 is generally interpreted as a reliable model (Nahm, 2022). The plotted AUROC curves were shown in Figure 3 and the baseline, of which AUROC is equal to 0.5, was denoted as the dotted diagonal line in the graph.

Based on the overall consideration of four evaluation metrics, SC-metabolism, SC-virulence, and SC-antibiotic resistance models were considered as models which could provide robust prediction and were selected for further identification of significant genes.

3.3 Significant genes enriched by source type

Twenty significant genes identified by the SC-metabolism and SC-virulence models, genes used as the predictors in the SC-antibiotic resistance model as well as the relevant information about their homologies in different databases and prevalence rates in the seafood and clinical groups were listed in Tables 2–4; and the related heatmaps were displayed as Figures 4–6.

As presented in the SC-metabolism model (shown in Table 2), the top 20 important genes were predicted as homologies to genes coding for proteins belonging to 13 different functional categories and intracellular trafficking, secretion, and vesicular transport, cell motility, as well as transcription were the most predominant categories. Most of the proportion of strains harboring the above genes (14 out of 20) were significantly greater in the clinical cohort than in the seafood group (displayed in Table 2; Figure 4).

According to the SC-virulence model (presented in Table 3), genes of great importance in characterizing the virulence profiles of seafood and clinical isolates belonged to six different functional categories and were primarily associated with exotoxin followed by effector delivery system. The proportion test revealed that the prevalence rates of 15 out of 20 important virulence genes differed significantly in seafood and clinical isolates. Among the genes with significantly different ubiquity, all of them were more encoded in clinical samples, other than '*flaD_1~~~flaD_3*' (*flaC*) (Table 3; Figure 5).

Gene clusters used as predictors in our SC-antibiotic resistance model were predicted to resist 12 different drug classes including three multidrug classes (Table 4; Figure 6), among which genes conferring tetracycline resistance, elfamycin resistance, as well as multi-drug resistance (*tet(35)*, *Ecol_EFTu_PLV*, and *MexS*) were the top three important genes. The most common antibiotic resistance genes in the seafood cohort were *macB* (macrolide resistance, 89.13%), *dfrA3* (diaminopyrimidine resistance, 89.13%), and *ugd* (peptide resistance, 80.51%), while the most common antibiotic resistance genes in the clinical cohort were *dfrA3* (diaminopyrimidine resistance, 97.11%), *macB* (macrolide resistance, 96.76%), and *Ecol_EFTu_PLV* (elfamycin resistance; 66.13%). On the other hand, five different antibiotic resistance mechanisms were involved in differentiating the antibiotic resistance of seafood and clinical samples, and efflux pump, as well as target site alteration, were the two major categories.

4 Discussion

The overarching goal of this study was to use the differences in the presence and absence of genes among *V. parahaemolyticus* isolates as ML input to (i) develop classification models that differentiate *V. parahaemolyticus* isolates from environmental, seafood, and clinical samples, based on the accessory genes they carry that encode critical functions (metabolism, virulence, and antibiotic resistance) and (ii) identify the specific genes underlying the differences. Understanding potential mechanisms involved in transmission, pathogenicity, and antibiotic resistance of *V. parahaemolyticus* along the seafood supply chain could inform new strategies for food safety control and public health surveillance. To our knowledge, this is the initial attempt to

TABLE 1 Model performance of the developed predictive models.

Models		Sensitivity	Specificity	Balanced accuracy	AUROC
ES	Metabolism	0.52	0.88	0.70	0.82
	Virulence	0.44	0.72	0.58	0.66
	Antibiotic resistance	0.52	0.76	0.64	0.70
SC	Metabolism	0.85	0.96	0.90	0.96
	Virulence	0.88	0.92	0.90	0.94
	Antibiotic resistance	0.73	0.87	0.80	0.87

ES, environmental vs. seafood; SC, seafood vs. clinical; AUROC, area under the receiver operating characteristics curve.

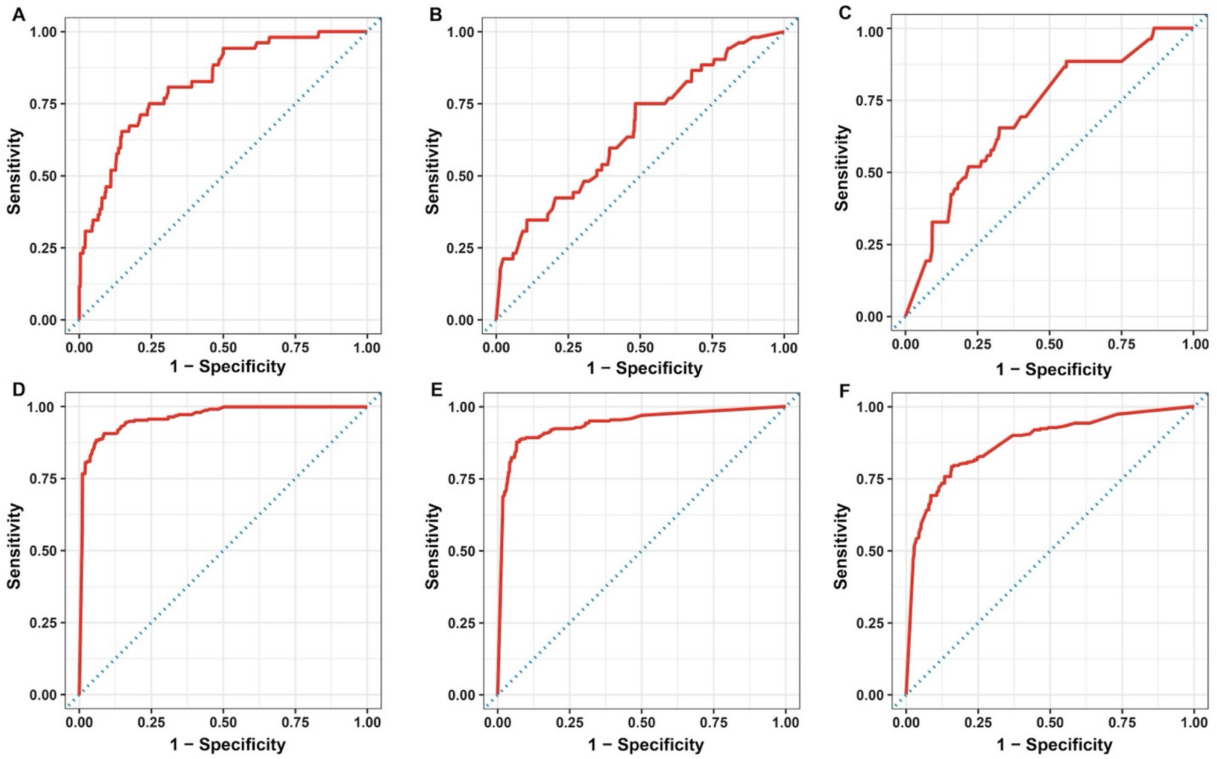


FIGURE 3
Area under the receiver operating characteristics curve for the developed RF models: ES-metabolism **(A)**, ES-virulence **(B)**, ES-antibiotic resistance **(C)**, SC-metabolism **(D)**, SC-virulence **(E)**, and SC-antibiotic resistance **(F)**. ES, environmental vs. seafood; SC, seafood vs. clinical.

adopt a bioinformatics workflow combined with ML to characterize differences in genetic diversity of *V. parahaemolyticus* strains across different isolation sources.

Our results showed that the three SC models could generate accurate predictions while the ES models did not perform as well. Therefore, only significant genes identified by SC models were analyzed and discussed. It is possible that compared with SC isolates, ES isolates were not that much different since these are all 'commensal, possibly pathogenic' strains recovered in monitoring while clinical strains are likely pathogens. However, limitations with the smaller sample size and data imbalance in the ES dataset may have affected the performance of ES models. In fact, significant biofilm formation was observed for *V. parahaemolyticus* in seafood compared with strains from the environment, implying the different lifestyles between environmental and seafood isolates (Rajkowski, 2009). Moreover, Feng et al. (2024) demonstrated that *V. parahaemolyticus* strains isolated from seawater and oyster were differently impacted by the same environmental parameters, indicating functional differences between certain environmental and seafood isolates as well. The inclusion of more environmental samples in the future should enable the model to capture and characterize the difference better.

In general, as shown in Table 2 and Figure 4, most of the top genes identified by our SC-metabolism model were more prevalent in clinical strains compared with seafood strains, indicating the more active metabolic activities occurring in clinical strains. This could be explained by the fact that the adaptative responses required to survive in the human body were more complicated than the ones associated with the seafood isolates due to the two distinguished conditions provided by the human body and seafood. When *V. parahaemolyticus* enters the human body, it could face various stresses such as thermal stress, acid stress, bile salts stress, and attack from the host cells, resulting in potential damage to different components of *V. parahaemolyticus* including cell membrane, DNA, and protein (Qadri et al., 2003; Pazhani et al., 2021). However, the stresses that seafood isolates may encounter are majorly associated with postharvest handling procedures such as cold stress caused by refrigeration storage and low salinity stress caused by washing (Huang and Wong, 2012; Tang et al., 2018). Thus, adaptive response of *Vibrio* along the processing and supply chain may become relevant for transmission and persistence that precedes consumption.

Specifically, the top two gene clusters ('*group_1266*' and '*group_5540*'), which were orthologous to cytoskeletal protein and superfamily II DNA or RNA helicase, were annotated as hypothetical proteins, pressing the need to study and reveal their functions and roles in the survival of *V. parahaemolyticus*. Intracellular trafficking, secretion, and vesicular, specifically, proteins associated with type III secretion system (T3SS), was one of the most predominant categories recognized by the SC-metabolism model and all the homologies (COG symbol: PulD, FliI, and EscV, ranked the third, fifth/eleventh, and sixth, respectively in the SC-metabolism model) belonging to this group were more prevalent in clinical isolates. Our findings were consistent with a previous study, in which the pangenome of *V. parahaemolyticus* was analyzed and significant enrichment of genes related to intracellular trafficking, secretion, and vesicular transport was observed for the clinical isolates (Pérez-Duque et al., 2021). This

TABLE 2 Twenty important genes identified by SC-metabolism model, information of their homologies from COG, and their prevalence rates.

Gene cluster from pangenome	COG annotation	COG symbol	COG name	COG functional category	Prevalence rate		*p* values
					Seafood	Clinical	
group_1266	Hypothetical protein VPA1391	RodZ	Cytoskeletal protein RodZ	D	12.10	87.98	< 0.001
group_5540	Hypothetical protein VPA1393	SSL2	Superfamily II DNA or RNA helicase	KL	12.10	87.98	< 0.001
sctC_5~~~sctC_3	Type III secretion system EscC protein	PulD	Type II secretory pathway component GspD/PulD (secretin)	U	3.69	60.69	< 0.001
legI_2~~~legI	N-acetylneuraminic acid synthetase	SpsE	Sialic acid synthase SpsE, contains C-terminal SAF domain	M	20.31	1.39	< 0.001
yscN_2~~~atpB_1	ATPase YscN	FliI	Flagellar biosynthesis/ type III secretory pathway ATPase FliI	NU	3.69	60.58	< 0.001
ssaV	Type III secretion system EscV protein	EscV	Type III secretory pathway, component EscV	U	3.69	60.58	< 0.001
group_268	Outer membrane protein	OmpA	Outer membrane protein OmpA and related peptidoglycan-associated (lipo) proteins	M	3.69	60.46	< 0.001
accA1_2~~~accA1	acyl-CoA carboxylase alpha chain	PccA	Acetyl/propionyl-CoA carboxylase, alpha subunit	I	9.13	0.58	< 0.001
yhfA_2~~~yhfA_1~~~yhfA_3	Hypothetical protein VP1807	YhfA	Uncharacterized OsmC-related protein	R	36.92	69.60	< 0.001
group_31591	Hypothetical protein VP1134	NhaC	Na+/H+ antiporter NhaC/MleN	C	32.62	4.51	< 0.001
hrcN	Hypothetical protein	FliI	Flagellar biosynthesis/ type III secretory pathway ATPase FliI	NU	20.10	32.37	< 0.001
group_999	Hypothetical protein VP1825	AF2118	Predicted transcriptional regulator, contains an XRE-type HTH domain (archaeal members contain CBS pair)	K	4.10	51.33	< 0.001
ureG	Urease accessory protein UreG	HypB	Hydrogenase/urease maturation factor HypB, Ni2 + −binding GTPase	O	20.41	32.37	< 0.001
group_965	Hypothetical protein VP2937	Dph6	Diphthamide synthase (EF-2-diphthine--ammonia ligase)	J	20.00	7.51	< 0.001
rnr_1~~~rnr_2	Virulence-associated protein VacB/Rnase R	VacB	Exoribonuclease R	K	62.36	62.77	0.892

(Continued)

TABLE 2 (Continued)

Gene cluster from pangenome	COG annotation	COG symbol	COG name	COG functional category	Prevalence rate		*p* values
					Seafood	Clinical	
group_4703	Hypothetical protein VPA0394	EmrA	Multidrug resistance efflux pump EmrA	V	57.74	87.51	< 0.001
icaA	Hypothetical protein VPA0393	BcsA	Glycosyltransferase, catalytic subunit of cellulose synthase and poly-beta-1,6-N-acetylglucosamine synthase	N	57.74	87.51	< 0.001
flhB_3~~~yscU_2	Type III secretion system EscU protein	FlhB	Flagellar biosynthesis protein FlhB	N	3.69	60.46	< 0.001
aaeB~~~aaeB_1	Hypothetical protein VP1358	YccC	Uncharacterized membrane protein YccC	S	72.51	95.95	< 0.001
tufB~~~tuf~~~tuf1~~~tufA_2~~~tufA~~~tufA_1	Elongation factor Tu	TufA	Translation elongation factor EF-Tu, a GTPase	J	61.33	60.92	0.895

COG, the Database of Clusters of Orthologous Genes. SC, seafood vs. clinical. C, Energy production and conversion; D, Cell cycle control, cell division, chromosome partitioning; I, Lipid transport and metabolism; J, Translation, ribosomal structure and biogenesis; K, Transcription; L, Replication, recombination and repair; M, Cell wall/membrane/envelope biogenesis; N, Cell motility; O, Posttranslational modification, protein turnover, chaperones; R, General function prediction only; S, Function unknown; U, Intracellular trafficking, secretion, and vesicular transport; V, Defense mechanisms.

observation may be attributed to the fact that T3SS is a key virulence factor of *V. parahaemolyticus* (Li et al., 2019). Cell motility was the other most predominant functional category and four genes (COG symbol: FliI, FliI, BcsA, and FlhB, ranked the fifth, eleventh, seventeenth, and eighteenth, respectively, in the SC-metabolism model) out of the 20 important genes were recognized as the homologies to genes coding for proteins belonging to this category, particularly the orthologous cluster of flagellar biosynthesis. Similar to PulD and EscV, genes associated with flagellar biosynthesis were more frequently detected in clinical samples, highlighting the significance of flagellar in helping the transmission and survival of *V. parahaemolyticus* and possibly contributing to infection. It has been reported that the formation of biofilm, which is one of the important survival strategies of *V. parahaemolyticus*, is achieved with the aid of a dual flagellar system (Zhang Y. et al., 2023). On the other hand, the significantly high prevalence of four genes (COG symbol: SpsE, PccA, Nhac, and Dph6, ranked the fourth, eighth, tenth, and fourteenth, respectively) in seafood isolates could be explained by the response of the strain to the environmental pressure caused by the postharvest treatment of seafood. For example, *V. parahaemolyticus* has been reported to increase the expression of acetyl-CoA carboxylase (COG symbol: PccA) to synthesize unsaturated fatty acids and increase cell membrane fluidity to adapt to high hydrostatic pressure conditions, which has been commonly applied to inactivate the pathogen and extend the shelf life of seafood (Liang et al., 2022).

According to the SC-virulence model, exotoxin was the most predicted functional category (presented in Table 3). Specifically, two gene clusters were predicted to be two different copies of *tdh* and ranked first and eighteenth, respectively. On the other hand, one gene cluster from the pangenome was recognized as homology to *trhX* (also known as *trh*) and ranked sixteenth. These three gene clusters were significantly more prevalent in the clinical group, though none of them were present in *all* clinical isolates. Similar results have been found in previous studies, indicating the contribution of other factors to the pathogenicity of *V. parahaemolyticus* (Chao et al., 2010; Velazquez-Roman et al., 2012; Zha et al., 2023). Apart from homologies to *tdh* and *trhX*, homologies to four different alpha-hemolysin coding genes (*hlyD*, *hlyC*, *hlyA*, and *hlyB*) belonging to the exotoxin category have been identified as the top 20 influential predictors and ranked the third, fourth, sixth, and fourteenth, respectively. Interestingly, only *hlyD* was profoundly enriched in the clinical cohort compared with the seafood group while no significant difference was found regarding the prevalence rate of *hlyA*, *hlyB*, and *hlyC* in seafood and clinical isolates. In fact, the presence of *hlyA*, *hlyB*, *hlyC*, and *hlyD* in *V. parahaemolyticus* was only reported in a study investigating the pathogenesis of *V. parahaemolyticus* 353 isolated seafood in China (Zha et al., 2023). More studies are needed to reveal how these alpha-hemolysins contributed to the pathogenicity of *V. parahaemolyticus*, which could aid in explaining why their prevalence between seafood and clinical strains was similar but still critical to differentiate these two groups.

Moreover, it has been revealed that effector delivery system, T3SS, played an important role in differentiating nonpathogenic and pathogenic (seafood and clinical) groups. Based on our SC-virulence model, five genes related to T3SS (*vscJ2*, *vscC2*, *vopB2*, *VP_RS21585*, and *vopD2*) were identified as important genes and ranked the second, fifth, seventh, tenth, and eleventh, respectively. *V. parahaemolyticus* possesses two sets of T3SS: T3SS1 and T3SS2, which are responsible for cytotoxicity and enterotoxicity, respectively (Li et al., 2019). All the effector delivery system genes identified by the SC-virulence model were associated with T3SS2, which could be explained by the fact that T3SS1 is commonly found in both nonpathogenic and pathogenic isolates while T3SS2 is exclusive to pathogenic/clinical isolates (Matsuda et al., 2020). Generally, the proteins of T3SS could be categorized into four classes: structural proteins, translocators, effector proteins, and molecular chaperones (Li et al., 2019). In this

10.3389/fmicb.2025.1549260

TABLE 3 Twenty important genes identified by SC-virulence model, information of their homologies from VFDB, and their prevalence rates.

Gene cluster from pangenome	Name of the homologous gene in VFDB	VFDB gene product	VFDB functional category	Prevalence rate		*p* values
				Seafood	Clinical	
tdh1_1~~~tdh3~~~tdh1~~~tdh2	*tdh*	Thermostable direct hemolysin A	Exotoxin	9.13	86.24	< 0.001
group_5343	*vscJ2*	Type III secretion system protein VscJ2	Effector delivery system	3.69	60.81	< 0.001
hlyD~~~hlyD_1~~~hlyD_2	*hlyD*	Hemolysin D	Exotoxin	14.36	30.17	< 0.001
hlyC_2	*hlyC*	Hemolysin C	Exotoxin	7.28	7.40	0.995
sctC_5~~~sctC_3	*vscC2*	Type III secretion system protein VscC2	Effector delivery system	3.69	60.69	< 0.001
hlyA~~~hlyA_2	*hlyA*	Hemolysin A	Exotoxin	7.18	7.51	0.853
group_9636	*vopB2*	Type III secretion system translocator protein VopB2	Effector delivery system	3.69	60.69	< 0.001
group_10785	*mshC*	MSHA pilin protein MshC	Adherence	35.18	54.34	< 0.001
epsL_1~~~epsL_2~~~pssY~~~epsL_3	*wbfU*	Sugar transferase	Immune modulation	27.18	36.18	< 0.001
group_6266	*VP_RS21585*	Putative type III secretion system protein	Effector delivery system	3.69	60.69	< 0.001
group_6750	*vopD2*	Type III secretion system translocator protein VopD2	Effector delivery system	3.69	60.69	< 0.001
flaD_4~~~flaD_2~~~flaD_5~~~flaD_1~~~flaD_3	*flaC*	Flagellin	Motility	60.92	56.18	0.044
tufB~~~tuf~~~tuf1~~~tufA_2~~~tufA~~~tufA_1	*tufA*	Elongation factor Tu	Adherence	61.33	60.92	0.895
hlyB~~~hlyB_2	*hlyB*	Hemolysin B	Exotoxin	8.21	7.40	0.578
group_10962	*VP_RS21705*	Hypothetical protein	Effector delivery system	3.69	60.69	< 0.001
tdh2~~~tdh2_1~~~tdh2_2	*trhX*	TDH-related hemolysin	Exotoxin	17.95	32.14	< 0.001
flaD_1~~~flaD_3	*flaC*	Flagellin	Motility	34.05	9.60	< 0.001
tdh3_2~~~tdh3~~~tdh1	*tdh*	Thermostable direct hemolysin A	Exotoxin	0.62	48.09	< 0.001
rffH_2~~~rffH	*rmlA*	Glucose-1-phosphate thymidylyltransferase RfbA	Immune modulation	5.33	23.24	< 0.001
hag	*lafA*	Lateral flagellin LafA	Biofilm	78.46	90.75	< 0.001

SC, seafood vs. clinical. VFDB, the Virulence Factor Database.

study, we observed two genes predicted to encode structural proteins of T3SS2 (*vscJ2* and *vscC2*), which contribute to the formation of the physical structure of T3SS2, particularly the assemble of the inner membrane of both the basal body and export apparatus (Deng et al., 2017). Genes *vopB2* and *vopD2*, coding for the translocator protein of T3SS2, have been reported to be responsible for creating the pathway, pores in the membrane of host cells, through which effectors could be delivered into the host cells (Paria et al., 2021). It should be noted that two of the T3SS-related genes were hypothetical/putative proteins, which presses the need to perform further research specifically on these genes to unveil their characterizations and roles in contributing to the pathogenicity of *V. parahaemolyticus*.

Among all gene clusters identified as important by our SC-virulence model, the homology to *flaC* (ranked the seventeenth) was the only one that was more prevalent in the seafood group. It has been reported that FlaC, which is one of the flagellin subunits of the filament of *V. parahaemolyticus* flagellum coded by *flaC*, can activate the immune protection function of shellfish (Chen et al., 2019). We suspect that *flaC*-activated immune protection could result in changes in the texture or appearance of shellfish, causing consumers to perceive it as unsafe to eat. In contrast, shellfish contaminated with *V. parahaemolyticus* lacking *flaC* may not exhibit such changes, which makes people consider it as safe for consumption. Consequently, shellfish contaminated with *V. parahaemolyticus* lacking *flaC* is more

TABLE 4 Genes used as predictors in SC-antibiotic resistance model, information of their homologies from CARD, and their prevalence rates.

