Degree project

UNIVERSITY OF SKÖVDE

GENOMIC CHARACTERIZATION AND BIOINFORMATIC ANALYSIS OF *CITROBACTER* SPECIES IN SEPSIS PATIENTS

Master Degree Project (120 credits) in Systems Biology
Second Cycle 45 credits
Spring term 2024

Student: Dhanya Kuzhippanathil Gopi
e-mail: b22dhaku@student.his.se

Supervisor: Magnus Fagerlind, University of Skovde
e-mail: magnus.fagerlind@his.se

Examiner: Bjorn Olsson, University of Skovde
e-mail: bjorn.olsson@his.se

Abstract

The enigmatic etiology of sepsis underscores the vital need for rapid and accurate diagnostic methods. In this context, bacterial whole-genome sequencing (WGS) emerges as a powerful tool at the forefront of clinical microbiology, holding significant promise for enhanced sepsis management. Prompt detection and early initiation of antibiotic therapy are crucial for optimal patient outcomes. This study aimed to evaluate and compare the performance of two bioinformatic pipelines for analysing WGS data from clinical *.* isolates associated with sepsis. The first approach employed an in-house developed pipeline, while the second utilized the Bacterial and Viral Bioinformatics Resource Center (BV-BRC) pipeline. Predicted results from both pipelines were compared to clinically derived data. This study analyzes WGS data generated on *Citrobacter* isolates collected within a prospective sepsis study at Skaraborg Hospital, Sweden, using the Illumina HiSeq X platform. Phenotypic antimicrobial susceptibility testing (AST) and species identification for these isolates were previously performed using MALDI-TOF MS analysis, adhering to EUCAST guidelines. WGS data analysis commenced with quality control and trimming of FastQ files, followed by de novo assembly and quality assessment of the assembled contigs. Isolates with lower genome coverage were excluded from further analysis to maintain study consistency. Species identification comparing both phenotypic and genotypic methods demonstrated complete agreement between the methods. To evaluate the concordance for antimicrobial resistance (AMR) determination, the predicted genotypic resistance profiles were subsequently compared to the phenotypic AST results generated by the clinical laboratory. WGS revealed widespread beta-lactam resistance in *Citrobacter* isolates, supported by the identification of diverse resistance genes and plasmids associated with multidrug resistance. Interestingly the study identified the *Yersinia pestis* plasmid (*IncFIIYp*) in *Citrobacter freundii* isolates. The BV-BRC pipeline identified the prevalence of virulence genes encoding functionalities associated with adhesion, iron uptake via siderophores, immune system evasion, secretion system invasion, endotoxin, enterotoxin, and serum resistance. This study identified potentially novel *Citrobacter* sequence types alongside geographically widespread STs suggesting global circulation of these pathogens and highlighting the necessity for further characterization to elucidate their distribution and virulence.

Popular scientific summary

Sepsis, a hidden danger triggered by the body's overreaction to infection, leaves doctors scrambling for answers. Blood cultures, the current method, take days to identify the culprit – bacteria, virus, or some other pathogen. This crucial delay forces physicians to rely on broad-spectrum antibiotics, a shotgun approach that wipes out both good and bad bacteria. But a new hope emerges with whole genome sequencing (WGS). This cutting-edge tool analyzes the entire genetic code of bacteria in a patient's blood within hours, providing a precise fingerprint of the enemy. This not only allows for rapid and accurate identification of the specific bacteria causing the infection but also reveals any genes that might confer antibiotic resistance. Armed with this detailed information, doctors can then choose the most effective antibiotic right away, minimizing the need for broad-spectrum treatment and potentially saving lives. This study investigated the potential of WGS to revolutionize the fight against sepsis, focusing on the bacterium *Citrobacter*. The research compared two bioinformatic workflows, each consisting of a series of computational tools and processes, for analyzing WGS data from *Citrobacter* isolates collected from sepsis patients. This comparison aimed to identify the most time-efficient and accurate approach for future clinical applications. The study further explored the growing public health threat of antimicrobial resistance (AMR). The overuse and misuse of antibiotics in both humans and animals are driving forces behind AMR. Unnecessary prescriptions for non-bacterial infections or incomplete treatment courses create a breeding ground for hardier bacteria that can reproduce and spread their resistance genes, making future infections more challenging to treat. The World Health Organization (WHO) recognizes AMR as a top global health threat, highlighting the critical need for responsible antibiotic use and the development of novel treatment options. This urgency is underscored by the widespread resistance to beta-lactam antibiotics observed in *Citrobacter* isolates within this study. Complicating matters further, the study identified concerning plasmids in *Citrobacter*. These mobile genetic elements, capable of horizontal gene transfer between bacteria, can harbour resistance genes and render antibiotics ineffective. Notably, the study found a plasmid typically associated with the plague bacterium (*Yersinia pestis*) in *Citrobacter freundii* isolates, raising questions about potential transmission pathways. Beyond just identifying *Citrobacter* species, the study explored its role in sepsis by investigating virulence factor genes. The identified genes for adhesion (sticking to host cells), iron uptake (essential for bacterial growth), and toxin production all play a role in sepsis progression. Understanding the specific *Citrobacter* sequence type (ST) can be informative. While virulence factor genes tell us about the ability of a microorganism to cause disease, sequence typing helps identify specific strains within a bacterial species. Studying the distribution of these STs helps us understand how resistance is spreading geographically and identify potential hotspots for outbreaks. Sequence typing can also be used to investigate the spread of plasmids containing antibiotic resistance genes. By comparing the genetic makeup of plasmids from different *Citrobacter* isolates, scientists can determine if they share a common origin, suggesting potential plasmid transfer events between bacteria. Further research is needed to understand the global distribution and virulence potential of these novel STs. By providing faster and more detailed information compared to traditional methods, WGS can significantly enhance treatment decisions and potentially save lives. However, challenges such as cost and data analysis complexity must be addressed for broader clinical adoption. This study contributes to the ongoing fight against sepsis by paving the way for more targeted therapies and improved antibiotic stewardship. Ultimately, these advancements benefit public health by reducing unnecessary antibiotic use and lowering mortality rates associated with sepsis. In conclusion, WGS represents a promising frontier in precision medicine, offering hope for better outcomes in sepsis management.

Table of Contents

Abbreviations

A/E	Attaching and effacing
AMR	Antimicrobial resistance
ANI	Average nucleotide identity
AST	Antibiotic susceptibility test
BANDAGE	Bioinformatics Application for Navigating De novo Assembly Graphs Easily
BV-BRC	Bacterial and Viral Bioinformatics Resource Center
CARD	Comprehensive Antibiotic Resistance Database
ESBL	Extended spectrum beta lactamase
EUCAST	European Committee on Antimicrobial Susceptibility Testing
HPI	High Pathogenicity Island
MALDI-TOF MS	Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry
MDR	Multi-drug resistant
MLST	Multilocus Sequence Typing
NCBI	National Centre for Biotechnology Information
NGS	Next-generation sequencing
NIAID	National Institute of Allergy and Infectious Diseases
qSOFA	quick Sequential Organ Failure Assessment
QUAST	Quality Assessment Tool for Genome Assemblies
SBS	Sequencing by Synthesis
SIRS	Systemic Inflammatory Response Syndrome
SOFA	Sequential Organ Failure Assessment
ST	Sequence type
VFDB	Virulence factor database
WGS	Whole genome sequencing
WHO	World Health Organization

Introduction

Sepsis

The Greek word "sepo," which means decay or decomposition, is where the word "Sepsis" originates (Geroulanos & Douka, 2006). In 1914, Hugo Schottmüller made the first modern definition for sepsis stating that sepsis is recognized when a site forms, continuously or intermittently releasing pathogenic bacteria into the bloodstream, leading to noticeable subjective and objective symptoms (Gül et al., 2017). Sepsis remains a global healthcare challenge with its high rates of morbidity and mortality, demanding a more comprehensive understanding of its causes and causative microorganisms (Rudd et al., 2018). The first definitions of sepsis were defined in 1991, and known as the sepsis-1 definition focuses on the idea that sepsis was caused by infection-induced systemic inflammatory response syndrome (SIRS) in the context of severe clinical criteria (Bone et al., 1992). The temperature of the body is either above 38°C or below 36°C, a heart rate exceeding 90 beats per minute, a respiratory rate which is greater than 20 breaths per minute or a $PaCO_2$ level that is lower than 32 mmHg, and a white blood cell count that is greater than or equal to 12,000/cu mm or less than 4,000/cu mm, or that has more than 10% immature forms are the SIRS criteria being evaluated (Gyawali et al., 2019). However, this was revised by the American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference in 2001. This indicates that sepsis-2 was defined as SIRS plus infection and included tachycardia, tachypnea, hypothermia, and leukopenia (Levy et al., 2003). New definitions of sepsis- "Sepsis-3," or the Third International Consensus Definitions for Sepsis and Septic Shock, had been established in 2016 (Gul et al., 2017). Sepsis is now defined as life-threatening organ failure induced by a dysregulated host response to infection (Thompson et al., 2019). This revised definition placed more importance on organ dysfunction, which is described as a sudden increase in the infection-related total Sequential Organ Failure Assessment (SOFA) score of two points or higher (Vincent et al., 1996). The revised sepsis criteria also include a bedside clinical test called "qSOFA" which refers to the "quick Sequential Organ Failure Assessment", which helps to quickly identify patients with suspected infections who are at risk of developing sepsis outside of critical care units. Three main factors taken into account by the qSOFA test include reduced blood pressure, elevated respiratory rate, and changed mental condition (Singer et al., 2016).