Gene cluster from pangenome	Name of the homologous gene in CARD	Drug class	AMR gene family	Resistance mechanism	Prevalence rate		*p* values
					Seafood	Clinical	
group_31591	*tet(35)*	Tetracycline antibiotic	ATP-binding cassette (ABC) antibiotic efflux pump	Efflux pump	32.62	4.51	< 0.001
tufA_1~~~tuf~~~tufA~~~tufB	*Ecol_EFTu_PLV*	Elfamycin antibiotic	elfamycin resistant EF-Tu	Target site alteration	36.51	66.13	< 0.001
group_5516	*MexS*	Phenicol antibiotic, diaminopyrimidine antibiotic, fluoroquinolone antibiotic	resistance-nodulation-cell division (RND) antibiotic efflux pump	Efflux pump	6.77	29.60	< 0.001
group_11708	*ErmY*	Streptogramin antibiotic, lincosamide antibiotic, macrolide antibiotic	Erm 23S ribosomal RNA methyltransferase	Target site alteration	77.03	62.43	< 0.001
group_8131	*Ctra_murA_FOF*	Phosphonic acid antibiotic	antibiotic-resistant murA transferase	Target site alteration	16.92	18.15	0.539
ugd~~~ugd_2~~~ugd_1	*ugd*	Peptide antibiotic	pmr phosphoethanolamine transferase	Target site alteration	80.51	64.97	< 0.001
macB_6~~~macB_4~~~macB_5~~~macB_3~~~macB_2	*macB*	Macrolide antibiotic	ABC antibiotic efflux pump	Efflux pump	89.13	96.76	< 0.001
ugd_1~~~ugd_2~~~ugd	*ugd*	Peptide antibiotic	pmr phosphoethanolamine transferase	Target site alteration	19.69	35.26	< 0.001
pse4	*CARB-23*	Penam	CARB beta-lactamase	Inactivation of antibiotic	27.59	4.74	< 0.001
tufB~~~tuf~~~tuf1~~~tufA_2~~~tufA~~~tufA_1	*Ecol_EFTu_PLV*	Elfamycin antibiotic	elfamycin resistant EF-Tu	Target site alteration	61.33	60.92	0.895
tufB~~~tufA~~~tufA_1~~~tufB_1~~~tuf1	*Ecol_EFTu_PLV*	Elfamycin antibiotic	elfamycin resistant EF-Tu	Target site alteration	21.54	25.66	0.042
dhfrIII	*dfrA3*	Diaminopyrimidine antibiotic	trimethoprim resistant dihydrofolate reductase dfr	Antibiotic target replacement	89.13	97.11	< 0.001
group_10971	*LpxA*	Peptide antibiotic	Acinetobacter mutant Lpx gene conferring resistance to colistin	Target site alteration	4.10	12.83	< 0.001
hns	*H-NS*	Tetracycline antibiotic, penam, cephamycin, cephalosporin, fluoroquinolone antibiotic, macrolide antibiotic	RND antibiotic efflux pump, major facilitator superfamily (MFS) antibiotic efflux pump	Efflux pump	15.79	2.43	< 0.001
cat_3~~~cat_2	*catB9*	Phenicol antibiotic	chloramphenicol acetyltransferase (CAT)	Inactivation of antibiotic	16.31	3.82	< 0.001

(Continued)

TABLE 4 (Continued)

Gene cluster from pangenome	Name of the homologous gene in CARD	Drug class	AMR gene family	Resistance mechanism	Prevalence rate		*p* values
					Seafood	Clinical	
group_31739	*qnrAS*	Fluoroquinolone antibiotic	quinolone resistance protein (qnr)	Target protection	13.95	3.58	< 0.001
acoR_2~~~qseF~~~dctD_1	*txR*	Tetracycline antibiotic	ABC antibiotic efflux pump	Efflux pump	9.64	1.85	< 0.001

AMR, antimicrobial resistance. CARD, the Comprehensive Antibiotic Resistance Database. SC, seafood vs. clinical.

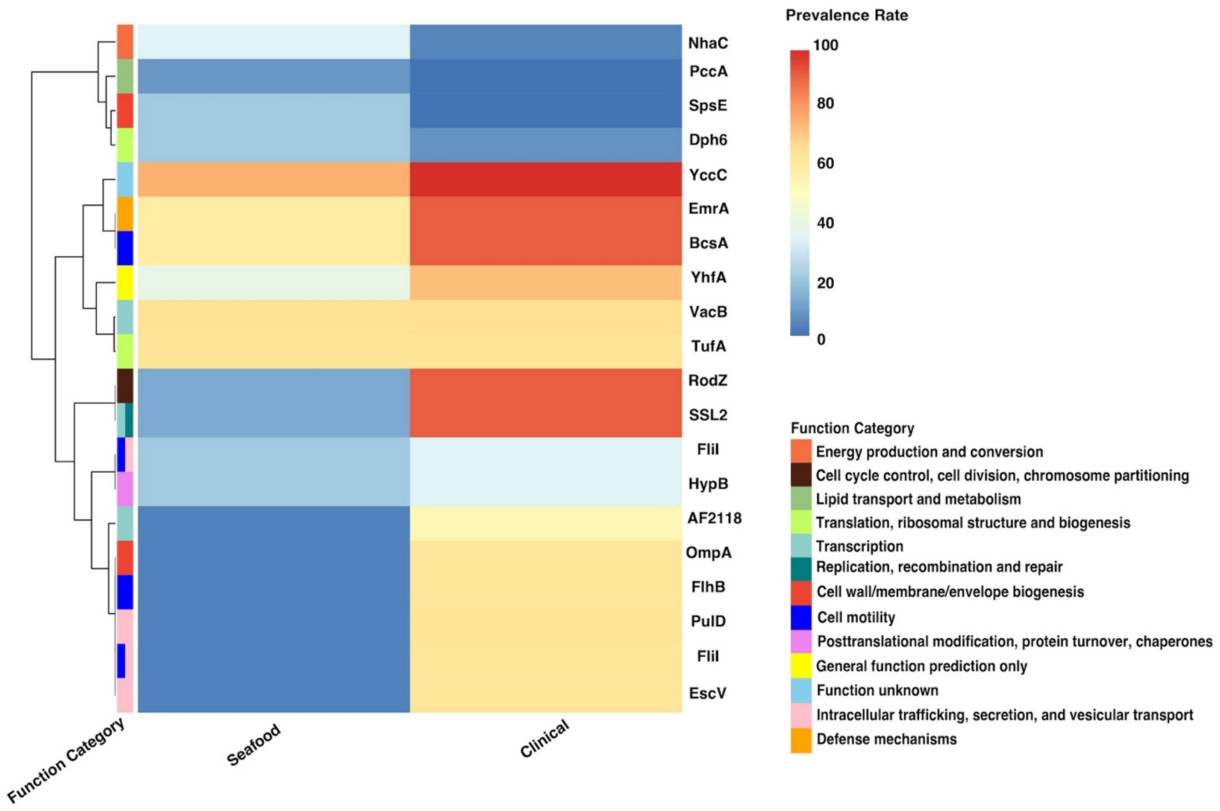


FIGURE 4
Heatmap for the prevalence of twenty important genes identified by the SC-metabolism model. SC, seafood vs. clinical.

likely to be eaten by consumers, which may explain the relatively lower prevalence of *flaC* in clinical isolates.

The gene *tet(35)*, which confers tetracycline resistance, was identified as the most important predictor in the SC-antibiotic resistance model with higher prevalence in seafood samples (shown in Table 4). Our results corresponded well with the frequently observed tetracycline resistance in seafood isolates worldwide (Elmahdi et al., 2016). *EFTu*, which confers to elfamycin resistance, ranked as the second among all the predictors in our SC-antibiotic resistance model. Several studies have described elfamycin resistance in pathogens obtained from various seafood and aquatic environments, which could be attributed to the usage of elfamycins as growth promoters for aquaculture (Behera et al., 2021; Liu et al., 2019; Zhang Q. et al., 2023). In addition, *MexS* (ranked the third), possessing multidrug resistance (phenicol antibiotic, diaminopyrimidine antibiotic, and fluoroquinolone antibiotic), were more predominantly found in the clinical group rather than the seafood cohort. The low prevalence rate of *MexS* in seafood samples (6.77%) observed in our study was consistent with previous research (Hanekamp and Bast, 2015; Obaidat et al., 2017; Lei et al., 2020; Kemp et al., 2021; Bondad-Reantaso et al., 2023).

Efflux pump and target set alteration were the most prevalent antibiotic resistance mechanisms associated with the predictors used by our SC-antibiotic resistance model (shown in Table 4). The presence of *tet(35)*, *MexS*, *macB*, *H-NS*, and *txR* (ranked the first, third, seventh, fourteenth, and seventeenth, respectively), which are related to ATP-binding cassette (ABC), resistance-nodulation-cell division (RND), and major facilitator superfamily (MFS) antibiotic efflux pump, could be indicative of the essential roles of ABC, RND, and MFS efflux pumps in differentiating antibiotic resistance profiles of seafood and clinical isolates and similar insights have been gained from prior studies (Pérez-Acosta et al., 2018; Lloyd et al., 2019; Stephen et al., 2022).

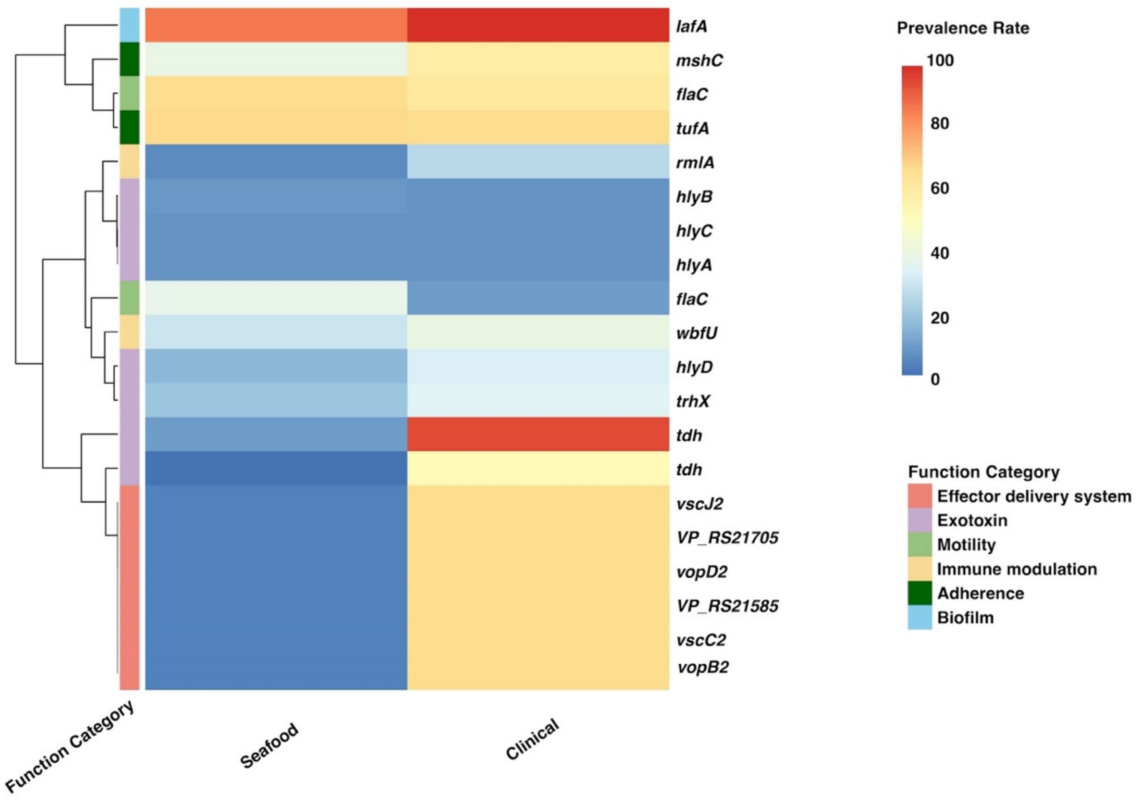


FIGURE 5
Heatmap for the prevalence of twenty important genes identified by the SC-virulence model. SC, seafood vs. clinical.

Though the target site alteration mechanisms of the listed important genes in *V. parahaemolyticus* (*Ecol_EFTu_PLV*, *ugd*, and *LpxA*) have not been extensively studied, the involvement of their related gene family in the antibiotic resistance have been demonstrated (Miele et al., 1994; Tracevska et al., 2002; Novović and Jovčić, 2023).

Additionally, the characterization of the individual pangenomes for the respective seafood and clinical isolates were summarized in Table 5. The total numbers of core genes and shell genes between the pangenomes of seafood and clinical isolates appeared similar, while the number of cloud genes for the seafood pangenome was about two-fold more than that for clinical pangenome, resulting in the drastic difference of the sizes of pangenome. The respective genes-per-genome by isolate source were consistent with this observation, indicating much greater genomic diversity of *V. parahaemolyticus* isolated from seafood samples. These differences may be attributed to the broader geographic distribution of isolation locations of the isolates from seafood samples compared to clinical samples. Horizontal gene transfer (HGT) of mobile genetic elements is commonly found in *V. parahaemolyticus* and has been proven that could greatly contribute to its genetic diversity (Xu et al., 2022). To be more specific, seafood isolates from diverse locations could obtain various genes through HGT, which explains the massive number of cloud genes in its pangenome.

Although some models developed and used in this study could predict the isolation sources accurately and provide useful insights, certain limitations have been recognized. The limited availability of environmental isolates, which resulted in a severe class imbalance for our ES models, has constituted an obvious limitation and affected the robustness of the model greatly in this study. Though ROSE has been applied to overcome the bias caused by the imbalanced class and has significantly improved the model performance compared with models built based on data without ROSE, the obtained ES models were still not capable of providing accurate predictions. The scarcity of *V. parahaemolyticus* strains isolated from environmental samples has also been described in several other studies (Turner et al., 2013; Ronholm et al., 2016; Obaidat et al., 2017; Yan et al., 2020). Therefore, in the future, times of sampling events and detections of *V. parahaemolyticus* in environmental samples should be increased to aid in comprehending the population features of environmental strains more representatively. Moreover, as the genome assemblies were downloaded from the NCBI database, potential bias or batch effects among different studies (e.g., sequencing platform, sequencing depth, assembler) may have contributed to variations we observed.

A great number of tools with different mechanisms are available for each bioinformatic analysis step in this study and alternative tools may be resourceful to find additional differences correlated with the metadata. Therefore, the choice of method for each step could potentially impact our results. Although Prokka and Panaroo were used in this study, future work will explore other bioinformatics tools, such as PGAP, Roary, and PIRATE, to better understand how method selection may impact the downstream analysis. Additionally, the cutoff values for query coverage and percent identity were set based on the number of predictors, potentially impacting the performance of our random forest models. To enhance the robustness of our

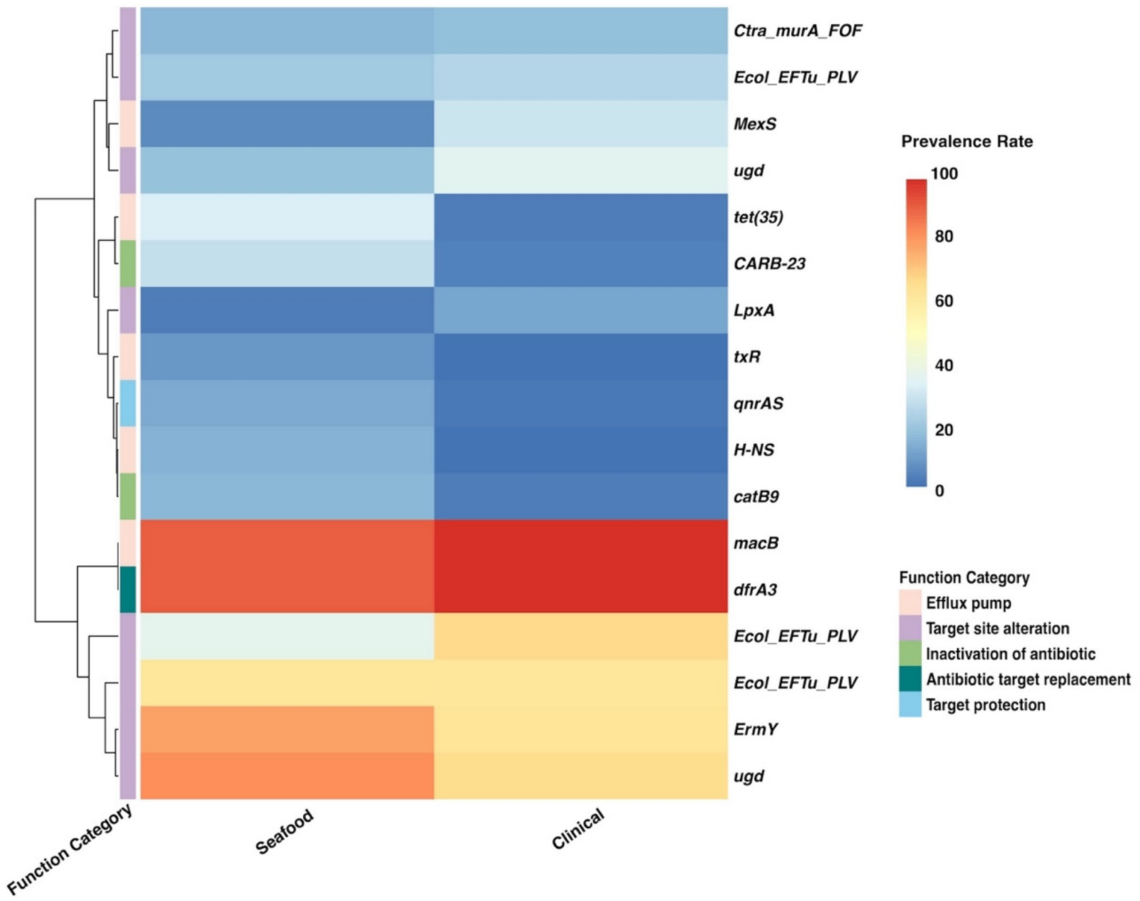


FIGURE 6
Heatmap for the prevalence of genes used as predictors in the SC-antibiotic resistance model. SC, seafood vs. clinical.

TABLE 5 Summary table for the pangenomes of seafood and clinical isolates.

Pangenome	Core genes	Shell genes	Cloud genes	Total genes	Genes per genome (mean ± SD)
Seafood	3,886	877	32,543	37,306	4,629 ± 195
Clinical	4,017	1,025	14,325	19,367	4,580 ± 84

models, we systematically tested various BLASTp thresholds for query coverage and percent identity (as described in 2.2 Bioinformatics analysis), identifying the thresholds that yielded the most reliable predictions. In future studies, higher cutoff values should be applied when more datasets become available, as this may reduce noise associated with lower cutoff thresholds. Further research is needed to thoroughly assess how the choice of different bioinformatics tools influences downstream analysis and to develop a standardized and most optimal workflow for bioinformatics-ML studies.

Moreover, the prediction of gene function was greatly restrained by the size and accuracy of databases (COG, VFDB, and CARD) used for performing BLASTp analysis. It has been noticed that models for metabolism and virulence outperformed models for antibiotic resistance, which could be explained by the relatively limited predictors available for antibiotic resistance models, as the size of CARD is smaller than COG and VFDB. Expanding and updating respective gene function databases when new genes and functions are identified could contribute to overcoming this bias in the future. On the other hand, combining multiple databases might improve the performance of our models as well by providing a more comprehensive input. However, the lack of standardization and the methodological discrepancies between databases hinder the application of the database combination. Improved harmonization across databases and a thorough evaluation of the associated analysis method in the future could help address these challenges and make the combined database a feasible approach for enhancing model performance.

5 Conclusion

In this study, the application of machine learning was used to analyze pangenomes of *V. parahaemolyticus* to identify important genes associated with different isolation sources (environmental, seafood, and clinical). Our study highlights the crucial role of the type

III secretion system in distinguishing metabolic and virulence accessory gene profiles of *Vibrio parahaemolyticus* seafood and clinical isolates. We also found that virulence-related genes encoding alpha-hemolysins were key in differentiating these groups. Among the top three most important predictors from our SC-antibiotic resistance model, gene conferring to tetracycline resistance was more prevalent in seafood isolates while genes confer to elfamycin, and multidrug (phenicol antibiotic, diaminopyrimidine antibiotic, fluoroquinolone antibiotic) resistance were greatly enriched in clinical isolates. These findings can help enhance risk management strategies along the seafood-to-consumer chain. However, the limited availability of environmental isolates significantly impacted the performance of our environmental-seafood model. Future research should focus on expanding sequencing databases for environmental samples and evaluating the impact of genomics workflow selection on analysis outcomes, providing a stronger scientific basis for selecting appropriate genomics tools.

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

SF: Conceptualization, Methodology, Data curation, Formal analysis, Investigation, Visualization, Writing – original draft. PR: Conceptualization, Methodology, Writing – review & editing, Data curation, Investigation. RB: Methodology, Writing – review & editing. AP: Methodology, Writing – review & editing, Conceptualization, Funding acquisition, Resources, Supervision.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported in part through a grant from the U.S. Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) Agriculture and Food Research Initiative (award number 2024-67021-42527).

Acknowledgments

The authors acknowledge the University of Maryland supercomputing resources (http://hpcc.umd.edu) made available for conducting the research reported in this paper.

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.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Author disclaimer

Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the USDA-NIFA.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2025.1549260/full#supplementary-material

References

Alcock, B. P., Huynh, W., Chalil, R., Smith, K. W., Raphenya, A. R., Wlodarski, M. A., et al. (2023). CARD 2023: expanded curation, support for machine learning, and resistome prediction at the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 51, D690–D699. doi: 10.1093/nar/gkac920

Behera, B. K., Dehury, B., Rout, A. K., Patra, B., Mantri, N., Chakraborty, H. J., et al. (2021). Metagenomics study in aquatic resource management: recent trends, applied methodologies and future needs. *Gene Reports* 25:101372. doi: 10.1016/j.genrep.2021.101372

Benefo, E. O., Karanth, S., and Pradhan, A. K. (2024a). A machine learning approach to identifying *Salmonella* stress response genes in isolates from poultry processing. *Food Res. Int.* 175:113635. doi: 10.1016/j.foodres.2023.113635

Benefo, E. O., Ramachandran, P., and Pradhan, A. K. (2024b). Genome-based machine learning for predicting antimicrobial resistance in *Salmonella* isolated from chicken. *LWT* 199:116122. doi: 10.1016/j.lwt.2024.116122

Biau, G., and Scornet, E. (2016). A random forest guided tour. *TEST* 25, 197–227. doi: 10.1007/s11749-016-0481-7

Blaustein, R. A., McFarland, A., Ben, S., Lopez, A., Castro-Wallace, S., and Hartmann, E. M. (2019). Pangenomic approach to understanding microbial adaptations within a model built environment, the international space station, relative to human hosts and soil. *mSystems* 4:4. doi: 10.1128/msystems.00281-18

Bondad-Reantaso, M. G., MacKinnon, B., Karunasagar, I., Fridman, S., Alday-Sanz, V., Brun, E., et al. (2023). Review of alternatives to antibiotic use in aquaculture. *Rev. Aquac.* 15, 1421–1451. doi: 10.1111/raq.12786

Brauge, T., Mougin, J., Ells, T., and Midelet, G. (2024). Sources and contamination routes of seafood with human pathogenic *Vibrio* spp.: a farm-to-fork approach. *Compr. Rev. Food Sci. Food Saf.* 23:e13283. doi: 10.1111/1541-4337.13283

Brown, E., Dessai, U., Mcgarry, S., and Gerner-Smidt, P. (2019). Use of whole-genome sequencing for food safety and public health in the United States. *Foodborne Path. Dis.* 16, 441–450. doi: 10.1089/fpd.2019.2662

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

Centers for Disease Control and Prevention (2013). *Vibrio* species causing vibriosis. Atlanta, GA: Centers for Disease Control and Prevention.