Diagnosis

Initial diagnosis for suspected sepsis includes basic lab tests, blood cultures, relevant imaging, and sepsis biomarkers like procalcitonin and lactate levels. Priority in early management is fluid resuscitation, administering an IV crystalloid at 30 mL per kg within the first three hours. There are different phases of manifestations, from sepsis to severe sepsis and ultimately to septic shock (Shankar et al., 2016). Sepsis occurs from SIRS, which can be identified by the signs and symptoms mentioned above marking the first stage of the disease (Levy et al., 2003). The diagnosis of acute organ dysfunction marks the onset of the second stage, sometimes known as severe sepsis. Figure 1 highlights the intricate interplay between the body's defense mechanisms and potential harmful consequences during severe sepsis. When sepsis is combined with hypotension (low blood pressure) or hypoperfusion (decreased blood flow through an organ), severe sepsis may also be diagnosed. The signs and symptoms of organ damage include low urine production, sudden mental shifts, low blood platelet counts, dyspnea, irregular heartbeat, and stomach troubles (Jarczak et al., 2021). The third and most severe stage of sepsis is septic shock, having a high mortality rate (Mahapatra et al., 2023). While vasopressor therapy might be needed for persistent high-risk sepsis with perfusion abnormalities, lactate elevation, and despite fluid resuscitation

(Gauer et al., 2020), the cornerstone of early management involves prompt antimicrobial therapy, ideally within the first hour according to the latest guidelines, although the exact timeframe remains under debate (Arora et al., 2023; Gauer et al., 2020). Close monitoring of vital signs is crucial. Researchers are actively exploring various biomarkers, including procalcitonin, C-reactive protein, and the neutrophil-lymphocyte ratio, to predict immune responses during sepsis, with NLR emerging as an independent indicator, but requiring combination with other markers due to limitations in sensitivity and specificity for guiding treatment decisions (Huang et al., 2020; Mas-Celis et al., 2021; Sinha et al., 2018).

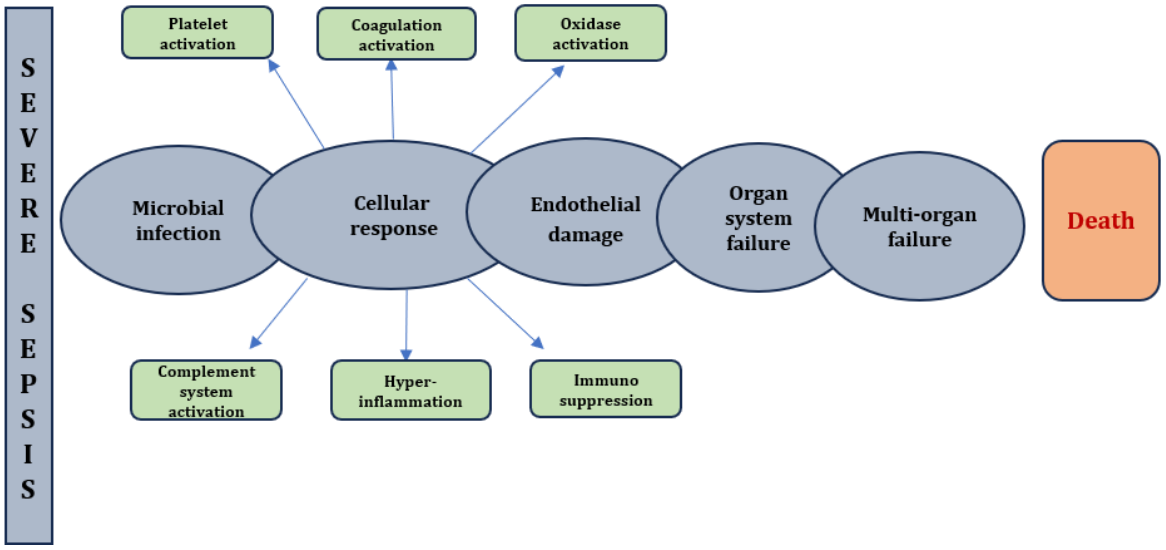

Figure 1| This flowchart illustrates the complex cascade of events in severe sepsis. It begins with a microbial infection that triggers a series of cellular responses, including platelet activation, oxidase activation (ROS production), coagulation activation (clotting), and hyperinflammation. The complement system also becomes activated, while the immune system can become suppressed. These events collectively lead to endothelial damage, compromising the integrity of blood vessels. Subsequently, organ failure can occur, potentially progressing to multi-organ failure and ultimately death.

Epidemiology and global burden

Sepsis, a leading cause of global morbidity and mortality, demands further investigation to precisely define its global burden (Leone et al., 2023). Highlighted on World Sepsis Day (September 13th), collaborative efforts focus on sepsis prevention, early detection, effective treatment, and knowledge sharing between the public, policymakers, and healthcare professionals (Schlapbach et al., 2020). This global health threat is estimated to cause 50 million cases and 11 million deaths annually, with a significant impact in the United States alone (Rhee et al., 2017). In Europe, the incidence of sepsis varies widely between countries, and there is no standardized method for estimating the burden of sepsis across the continent. However, a population-based study in Sweden estimated that approximately 3.4 million people in Europe suffer from sepsis each year, resulting in 680,000 deaths (Mellhammar et al., 2016). According to predictions, low- and middle-income nations account for the bulk of sepsis cases and fatalities. The World Health Organization (WHO) has released a study that emphasizes the impact of sepsis on public health, focusing on specific demographics and those seeking medical attention (Rudd et al., 2018). Compared to prevalent conditions including heart failure, pneumonia, and chronic

obstructive pulmonary disease, individuals with sepsis are more likely to experience post-discharge consequences, especially persistent immunological dysfunction and systemic inflammation (Wardi et al., 2023). However, because sepsis incorporates symptoms and indications with various health ailments, data-driven early detection of sepsis is challenging (Jazayeri et al., 2023). Sepsis may be prevented, and its future outlook is highly dependent on the effectiveness and timeliness of medical treatment. Studying sepsis becomes essential as it may result in new insights into the pathophysiology of the condition, the discovery of novel therapeutic targets, and the development of more effective treatments (Yatabe et al., 2022).