Chao, G., Jiao, X., Zhou, X., Wang, F., Yang, Z., Huang, J., et al. (2010). Distribution of genes encoding four Pathogenicity Islands (VPaIs), T6SS, biofilm, and type I pilus in food and clinical strains of *Vibrio Parahaemolyticus* in China. *Foodborne Pathog. Dis.* 7, 649–658. doi: 10.1089/fpd.2009.0441

Chen, C., Kang, C., Rong, N., Wu, N., Chen, C., Wu, S., et al. (2019). Evaluation of immunogenicity, protective immunity on aquaculture pathogenic *Vibrio* and fermentation of *Vibrio Alginolyticus* Flagellin Flac protein. *Iran. J. Biotechnol.* 17, e2628–e2642. doi: 10.29252/ijb.2628

D'Agostino, R. B., Pencina, M. J., Massaro, J. M., and Coady, S. (2013). Cardiovascular disease risk assessment: insights from Framingham. *Glob. Heart* 8, 11–23. doi: 10.1016/j.gheart.2013.01.001

Deng, W., Marshall, N. C., Rowland, J. L., McCoy, J. M., Worrall, L. J., Santos, A. S., et al. (2017). Assembly, structure, function and regulation of type III secretion systems. *Nat. Rev. Microbiol.* 15, 323–337. doi: 10.1038/nrmicro.2017.20

DePaola, A., Ulaszek, J., Kaysner, C. A., Tenge, B. J., Nordstrom, J. L., Wells, J., et al. (2003). Molecular, serological, and virulence characteristics of *Vibrio Parahaemolyticus* isolated from environmental, food, and clinical sources in North America and Asia. *Appl. Environ. Microbiol.* 69, 3999–4005. doi: 10.1128/AEM.69.7.3999-4005.2003

Elmahdi, S., DaSilva, L. V., and Parveen, S. (2016). Antibiotic resistance of *Vibrio parahaemolyticus* and *Vibrio vulnificus* in various countries: a review. *Food Microbiol.* 57, 128–134. doi: 10.1016/j.fm.2016.02.008

Feng, S., Karanth, S., Almuhaideb, E., Parveen, S., and Pradhan, A. K. (2024). Machine learning to predict the relationship between *Vibrio* spp. concentrations in seawater and oysters and prevalent environmental conditions. *Food Res. Int.* 188:114464. doi: 10.1016/j.foodres.2024.114464

Gmeiner, A., Njage, P. M. K., Hansen, L. T., Aarestrup, F. M., and Leekitcharoenphon, P. (2024). Predicting *Listeria monocytogenes* virulence potential using whole genome sequencing and machine learning. *Int. J. Food Microbiol.* 410:110491. doi: 10.1016/j.ijfoodmicro.2023.110491

Hanekamp, J. C., and Bast, A. (2015). Antibiotics exposure and health risks: chloramphenicol. *Environ. Toxicol. Pharmacol.* 39, 213–220. doi: 10.1016/j.etap.2014.11.016

Huang, W. S., and Wong, H. C. (2012). Characterization of low salinity stress in *Vibrio Parahaemolyticus*. *J. Food Prot.* 75, 231–237. doi: 10.4315/0362-028X.JFP-11-321

Hyatt, D., Chen, G., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Karanth, S., Tanui, C. K., Meng, J., and Pradhan, A. K. (2022). Exploring the predictive capability of advanced machine learning in identifying severe disease phenotype in *Salmonella enterica*. *Food Res. Int.* 151:110817. doi: 10.1016/j.foodres.2021.110817

Kemp, J. O. G., Taylor, J. J., Kelly, L. A., Larocque, R., Heriazon, A., Tiessen, K. H. D., et al. (2021). Antibiotic resistance genes in the aquaculture sector: global reports and research gaps. *Environ. Rev.* 29, 300–314. doi: 10.1139/er-2020-0087

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Available online at: http://robotics.stanford.edu/~ronnyk

Kuhn, M. (2019). The caret package. Available online at: https://Topepo.Github.Io/Caret/

Lei, T., Jiang, F., He, M., Zhang, J., Zeng, H., Chen, M., et al. (2020). Prevalence, virulence, antimicrobial resistance, and molecular characterization of fluoroquinolone resistance of *Vibrio parahaemolyticus* from different types of food samples in China. *Int. J. Food Microbiol.* 317:108461. doi: 10.1016/j.ijfoodmicro.2019.108461

Letchumanan, V., Ser, H. L., Chan, K. G., Goh, B. H., and Lee, L. H. (2016). Genome sequence of *Vibrio Parahaemolyticus* VP103 strain isolated from shrimp in Malaysia. *Front. Microbiol.* 7:01496. doi: 10.3389/fmicb.2016.01496

Letchumanan, V., Yin, W. F., Lee, L. H., and Chan, K. G. (2015). Prevalence and antimicrobial susceptibility of *Vibrio parahaemolyticus* isolated from retail shrimps in Malaysia. *Front. Microbiol.* 6:33. doi: 10.3389/fmicb.2015.00033

Li, L., Meng, H., Gu, D., Li, Y., and Jia, M. (2019). Molecular mechanisms of *Vibrio parahaemolyticus* pathogenesis. *Microbiol. Res.* 222, 43–51. doi: 10.1016/j.micres.2019.03.003

Liang, S., Zhang, T., Liu, Z., Wang, J., Zhu, C., Kong, Q., et al. (2022). Response mechanism of *Vibrio parahaemolyticus* at high pressure revealed by transcriptomic analysis. *Appl. Microbiol. Biotechnol.* 106, 5615–5628. doi: 10.1007/s00253-022-12082-y

Liu, K., Han, J., Li, S., Liu, L., Lin, W., and Luo, J. (2019). Insight into the diversity of antibiotic resistance genes in the intestinal bacteria of shrimp *Penaeus vannamei* by culture-dependent and independent approaches. *Ecotoxicol. Environ. Saf.* 172, 451–459. doi: 10.1016/j.ecoenv.2019.01.109

Liu, B., Zheng, D., Zhou, S., Chen, L., and Yang, J. (2022). VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res.* 50, D912–D917. doi: 10.1093/nar/gkab1107

Livingstone, P. G., Morphew, R. M., and Whitworth, D. E. (2018). Genome sequencing and Pan-genome analysis of 23 *Corallococcus spp.* strains reveal unexpected diversity, with particular plasticity of predatory gene sets. *Front. Microbiol.* 9:3187. doi: 10.3389/fmicb.2018.03187

Lloyd, N. A., Nazaret, S., and Barkay, T. (2019). Genome-facilitated discovery of RND efflux pump-mediated resistance to cephalosporins in *Vibrio* spp. isolated from the mummichog fish gut. *J. Glob. Antimicrob. Resist.* 19, 294–300. doi: 10.1016/j.jgar.2019.05.006

Loo, K. Y., Letchumanan, V., Law, J. W. F., Pusparajah, P., Goh, B. H., Mutalib, N. S. A., et al. (2020). Incidence of antibiotic resistance in *Vibrio* spp. *Rev. Aquac.* 12, 2590–2608. doi: 10.1111/raq.12460

Lunardon, N., Menardi, G., and Torelli, N. (2014). ROSE: a package for binary imbalanced learning. *The R J.* 6, 79–89. doi: 10.32614/RJ-2014-008

Matsuda, S., Hiyoshi, H., Tandhavanant, S., and Kodama, T. (2020). Advances on *Vibrio parahaemolyticus* research in thepostgenomic era. *Microbiol. Immunol.* 64, 167–181. doi: 10.1111/1348-0421.12767

Miele, A., Goldstein, B. P., Bandera, M., Jarvis, C., Resconi, A., and Williams, R. J. (1994). Differential susceptibilities of Enterococcal species to Elfamycin antibiotics. *J. Clin. Microbiol.* 32, 2016–2018. doi: 10.1128/jcm.32.8.2016-2018.1994

Nahm, F. S. (2022). Receiver operating characteristic curve: overview and practical use for clinicians. *Korean J. Anesthesiol.* 75, 25–36. doi: 10.4097/kja.21209

Ndraha, N., Hsiao, H. I., Hsieh, Y. Z., and Pradhan, A. K. (2021). Predictive models for the effect of environmental factors on the abundance of *Vibrio Parahaemolyticus* in oyster farms in Taiwan using extreme gradient boosting. *Food Control* 130:108353. doi: 10.1016/j.foodcont.2021.108353

Novović, K., and Jovčić, B. (2023). Colistin resistance in *Acinetobacter Baumannii*: molecular mechanisms and epidemiology. *Antibiot.* 12:516. doi: 10.3390/antibiotics12030516

Obaidat, M. M., Salman, A. E. B., and Roess, A. A. (2017). Virulence and antibiotic resistance of *Vibrio Parahaemolyticus* isolates from seafood from three developing countries and of worldwide environmental, seafood, and clinical isolates from 2000 to 2017. *J. Food Prot.* 80, 2060–2067. doi: 10.4315/0362-028X.JFP-17-156

Paria, P., Behera, B. K., Mohapatra, P. K. D., and Parida, P. K. (2021). Virulence factor genes and comparative pathogenicity study of *tdh*, *trh* and *tlh* positive *Vibrio parahaemolyticus* strains isolated from Whiteleg shrimp, *Litopenaeus Vannamei* (Boone, 1931) in India. *Infect. Genet. Evol.* 95:105083. doi: 10.1016/j.meegid.2021.105083

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114

Pazhani, G. P., Chowdhury, G., and Ramamurthy, T. (2021). Adaptations of *Vibrio parahaemolyticus* to stress during environmental survival, host colonization, and infection. *Front. Microbiol.* 12:737299. doi: 10.3389/fmicb.2021.737299

Pérez-Acosta, J. A., Martínez-Porchas, M., Elizalde-Contreras, J. M., Leyva, J. M., Ruiz-May, E., Gollas-Galván, T., et al. (2018). Proteomic profiling of integral membrane proteins associated to pathogenicity in *Vibrio Parahaemolyticus* strains. *Microbiol. Immunol.* 62, 14–23. doi: 10.1111/1348-0421.12556

Pérez-Duque, A., Gonzalez-Muñoz, A., Arboleda-Valencia, J., Vivas-Aguas, L. J., Córdoba-Meza, T., Rodriguez-Rey, G. T., et al. (2021). Comparative genomics of clinical and environmental isolates of *Vibrio* Spp. of Colombia: implications of traits associated with virulence and resistance. *Pathogens* 10:1605. doi: 10.3390/pathogens10121605

Qadri, F., Alam, M. S., Nishibuchi, M., Rahman, T., Alam, N. H., Chisti, J., et al. (2003). Adaptive and inflammatory immune responses in patients infected with strains of *Vibrio parahaemolyticus*. *J. Infect. Dis.* 187, 1085–1096. doi: 10.1086/368257

Raghunath, P. (2015). Roles of thermostable direct hemolysin (TDH) and TDH-related hemolysin (TRH) in Vibrio parahaemolyticus. *Front. Microbiol.* 5:805. doi: 10.3389/fmicb.2014.00805

Rajkowski, K. T. (2009). "Biofilms in fish processing" in Biofilms in the Food and Beverage Industries. eds. P. M. Fratamico, B. A. Annous and N. W. Guenther (Amsterdam: Elsevier).

Ronholm, J., Petronella, N., Leung, C. C., Pightling, A. W., and Banerjeea, S. K. (2016). Genomic features of environmental and clinical *Vibrio parahaemolyticus* isolates lacking recognized virulence factors are dissimilar. *Appl. Environ. Microbiol.* 82, 1102–1113. doi: 10.1128/AEM.03465-15

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

Sidey-Gibbons, J. A. M., and Sidey-Gibbons, C. J. (2019). Machine learning in medicine: a practical introduction. *BMC Med. Res. Methodol.* 19:64. doi: 10.1186/s12874-019-0681-4

Stephen, J., Lekshmi, M., Ammini, P., Kumar, S. H., and Varela, M. F. (2022). Membrane efflux pumps of pathogenic *Vibrio* species: role in antimicrobial resistance and virulence. *Microorganisms.* 10:382. doi: 10.3390/microorganisms10020382

Su, Y. C., and Liu, C. (2007). *Vibrio parahaemolyticus*: a concern of seafood safety. *Food Microbiol.* 24, 549–558. doi: 10.1016/j.fm.2007.01.005

Tang, J., Jia, J., Chen, Y., Huang, X., Zhang, X., Zhao, L., et al. (2018). Proteomic analysis of *Vibrio parahaemolyticus* under cold stress. *Curr. Microbiol.* 75, 20–26. doi: 10.1007/s00284-017-1345-4

Tanui, C. K., Benefo, E. O., Karanth, S., and Pradhan, A. K. (2022a). A machine learning model for food source attribution of *Listeria Monocytogenes*. *PathoGenetics* 11:691. doi: 10.3390/pathogens11060691

Tanui, C. K., Karanth, S., Njage, P. M. K., Meng, J., and Pradhan, A. K. (2022b). Machine learning-based predictive modeling to identify genotypic traits associated with *Salmonella enterica* disease endpoints in isolates from ground chicken. *LWT.* 154:112701. doi: 10.1016/j.lwt.2021.112701

Thölke, P., Mantilla-Ramos, Y. J., Abdelhedi, H., Maschke, C., Dehgan, A., Harel, H., et al. (2023). Class imbalance should not throw you off balance: choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *NeuroImage* 277:120253. doi: 10.1016/j.neuroimage.2023.120253

Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J. A., et al. (2020). Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* 21:180. doi: 10.1186/s13059-020-02090-4

Tracevska, T., Jansone, I., Broka, L., Marga, O., and Baumanis, V. (2002). Mutations in the *rpoB* and *katG* genes leading to drug resistance in *Mycobacterium tuberculosis* in Latvia. *J. Clin. Microbiol.* 40, 3789–3792. doi: 10.1128/JCM.40.10.3789-3792.2002

Turner, J. W., Paranjpye, R. N., Landis, E. D., Biryukov, S. V., González-Escalona, N., Nilsson, W. B., et al. (2013). Population structure of clinical and environmental *Vibrio parahaemolyticus* from the Pacific northwest coast of the United States. *PLoS One* 8:e55726. doi: 10.1371/journal.pone.0055726

Unrath, N., McCabe, E., Macori, G., and Fanning, S. (2021). Application of whole genome sequencing to aid in deciphering the persistence potential of *Listeria monocytogenes* in food production environments. *Microorganisms* 9:1856. doi: 10.3390/microorganisms9091856

Velazquez-Roman, J., León-Sicairos, N., Flores-Villaseñor, H., Villafaña-Rauda, S., and Canizalez-Roman, A. (2012). Association of pandemic *Vibrio parahaemolyticus* O3:K6 present in the coastal environment of Northwest Mexico with cases of recurrent diarrhea between 2004 and 2010. *Appl. Environ. Microbiol.* 78, 1794–1803. doi: 10.1128/AEM.06953-11

Xu, D., Peng, X., Xie, L., and Chen, L. (2022). Survival and genome diversity of *Vibrio parahaemolyticus* isolated from edible aquatic animals. *Diversity* 14:350. doi: 10.3390/d14050350

Yan, W., Ji, L., Xu, D., Chen, L., and Wu, X. (2020). Molecular characterization of clinical and environmental *Vibrio parahaemolyticus* isolates in Huzhou, China. *PLoS One* 15:e0240143. doi: 10.1371/journal.pone.0240143

Zha, F., Pang, R., Huang, S., Zhang, J., Wang, J., Chen, M., et al. (2023). Evaluation of the pathogenesis of non-typical strain with α-Hemolysin, *Vibrio parahaemolyticus* 353, isolated from Chinese seafood through comparative genome and transcriptome analysis. *Mar. Pollut. Bull.* 186:114276. doi: 10.1016/j.marpolbul.2022.114276

Zhang, Q., Alter, T., Strauch, E., Hammerl, J. A., Schwartz, K., Borowiak, M., et al. (2023). Genetic and phenotypic virulence potential of non-O1/non-O139 *Vibrio cholerae* isolated from German retail seafood. *Microorganisms* 11:2751. doi: 10.3390/microorganisms11112751

Zhang, W., Chen, K., Zhang, L., Zhang, X., Zhu, B., Lv, N., et al. (2023). The impact of global warming on the signature virulence gene, Thermolabile Hemolysin, of *Vibrio parahaemolyticus*. *Microbiol. Spectr.* 11:e0150223. doi: 10.1128/spectrum.01502-23

Zhang, Y., Zhang, T., Qiu, Y., Zhang, M., Lu, X., Yang, W., et al. (2023). Transcriptomic profiles of *Vibrio parahaemolyticus* during biofilm formation. *Curr. Microbiol.* 80:371. doi: 10.1007/s00284-023-03425-7

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 16 April 2025
DOI 10.3389/fmicb.2025.1546744

OPEN ACCESS

EDITED BY
Renmao "Tim" Tian,
Illinois Institute of Technology, United States

REVIEWED BY
Joshua Mbanga,
National University of Science and Technology, Zimbabwe
Muhammed Duman,
Bursa Uludağ University, Türkiye
Marta Laranjo,
University of Evora, Portugal

*CORRESPONDENCE
Yong Hu
✉ huyong1979@gmc.edu.cn
Shijun Li
✉ zjumedjun@163.com

RECEIVED 17 December 2024
ACCEPTED 26 March 2025
PUBLISHED 16 April 2025

CITATION
Zhou J, Liu Y, Gu T, Zhou J, Chen F, Hu Y and Li S (2025) Whole-genome analysis and antimicrobial resistance phenotype of *Vagococcus fluvialis* isolated from wild *Niviventer*.
Front. Microbiol. 16:1546744.
doi: 10.3389/fmicb.2025.1546744

Whole-genome analysis and antimicrobial resistance phenotype of *Vagococcus fluvialis* isolated from wild *Niviventer*

Jian Zhou[1,2], Ying Liu[2], Tao Gu[1], Jingzhu Zhou[2], Fengming Chen[2], Yong Hu[1]* and Shijun Li[1,2]*

[1]The Key Laboratory of Environmental Pollution Monitoring and Disease Control, School of Public Health, Ministry of Education, Guizhou Medical University, Guiyang, China, [2]Key Laboratory of Microbio and Infectious Disease Prevention and Control in Guizhou Province, Guizhou Center for Disease Control and Prevention, Guiyang, China

Vagococcus fluvialis (*V. fluvialis*), a Gram-positive bacterium belonging to the *Enterococcaceae* family, has been associated with human infections, including bacteremia and endocarditis. Its zoonotic potential raises concerns for public health, yet research on its antimicrobial resistance and pathogenicity is still limited. This study aimed to isolate and characterize *V. fluvialis* from wild *Niviventer*, analyze its genomic features (including antimicrobial resistance and virulence genes), and evaluate its antibiotic susceptibility profile to assess potential public health risks. We first isolated *V. fluvialis* (strain 25C42) from the rectum of wild *Niviventer*, confirmed through Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) and 16S rRNA gene sequencing. Whole-genome sequencing (WGS) was performed using second-and third-generation technologies, with subsequent quality control and assembly. Six databases including KEGG, COG, CARD and VFDB were used for genome annotation. Antibiotic susceptibility was evaluated according to Clinical and Laboratory Standards Institute (CLSI) guidelines, determining the minimum inhibitory concentrations (MIC) for 16 antibiotics. Strain 25C42 was identified as *V. fluvialis*, confirmed by MALDI-TOF MS and 16S rRNA sequencing. WGS revealed a genome length of 2,720,341 bp, GC content of 32.57%. Functional genomic analysis identified 2,268 genes in the COG database and 2,023 genes in KEGG, highlighting key metabolic and cellular processes. Notably, 119 virulence genes and 65 antimicrobial resistance genes were found, indicating significant resistance potential. Phylogenetic analysis demonstrated a close relationship with other *Vagococcus* species, particularly *V. fluvialis* (ANI 98.57%, DDH 88.6%). Antibiotic susceptibility tests indicated strain 25C42 was resistant to clindamycin, tetracycline, rifampicin, cefoxitin and levofloxacin. Our findings reveal that the wild rodent-derived *V. fluvialis* strain 25C42 harbors clinically relevant antimicrobial resistance determinants and virulence-associated genes. The high genomic integrity and extensive functional gene annotation underscore its metabolic versatility. Notably, strain 25C42 exhibits significant antimicrobial resistance, necessitating ongoing surveillance and research to understand its implications for public health and environmental monitoring, as well as strategies for effective therapeutic intervention.

KEYWORDS

***Vagococcus fluvialis*, *Niviventer*, whole genome sequencing, genes annotation, drug resistance**

1 Introduction

Vagococcus fluvialis (*V. fluvialis*) is a Gram-positive bacterium classified within the genus *Vagococcus* and the family *Enterococcaceae*, closely related to other *enterococci*, such as *Enterococcus*. Originally isolated from chicken manure and river water, *V. fluvialis* was first characterized by Hashimoto et al. (Collins et al., 1989; Schleifer et al., 1985). The specific host species of *V. fluvialis* in animals remains uncertain, however, it was identified in bats by Qin et al. (2021) in China, followed by its detection in bovine urine by Giannattasio-Ferraz et al. (2021) in the United States. Notably, *Niviventer*, a small rodent prevalent in mountainous and forested regions of Asia, particularly in the Yunnan and Guizhou provinces of China, has been recognized as a host for a diverse array of pathogenic bacteria and parasites (Yan, 2021; Tian et al., 2020; Lu et al., 2022).

In recent years, the isolation of *V. fluvialis* from human patients has become increasingly common. Matsuo et al. (2021) reported a case of bacteremia and pressure ulcers associated with *V. fluvialis*, while Jadhav and Pai (2019) documented endocarditis linked to this bacterium. Furthermore, *V. fluvialis* has been isolated from the bile of patients suffering from cholecystitis and from peritoneal fluid in individuals with cirrhosis (Zhang et al., 2023; Kucuk et al., 2022). In China, Ting et al. (2019) isolated *V. fluvialis* from postoperative infected puncture fluid of the lower left femur for the first time. Additionally, this bacterium has been found in human urine on multiple occasions, demonstrating its potential to cause harm to patients (Chen et al., 2024; Kitano et al., 2024).

Researchers have raised concerns regarding the zoonotic transmission of *V. fluvialis* from animals to humans, leading to heightened clinical interest in this organism. Despite an increasing number of reported clinical cases, the direct clinical implications of *V. fluvialis* on human health remain unclear. Moreover, the spectrum of drug resistance and the underlying mechanisms of *V. fluvialis* are yet to be thoroughly investigated. Current studies indicate a potential pathogenicity, however, compared to other common pathogens, research on this bacterium is relatively limited. Other known species within the *Vagococcus* genus exhibit certain pathogenic characteristics and drug resistance profiles (Racero et al., 2021), further accentuating the clinical importance of *V. fluvialis*.

Consequently, it is reasonable to postulate that *V. fluvialis* may pose a risk to human health, however, our understanding of this organism is still inadequate. We report here the first isolation of *V. fluvialis* from the gastrointestinal tract of *Niviventer*. Although previous studies have provided insights into the biological characteristics of this strain (Collins et al., 1989; Schleifer et al., 1985), there is a notable paucity of research focusing on its antimicrobial resistance and pathogenicity, which are critical for clinical diagnosis and treatment. The discovery of *V. fluvialis* in diverse hosts is of paramount significance for elucidating its transmission mechanisms.

To address these gaps, The objectives of this study were: (1) to isolate and identify *V. fluvialis* from wild *Niviventer* using molecular and phenotypic methods; (2) to perform whole-genome sequencing and functional annotation to uncover genomic traits, including antimicrobial resistance and virulence determinants; (3) to determine the antibiotic resistance profile of the isolated strain; and (4) to assess the potential implications of these findings for zoonotic transmission and public health surveillance.

2 Methods and materials

2.1 Ethics approval

Our study was approved by the Ethics Committee of Guizhou Center for Disease Control and Prevention, approval number: G2019-01. The Ethics Committee agreed that the research was in accordance with the Helsinki Declaration and the Guidelines for the Good Treatment of Animals.

2.2 Source of strains

A collection of wild rats were conducted in Jinping County, Qiandongnan City, Guizhou Province. Traps were set overnight and retrieved the following morning. Live mice were euthanized through cervical dislocation under ether anesthesia to ensure a rapid and painless process, this method of euthanasia is recommended in animal euthanasia guidelines in China and the United States [Lu et al., 2021; National Technical Committee for Standardization of Laboratory Animals (SAC/TC 281), 2021]. This research complied with all relevant wildlife protection laws, ensuring the welfare of the captured animals. The species, size, and sex of the specimens were identified according to the Handbook of Important Medical Animals in China (Lu, 1982). After identification, the mice were promptly dissected in the local disease prevention and control center's laboratory under sterile conditions. Lung tissues (approximately 500 mg) were excised using surgical scissors and immediately inoculated into 1.5 mL of brain heart infusion liquid containing 20% glycerol for preservation. The specimens were homogenized and plated onto Columbia blood agar plates, which were then incubated at 37°C for 48 h to observe microbial growth. Single colonies were isolated for further purification on additional Columbia blood agar plates. The initial identification of the purified colonies was conducted using Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS), targeting the isolation of potentially significant microbial strains for subsequent identification and analysis. Finally, a strain suspected of *V. fluvialis* (Strain 25C42) was cultured in a rectal specimen of a Niviventer, and the identification score of MALDI-TOF MS was 8.285.

2.3 Identification of 16S rRNA gene

Strain 25C42 was isolated and subsequently purified. DNA extraction from the strain was performed using a nucleic acid extraction kit (Hangzhou Baiyi Technology Co., Ltd.). The 16S rRNA gene was amplified via PCR using the primers 27F (5′-AGTTTG ATCMTGGCTCAG-3′) and 1492R (5′-AGTTTGATCMTGGCTC AG-3′). Reaction system total 50 ul: Premix Taq 25 ul; Primer 27F 2 ul; Primer 1492R 2 ul; Template DNA 2 ul; Water 19 ul. The reaction condition of PCR amplification was 94°C for 10 min. 94°C 45 s, 52°C 60 s, 72°C 60 s, 40 cycles; 72°C for 10 min. The resultant PCR products were sequenced, and the obtained sequences were subjected to a comparative analysis against the NCBI BLAST database. Homologous sequences with high similarity from GenBank were downloaded for evolutionary analysis.

Phylogenetic reconstruction based on 16S rRNA gene sequences was performed using MEGA 11 software. Four distinct phylogenetic trees were generated through the Clustal_W alignment module, employing the following algorithms: Neighbor-Joining (NJ), Maximum Likelihood (ML), Maximum Parsimony (MP), and Minimum Evolution (ME). Branch node confidence was assessed through 1,000 bootstrap replications, with support values expressed as percentages.

2.4 Whole-genome sequencing

A hybrid assembly strategy combining second-generation and third-generation (Oxford Nanopore) sequencing data was employed to leverage the high accuracy of second-generation sequencing and the long-read capabilities of third-generation sequencing, yielding a complete and accurate genome assembly of the strain.

Second-Generation Sequencing was performed using advanced technologies provided by the Beijing Genomics Institute. The process began with DNA fragmentation via enzymatic digestion to generate fragments of 200–500 bp in length suitable for sequencing. Adapter ligation was then performed on both ends of the DNA fragments. The adapters consist of a DNA sequence that includes amplification primers, sequencing primers, and barcode sequences. The barcode allows for distinguishing between different samples and ensuring sequencing accuracy. Following adapter ligation, PCR amplification was carried out to construct the sequencing library. The amplification was performed using primers with dual-barcode sequences, and the products were purified before sequencing. The sequencing was conducted on the BGISEQ platform, which utilizes DNA nanoball (DNB) amplification technology. This method forms DNA nanoballs through rolling circle amplification and generates high-quality raw sequencing data in FASTQ format.

Third-generation sequencing was performed using Oxford Nanopore's platform, following the proprietary library preparation protocols. DNA fragmentation was carried out using Oxford Nanopore's DNA fragmentation method, and adapters were ligated to the DNA fragments. Library quality control was conducted using the Agilent Bioanalyzer to assess fragment size distribution, ensuring the library was suitable for sequencing. Sequencing was performed on the Oxford Nanopore platform using electrophoretic migration to load DNA samples into nanopore chips for real-time single-molecule sequencing. Data acquisition and preliminary processing were carried out using the Oxford Nanopore MinKNOW software, and raw sequencing data were obtained in FASTQ format.

The hybrid assembly strategy combined both second-and third-generation sequencing data using the Micro IBS Analyzer software (Beijing MicroFuture Technology Co., Ltd.) for pathogen genome assembly and identification. First, second-generation sequencing data underwent quality control using Trimmomatic (v0.39) to remove low-quality reads. Third-generation sequencing data were filtered and corrected using the PBJelly tool. Next, an initial assembly of the second-generation data was performed using SPAdes, while the third-generation data were initially assembled with Canu (v2.1.1). Finally, the hybrid assembly was refined by integrating both datasets with Pilon (v1.23), which allowed for genome correction and the generation of high-quality, clean data in FASTA format.

2.5 Genetic analysis

Genome annotation was executed using nine databases, including the Comprehensive Antibiotic Resistance Database (CARD), Kyoto Encyclopedia of Genes and Genomes (KEGG), Clusters of Orthologous Groups (COG), Non-Redundant Protein Database (NR), Pathogen-Host Interaction (PHI), Swiss-Prot, Virulence Factor Database (VFDB). These databases were accessed and annotated through the Beijing Micro Future Pathogen Microbiological Information Analysis System platform. Circular genome maps were generated utilizing CGView online software (available at https://stothardresearch.ca/cgview/). Phylogenetic relationships and genomic homology were assessed using Average Nucleotide Identity (ANI) to compare the genomic sequences of the bacteria against reference databases, thus providing a measure of genetic relatedness (available at https://www.ezbiocloud.net/). Furthermore, DNA–DNA hybridization (DDH) was performed to quantitatively assess genomic similarity between the unknown strain and reference strains (accessible at https://ggdc.dsmz.de/).

2.6 Drug resistance phenotype

Antimicrobial susceptibility testing was performed using the CHN5GOVF customized panel (broth microdilution method, National Pathogen Identification Network, China) following the Clinical and Laboratory Standards Institute (CLSI) M100 guidelines (Humphries et al., 2021). The procedure included: Bacterial suspension preparation: Pure colonies of *Vagococcus fluvialis* were adjusted to a 0.5 McFarland standard (~1×10^8 CFU/mL) in sterile saline and further diluted to a final concentration of 1×10^6 CFU/mL. Inoculation and incubation: The bacterial suspension was dispensed into the panel wells and incubated aerobically at 35°C for 18–24 h. MIC determination: The minimum inhibitory concentration (MIC) was defined as the lowest drug concentration showing no visible growth. Susceptibility (S), intermediate (I), or resistance (R) were interpreted according to CLSI breakpoints. Quality control: *Escherichia coli* ATCC 25922 was included as a quality control strain in each run.

2.7 Motility assay

Motility was assessed by inoculating strains into semi-solid agar and performing a stab-inoculation method. After 24 h of incubation at 37°C, motility was determined by the diffusion of the inoculation line: non-diffusive growth indicated negative motility, while a blurred, diffused line indicated positive motility.

3 Results

3.1 Morphological characterization of strain 25C42

Following Gram staining, strain 25C42 exhibited a positive result, and the colony morphology was cocci (Supplementary Figure S1A). Motility assays indicated that strain 25C42 was non-motile, contrasting with the positive motility observed in the control strain, *Escherichia coli* ATCC25922 (Supplementary Figure S1B).

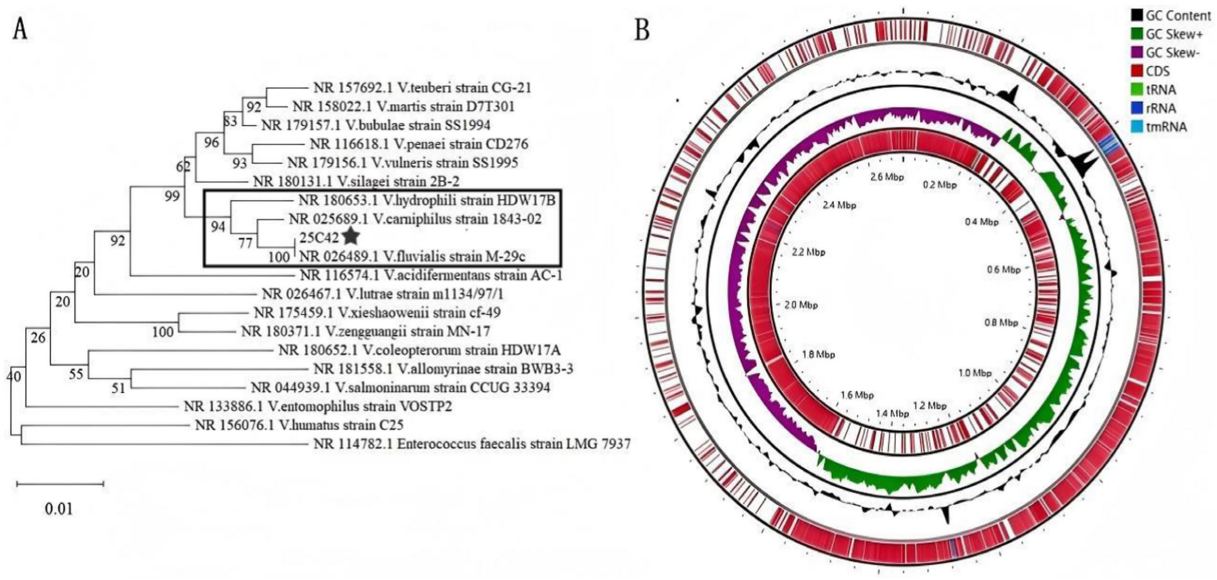


FIGURE 1
(A) Phylogenetic tree with 16S rRNA gene sequences of strains 25C42 and 18 other Vagococcus species. **(B)** Complete gene map of strain 25C42.

3.2 Phylogenetic relationships of 16S rRNA gene sequences

The identification of strain 25C42 through MALDI-TOF MS yielded a definitive classification as *V. fluvialis* with a high identification score of 8.285. Further confirmation was obtained via 16S rRNA gene sequencing, which showed 100.00% identity with *V. fluvialis* (NCBI accession number: NR_026489.1). The 16S rRNA gene sequence of strain 25C42 encompass 1,549 base pairs, we have uploaded the sequence to a public database.[1] We downloaded the 16S rRNA sequences of all *Vagococcus* species similar to strain 25C42 from the NCBI database, totaling 18 species, the phylogenetic tree was rooted using *Enterococcus faecalis* strain LMG 7937. The NJ phylogenetic tree demonstrated that strain 25C42 formed a robust monophyletic cluster with *V. fluvialis* strain M-29c (NR_026489.1), supported by a maximum bootstrap value of 100% (Figure 1A). Additionally, strain 25C42 demonstrated a close phylogenetic relationship with *V. hydrophili* strain HDW17B (NR_180653.1) and *V. carnipilus* strain 1843-02 (NR_025689.1), both of which showed high bootstrap values. This topology effectively resolved the evolutionary relationships among Vagococcus species members. The observed phylogenetic patterns were consistently validated by three complementary reconstruction methods (ML, ME, and MP), as illustrated in Supplementary Figures S2–S4.

3.3 Complete gene map of strain 25C42

By integrating both second and third-generation WGS data for assembly and gene prediction, we achieved a genome assembly integrity of 99.92%, maintaining a splicing quality of 100% and a reduced contamination level of 1.38%. The annotated genome comprises 77 tRNA genes and 7 copies each of the 23S, 16S, and 5S rRNA genes. The total sequence length is 2,720,341 bp, characterized by a GC content of 32.57%. The assembly yielded a single contig with no gaps, exhibiting N50 and N75 values both at 2,720,341 bp. We've uploaded the sequence to a public database.[2] Based on the criteria established by Bowers et al. (2017) and Duan et al. (2020), this genome sequence was classified as complete and of high quality. The complete gene map features a circular representation of the coding regions, color-coded according to functional categories, and includes information on non-coding RNAs (tmRNA, tRNA, rRNA), GC content, and coding DNA sequences (CDS) (Figure 1B).

3.4 Functional gene analysis

The complete gene sequence of strain 25C42 underwent annotation across six databases, yielding 2,268 genes in the COG database, 2,023 in KEGG, 133 in PHI, 2,610 in NR, 119 in VFDB, 65 in CARD.

COG analysis classified the 2,268 annotated genes into four primary categories and 23 functional groups. The major categories included Cellular Processes and Signaling, Information Storage and Processing, Metabolism, and Poorly Characterized functions. Notably, the most prevalent annotations were found in "Translation, ribosomal structure and biogenesis" (238 genes), "Carbohydrate transport and metabolism" (231 genes), and "Transcription" (205 genes). Other significant functional annotations encompassed cell wall/membrane/envelope biogenesis, inorganic ion transport and metabolism, and defense mechanisms (Figure 2A).

1 https://bigd.big.ac.cn/gsa/browse/CRA021120

2 https://bigd.big.ac.cn/gsa/browse/CRA021105

A total of 2,023 orthologous protein-coding genes were mapped to 32 KEGG metabolic pathways, categorized into Cellular Processes, Environmental Information Processing, Genetic Information Processing, Human Diseases, Metabolism, and Organismal Systems. The predominant pathways identified were related to Metabolism, including "Global and overview maps" (750 genes), "Carbohydrate metabolism" (223 genes), and "Amino acid metabolism" (122 genes). Additionally, Environmental Information Processing pathways, such as "Membrane transport" (114 genes), and pathways associated with Human Diseases, including "Drug resistance: Antimicrobial" (29 genes) and "Infectious diseases: Bacterial" (13 genes), were also noted (Figure 2B).

Pathogen-host interaction-related genes were examined with 133 genes annotated in the PHI database. These were categorized into 7 groups, with "Reduced virulence" (88 genes) being the most prominent, followed by "Unaffected pathogenicity" (24 genes), "Lethal" (8 genes), "Increased virulence (hypervirulence)" (7 genes), "Loss of pathogenicity" (4 genes), "Effector (plant avirulence determinant)" (1 genes), and "Chemistry target: sensitivity to chemical" (1 gene) (Figure 3A).

In the NR database, the gene sequence of strain 25C42 was translated into amino acid sequences, revealing 2,610 annotated genes. Notably, *V. fluvialis* was the most frequently annotated, comprising 86.59% of the annotations, followed by *Enterococcus* sp. at 8.62%, and *V.* sp. accounts for 0.84% (Figure 3B).

Virulence factors were classified into 6 functional categories: Invasion (3 genes), Secretion system (3 genes), Adherence (6 genes), Iron uptake system (14 genes), Stress protein (22 genes), and Toxin (23genes). DIAMOND analysis against the VFDB identified 119 putative virulence genes in strain 25C42, positive results were accepted with at least 40% identity (Figure 4A). There were 13 virulence genes with more than 60% identity (Supplementary Table 1). With the highest number of annotated genes associated with carbohydrate ABC transporter ATP-binding (sugC) and peptide/nickel transport system substrate-binding protein (oppA) in Toxin (Figure 4B). With the highest number of annotated genes associated with flagellar biosynthesis protein FlhA (flhA), capsular exopolysaccharide family protein (BCE_5400) and putative alcohol-acetaldehyde dehydrogenase (lap) in Stress protein (Figure 4C). With the highest number of annotated genes associated with DNA-binding response regulator RegX3 (regX3), e1-beta chain (pdhB) and Coxiella Dot/Icm type IVB secretion system translocated effector (CBU_1566) in Iron uptake system (Figure 4D). In addition, Secretion system includes bactoprenol glucosyl transferase (gtrB), capsular polysaccharide biosynthesis protein Cap5H (cap5H), Secretion System Chemotaxi-specific methylesterase (cheB). Invasion contains UDP-glucose pyrophosphorylase (hasC), chemotaxis protein CheA (cheA), type IV pilus response regulator PilG (pilG). Adherence includes Elongation factor Tu (tuf), chaperonin GroEL (groEL), autolysin (lytA), endocarditis specific antigen (efaA), molecular chaperone DnaK (CT396), putative lipoate protein ligase A (lplA1).

The CARD database annotated 65 antimicrobial resistance genes from strain 25C42 (identity >40%) (Supplementary Table 2). Identifying five primary resistance mechanisms: Antibiotic efflux, Antibiotic inactivation, Antibiotic target alteration, Antibiotic target protection, and Antibiotic target replacement. The resistance gene families included 21 distinct types, notably the ABC-F ATP-binding cassette ribosomal protection protein, major facilitator superfamily (MFS) antibiotic efflux pump, and glycopeptide resistance gene cluster. Predictions of drug resistance indicated the potential for resistance to a wide range of antibiotics, including acridine dyes, fluoroquinolone antibiotics, and others (Figure 5).

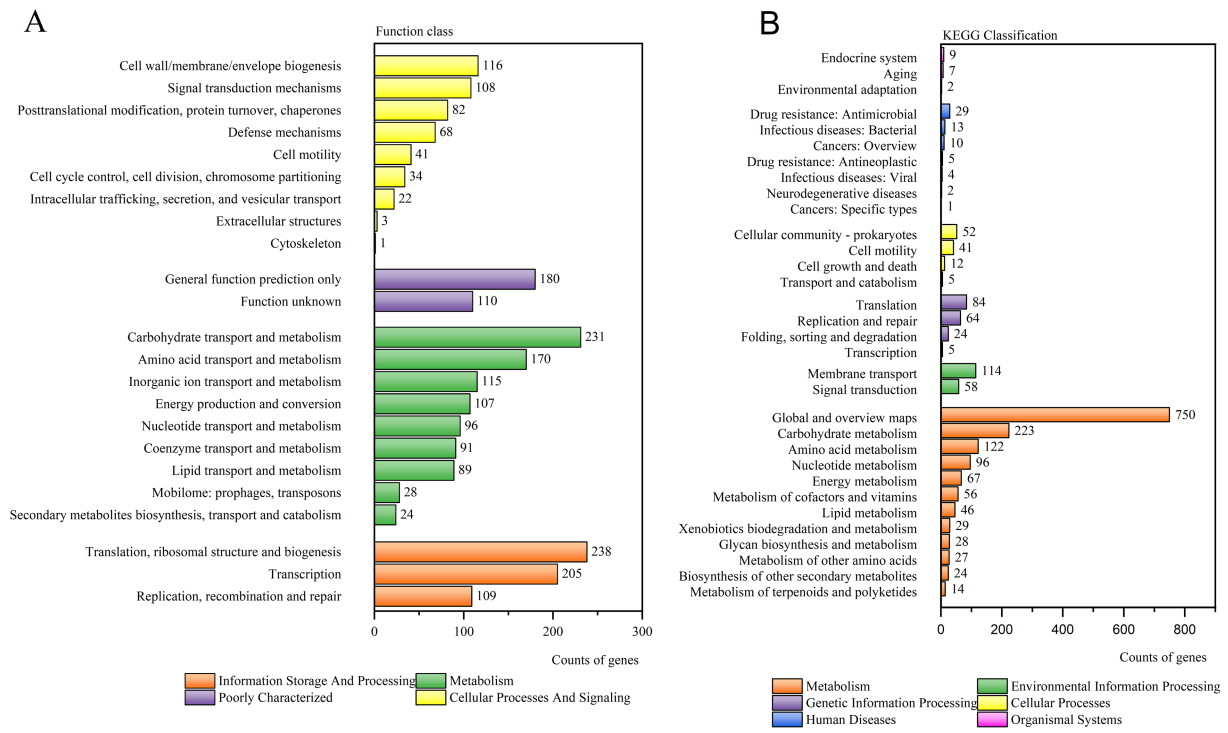


FIGURE 2
(A) COG analysis of strain 25C42. **(B)** KEGG analysis of strain 25C42.

10.3389/fmicb.2025.1546744

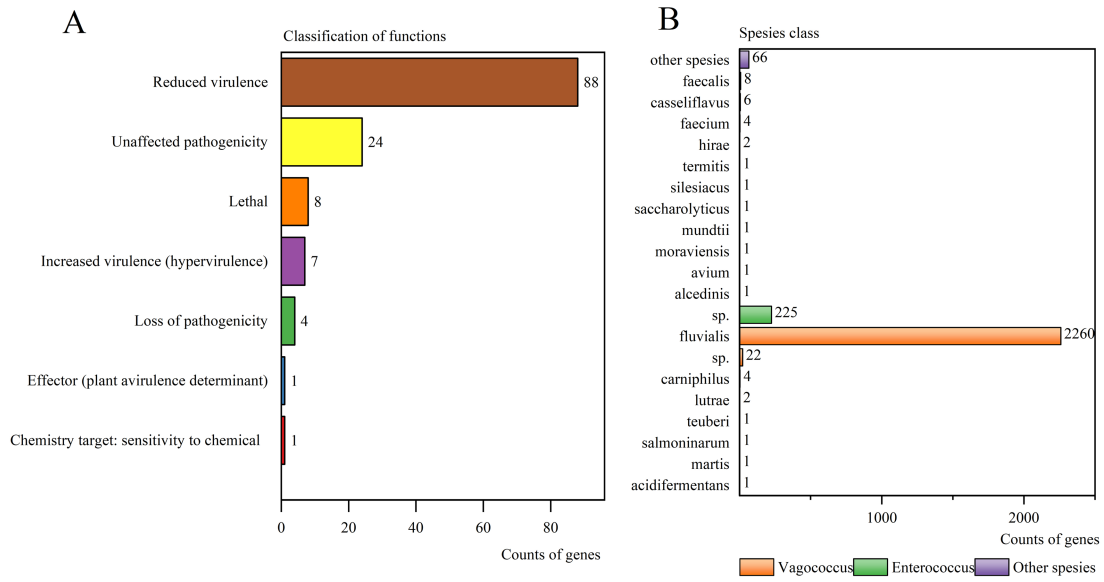


FIGURE 3
(A) Types distribution of PHI genes of strain 25C42. **(B)** NR analysis of strain 25C42.