Microbial etiology and antibiotic resistance

The most common pathogens responsible for sepsis include viruses, fungi, and bacteria. The majority of sepsis is caused by bacterial infections, which usually account for more than 70% of sepsis cases that have been reported (Lin et al., 2018). The bacterial pathogens *Escherichia coli, Klebsiella pneumoniae, Staphylococcus aureus, Pseudomonas aeruginosa,* and *Streptococcus pyogenes,* as well as the fungal pathogens of the *Candida* species, are some of the common pathogens that cause sepsis (Gatica et al., 2023). Among them, Gram-negative bacteria are one of the most common sources of sepsis that are found (Ljungström et al., 2019). Gram-negative bacteria are a group of microorganisms known for their significant impact on infections in humans. They are characterized by the presence of distinctive cell wall structures that can adhere to host cells during infection and form resilient biofilms. The production of lipopolysaccharide toxins that can damage host cells and induce immune response is a peculiar feature of these bacteria. They often damage the immune system by secreting virulence genes, immune invasion etc. Furthermore, the innate resistance of Gram-negative bacteria to antibiotics, which is linked to the development of resistance genes and a protective outer layer, creates significant challenges in treatment. This may also lead to prolonged illness, elevated healthcare expenses, and increased mortality rates (Hirakawa et al., 2021).

New strains of bacteria, viruses, or fungi are examples of emerging microorganisms that can have unique pathogenic characteristics, resistance to antibiotics, or transmission patterns that might affect sepsis epidemiology and clinical consequences (Van Engelen et al., 2018). Because of the high mortality risk associated with Gram-negative sepsis, timely and focused antibiotic treatment is important (Pant et al., 2021). Effective sepsis care requires attention and adaptability in detecting AMR and dealing with these new infectious pathogens (Schwartz, 2021). Antimicrobial resistance is the ability of microorganisms, such as bacteria, viruses, fungi, and parasites, to resist the impact of various antibiotics (Tang et al., 2023). Resistance strains emerge and grow as a result of both the overuse and misuse of antibiotics as well as inadequate infection control measures. Continued surveillance of antibiotic resistance in clinical studies is recommended since horizontal gene transfer could play a significant role in the spread of virulence genes and antibiotic resistance genes (Zhang et al., 2023). Treatment plans are made more difficult by the emergence of more antibiotic-resistant forms of Gram-negative bacteria, which emphasizes the necessity of meticulous antimicrobial stewardship and innovative techniques for treating sepsis brought about by these bacterial infections (McLeod et al., 2019). Research must continue to keep up with the constantly changing environment of microbes and enhance our ability to treat sepsis caused by newly discovered infectious agents (Meckawy et al., 2022).

Citrobacter Species

Citrobacter species are presently gaining importance as a clinical, multidrug-resistant bacteria that cause different nosocomial and community-based infections (Yao et al., 2021). Werkman and Gillen first isolated *Citrobacter* organisms in 1932, identifying seven species and proposing the general name *Citrobacter* (Borenshtein & Schauer, 2006). *Citrobacter* species are prokaryotic, Gram-negative, facultative anaerobic bacteria, that are members of the Enterobacteriaceae family (Kumar et al., 2021). They are a common element of the environment and the usual microbiota of the human digestive system. *Citrobacter* species are usually isolated from soil, water, animals, blood, cerebrospinal fluid, wound, faecal matter, and respiratory and urogenital tracts of humans (Kolínská et al., 2015). There are 15 species in the genus *Citrobacter*; the most clinically significant ones are *Citrobacter freundii*, *Citrobacter koseri,* and *Citrobacter braakii.* Multiple diseases, such as bloodstream infections, respiratory tract infections, urinary tract infections, and other extraintestinal infections, can be caused by these bacteria (Liu et al., 2021).

Citrobacter species inclusive of *Citrobacter freundii, Citrobacter koseri,* and *Citrobacter amalonaticus* are opportunistic pathogens in humans that can cause invasive diseases such as infections of the skin, soft tissues, central nervous system, respiratory tract, and urinary tract infections. Particularly in newborns and immunocompromised hosts, the bacteria can cause intra-abdominal infections, suppurative arthritis, bacteremia, endocarditis, and endophthalmitis (Yuan et al., 2019). In sepsis, inflammatory mediators such as cytokines and endotoxins can be produced due to the interaction between the host's immune system and Gram-negative bacteria (Mitchell et al., 2019). A significant portion of strains of *Citrobacter* species included the Yersinia high-pathogenicity island, related to higher virulence and resistance to antibiotics (Carniel, 1999). *Citrobacter* species are categorized by their genetic and phenotypic traits, and because of their potential virulence and correlation with antibiotic resistance, their clinical relevance has been accepted more and more (Liu et al., 2021).