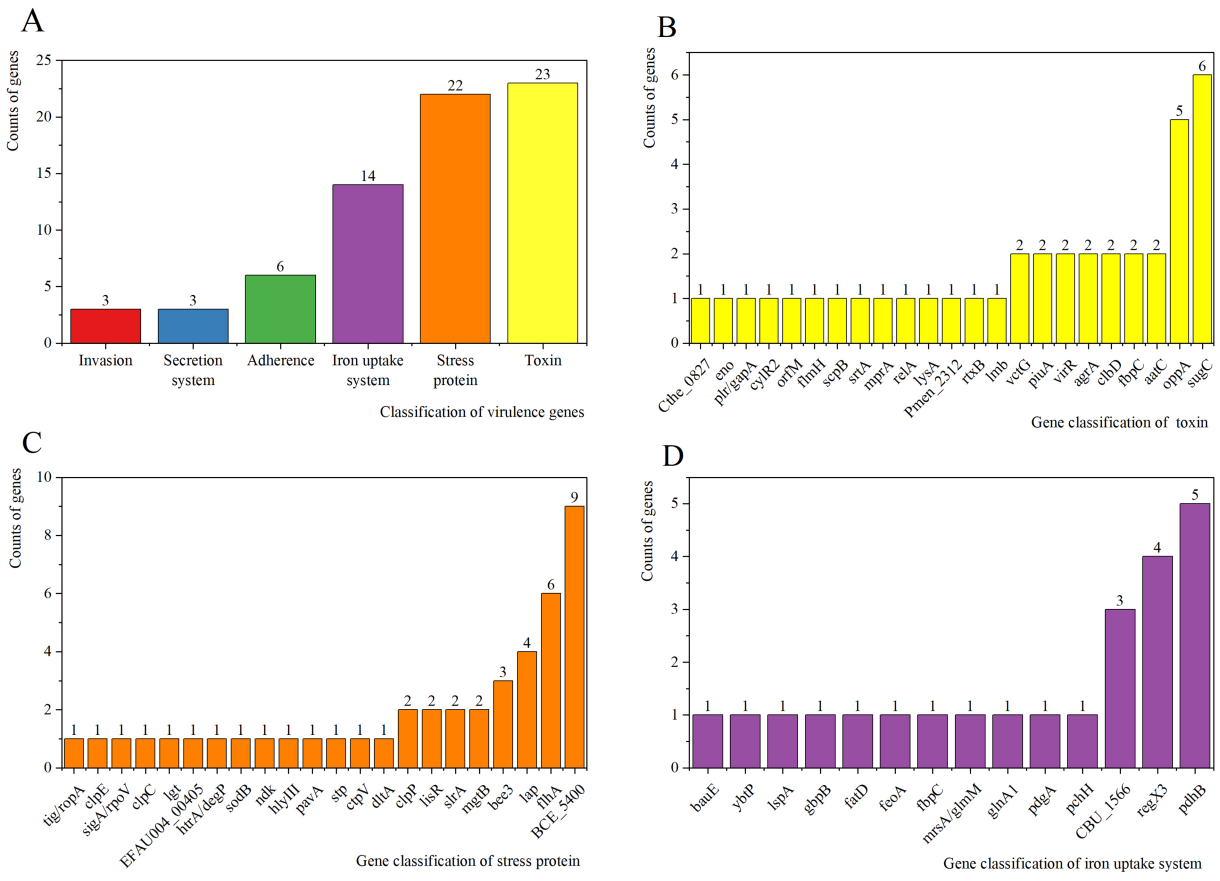


FIGURE 4
(A) VFDB analysis of strain 25C42. **(B)** Results of virulence gene distribution in toxin. **(C)** Results of virulence gene distribution in stress protein. **(D)** Results of virulence gene distribution in iron uptake system.

3.5 Whole genome similarity comparisons

For comparative genomic analysis, strain 25C42 was selected for ANI and DDH assessment against 18 other *Vagococcus* strains with complete genome sequences and closely related phylogenetic backgrounds. The ANI value between strain 25C42 and *V. fluvialis* (GCF_020628455.1) was determined to be 98.57%, exceeding the 95% threshold indicative of significant genomic similarity. Furthermore, strain 25C42 clustered closely with *V. hydrophili* (GCF_011304195.1) and *V. carnipilus* (GCF_014397115.1). Additionally, the ANI value between *V. Martis* (GCF_002026305.1) and *V. teuberi* (GCF_001870205.1) was calculated at 95.77% (Figure 6). For the DDH analysis, strain 25C42 was compared with 18 closely related *Vagococcus* strains using all three default calculation formulas. The highest DDH value obtained was 88.6% (DDH $\geq$ 70%), recorded in comparison with *V. fluvialis* (GCF_020628455.1), yielding a probability of 95.32% (Supplementary Table 3).

3.6 Antimicrobial resistance profile

The strain exhibited resistance to clindamycin, tetracycline, rifampin, cefoxitin, and levofloxacin (Table 1). Additionally, penicillin resistance may be related to the activity of natural β-lactamase (Rodriguez Jimenez et al., 2022). The strain was sensitive to erythromycin, while resistant to clindamycin, which could suggest erm-mediated inducible resistance, requiring confirmation through a D-test (Hollenbeck and Rice, 2012).

4 Discussion

The emergence of *V. fluvialis* as a potential human pathogen poses significant challenges within microbiology and clinical medicine. Historically regarded as a less prominent member of the *Enterococcaceae* family, recent findings indicate its isolation from a diverse range of hosts, including humans, bats, and livestock. This suggests a broader ecological niche than previously recognized. Our study reports the first isolation of the more resistant *V. fluvialis* in wild *Niviventer*, contributing to the growing understanding of this bacterium's presence in mammals. By investigating the resistance and pathogenicity of strains isolated from novel hosts, we aim to elucidate their potential public health implications and provide essential data for future environmental monitoring and the development of antimicrobial therapeutics. These findings not only enhance our comprehension of the transmission and adaptation mechanisms of *V. fluvialis* across different ecosystems but also offer a new avenue for research in the prevention and control of related infections.

The identification of strain 25C42 as *V. fluvialis* was substantiated through comprehensive analyses employing both MALDI-TOF mass spectrometry and 16S rRNA gene sequencing, achieving a 100% identity match with established *V. fluvialis* sequences (NCBI number:

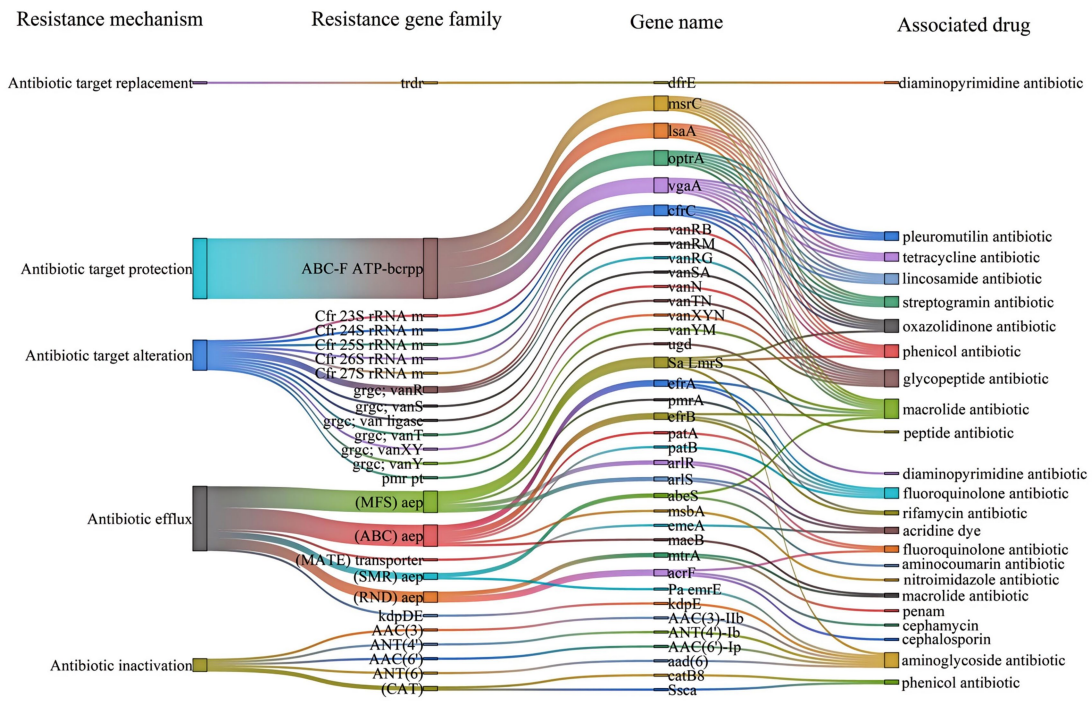


FIGURE 5
Drug-resistant genes of strain 25C42 annotated in the CARD database (identity > 40%). ABC-F, ABC-binding cassette ribosomal protection protein; ABC, ATP-binding cassette; Cfr 23S rRNA m, Cfr 23S ribosomal RNA methyltransferase; CAT, chloramphenicol acetyltransferase; grgc, glycopeptide resistance gene cluster; aep, antibiotic efflux pump; MFS, major facilitator superfamily; MATE, multidrug and toxic compound extrusion transporter; pmr pt., pmr phosphoethanolamine transferase; RND, resistance-nodulation-cell division; SMR, small multidrug resistance; trdr, trimethoprim-resistant dihydrofolate reductase (dfr); Pa emrE, *Pseudomonas aeruginosa* emrE; Sa LmrS, *Staphylococcus aureus* LmrS; Ssca, *Streptococcus suis* chloramphenicol acetyltransferase.

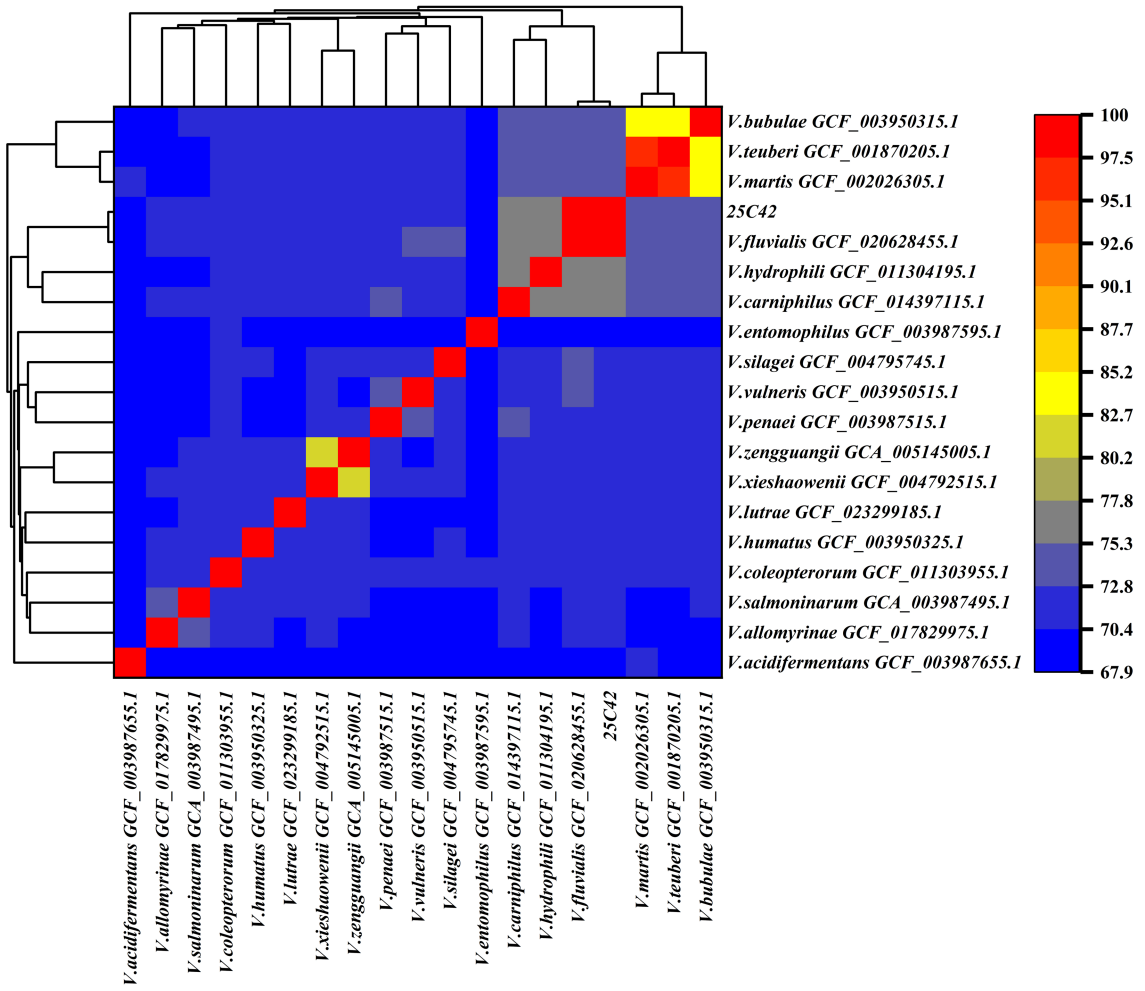


FIGURE 6
Heat map of ANI between strain 25C42 and 18 phylogentically close *Vagococcus* species.

NR_026489.1). This underscores the reliability of these methodologies for accurate identification, aligning with findings from prior studies (Chen et al., 2024; Kitano et al., 2024).

Our sequencing efforts yielded a high-quality genome assembly with an integrity of 99.92% and a contamination level of 1.38%, demonstrating the efficacy of integrating second-and third-generation sequencing technologies for thorough genomic analysis. The assembled genome spans a total length of 2,720,341 bp with a GC content of 32.57%, further affirming its classification within the *Vagococcus* genus, as similar GC content ranges have been reported in closely related species (Lewis and Jorgensen, 2005).

Gene annotation via Prokka unveiled a diverse array of functional genes categorized into multiple biological processes. Notably, the predominant functions identified encompassed translation, ribosomal structure, carbohydrate transport, and amino acid metabolism. These results are consistent with the observations of Rodriguez Jimenez et al. (2022), who highlighted metabolic versatility as a defining trait of *Vagococcus* species, enabling adaptation to fluctuating environmental conditions. The comprehensive mapping of 2,268 genes to COG functional categories illustrates a complex regulatory and metabolic network, which is essential for the organism's survival and pathogenic potential.

The pathogenicity potential of strain 25C42 is further accentuated by the identification of 133 pathogen-host interaction-related genes, predominantly linked to reduced virulence. This observation suggests that many environmental *Vagococcus* species harbor genes that may modulate virulence without necessarily conferring hypervirulence. Furthermore, the identification of virulence factors, including secretion systems and iron uptake mechanisms, corroborates the findings of Jimenez Ana et al., highlighting the significance of these factors in the pathogenicity of *Vagococcus* species (Lewis and Jorgensen, 2005).

This study is the first to report the antibiotic susceptibility profile of *Vagococcus fluvialis* to 16 antibiotics in the CHN5GOVF panel. The antimicrobial resistance profile of strain 25C42 is particularly concerning. Resistance to clindamycin, tetracycline, rifampicin, and cefoxitin reflects an alarming trend in *V. fluvialis*, as noted in studies where strains isolated from infected patients exhibited poor antibacterial efficacy against clindamycin, oxacillin, and sulfamethoxazole/trimethoprim (Ting et al., 2019). Additionally, *V. fluvialis* isolated from bats demonstrated potential resistance to macrolides (Qin et al., 2021). The results indicated multi-drug resistance, with the strain showing resistance to clindamycin, tetracycline, rifampin, cefoxitin, and levofloxacin, suggesting the presence of multiple resistance genes, such as erm, tet, and gyrA

TABLE 1 Antimicrobial susceptibility profile of strain 25C42.

Class	Antibiotic	Abbreviation	MIC (μg/mL)	Interpretation
β-Lactams	Penicillin	PEN	0.5	R*
	Cefoxitin	CFX	>8	R
Macrolides/Lincosamides	Clindamycin	CLI	>8	R
	Erythromycin	ERY	≤0.5	S
	Erythromycin/Clindamycin	ERY/CLI	≤4/0.5	S/R†
Tetracyclines	Tetracycline	TET	>16	R
Rifamycins	Rifampicin	RIF	>4	R
Fluoroquinolones	Levofloxacin	LEV	4	R
Glycopeptides	Vancomycin	VAN	1	S
	Teicoplanin	TEC	≤0.5	S
Oxazolidinones	Linezolid	LZD	≤1	S
Lipopeptides	Daptomycin	DAP	0.5	S
Aminoglycosides	Gentamicin	GEN	0.5	S
Folate pathway inhibitors	Cotrimoxazole	SXT	≤0.25	S
Nitrofurans	Nitrofurantoin	NIT	≤16	S
Other	Oxacillin	OXC	≤0.12	S

"R" indicates resistant; "S" indicates sensitive; R*, Penicillin resistance may be related to natural beta-lactamase activity; S/R†, Erythromycin sensitive but clindamycin resistant suggests potential induced resistance (D test required).

mutations (RDarby et al., 2023). Susceptible antibiotics included glycopeptides (vancomycin, teicoplanin), aminoglycosides (gentamicin), and linezolid (Arias and Murray, 2012), offering alternative treatment options. The minimum inhibitory concentration (MIC) of penicillin was 0.5 μg/mL (determined resistant by CLSI), which may reflect *β*-lactamase activity or mutations in penicillin-binding proteins (PBPs) (Queenan and Bush, 2007), and should be further verified by molecular testing (e.g., blaZ gene detection). The sensitivity to erythromycin and resistance to clindamycin may indicate erm-mediated inducible resistance, which needs to be confirmed by a D-test (Hollenbeck and Rice, 2012). Future studies are warranted to determine whether different sources of *V. fluvialis* exhibit distinct resistance patterns or resistance genes.

Genomic analysis revealed 65 antibacterial resistance genes in strain 25C42, many of which correlate with its phenotypic resistance. For example, lsa (A) and msrC likely mediate clindamycin resistance through ribosomal protection and efflux, while tet (M) and tet (L) synergistically confer tetracycline resistance. Similarly, fluoroquinolone resistance (levofloxacin) aligns with efflux pumps (pmrA, qacA) and potential gyrA mutations. However, some discrepancies exist: rifampicin resistance (MIC > 4 μg/mL) may stem from rpoB mutations not captured by CARD, and cefoxitin resistance could involve unannotated β-lactamases or PBPs. Notably, erythromycin sensitivity despite erm (B) presence suggests inducible resistance, requiring phenotypic validation via D-test. These findings underscore the importance of integrating genomic predictions with phenotypic assays to fully resolve resistance mechanisms, particularly for emerging pathogens like *V. fluvialis*.

Identified mechanisms of resistance, such as antibiotic efflux and target alteration, mirror patterns observed in other pathogenic bacteria, indicating shared evolutionary pathways (Jiang et al., 2018; Guitor and Wright, 2018). This finding emphasizes the critical need for ongoing surveillance of antimicrobial resistance patterns in *V. fluvialis* to mitigate the public health risks associated with these emerging pathogens.

Phylogenetic analyses revealed that strain 25C42 is closely related to *V. fluvialis* strain M-29c, with a high Average Nucleotide Identity (ANI) value of 98.57% and robust DNA–DNA Hybridization (DDH) values further supporting its classification. Collectively, these findings illuminate the intricate interplay between environmental adaptation, pathogenicity, and antimicrobial resistance in *V. fluvialis*, underscoring the necessity for further investigations to unravel the implications of these traits in clinical and environmental contexts.

The limitations of this study include the lack of detection of resistance genes and the absence of animal model validation.

5 Conclusion

Our findings highlight the emergent role of *V. fluvialis* as a potential human pathogen, revealing its resistance mechanisms and pathogenicity in diverse hosts. The high genomic integrity and extensive functional gene annotation underscore its metabolic versatility. Notably, strain 25C42 exhibits significant antimicrobial resistance, necessitating ongoing surveillance and research to understand its implications for public health and environmental monitoring, as well as strategies for effective therapeutic intervention.

Data availability statement

The datasets generated and/or analyzed during the current study are available in the Genome Sequence Archive (GSA) repository, 16S rRNA gene sequence: https://bigd.big.ac.cn/gsa/browse/CRA021120 (accession number: CRA021120); whole genome sequences: https://bigd.big.ac.cn/gsa/browse/CRA021105 (accession number: CRA021105).

Ethics statement

The animal study was approved by the Ethics Committee of Guizhou Center for Disease Control and Prevention. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

JiaZ: Data curation, Investigation, Methodology, Software, Writing – original draft. YL: Investigation, Methodology, Writing – original draft. TG: Methodology, Writing – original draft. JinZ: Resources, Writing – original draft. FC: Investigation, Writing – original draft. YH: Conceptualization, Data curation, Formal analysis, Supervision, Validation, Writing – review & editing. SL: Funding acquisition, Project administration, Supervision, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by Project for Public Health Talent Cultivation of China (grant no. Guo Jikong Zong Ren Han [2024]122); Guizhou Infectious Disease Talent Training Base project "Difficult bacteria identification Research team" (no. RCJD2102).

Acknowledgments

The authors are very grateful to the staff of Jinping CDC who participated in 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.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2025.1546744/full#supplementary-material

SUPPLEMENTARY TABLE 1
VFDB.

SUPPLEMENTARY TABLE 2
CARD.

SUPPLEMENTARY TABLE 3
DDH values between 25C42 and 18 phylogenetically close Vagococcus species.

SUPPLEMENTARY FIGURE 1
(A) Morphology of strain 25C42 under microscope (100x oil lens). **(B)** Motility assays observation (*Escherichia coli* ATCC25922 was the control group).

SUPPLEMENTARY FIGURE 2
Phylogenetic tree(ML) with 16S rRNA gene sequences of strains 25C42 and 18 other *Vagococcus* species.

SUPPLEMENTARY FIGURE 3
Phylogenetic tree(ME) with 16S rRNA gene sequences of strains 25C42 and 18 other *Vagococcus* species.

SUPPLEMENTARY FIGURE 4
Phylogenetic tree(MP) with 16S rRNA gene sequences of strains 25C42 and 18 other *Vagococcus* species.

References

Arias, C. A., and Murray, B. E. (2012). The rise of the Enterococcus: beyond vancomycin resistance. *Nat. Rev. Microbiol.* 10, 266–278. doi: 10.1038/nrmicro2761

Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., et al. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* 35, 725–731. doi: 10.1038/nbt.3893

Chen, Q., Tan, S., Long, S., Wang, K., and Liu, Q. (2024). *V. fluvialis* isolation from the urine of a bladder cancer patient: a case report. *BMC Infect. Dis.* 24:261. doi: 10.1186/s12879-024-09082-w

Collins, M. D., Ash, C., Farrow, J. A., Wallbanks, S., and Williams, A. M. (1989). 16S ribosomal ribonucleic acid sequence analyses of lactococci and related taxa. Description of *Vagococcus fluvialis* gen. Nov., sp. nov. *J. Appl. Bacteriol.* 67, 453–460. doi: 10.1111/j.1365-2672.1989.tb02516.x

Duan, Y. F., Wang, S. Y., Chen, Y. B., Yang, R. F., Li, H. K., Zhu, H. Q., et al. (2020). Expert consensus on microbiome sequencing and analysis. *Biol Engineer J* 36, 2516–2524. doi: 10.13345/j.cjb.200386

Giannattasio-Ferraz, S., Ene, A., Maskeri, L., Oliveira, A. P., Barbosa-Stancioli, E. F., and Putonti, C. (2021). *Vagococcus fluvialis* isolation and sequencing from urine of healthy cattle. *G3* 11:jkaa034. doi: 10.1093/g3journal/jkaa034

Guitor, A. K., and Wright, G. D. (2018). Antimicrobial resistance and respiratory infections. *Chest* 154, 1202–1212. doi: 10.1016/j.chest.2018.06.019

Hollenbeck, B. L., and Rice, L. B. (2012). Intrinsic and acquired resistance mechanisms in enterococcus. *Virulence* 3, 421–433. doi: 10.4161/viru.21282

Humphries, R., Bobenchik, A. M., Hindler, J. A., and Schuetz, A. N. (2021). Overview of changes to the clinical and laboratory standards institute performance standards for antimicrobial susceptibility testing, M100, 31st edition. *J. Clin. Microbiol.* 59:e0021321. doi: 10.1128/JCM.00213-21

Jadhav, K. P., and Pai, P. G. (2019). A rare infective endocarditis caused by *V. fluvialis*. *J. Cardiol. Cases* 20, 129–131. doi: 10.1016/j.jccase.2019.07.001

Jiang, S., Zeng, J., Zhou, X., and Li, Y. (2018). Drug resistance and gene transfer mechanisms in respiratory/Oral Bacteria. *J. Dent. Res.* 97, 1092–1099. doi: 10.1177/0022034518782659

Kitano, H., Kitagawa, H., Tadera, K., Saito, K., Kohada, Y., Takemoto, K., et al. (2024). First reported human case of isolation of *V. fluvialis* from the urine of a former zoo clerk in Japan: a case report. *BMC Infect. Dis.* 24:341. doi: 10.1186/s12879-024-09193-4

Kucuk, I., Daldaban Dincer, S., Yazgan, Y., Secen, B. S., Yildirim, I., and Kaplan, M. (2022). *V. fluvialis* first isolated from the ascitic fluid of a patient with cirrhosis. *Rev. Esp. Enferm. Dig.* 114:679. doi: 10.17235/reed.2022.8852/2022

Lewis, J. S., and Jorgensen, J. H. (2005). Inducible clindamycin resistance in staphylococci: should clinicians and microbiologists be concerned? *Clin. Infect. Dis.* 40, 280–285. doi: 10.1086/426894

Lu, B. L. (1982). Chinese important medical animal identification manual. Guizhou, China: People's Publishing House.

Lu, M., Tang, G., Ren, Z., Zhang, J., Wang, W., Qin, X., et al. (2022). Ehrlichia, Coxiella and Bartonella infections in rodents from Guizhou Province, Southwest China. *Ticks Tick Borne Dis.* 13:101974. doi: 10.1016/j.ttbdis.2022.101974

Lu, J., Zhang, Y., Pan, X. Y., Wang, J., Yan, G. F., Zhou, J., et al. (2021). Analysis of the 2020 American Veterinary Medical Association animal euthanasia guidelines. *Lab Anim Compar Med* 41, 195–206. doi: 10.12300/j.issn.1674-5817.2021.086

Matsuo, T., Mori, N., Kawai, F., Sakurai, A., Toyoda, M., Mikami, Y., et al. (2021). *V. fluvialis* as a causative pathogen of bloodstream and decubitus ulcer infection: case report and systematic review of the literature. *J. Infect. Chemother.* 27, 359–363. doi: 10.1016/j.jiac.2020.09.019

National Technical Committee for Standardization of Laboratory Animals (SAC/TC 281) (2021). Guide to Euthanasia of Laboratory Animals. Guizhou, China: National Technical Committee for Standardization of Laboratory Animals (SAC/TC 281).

Qin, T., Jiang, L. X., Ren, H. Y., Jia, X. Y., Nie, X. D., and Li, Y. N. (2021). Isolation and characterization of *V. fluvialis* from bats. *Biomed. Environ. Sci.* 34, 834–837. doi: 10.3967/bes2021.114

Queenan, A. M., and Bush, K. (2007). Carbapenemases: the versatile β-lactamases. *Clin. Microbiol. Rev.* 20, 440–458. doi: 10.1128/CMR.00001-07

Racero, L., Barberis, C., Traglia, G., Loza, M. S., Vay, C., and Almuzara, M. (2021). Infections due to Vagococcus spp. microbiological and clinical aspects and literature review. *Enferm Infecc Microbiol Clin* 39, 335–339. doi: 10.1016/j.eimce.2021.05.002

RDarby, E. M., Trampari, E., Siasat, P., Gaya, M. S., Alav, I., Webber, M. A., et al. (2023). Molecular mechanisms of antibiotic resistance revisited. *Nat. Rev. Microbiol.* 21, 280–295. doi: 10.1038/s41579-022-00820-y

Rodriguez Jimenez, A., Guiglielmoni, N., Goetghebuer, L., Dechamps, E., George, I. F., and Flot, J. F. (2022). Comparative genome analysis of *Vagococcus fluvialis* reveals abundance of mobile genetic elements in sponge-isolated strains. *BMC Genomics* 23:618. doi: 10.1186/s12864-022-08842-9

Schleifer, K. H., Kraus, J., Dvorak, C., Kilpper-Bälz, 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

Tian, H., Tie, W. F., Li, H., Hu, X., Xie, G. C., Du, L. Y., et al. (2020). Orthohantaviruses infections in humans and rodents in Baoji, China. *PLoS Negl. Trop. Dis.* 14:e0008778. doi: 10.1371/journal.pntd.0008778

Ting, Z., Cao, Y., Xiaolei, D., et al. (2019). Micrococcus fluvialis was isolated for the first time from infected puncture fluid of the lower left femur after surgery. *Chin J Nosocom Infect* 29, 2825–2829. doi: 10.11816/cn.ni.2019-190582

Yan, L. (2021). Fiber chigger field in Southwest China, and other important chigger distribution research. Guizhou, China: Guizhou University.

Zhang, D., Wang, X., Yu, J., Dai, Z., Li, Q., and Zhang, L. (2023). A case of *V. fluvialis* isolated from the bile of a patient with calculous cholecystitis. *BMC Infect. Dis.* 23:689. doi: 10.1186/s12879-023-08696-w

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 24 April 2025
DOI 10.3389/fmicb.2025.1546594

OPEN ACCESS

EDITED BY
Renmao "Tim" Tian,
Illinois Institute of Technology, United States

REVIEWED BY
Ulises Garza-Ramos,
National Institute of Public Health, Mexico
Diana Tilevik,
University of Skövde, Sweden
Samiksha Garse,
DY Patil Deemed to be University, India

*CORRESPONDENCE
Yuan Fang
✉ yuan.fang@genoxor.com
Jingbo Chen
✉ jimmychen777@smmu.edu.cn

RECEIVED 17 December 2024
ACCEPTED 02 April 2025
PUBLISHED 24 April 2025

CITATION
Sun Y, Cai Q, Li T, Chen J and Fang Y (2025) Genome assembly of *Klebsiella michiganensis* based on metagenomic next-generation sequencing reveals its genomic characteristics in population genetics and molecular epidemiology. *Front. Microbiol.* 16:1546594. doi: 10.3389/fmicb.2025.1546594

Genome assembly of *Klebsiella michiganensis* based on metagenomic next-generation sequencing reveals its genomic characteristics in population genetics and molecular epidemiology

Yong Sun[1], Qingqing Cai[2], Tianyu Li[2], Jingbo Chen[3]* and Yuan Fang[2]*

[1]Department of Respiratory and Critical Care Medicine, Anqing Medical Center of Anhui Medical University, Anqing Municipal Hospital, Anqing, China, [2]Genoxor Medical Science and Technology Inc., Shanghai, China, [3]Faculty of Naval Medicine, Navy Medical University, Shanghai, China

Introduction: *Klebsiella michiganensis*, a significant member of the *Klebsiella oxytoca* complex, has emerged as a potential pathogen in clinical settings. Despite extensive research on the *Klebsiella pneumoniae* complex, the pathogenicity and drug resistance of the *K. oxytoca* complex remain understudied, particularly regarding the reconstruction of whole genomes from metagenomic next-generation sequencing (mNGS) data.

Methods: In this study, bronchoalveolar lavage fluid (BALF) from a 55-year-old woman with a suspected right lung infection in Anhui Province, China, was analyzed using mNGS.

Results: Three distinct assembly strategies were employed to reconstruct the genome of *K. michiganensis*, leading to the identification of a novel ST452 strain, KMLRT2206. Comprehensive genomic analysis of this strain and 206 clinical isolates (genomes downloaded from public databases) revealed the population structure, distribution of drug resistance genes, and virulence factors of *K. michiganensis*. The results demonstrated significant genetic diversity, with the species divided into three major clades, each exhibiting distinct patterns of drug resistance and virulence genes. Notably, 38.6% of the strains harbored the $bla_{OXY-1-1}$ gene, highlighting a potential threat of drug resistance. While virulence gene distribution was not correlated with sequence type (ST), significant differences were observed among clades.

Conclusion: This study underscores the value of mNGS combined with optimized assembly strategies for accurate species identification within the *K. oxytoca* complex, providing critical insights for clinical pathogen detection and epidemiological surveillance.

KEYWORDS

***Klebsiella michiganensis*, genome assembly, metagenomic next-generation sequencing, genomic characteristics, epidemiology**

Introduction

Klebsiella is a gram-negative, rod-shaped, facultative anaerobic organism. Known for its capability to produce polysaccharide capsules, it possesses a defense mechanism against the host immune responses. *Klebsiella* has been isolated and characterized from diverse environmental habitats, including soil, food, plants, insects, and water. However, it is primarily recognized for its associated with hospital-acquired infections (Dantur et al., 2018; Hu et al., 2022). *Klebsiella michiganensis* is an important member of the *Klebsiella oxytoca* complex within the *Klebsiella* genus. Strains of the *K. oxytoca* complex are non-spore-forming and non-motile, forming smooth, round, dome-shaped, and glistening colonies on agar plates (Merla et al., 2019). *K. michiganensis* was first discovered in 2012 in a toothbrush holder in a Michigan household and exhibits biochemical characteristics consistent with the *Klebsiella* genus (Saha et al., 2013). This emerging pathogen has since been reported in clinical settings worldwide (Kula et al., 2024; López-Camacho et al., 2025; Yamada et al., 2024). Drug-resistant strains of *K. michiganensis* have been isolated from the abdominal fistula, sputum, blood, and rectal swabs (Chapman et al., 2020). The species demonstrates various resistance mechanisms similar to other pathogenic *Klebsiella* species, including genome- and plasmid-mediated extended-spectrum β-lactamases, carbapenemases, and drug efflux pumps (Abed et al., 2021; Flerlage et al., 2020; Zhang et al., 2021).

Next-generation sequencing (NGS) provides valuable data for epidemiological monitoring due to its high-throughput capabilities, particularly in reconstructing entire target genome from metagenomic NGS (mNGS) data. This information is crucial for clinical practice, enabling the identification of pathogenic strains and understanding their characteristics. Advances in NGS technology have revealed mechanisms by which *Klebsiella* invades the host immune system and causes disease, including studies on lipopolysaccharide, fimbriae, outer membrane proteins, siderophores, and allantoin metabolism related to virulence and genomics in *K. pneumoniae* (Hu et al., 2022; Igo and Schaffner, 2022).

K. michiganensis shares the closest phylogenetic relationship with *K. oxytoca*, with 99% similarity in the 16S rRNA gene sequences of two species (Yang et al., 2022). Despite the use of advanced technologies such as matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and 16S rRNA gene sequencing in clinical laboratories, *K. michiganensis* is often misidentified as *K. oxytoca*. In 2013, MALDI-TOF MS misidentified *K. michiganensis*, likely because this species, first discovered in 2012, had not yet been included in the database. With updated dataset, MALDI-TOF MS can now accurately identify the species within the *K. oxytoca* complex (Merla et al., 2019). For example, strain 12084 and strain K210011, isolated from patients' sputum samples and rectal swabs were initially identified as *K. oxytoca* based on MALDI-TOF/MS. Subsequent whole-genome analysis confirmed these strains as *K. michiganensis* (Li et al., 2022). Despite recent publications on the genome of *K. michiganensis*, there remains a lack of research on the direct reconstruction of high-quality genomes of *K. michiganensis* from mNGS data. Therefore, our study focused on the genomic reconstruction of *K. michiganensis* from clinical metagenomic sequencing data and conducted a comprehensive genomic analysis of global clinical isolates to address critical gaps in understanding its pathogenicity, drug resistance, and population structure, while also providing valuable insights for clinical pathogen detection and epidemiological surveillance.

Materials and methods

Clinical specimen

Figure 1 depicts the analytical workflow of this study. A BALF sample was collected from a 55-year-old woman with a suspected right lung infection in Anqing Medical Center of Anhui Medical University, China, in May 2022. Upon admission, the patient received empirical antibiotic treatment. During the diagnostic process, bronchoscopy and pathogen culture were performed. With patient's consent, the BALF sample was sent to Genoxor Medical & Science Technology Inc. (Shanghai, China) for mNGS test. After 48 h, the BALF cultivation test returned negative results, while the mNGS test identified numerous *Klebsiella* sp. sequences, primarily including *Klebsiella michiganensis*, *Klebsiella oxytoca*, *Klebsiella pneumoniae*, and *Klebsiella* aerogenes. Following treatment with Piperacillin Sodium and Tazobactam Sodium (4.5 g, IV, tid), the patient was discharged three days later with improved symptoms.

Metagenomic sequencing and taxonomic profiling

DNA was extracted from the BALF sample using a TIANamp Micro DNA Kit (DP710-T2A, TIANGEN BIOTECH) according to the manufacturer's instructions. Metagenomic libraries were then constructed using the Hieff NGS OnePot Pro DNA Library Prep Kit for Illumina (Yeasen Biotech, China). Sequencing was performed in a 75-bp single-end mode on a NextSeq 550 system (Illumina Inc., USA). To ensure high-quality sequencing data, low-quality and short reads were removed (Chen et al., 2018). A quality cutoff value of Q20 was set, and reads with base quality values below Q15 at any position were discarded. Additionally, reads shorter than 50 bp were filtered out. Human sequence data were identified and removed by mapping the data to the human reference genome (hg19) using Bowtie v2.2.6 (Langmead and Salzberg, 2012) with the parameters –threads 24 –seed 100 –no-unal. The remaining data were classified using the NCBI Microbial Genome Database. To estimate the relative abundance of each species reads in the bacterial community, the taxonomy of species abundance was calculated using Kraken v2.0.9 (Wood et al., 2019). Subsequently, metaMLST (Zolfo et al., 2017) was applied for strain-level ST typing and identification of the metagenomic data. No-template control (NTC) samples were sequenced simultaneously to monitor contamination during the experiments.

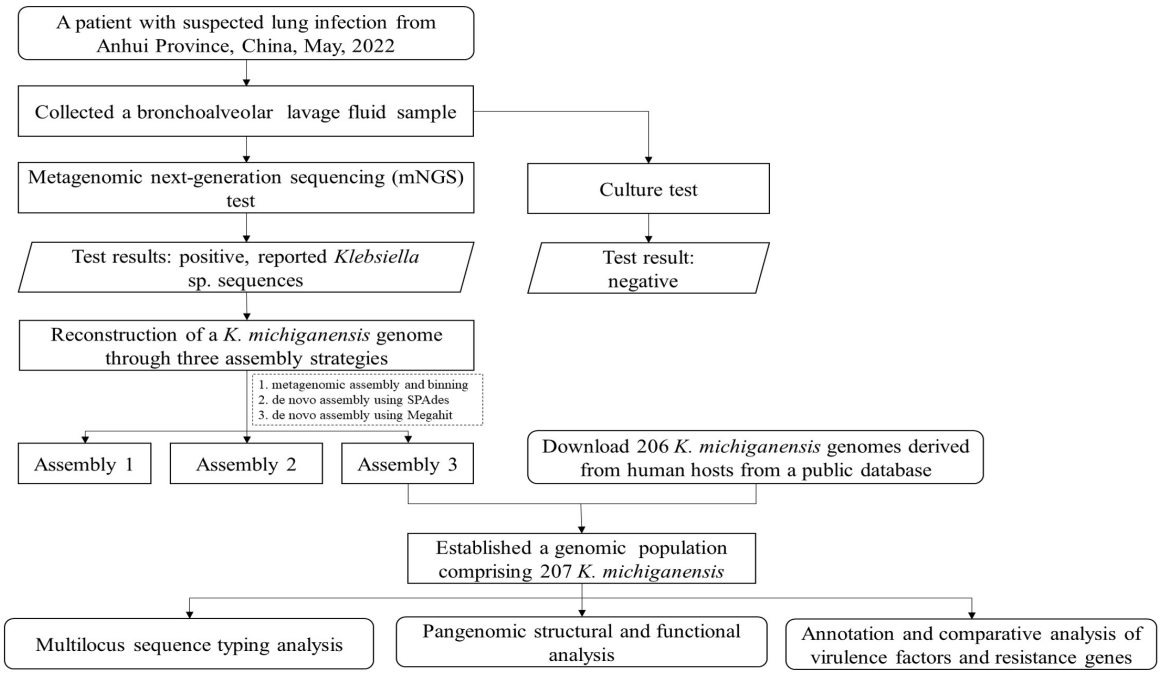


FIGURE 1
Schematic workflow of the research methodology.

Genome assembly and characterization of the targeted pathogen

Three strategies (designated as Assembly 1–3) were employed for the genomic reconstruction of *K. michiganensis* from metagenomic reads. For Assembly 1, MetaSpades v3.15.4 (Nurk et al., 2017) was used for assembly, followed by genome binning using the binning and bin refinement modules in metaWRAP v1.3.2 (Uritskiy et al., 2018). Contig binning was performed using MetaBAT and Maxbin2, and the resulting bins were optimized and combined to generate a draft genome for a single strain. The quality of g draft genome (Assembly 1) was assessed using CheckM (Parks et al., 2015), which evaluated completeness and contamination. Taxonomic classification was performed using GTDB-tk (Chaumeil et al., 2022), confirming the draft genome as *K. michiganensis*. For Assembly 2 and Assembly 3, RefSeq genomes of 384 *K. michiganensis* isolates (NCBI TaxID: 1134687) were retrieved from the NCBI Assembly database. Reads affiliated with *K. michiganensis* were captured using BBmap and assembled using SPAdes (parameters: -t 24 -m 128 –cov-cutoff auto –isolate) (Gurevich et al., 2013) and Megahit software (Li et al., 2015), respectively. These three assembly strategies were selected because Assembly 1 is optimized for complex metagenomic data processing, Assembly 2 is tailored for specific analysis of the target strain, and Assembly 3 can efficiently handles large-scale data, enabling a more comprehensive reconstruction of *K. michiganensis* genome.

The quality of genome assemblies was evaluated using QUAST (Nurk et al., 2017) with the reference genome of the THO-011 Strain (GenBank accession: ASM1513957v1), while contamination and completeness were evaluated using CheckM (Parks et al., 2015). Then, the Pyani v0.2.11 (Leighton Pritchard et al., 2016) software was employed to accurately evaluate the Average Nucleotide Identity (ANI) among the optimal assembly genomes, its closest relatives, and the type strains of the species. For each species, two or three complete genomes retrieved from NCBI Assembly database are used for the comparative analysis. An ANI score of >95% was adopted as the criterion for species delineation (Jain et al., 2018), confirming the genome as *K. michiganensis*.

Genome analysis of clinically related genotypes

Multilocus sequence typing (MLST) was performed using FastMLST v0.0.15 (Guerrero-Araya et al., 2021) and pubMLST database[1] to assign the alleles number and STs. The analysis was conducted on the optimal assembled genome and 206 additional *K. michiganensis* genomes with an assembly level of scaffold or higher downloaded from the NCBI Assembly database (updated on April, 2023). These genomes were sourced from sapiens hosts and included seven conserved alleles of *K. michiganensis*, gapA, infB, mdh, pgi, phoE, rpoB, and tonB. Kleborate (Lam et al., 2021) software was used to invoke the software Kaptive (Lam et al., 2022) for K and O antigen typing. Virulence factors (VFs) genes were detected by performing a BLASTp against the VFDB database (Chen et al., 2005), with putative virulence genes screened using an e-value threshold of $1e^{-20}$. Additionally, resistance genes were annotated using the Resistance Gene Identifier (RGI) (Alcock et al., 2020) software, referencing the Comprehensive Antibiotic Resistance Database (CARD).

1 https://pubmlst.org

Phylogenetic analysis and gene annotation

For the pan-genome analysis, all *K. michiganensis* genomes were uniformly annotated using Prokka v1.14.6 (Seemann, 2014). Roary v3.13.0 (Page et al., 2015) was used to detect and cluster orthologous genes (OGs) with the following parameters: -p 36 -i 90 -e -n -t 11 -s -cd 100 -a -v -z. The core gene supergene was developed through the following process. After obtaining the OG clusters, we selected the representative genes from each OG. These representative genes were then concatenated to form the core gene supergene. This supergene represents the conserved genetic information shared by all *K. michiganensis* strains and can be used for high–resolution phylogenetic analysis.

A phylogenetic tree for 207 *K michiganensis* strains was constructed using FastTree V2.1.10 and visualized using Evolview-v3 (Subramanian et al., 2019). To further understand the functions of the genes in the OGs, we extracted the amino acid sequences of the representative genes from each OG. These sequences were then annotated based on the best hits and an *e*-value threshold of $1e^{-20}$, by Blastp v2.9.0 + against the COG database.

Results

Genomic analysis of *K. michiganensis* KMLRT2206 strain

After quality control and removal of host sequences from the mNGS raw data, a total of 6,128,944 clean reads were obtained. Among these, 4,476,090 reads were aligned to all microbial sequences by Kraken. Within these aligned reads, 2,786,596 were assigned to the genus *Klebsiella*. However, 96.5% of these sequences could not be further classified to a specific species. MetaMLST detected *K. oxytoca* in the metagenomic data and reported a novel ST, namely ST 100001, with a confidence of 100%. Supplementary Table 1 is the New *K. oxytoca* ST table, which includes all the known STs as well as the new ST detected in the sample.

To identify the pathogenicity of *Klebsiella* sp., three assembly strategies were employed to reconstruct its genome. Table 1 summarizes the quality metrics for the genome assemblies. Based on total contig size and N50, the optimal genome assembly was generated by Assembly 3, which produced 125 contigs longer than 500 bp, with a total length of approximately 5.86 Mb and an N50 of 118,958 bp. Assembly 3 outperformed Assembly 1 and 2 in terms of N50 and Largest contig metrics. With a GC content of 56.06%, Assembly 3 was selected for subsequent phylogenomic and pan-genomic analyses, and the strain was designated KMLRT2206. The KMLRT2206 genome contains 5,422 protein-coding sequences and 15 rRNA genes.

Genome-wide nucleotide sequence identity analysis was performed to infer the organismal origin of the metagenome strain. As shown in Figure 2, KMLRT2206 shared 98.79%–99.34% ANI with of *K. michiganensis* genomes, while displaying low ANI values (83.19%–93.19%) with other *Klebsiella* species (Supplementary Table 2). Since the 95% ANI threshold for species delineation, strain KMLRT2206 was reclassified as *K. michiganensis*.

TABLE 1 Comparison of genome assemblies of an *K. michiganensis* strain recovered from the metagenome data.

	Assembly 1	Assembly 2	Assembly 3
No. of contigs > 500 bp	233	305	125
No. of contigs > 10,000 bp	150	156	71
N50 (bp)	38,615	34,389	118,958
NGA50 (bp)	281,23	28,123	70,564
Largest contig (bp)	186,901	186,901	424,791
Total length (bp)	5,700,083	5,850,327	5,860,284
Average coverage depth	117.19	118.99	118.40
GC (%)	56.11	56.04	56.06
Genome fraction (%)	88.74	90.048	90.232
Completeness (%)	98.81	100	99.70
Contamination (%)	0.22	0.22	0.22
No. of CDSs	5,283	5,346	5,422
No. of rRNAs	11	17	15

No., number; bp, base pair; N50, the length of the shortest contig when the cumulative length reaches 50% of the total genome length; NGA50, calculated based on the genome-adjusted; CDSs, coding DNA sequences; rRNAs, ribosomal RNAs.