Due to the high mortality risk associated with Gram-negative sepsis, prompt and specific antibiotic treatment is vital (Mitchell et al., 2019). The rise of antibiotic-resistant strains of Gram-negative bacteria further complicates treatment techniques. This emphasizes the need for careful monitoring of antibiotics and creative approaches to controlling sepsis caused by these infections (Avatsingh et al., 2023). *Citrobacter* species have exhibited resistance to various antibiotics, encompassing aminoglycosides, fluoroquinolones, lincosamide antibiotics, cephalosporin, cephamycin, and penam. A comprehensive analysis of 129 *Citrobacter* genomes revealed the presence of resistance genes across all species, including those responsible for different types of antibiotic efflux pumps. The rise of antibiotic-resistant *Citrobacter* strains is a growing concern, underscoring the urgency for the development of novel treatment strategies. Notably, *Citrobacter koseri* displayed lower susceptibility to several antibiotics compared to other *Citrobacter* species (Yuan et al., 2019). The rise of multidrug-resistant (MDR) strains of *Citrobacter freundii* has generated concerns over the antibiotic resistance pattern of this species. *Citrobacter freundii* is known to be resistant to several antibiotics, including piperacillin, ceftriaxone, ceftazidime, and piperacillin/tazobactam and is also associated with mutations in the *gyrA* gene's quinolone resistance-determining region (Liu et al., 2021; Samonis et al., 2009). Numerous global surveillance systems have documented *Citrobacter* species resistance to β-lactam antibiotics, including the generation of β-lactamases such as Amp-C, broad-spectrum β-lactamase, extended-spectrum β-lactamase (ESBL), and carbapenemase (Liu et al., 2021).

The genetic diversity of *Citrobacter* species has been associated with variations in antibiotic resistance patterns and virulence properties. The genetic range of *Citrobacter* species could be useful for interpreting their epidemiology, antibiotic resistance, and virulence traits, and may also provide important information for medical management and infection prevention. By studying the seven housekeeping genes *(aspC, clpX, fadD, mdh, arcA, dnaG,* and *lysP)* using multilocus sequence typing (MLST) genetic diversity of sequence types (STs) and clonal complexes (CCs) of *Citrobacter* species are revealed (Liu et al., 2021). The difficulties in fighting antibiotic resistance are further complicated by the spread of antibiotic-resistant genes and the threat of horizontal gene transfer among bacterial populations (Luo et al., 2021). Treatment techniques for *Citrobacter* species antibiotic resistance include optimization of using antibiotics, reduction of unnecessary prescriptions, prevention of the spread of resistant strains, conducting programs related to the education of healthcare providers and implementation of new guidelines for proper antibiotic use (Bassetti et al., 2016; Baur et al., 2017).

Genomic Characterization and Whole Genome Sequencing

The process of thoroughly examining an organism's genome, which includes its genetic composition, structure, function, and organization, is known as genomic characterization. The process includes identifying and understanding the genetic components included in an organism's DNA, including genes, regulatory sequences, and non-coding regions (Emerson et al., 2008). WGS, genetic and genomic analysis, and plasmid conjugation experiments are some of the techniques used in genomic characterization that provide insights into the genetic basis for specific traits or characteristics, such as antibiotic resistance (Lalaoui et al., 2019). The entire spectrum of true somatic genomic changes including copy number alterations, microbial infections, indels, rearrangements of repetitive elements, nucleotide substitution mutations, inversions, translocations, and complex whole-genome rearrangements, can all be identified using WGS. Also, WGS allows the identification of genes responsible for virulence factors and biochemical pathways as well as to genetically manipulate the genome to study phylogenetic and phenotypic characters (Massilamany et al., 2016). For appropriate diagnosis and early decision of adequate treatment among critically ill patients, host immunological responses, assessment of severity of ailments, risk factors for antibiotic-resistant pathogens, and identifying the exact causative pathogen are crucial (Cillóniz et al., 2021). Over the last ten years, tremendous improvements in genomic technology, such as next-generation sequencing (NGS), coupled with enhancements in analytical techniques have resulted in a better understanding of molecular mechanisms (Tuna & Amos, 2013).

Species identification and next-generation sequencing

The timely and accurate identification of the causative organism is very important in the case of sepsis as it plays an important role in patient survival (Patel, 2015). However, culture-based techniques can lead to inefficient treatment in time and overtreatment with antibiotics ending with patients at risk. Over the past few decades, sequencing technologies have developed quickly that improved efficiency in diagnosis in time. This high-throughput technology and bioinformatics can not only identify pathogens more quickly, but can also provide novel information on virulence, antibiotic resistance, and the transmission of illness. This was made possible by the development of next generation sequencing (NGS) technologies including Illumina (Kozarewa et al., 2009), Ion Torrent (Quail et al., 2012; Wick et al., 2023), PacBio and Nanopore (Rayamajhi et al., 2022). These new DNA sequencing platforms allowed for the sequencing of millions of DNA fragments simultaneously, with detailed insights into genome