Phylogeny of the KMLRT2206 genome

MLST analysis revealed that strain KMLRT2206 possesses a previously unreported allelic profile of housekeeping genes [*gapA* (3), *infB* (8), *mdh* (16), *pgi* (21), *phoE* (107), *rpoB* (19), and *tonB* (~18)]. Upon submission of these sequences to the PubMLST database, the strain was assigned to a novel ST, ST452. Then, MLST analysis of the remaining 206 strains classified 165 strains, with the top 10 STs being ST27 (13 strains), ST50 (10), ST85 (10), ST43 (9), ST11 (8), ST213 (8), ST29 (6), ST231 (5), ST138 (5), ST35 (4), ST84 (4), and ST205 (4). Thirty-six strains represented novel ST types, whereas six strains could not be classified due to the missing housekeeping gene fragments (Supplementary Table 3).

A maximum-likelihood phylogenetic tree was constructed using the 206 *K. michiganensis* strains and KMLRT2206 (Figure 3). The tree revealed three major clades: Clade 1, Clade 2, and Clade 3. Clade 1 comprised 28 strains, including ST29 (6 strains) and 8 novel ST strains. Clade 2 contained 79 strains, with ST27, ST43, ST85, ST11, and ST213 exclusively distributed within this clade. Clade 3 included 100 strains (48.3% of the total), with ST50, ST32, and ST321 uniquely present in Clade 3. Additionally, the strain KMLRT2206 was located in Clade 3, exhibiting the closest genetic resemblance to ST32 strains (GCF010590665, GCF015721285, and

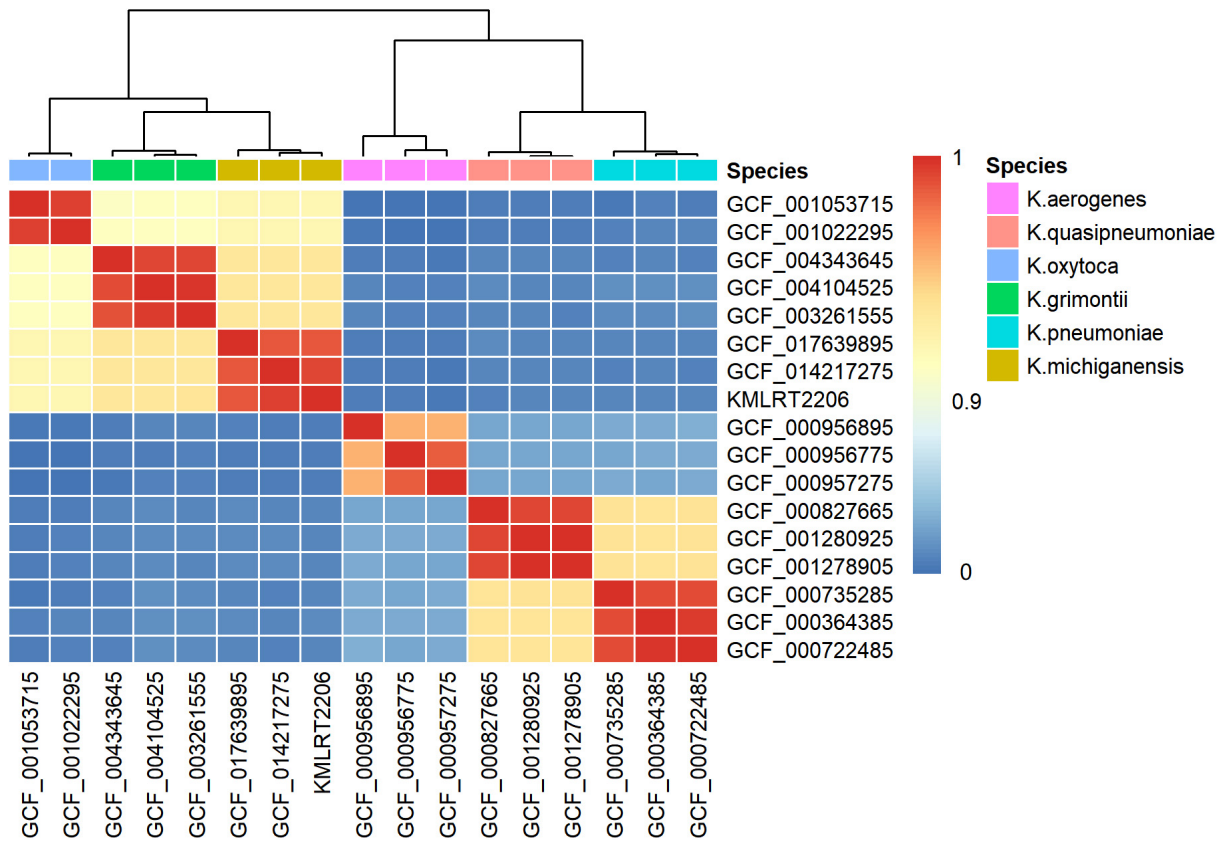


FIGURE 2
Species identification of the *K. michiganensis* strain KMLRT2206 genome recovered from clinical metagenomics data. The heatmap displays color-coded average nucleotide identity (ANI) values between pairwise comparisons of genomes from six *Klebsiella* species. The color bar (on the right) indicates the degree of similarity between species, with values ranging from 0 to 1. Red represents high similarity, while blue denotes low similarity.

GCF015721325) isolated from the United States. These findings indicate significant genetic diversity and non-clonal population structure among *K. michiganensis* strains.

Further analysis of the relationship between 207 *K. michiganensis* strains and their isolation locations and times revealed no clustering based on these factors (Supplementary Figure 1), suggesting that the evolutionary process of *K. michiganensis* is not influenced by geography or time. In terms of geographical distribution, the top six countries of isolation were the United States (53 strains), Switzerland (61), China (29), Germany (22), Australia (15) and the United Kingdom (5).

Pan-genomic composition of *K. michiganensis*

Pan-genome analysis was performed using 26,966 protein-coding sequences (CDSs) across the 207 *K. michiganensis* genomes. The analysis identified 2,587 (9.6%) core OGs, 16,155 (60%) accessory genes, and 8,224 (30%) strain-specific genes, which may contribute to unique phenotypes of individual strains. The pan-genomic asymptotic curve did not reach a platform (Figure 4A), and the exponential value of the mathematical function derived from the curve exceeded 0.5 (Figure 4B), indicating that *K. michiganensis* has an open pan-genome. This suggests a high rate of gene exchange within the *K. michiganensis* species and a strong capacity to acquire novel genes from the environment or other species. The distribution of core, accessory, and unique gene sets across COG functional categories is shown in Figure 4C. Among the accessory genes, 766 genes were identified as mobile genetic elements, which play a crucial role in the transmission of virulence factors such as exotoxins and extracellular enzymes.

Patterns of drug-resistance genes in *K. michiganensis*

A total of 37 drug-resistance genes were identified in KMLRT2206, including genes conferring resistance to aminocoumarin antibiotic [mdtC, mdtB, APH (3′)-Ia, acrD, APH (3″)-Ib, APH (6)-Id], fluoroquinolone (emrR, emrB, QnrS1), diaminopyrimidine (dfrA14), glycopeptide (vanG), nitroimidazole antibiotic (msbA), peptide antibiotic (ArnT, PmrF, eptB, bacA), tetracycline antibiotic [tet(A)] and multiple-resistant genes (baeR, LptD, acrB, marR, rsmA, oqxA, oqxB, marA, KpnE, KpnF, KpnG, CRP, OmpA, bla$_{OXY-1-1}$). Analysis of 207 *K. michiganensis* strains identified 194 drug-resistance genes (Supplementary Table 4).

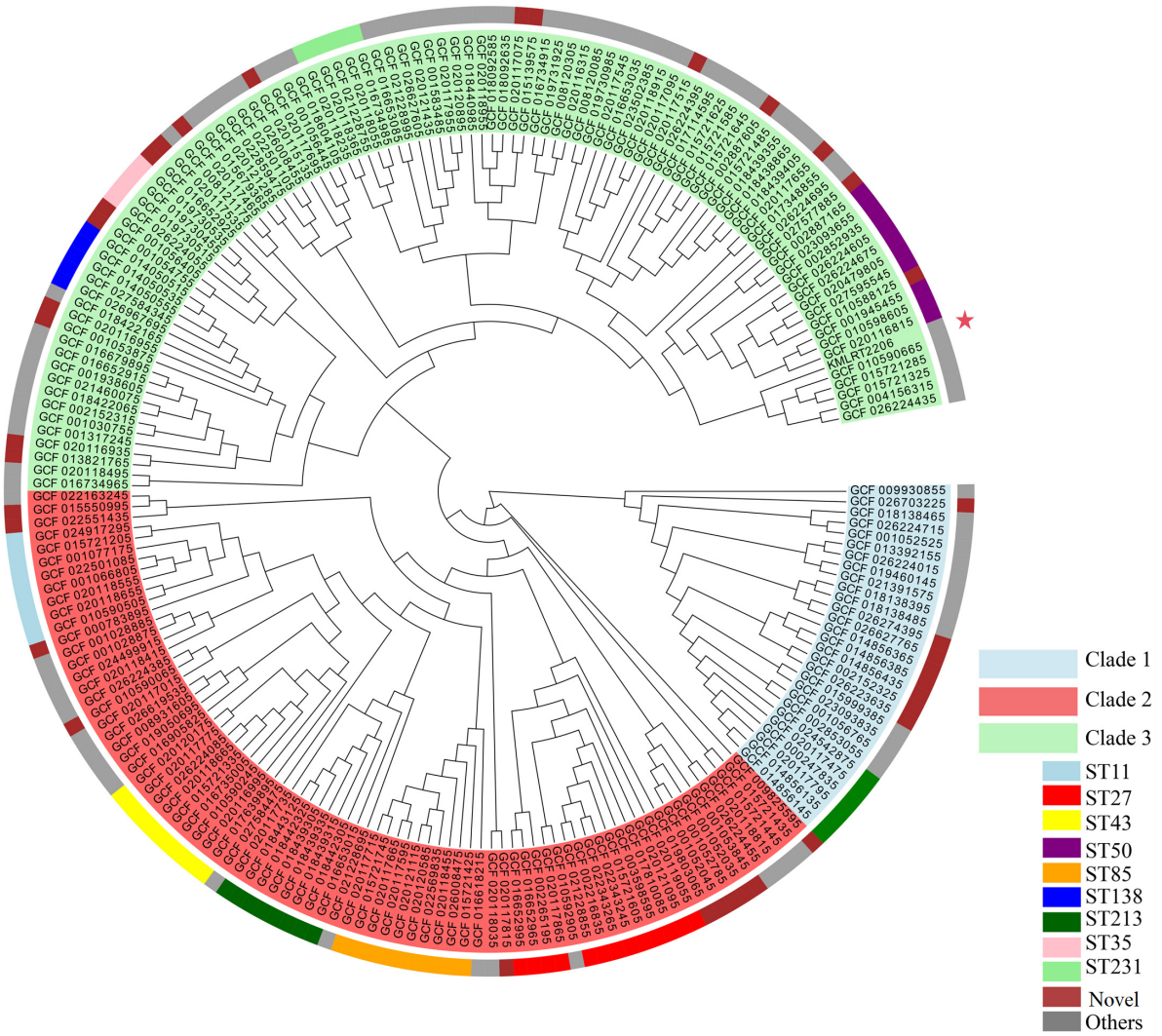


FIGURE 3
Maximum-likelihood phylogenetic tree of *K. michiganensis*. The outer circle indicates the STs of the isolates, while the inner circle represents the different clades of *K. michiganensis*. The strain KMLRT2206 is marked with a red five-pointed star.

Common resistance genes included those for aminocoumarin antibiotic (mdtB, mdtC), aminoglycoside antibiotic (acrD), peptide antibiotic (eptB), nitroimidazole antibiotic (msbA), and multiple resistance genes (PBP3 conferring resistance to β-lactam antibiotics, acrB, rsmA, KpnF, KpnG). Among the 60 β-lactamase resistance genes identified, bla$_{OXY-1-1}$ gene was present in 38.6% (80/207) of strains. In addition, a variety of carbapenemase genes were detected in 60 isolates, including 14 bla$_{NDM}$ genes, 12 bla$_{VIM}$ genes, 5 bla$_{IMP}$ genes, and 46 bla$_{KPC}$ genes (Supplementary Table 4). Among them, bla$_{KPC-2}$ was the most prevalent variant, accounting for 31 isolates. The co-occurrence of different carbapenemase genes was detected in two clinical isolates. We further investigated the correlation between resistance genes and evolutionary clades. The results indicated no obvious difference in the number of drug-resistance genes among different evolutionary clades (Supplementary Figure 2). However, bla$_{OXY-5-1}$ was present in all 27 strains of Clade 1, while bla$_{OXY-1-1}$ and bla$_{OXY-1-2}$ were absent in this clade. bla$_{OXY-1-4}$ was exclusively found in some strains of Clade 3 (Supplementary Figure 2).

Patterns of virulence genes in *K. michiganensis*

Fifteen categories of virulence genes were identified in the KMLRT2206 genome, including Nutritional/Metabolic factors (90 genes), immune regulation (75), inheritance (48), effect delivery system (46), biofilm (27), regulation (18), exoxin (15), stress survival (15), motility (10), and others. Key genes in the nutritional/metabolic factors category included those related to iron carriers, such as *Yersinia* (*ybtX*, *ybtU*, *ybtT*, *ybtS*, *ybtQ*, *ybtP*, *ybtE*, *ybtA*, *irp1*) and *Enterobacterium* (*entS*, *entF*, *entE*, *entD*, *entC*, *entB*, *entA*, *fes*, *fepG*, *fepD*, *fepC*, *fepB*, *fepA*) gene clusters (Supplementary Table 5). In the immune regulation category, 26 genes were involved in lipopolysaccharides (LPS) synthesis, such as *rfbA*, *rfbB*, *rfbD*, *waaC*, and *waaF* (Supplementary Table 5). LPS,

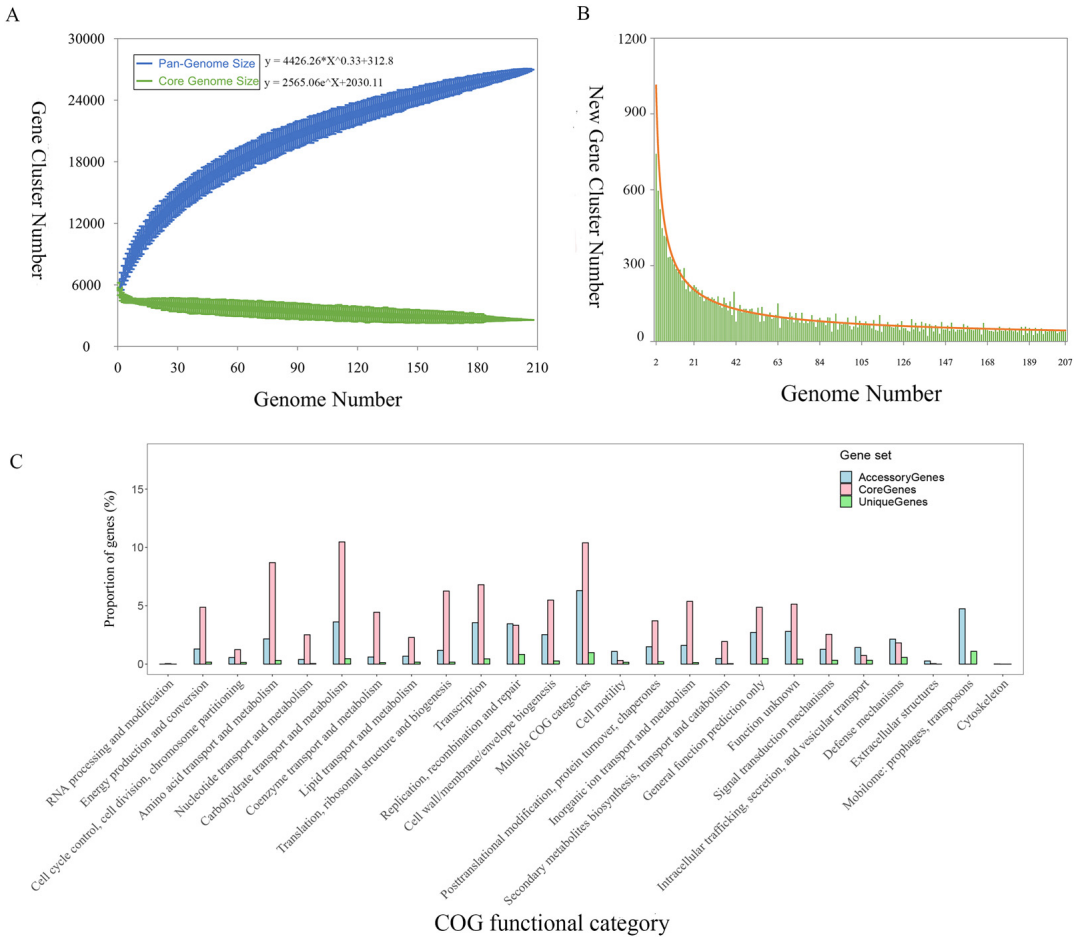


FIGURE 4
The pan-genomic structure and functional analysis of *K. michiganensis*. **(A)** Sizes of the pan-genome and core-genome. **(B)** Number of strain-specific genes per genome. **(C)** Distribution of the core genes, accessory genes, and strain-specific genes across individual COG functional categories.

a toxin of extracorporeal alveolar macrophages, can significantly inhibit phagocytosis, alter the host immune system, lead to pathogenic and physiological changes, and interfere with other antigens. Additionally, 21 genes related to capsular polysaccharide (CPS) were identified, which help the strain evade phagocytosis and enhance its survival. CPS and LPS are key pathogenicity factors and potential targets for novel control strategies.

Annotation of 207 genomes through the VFDB database identified 510 virulence genes, with 211 common to all strains. No correlation was observed between virulence genes and ST (Supplementary Figure 3). Notably, 27 strains in Clade 1 lacked *Yersiniabactin* virulence genes, 26 strains in Clade 1 lacked *cytochrome C* virulence genes, and 23 strains in Clade 3 lacked *Yersiniabactin* virulence genes. All strains in Clade 2 possessed both *Yersiniabactin* and *cytochrome C* virulence genes (Supplementary Figure 3).

Among the 207 strains, 161 had accurately identified O antigens, including O1 (127 strains), O2a (33), and O2ac (1). The remaining 46 strains could not be classified. KMLRT2206 was identified as O1. For K antigens, 32 strains were classified into six types: K70 (11), K43 (11), K26 (4), K74 (3), and K41 (1). The remaining 175 strains, including KMLRT2206, could not be classified, with the best match for KMLRT2206 being KL107 (75.85% identity). Further analysis indicated that all ST213 and ST32 strains were identified as O2a, and the K43 was closely associated with ST43 (Supplementary Figure 3).

Discussion

In this study, we report the genomic characteristics of *K. michiganensis* recovered from patient with pulmonary infection using mNGS technology in a hospital in Anhui Province, China, and place it within a broader range of clinical samples from 20 countries. At these two geographical scales, the overall *K. michiganensis* population exhibited rich genetic diversity, with the emergence of a large number of new STs. ST27, ST50, and ST85 are widely distributed, and strains of different STs showed differences in resistance genes, virulence genes, and antigen types.

Traditional tests based on phenotypes, biochemistry, and PCR for differentiating members of the *K. oxytoca* complex have certain limitations in sensitivity, accuracy, and speed. Culture method is the gold standard for bacterial identification, but the probability of false–negative results increases after empirical

drug use (Li et al., 2021). This is consistent with the results observed in this study. 16S rRNA gene sequencing is another commonly used bacterial identification technique. Due to its low taxonomic resolution and the close genetic distance among members of the *K. oxytoca* complex, it is difficult to accurately identify species (Pan et al., 2023). Whole-genome sequencing and analysis provide the highest level of resolution for accurate bacterial species identification (Konstantinidis and Tiedje, 2005), but it is time-consuming, culture-dependent, and cannot quickly obtain all potential pathogens in clinical samples. In contrast, in this study, mNGS combined with an optimal assembly strategy was used to simultaneously conduct rapid and extensive screening of pathogens, identification of species subgroups, and determination of resistance genes without relying on culture. This not only helps clinicians to identify pathogens earlier, reduce unnecessary use of broad-spectrum antibiotics, and lower the risk of drug resistance, but also provides important reference for deciphering *K. michiganensis*.

Accurate identification of bacterial species is not only crucial for improving patient care, as it may affect the interpretation of drug susceptibility tests and thus impact treatment outcomes, but also for enhancing our understanding of bacterial epidemiology (Saxenborn et al., 2021; Voellmy et al., 2022; Wu et al., 2021). Previous study has demonstrated that the use of a core-genome MLST scheme can provide fine-scale resolution for strain discrimination of *K. pneumoniae* (Raj et al., 2022). Ikhimiukor et al. (2023) analyzed a global *K. oxytoca* dataset spanning 15 years and 18 countries, and the results showed that ST2, ST176, and ST199 were prevalent and widely spread in clinical infections. We performed MLST analysis on 207 *K. michiganensis* strains obtained from human hosts. A large number of new STs (36 strains) and unclassifiable strains (6 strains) were discovered, which fully indicates the rich genetic diversity of *K. michiganensis* (Figure 3). The phylogenetic tree divided *K. michiganensis* into three major clades, and this distribution may be related to the origin and divergence history of each clade. ST27 was the most common ST (13/207, 6.3%), followed by ST50 and ST80 (Figure 3). This result differs from that of Li et al. (2022) in their analysis of 275 *K. michiganensis* strains from human, animal, and environmental sources, where they reported that ST29 (12/275) was the most common ST. Additionally, ST29, ST43 and ST92 have been reported as predominant STs of carbapenem-resistant *K. michiganensis* (Li H. et al., 2024; Wan et al., 2023). These differences may be due to variations in sample sources, geographical regions, and sample collection times. The high prevalence of ST27 indicates its relatively frequent occurrence in the local clinical setting compared to other STs. In addition, the ST27 *K. michiganensis* strains had the highest number of resistance genes (Li et al., 2022). Consequently, enhanced surveillance of this strain is warranted to mitigate potential public health risks. There was no obvious clustering relationship between the distribution of *K. michiganensis* strains and the isolation location and time (Supplementary Figure 1). This conclusion is different from the traditional view that microbial evolution is greatly influenced by geographical and temporal factors, which may imply that this species has a special transmission mechanism or evolutionary driving force.

A correlation between β-lactamase genes and species has previously been reported, where *K. pneumoniae* is associated with the bla_{SHV} gene, *K. quasipneumoniae* with the bla_{OKP} gene, and *K. variicola* with the bla_{LEN} gene (Haeggman et al., 2004). In the analysis of resistance genes in this study, the $bla_{OXY-1-1}$ gene was detected in 38.6% (80/207) of *K. michiganensis* strains, and it coexisted with other β-lactamase resistance genes (Supplementary Table 4), which poses a potential threat to public health. Recent study has demonstrated that bla_{OXY-5}-carrying *K. michiganensis* exhibits higher antibiotic resistance rates, representing a highly resistant subpopulation (Li Y. et al., 2024). In our study, the clade-specific distribution pattern of $bla_{OXY-5-1}$ gene was observed (100% prevalence in Clade 1) (Supplementary Table 4), suggesting a potential evolutionary link between phylogenetic divergence and antimicrobial resistance in *K. michiganensis* that warrants further investigation. It is known that sequence variations of the chromosomally encoded β-lactamase gene bla_{OXY} can assign the *K. oxytoca* complex to phylogroups (Brisse et al., 2016). Nine phylogroups, Ko1 to Ko9, were assigned to reflect the bla_{OXY} variants they carried (bla_{OXY-1} to bla_{OXY-9}) (Yang et al., 2022). Previous study has used the bla_{OXY} variants of bla_{OXY-1} and bla_{OXY-5} to help distinguish between phylogroup Ko1 and sub-phylogroup Ko5 (Fevre et al., 2005), and the results of our article are consistent with this (Supplementary Figure 2). Cosic et al. (2021) proposed that combining bla_{OXY} genotyping with auxiliary gene markers enables species discrimination within the environmental *K. oxytoca* species complex. This suggests that variations in β-lactamase genes may reveal lineage divergence more rapidly than variations in housekeeping genes. In the *Klebsiella* genus, the predominant carbapenem resistance gene is class A serine β-lactamase bla_{KPC-2} (Chatzidimitriou et al., 2024), which is consistent with our findings (Supplementary Table 4). Notably, this distribution pattern differs from Spanish epidemiological reports documenting bla_{VIM-1} and bla_{OXA-48} as the predominant carbapenemases in *K. oxytoca* populations (Pérez-Vazquez et al., 2019). Findings revealed that *K. michiganensis* strains can co-harbor multiple carbapenemase genes (Founou et al., 2018; Sun et al., 2024; Zhang et al., 2022), demonstrating this species' potential for accumulating antimicrobial resistance determinants. Although only a limited number of isolates carrying multiple carbapenem resistance genes were detected in our study (Supplementary Table 4).

In the analysis of antigen types, we found that the CPS K antigens of most *K. michiganensis* isolates in the population were unknown (Supplementary Table 5), which is consistent with a study from Australia (Stewart et al., 2022). Additionally, the CPS K antigen locus exhibited low prevalence (Supplementary Table 5), consistent with findings from a previous study (Biedrzycka et al., 2023). This indicates that the K antigens of most *K. michiganensis* strains remain uncharacterized, which may impede clinical serological analysis. Both ST213 and ST32 strains were identified as serotype O2a, and K43 was closely associated with ST43 (Supplementary Figure 3), suggesting that there is a certain correlation between the antigen type and the ST of the strains. Li Y. et al. (2024) have revealed the distinct distribution patterns of K and O loci between *K. michiganensis* and *K. oxytoca*. This correlation may contribute to more accurate classification and tracing of *K. michiganensis*. In the analysis of virulence genes, there was no obvious correlation between virulence genes and ST (Supplementary Figure 3), which means that the composition of virulence genes of strains cannot be inferred solely based on

ST. However, the distribution of virulence genes differed among different clades of *K. michiganensis*. The *Yersiniabactin* virulence genes were not detected in any of the strains within Clade 1, but were found in Clade 2 and Clade 3 (Supplementary Figure 3). It is known that *Yersiniabactin* virulence genes can block the production of reactive oxygen species in the respiratory system, indirectly reducing the bactericidal ability of host innate immune cells and increasing the virulence of clinical infections (Paauw et al., 2009). This implies that different clades may exhibit differences in pathogenic characteristics, which may be attributed to different selective pressures experienced during the evolutionary process, leading to the acquisition or loss of virulence genes.

Currently, there are still many research gaps in our understanding of the *K. oxytoca* complex. The proportion of *K. michiganensis* strains that are insensitive to carbapenems and cephalosporins has been increasing year by year. Without careful monitoring, it is likely to pose greater challenges to treatment and infection control in the future. This study pioneered the proposal of the value of mNGS combined with the optimal assembly strategy in the accurate identification of species within the *K. oxytoca* complex. However, the reliability and broad applicability of mNGS in practical clinical applications remain to be further verified. Recent studies have shown that MALDI-TOF also has the potential to identify subgroups within bacterial species for epidemiological assessment (Cuénod et al., 2021; Giraud-Gatineau et al., 2021). Therefore, performance comparison with MALDI-TOF technology is necessary in the future. At the sample level, this study only reconstructed one strain of *K. michiganensis* from a clinical specimen collected at a hospital in Anhui Province, China, while the other isolates were derived from global clinical samples in public database. There is still room for improvement in geographical coverage and sample diversity, and it may not comprehensively reflect the genetic characteristics and distribution patterns of *K. michiganensis* under different environmental and host backgrounds. Secondly, in terms of clinical significance, since the colonization incidence, pathogenicity, as well as clinical manifestations, severity, and prognosis of species within the *K. oxytoca* complex in patients are still largely unknown, it remains to be elucidated whether the accurate species identification of *K. michiganensis* or the differentiation of the *K. oxytoca* complex has an impact on patient treatment, prognosis prediction, epidemiological monitoring, and infection control. We believe that currently, the accurate identification of species within the *K. oxytoca* complex is necessary for scientific research. This helps to improve epidemiological monitoring, precisely monitor the situation of drug resistance, and thus formulate targeted prevention and control strategies. However, it may not be necessary for routine clinical practice.

Conclusion

In this study, mNGS was performed on a BALF sample obtained from a patient, enabling rapid and accurate pathogen identification and a comprehensive analysis of the strain's genomic characteristics. This provided reliable information for clinicians and epidemiologists. Additionally, we conducted an in-depth analysis of population genetics and molecular epidemiological features using 206 published *K. michiganensis* genomes from human hosts, establishing a critical framework for understanding this species. Among the 207 *K. michiganensis* strains analyzed, the predominant O- antigen was O1, and the most common ST was 27. The strain KMLRT2206, identified as a novel ST452 strain, carries multiple virulence and drug-resistance genes. It belongs to Clade 3 and encodes genes for the carbapenemase and the yersiniabactin, indicating its potential pathogenicity and warranting further investigation. A limitation of this study is the potential sampling bias in genome sequencing. Future research should focus on systematically exploring the prevalence, pathogenesis, antibiotic resistance, and transmission dynamics of *K. michiganensis*.

Data availability statement

The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA1021007.

Ethics statement

The studies involving humans were approved by the Ethics Committee of Anqing Municipal Hospital. 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. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

YS: Conceptualization, Formal Analysis, Methodology, Resources, Writing – original draft. QC: Conceptualization, Data curation, Formal Analysis, Methodology, Writing – original draft. TL: Data curation, Investigation, Writing – original draft, Visualization. JC: Project administration, Supervision, Writing – review and editing. YF: Conceptualization, Methodology, Project administration, Writing – review and editing.

Funding

The authors declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

QC, TL, YF were employed by Genoxor Medical Science and Technology Inc.

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.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

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.2025.1546594/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Maximum-likelihood phylogenetic tree of *K. michiganensis* with country and collective date information.

SUPPLEMENTARY FIGURE 2
Maximum-likelihood phylogenetic tree of *K. michiganensis* with a heatmap of antibiotic resistance gene.

SUPPLEMENTARY FIGURE 3
Maximum-likelihood phylogenetic tree of *K. michiganensis* with a heatmap of virulence gene.

SUPPLEMENTARY TABLE 1
Results of MetaML ST of BALF metagenomes.

SUPPLEMENTARY TABLE 2
Pairwise average nucleotide identity (ANI) values between genomes of distinct *Klebsiella* species.

SUPPLEMENTARY TABLE 3
Metadata of epidemiological and genomic features of all *K. michiganensis* strains used in this study.

SUPPLEMENTARY TABLE 4
Antibiotic resistance encoding genes in *K. michiganensis*.

SUPPLEMENTARY TABLE 5
Genes encoding virulence factors in *K. michiganensis*.

References

Abed, J., Déraspe, M., Bérubé, È, Dewar, K., Boissinot, M., et al. (2021). Complete genome sequences of Klebsiella michiganensis and Citrobacter farmeri, KPC-2-producers serially isolated from a single patient. *D'Iorio M* 10:1408. doi: 10.3390/antibiotics10111408

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

Biedrzycka, M., Urbanowicz, P., Żabicka, D., Hryniewicz, W., Gniadkowski, M., and Izdebski, R. (2023). Country-wide expansion of a VIM-1 carbapenemase-producing Klebsiella oxytoca ST145 lineage in Poland, 2009-2019. *Eur. J. Clin. Microbiol. Infect. Dis.* 42, 1449–1457. doi: 10.1007/s10096-023-04682-x

Brisse, S., Grimont, F., and Grimont, P. (2016). "The genus Klebsiella," in *The Prokaryotes: A Handbook on The Biology of Bacteria, vol 6. Proteobacteria: Gamma Subclass*, eds D. Martin, F. Stanley, R. Eugene, S. Karl-Heinz, and S. Erko (New York, NY: Springer), 159–196.

Chapman, P., Forde, B., Roberts, L., Bergh, H., Vesey, D., Jennison, A., et al. (2020). Genomic investigation reveals contaminated detergent as the source of an extended-spectrum-β-lactamase-producing klebsiella michiganensis outbreak in a neonatal unit. *J. Clin. Microbiol.* 58:e01980-19. doi: 10.1128/JCM.01980-19

Chatzidimitriou, M., Kavvada, A., Kavvadas, D., Kyriazidi, M., Eleftheriadis, K., Varlamis, S., et al. (2024). Carbapenem-resistant *Klebsiella pneumoniae* in the Balkans: Clonal distribution and associated resistance determinants. *Acta Microbiol. Immunol. Hung.* 71, 10–24. doi: 10.1556/030.2024.02230

Chaumeil, P., Mussig, A., Hugenholtz, P., and Parks, D. H. (2022). GTDB-Tk v2: Memory friendly classification with the genome taxonomy database. *Bioinformatics* 38, 5315–5316. doi: 10.1093/bioinformatics/btac672

Chen, L., Yang, J., Yu, J., Yao, Z., Sun, L., Shen, Y., et al. (2005). VFDB: A reference database for bacterial virulence factors. *Nucleic Acids Res.* 33, D325–D328. doi: 10.1093/nar/gki008

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

Cosic, A., Leitner, E., Petternel, C., Galler, H., Reinthaler, F., Herzog-Obereder, K., et al. (2021). Variation in accessory genes within the Klebsiella oxytoca species complex delineates monophyletic members and simplifies coherent genotyping. *Front. Microbiol.* 12:692453. doi: 10.3389/fmicb.2021.692453

Cuénod, A., Wüthrich, D., Seth-Smith, H., Ott, C., Gehringer, C., Foucault, F., et al. (2021). Whole-genome sequence-informed MALDI-TOF MS diagnostics reveal importance of Klebsiella oxytoca group in invasive infections: A retrospective clinical study. *Genome Med.* 13:150. doi: 10.1186/s13073-021-00960-5

Dantur, K., Chalfoun, N., Claps, M., Tórtora, M., Silva, C., Jure, Á, et al. (2018). The Endophytic strain Klebsiella michiganensis Kd70 lacks pathogenic island-like regions in its genome and is incapable of infecting the urinary tract in mice. *Front. Microbiol.* 9:1548. doi: 10.3389/fmicb.2018.01548

Fevre, C., Jbel, M., Passet, V., Weill, F., Grimont, P., and Brisse, S. (2005). Six groups of the OXY beta-Lactamase evolved over millions of years in Klebsiella oxytoca. *Antimicrob. Agents Chemother.* 49, 3453–3462. doi: 10.1128/AAC.49.8.3453-3462.2005

Flerlage, T., Brazelton de Cardenas, J. N., Garner, C. D., Hasan, N. A., Karathia, H., Qudeimat, A., et al. (2020). Multiple NDM-5-Expressing *Escherichia Coli* isolates from an immunocompromised pediatric host. *Open Forum Infect. Dis.* 7:ofaa018. doi: 10.1093/ofid/ofaa018

Founou, R., Founou, L., Allam, M., Ismail, A., and Essack, S. (2018). Genomic characterisation of Klebsiella michiganensis co-producing OXA-181 and NDM-1 carbapenemases isolated from a cancer patient in uMgungundlovu district, KwaZulu-Natal Province, South Africa. *S. Afr. Med. J.* 109, 7–8. doi: 10.7196/SAMJ.2018.v109i1.13696

Giraud-Gatineau, A., Texier, G., Fournier, P., Raoult, D., and Chaudet, H. (2021). Using MALDI-TOF spectra in epidemiological surveillance for the detection of bacterial subgroups with a possible epidemic potential. *BMC Infect. Dis.* 21:1109. doi: 10.1186/s12879-021-06803-3

Guerrero-Araya, E., Muñoz, M., Rodríguez, C., and Paredes-Sabja, D. (2021). FastMLST: A multi-core tool for multilocus sequence typing of draft genome assemblies. *Bioinform. Biol. Insights* 15:11779322211059238. doi: 10.1177/11779322211059238

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

Haeggman, S., Löfdahl, S., Paauw, A., Verhoef, J., and Brisse, S. (2004). Diversity and evolution of the class A chromosomal beta-lactamase gene in *Klebsiella pneumoniae*. *Antimicrob. Agents Chemother.* 48, 2400–2408. doi: 10.1128/AAC.48.7.2400-2408.2004

Hu, L., Zhang, X., Zeng, X., Chen, Y., Liu, L., and Li, G. (2022). Genomic characterization of a carbapenemase-producing, extensively drug-resistant klebsiella michiganensis strain from a renal abscess patient. *Microbiol. Resour. Announc.* 11:e0082522. doi: 10.1128/mra.00825-22

Igo, M., and Schaffner, D. (2022). Draft genome sequence for Klebsiella michiganensis B199A, originally identified as *Enterobacter* aerogenes. *Microbiol. Resour. Announc.* 11:e0031022. doi: 10.1128/mra.00310-22

Ikhimiukor, O., Souza, S., Akintayo, I., Marcovici, M., Workman, A., Martin, I., et al. (2023). Phylogenetic lineages and antimicrobial resistance determinants of clinical Klebsiella oxytoca spanning local to global scales. *Microbiol. Spectr.* 11:e0054923. doi: 10.1128/spectrum.00549-23

Jain, C., Rodriguez-R, L., Phillippy, A., Konstantinidis, K., 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

Konstantinidis, K., and Tiedje, J. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U S A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102

Kula, A., Arman, M., Appleberry, H., Wolfe, A., and Putonti, C. (2024). Draft genomes of Klebsiella aerogenes, Klebsiella huaxiensis, and Klebsiella michiganensis isolates from the urinary tract. *Microbiol. Resour. Announc.* 13:e0049224. doi: 10.1128/mra.00492-24

Lam, M., Wick, R., Judd, L., Holt, K., and Wyres, K. (2022). Kaptive 2.0: Updated capsule and lipopolysaccharide locus typing for the *Klebsiella pneumoniae* species complex. *Microb. Genom.* 8:000800. doi: 10.1099/mgen.0.000800

Lam, M., Wick, R., Watts, S., Cerdeira, L., Wyres, K., 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

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

Leighton Pritchard, R., Sonia, H., and John, G. (2016). Elphinstoneb and Ian K. Tothc Genomics and taxonomy in diagnostics for food security: Soft-rotting *enterobacterial* plant pathogens. *Anal. Methods* 8, 12–24. doi: 10.1039/c5ay02550h

Li, D., Liu, C., Luo, R., Sadakane, K., and Lam, T. W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033

Li, H., Dong, W., Liu, Y., Ma, J., and Liu, X. (2024). Whole-genome sequencing of clinical isolates of Klebsiella michiganensi in China carrying blaIPM-4 and blaNDM-1. *Microb. Pathog.* 197:107070. doi: 10.1016/j.micpath.2024.107070

Li, N., Cai, Q., Miao, Q., Song, Z., Fang, Y., and Hu, B. (2021). High-throughput metagenomics for identification of pathogens in the clinical settings. *Small Methods* 5:2000792. doi: 10.1002/smtd.202000792

Li, S., Jiang, X., Li, C., Ju, Y., Yue, L., Chen, F., et al. (2022). A bla SIM-1 and mcr-9.2 harboring Klebsiella michiganensis strain reported and genomic characteristics of Klebsiella michiganensis. *Front. Cell. Infect. Microbiol.* 12:973901. doi: 10.3389/fcimb.2022.973901

Li, Y., Wu, Y., Li, D., Du, L., Zhao, L., Wang, R., et al. (2024). Multicenter comparative genomic study of Klebsiella oxytoca complex reveals a highly antibiotic-resistant subspecies of Klebsiellamichiganensis. *J Microbiol. Immunol. Infect.* 57, 138–147. doi: 10.1016/j.jmii.2023.10.014

López-Camacho, E., Aguilera-Alonso, D., Buenestado-Serrano, S., Marín, M., Molero-Salinas, A., López Fresneña, N., et al. (2025). Genomically-supported Redefinition of an outbreak in a pediatric unit caused by blaVIM -harboring Klebsiella michiganensis. *Pediatr. Infect. Dis. J.* 44, 166–173. doi: 10.1097/INF.0000000000004571

Merla, C., Rodrigues, C., Passet, V., Corbella, M., Thorpe, H., Kallonen, T., et al. (2019). Description of Klebsiella spallanzanii sp. nov. and of Klebsiella pasteurii sp. nov. *Front. Microbiol.* 10:2360. doi: 10.3389/fmicb.2019.02360

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. (2017). metaSPAdes: A new versatile metagenomic assembler. *Genome Res.* 27, 824–834. doi: 10.1101/gr.213959.116

Paauw, A., Leverstein-van Hall, M., van Kessel, K., Verhoef, J., and Fluit, A. (2009). Yersiniabactin reduces the respiratory oxidative stress response of innate immune cells. *PLoS One* 4:e8240. doi: 10.1371/journal.pone.0008240

Page, A., Cummins, C., Hunt, M., Wong, V., Reuter, S., Holden, M., et al. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421

Pan, P., Gu, Y., Sun, D., Wu, Q., and Zhou, N. (2023). Microbial diversity biased estimation caused by intragenomic heterogeneity and interspecific conservation of 16S rRNA genes. *Appl. Environ. Microbiol.* 89:e0210822. doi: 10.1128/aem.02108-22

Parks, D., Imelfort, M., Skennerton, C., Hugenholtz, P., and Tyson, G. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114

Pérez-Vazquez, M., Oteo-Iglesias, J., Sola-Campoy, P., Carrizo-Manzoni, H., Bautista, V., Lara, N., et al. (2019). Characterization of carbapenemase-producing Klebsiella oxytoca in Spain, 2016-2017. *Antimicrob. Agents Chemother.* 63:e02529-18. doi: 10.1128/AAC.02529-18

Raj, S., Sharma, T., Pradhan, D., Tyagi, S., Gautam, H., Singh, H., et al. (2022). Comparative analysis of clinical and genomic characteristics of hypervirulent *Klebsiella pneumoniae* from hospital and community settings: Experience from a tertiary healthcare center in India. *Microbiol. Spectr.* 10:e0037622. doi: 10.1128/spectrum.00376-22

Saha, R., Farrance, C., Verghese, B., Hong, S., and Donofrio, R. (2013). Klebsiella michiganensis sp. nov., a new bacterium isolated from a tooth brush holder. *Curr. Microbiol.* 66, 72–78. doi: 10.1007/s00284-012-0245-x

Saxenborn, P., Baxter, J., Tilevik, A., Fagerlind, M., Dyrkell, F., Pernestig, A., et al. (2021). Genotypic characterization of clinical Klebsiella spp. isolates collected from patients with suspected community-onset sepsis, Sweden. *Front. Microbiol.* 12:640408. doi: 10.3389/fmicb.2021.640408

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

Stewart, J., Judd, L., Jenney, A., Holt, K., Wyres, K., and Hawkey, J. (2022). Epidemiology and genomic analysis of Klebsiella oxytoca from a single hospital network in Australia. *BMC Infect. Dis.* 22:704. doi: 10.1186/s12879-022-07687-7

Subramanian, B., Gao, S., Lercher, M., Hu, S., and Chen, W. (2019). Evolview v3: A webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res.* 47, W270–W275. doi: 10.1093/nar/gkz357

Sun, M., Xiao, W., and Xu, Q. (2024). Molecular characterization of a KPC-2- and NDM-1-producing Klebsiella michiganensis clinical isolate in cerebrospinal fluid. *Infect. Drug Resist.* 17, 3569–3578. doi: 10.2147/IDR.S468895

Uritskiy, G., DiRuggiero, J., and Taylor, J. (2018). MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6:158. doi: 10.1186/s40168-018-0541-1

Voellmy, I., Lang, C., Gasser, M., and Kronenberg, A. (2022). Antibiotic resistance surveillance of *Klebsiella pneumoniae* complex is affected by refined MALDI-TOF identification, Swiss data, 2017 to 2022. *Euro Surveill.* 27:2200104. doi: 10.2807/1560-7917.ES.2022.27.45.2200104

Wan, W., Yang, X., Yu, H., Wang, M., Jia, W., Huang, B., et al. (2023). Genomic characterization of carbapenem-resistant Klebsiella oxytoca complex in China: A multi-center study. *Front. Microbiol.* 14:1153781. doi: 10.3389/fmicb.2023.1153781

Wood, D., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20:257. doi: 10.1186/s13059-019-1891-0

Wu, W., Wei, L., Feng, Y., Xie, Y., and Zong, Z. (2021). Precise species identification by whole-genome sequencing of *Enterobacter* bloodstream infection. *China. Emerg. Infect. Dis.* 27, 161–169. doi: 10.3201/eid2701.190154

Yamada, A., Souza, A., Bertani, A., Campos, K., Sacchi, C., Assis, D., et al. (2024). Genomic characterization of a clinical NDM-1-producing klebsiella michiganensis from Brazil. *Microorganisms* 12:1408. doi: 10.3390/microorganisms12071408

Yang, J., Long, H., Hu, Y., Feng, Y., McNally, A., and Zong, Z. (2022). Klebsiella oxytoca complex: Update on taxonomy, antimicrobial resistance, and virulence. *Clin. Microbiol. Rev.* 35:e0000621. doi: 10.1128/CMR.00006-21

Zhang, Y., Gu, D., Yang, X., Wu, Y., Liu, C., Shen, Z., et al. (2021). Emergence and genomic characterization of a KPC- 2-, NDM- 1-, and IMP-4-producing Klebsiella michiganensis Isolate. *Front. Microbiol.* 12:762509. doi: 10.3389/fmicb.2021.762509

Zhang, Y., Gu, D., Yang, X., Wu, Y., Liu, C., Shen, Z., et al. (2022). Emergence and genomic characterization of a KPC- 2-, NDM- 1-, and IMP-4-producing Klebsiella michiganensis isolate. *Front. Microbiol.* 12:762509. doi: 10.3389/fmicb.2021.762509

Zolfo, M., Tett, A., Jousson, O., Donati, C., and Segata, N. (2017). MetaMLST: Multi-locus strain-level bacterial typing from metagenomic samples. *Nucleic Acids Res.* 45:e7. doi: 10.1093/nar/gkw837

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