structure, genetic variations, gene expression profiles, and epigenetic modifications (Pervez et al., 2022). To delve deeper into the sequencing method employed in this study, Illumina sequencing, a prominent technology within NGS utilizes the reversible termination sequencing by synthesis (SBS) technique to generate short reads (Hu et al., 2021; Pereira et al., 2020). Figure 2 illustrates the key steps involved in the SBS process. NGS data processing and interpretation rely heavily on bioinformatic techniques to unlock valuable biological insights from the massive datasets generated (Wang et al., 2015). This versatility, accuracy, cost-effectiveness, speed, and high throughput of NGS have revolutionized genomics research, including sepsis studies (Behjati & Tarpey, 2013). Species identification is another important stage in treatment that provides additional information on that particular microorganism (Guo et al., 2014). Proper treatment requires an accurate diagnosis of the causative organism of an infection. Phylogenetic analyses consistently position *Citrobacter* within the Enterobacteriaceae family, exhibiting close evolutionary relationships with *Salmonella* and *Escherichia coli* (Jabeen et al., 2023 ;Shih et al., 1996).

The Bacterial and Viral Bioinformatics Resource Center (BV-BRC) is a web-based information system used to support research on bacterial and viral infectious diseases. The National Institute of Allergy and Infectious Diseases (NIAID) merged three existing bioinformatics resources – PATRIC, IRD, and ViPR – to create the Bacterial and Viral Bioinformatics Resource Center (BV-BRC). This consolidation aims to streamline data analysis for researchers by offering a unified data model, improved web-based visualization tools, and a comprehensive suite of services for both bacterial and viral research(Olson et al., 2023). BV-BRC facilitates comprehensive data analysis for infectious diseases by offering automated and manual curation of high-value data and metadata, integrated multi-omics systems biology analysis with visual aids, private user workspaces for data exploration, sharing, and publication, and online outreach and educational materials (https://www.bv-brc.org/). This thesis aimed to study the *Citrobacter* genome data from clinical isolates of sepsis patients to determine the genetic diversity, the antimicrobial resistance profiles and virulence properties. An in-house bioinformatics pipeline developed by the Systems Biology Department of the University of Skövde and the BV-BRC pipeline are used to analyze the genome of *Citrobacter* isolates.

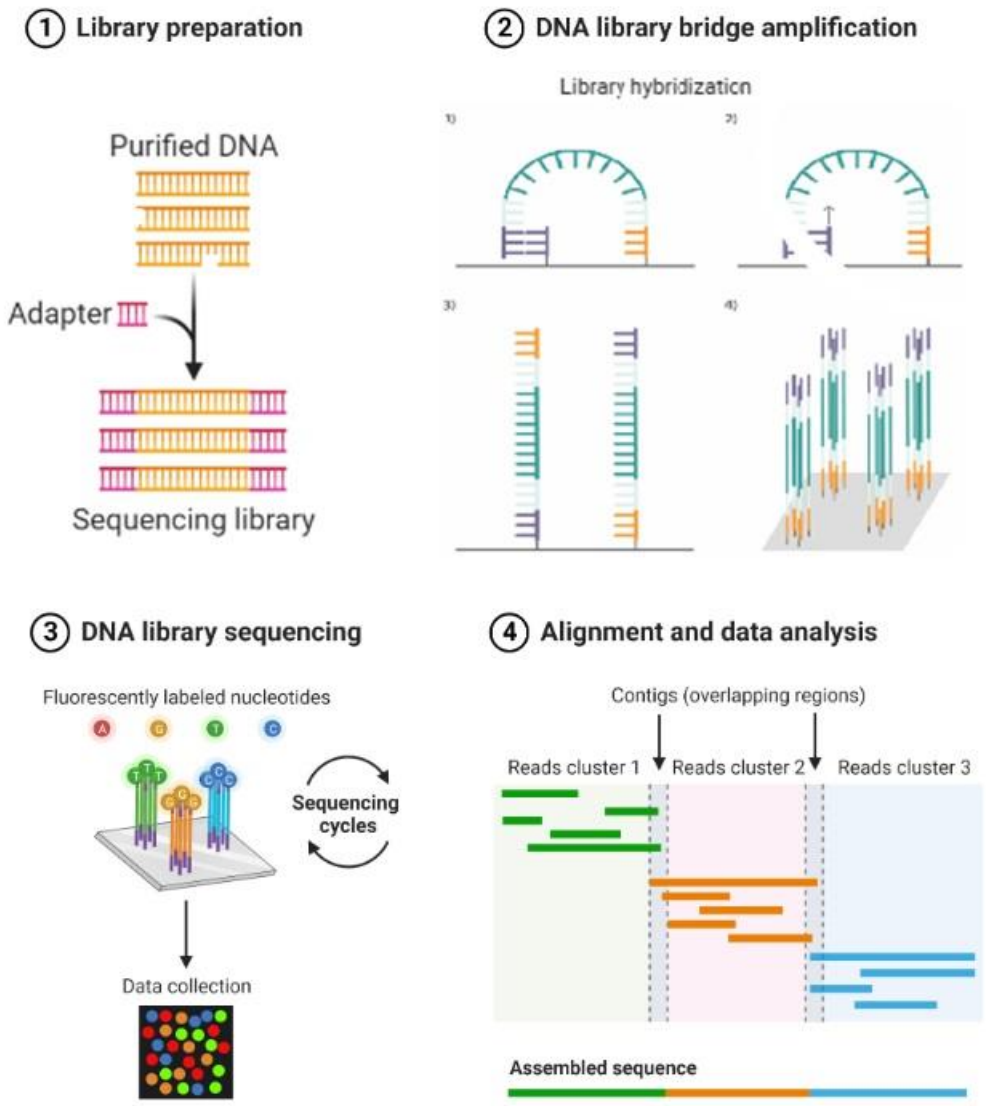

Figure 2 | Depicts the Illumina SBS workflow for whole genome sequencing. The process commences with the isolation of genomic DNA from various sources, fragmented and adorned with unique barcode adapters to create a library (step1). The sequence information of the library is subsequently obtained by synthesizing a complementary DNA strand on a flow cell. Bridge amplification, akin to PCR, enriches the template DNA and intensifies the sequencing signal (step 2). This step, crucial for cluster generation, ensures sufficient material for downstream sequencing. The fluorescently labelled nucleotides are sequentially incorporated into the growing DNA strand during sequencing cycles, with their identities recorded (step 3). Finally, the short sequencing reads are assembled into longer contigs and potentially grouped into clusters to generate final DNA sequence (step 4). The figure was created using the Bio Render online tool icons and PowerPoint.

Aim

The study attempts to elucidate the biodiversity and pathogenicity profiles of *Citrobacter* species isolated from adult patients suspected of community-onset sepsis. The scientific inquiry centered on identifying unique genomic signatures to differentiate *Citrobacter* species accurately, thereby mitigating the risk of misidentification and unnecessary antibiotic administration. The primary aim was to analyze WGS data of *Citrobacter* isolates using an in-house bioinformatic pipeline and compare it with the BV-BRC pipeline. Specifically, the study compared the performance of different assemblers within the BV-BRC pipeline, assessing their effectiveness in reconstructing *Citrobacter* genomes. This comparison aids in evaluating the strengths and weaknesses of different bioinformatic pipelines for future research endeavours. Characterizing genetic determinants of virulence genes, plasmid replicons and antimicrobial resistance in *Citrobacter* isolates through WGS analysis, and correlating these findings with phenotypic AST data to validate in silico predictions, constitute the key objectives of this research. Distinguishing between *Citrobacter* species enables more precise outbreak tracking and antimicrobial resistance pattern monitoring within specific populations, minimizing the risk of unnecessary broad-spectrum antibiotic usage and providing insight into the emergence of antibiotic resistance.

Materials and Methods

Collection of isolates

The study used a dataset that comprises whole-genome sequenced data of clinical isolates of *Citrobacter spp* identified using the routine culture method and MALDI-TOF MS (DB-4110, Bruker Daltonics, Germany) for phenotypic classification. The dataset used for analysis was obtained from previous research done by experts at multiple centers, as detailed in Figure 3. These isolates were obtained as part of a prospective observational sepsis study named "Sepsis study Skaraborg" conducted in the region of southwest Sweden, involving adult patients with community-onset sepsis. The study was carried out at Skaraborg Hospital in collaboration with the Clinical Microbiology Laboratory at Unilabs, Skövde, and the University of Skövde, in September 2011 and June 2012 (Ljungström L., 2017). A study population of 2,475 individuals who presented with suspected sepsis was included in the study. A variety of clinical specimens, including blood, serum, urine, and nasopharynx samples, were systematically collected and subsequently preserved at -80°C in Microbank™ vials (Pro-Lab Diagnostics, Ontario, Canada for analysis (Ljungström L., 2017).

Pre-analytical processing and sequencing (previously performed)

Approximately 1,800 bacterial isolates were obtained from these specimens and from that a total of 21 bacterial isolates were used for this study. It included *Citrobacter braakii* (n=1), *Citrobacter freundii* (n=7), *Citrobacter koseri* (n=11), *Enterobacter bugandensis* (n=1) and *Citrobacter farmeri* (n=1), as species identification done by MALDI-TOF MS. Among the isolates, blood samples (aerobic n=2; anaerobic n=1), urine cultures (n=13), upper respiratory tract culture (n=1) and wound culture (n=3) were present respectively. Data on phenotypic AST, tested using a disc diffusion method following EUCAST (www.eucast.org) guidelines were also provided.

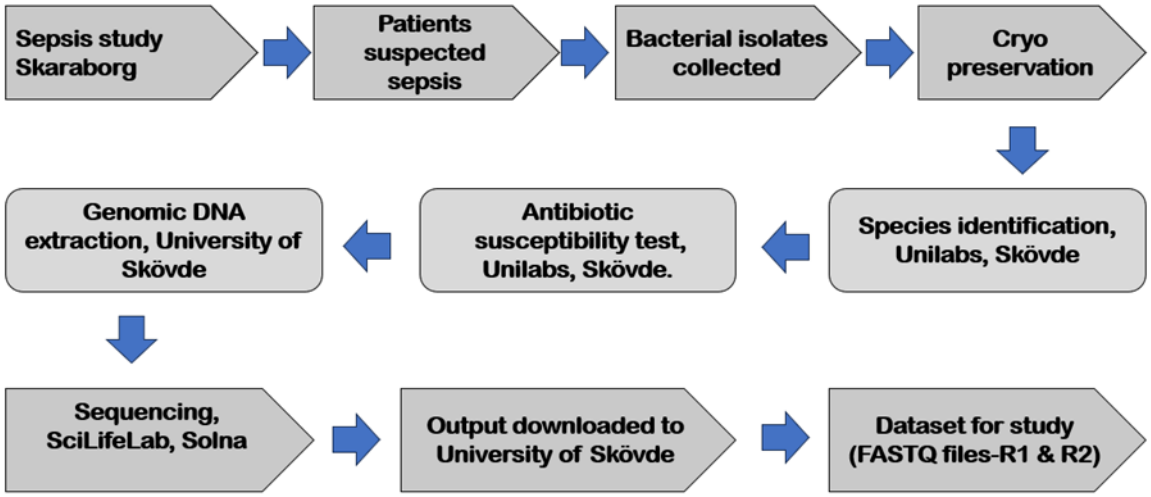

Figure 3 | The flowchart depicting the process of obtaining the dataset used for analysis. The process involved several key steps: (1) Patient screening for suspected sepsis as part of sepsis study Skaraborg, (2) Collection of different bacterial isolates, (3) Cryopreservation of these isolates , (4) Antibiotic susceptibility testing and (5) Species identification at Unilabs in Skövde, (6) Genomic DNA extraction at the University of Skövde, (7) Sequencing of the extracted DNA at SciLifeLab in Solna, and (8) Downloaded the resulting data back to the University of Skövde for further analysis.

Citrobacter isolates were subjected to genomic DNA extraction using the QIAamp DNA Mini and Blood Mini protocol on an automated QIAcube instrument (Qiagen, Germany) at the University of Skövde, Sweden. To eliminate PCR inhibitors, subsequent purification was performed using the DNeasy® PowerClean® Pro Cleanup Kit (Qiagen, Germany) on the same QIAcube platform. The concentration of DNA in each sample was quantified before and after cleanup with a Qubit 4.0 Fluorometer (Thermo Fisher Scientific, USA) employing the Qubit® dsDNA BR Assay kit. Samples with DNA concentrations below 4 ng/µl were excluded from downstream sequencing analyses and the samples were stored at -20°C.

At SciLifeLab, Solna, library preparation and Illumina sequencing were performed. The libraries were prepared with enzymatic tagmentation, PCR-clean-up and index ligations following Illumina's protocol for Nextera XT DNA sample preparation. To ensure that the libraries prepared are of appropriate size and concentration for sequencing fragment analysis was performed using a Bioanalyzer (Agilent Technologies, USA). Sequencing was carried out using the Illumina HiSeq 2500 platform. The raw FastQ files generated after sequencing were then downloaded to the University of Skövde from SciLifeLab, Solna. The sequenced data (FastQ files) included details such as the sequencer identifier, the actual DNA sequences (reads), and quality scores. The FastQ-files containing sequencing data (raw reads) downloaded from SciLifeLab to the University of Skövde were utilized as the dataset for this study.

In-house pipeline

The analysis of *Citrobacter* isolates raw data involved two pipelines, with the tools and methods used for the study illustrated in Figure 4. In the in-house pipeline, initial quality assessment and subsequent quality re-evaluation were conducted using FastQC (Andrews, 2010). Trimming of reads was performed with Trimmomatic (Bolger et al., 2014) followed by generating a MultiQC

report of the FastQ files. The unicycler assembler was used to perform *de novo* assembly, that uses SPAdes to assemble short reads (Wick et al., 2017). The quality of the assembly was assessed using QUAST (Gurevich et al., 2013) with selected reference genomes given in Appendix 1, Table 1. Further analysis utilized tools from the Center for Genomic Epidemiology (CGE). PlasmidFinder was employed for the identification of plasmids (Carattoli et al., 2014), while ResFinder was used to identify antibiotic-resistance genes (Zankari et al., 2012). Species identification was carried out using the JSpeciesWS tool (Richter et al., 2016) and multi-locus sequence typing of isolates was performed using the PubMLST tool (Jolley et al., 2018).

BV-BRC pipeline

The Bacterial and Viral Bioinformatics Resource Center (BV-BRC) is a comprehensive network designed to support the biomedical research community in investigating bacterial and viral infectious illnesses. It serves as a hub for researchers by integrating relevant pathogen data with a variety of analytical tools. BV-BRC offers readily accessible open-source resources for data analysis, genomic annotation, comprehensive genome analysis etc. To compare the in-house pipeline with the BV-BRC pipeline, tools and services in the BV-BRC database were utilized. These include the FastQ utilities service that offers independent or combined functionalities for read quality assessment, trimming, and reference genome alignment (FastQC and TrimGalore), the genome assembly service (Unicycler and SPAdes), the taxonomic classification service (Kraken2), the genome annotation service uses Comprehensive Antibiotic Resistance Database (CARD), Virulence factor database (VFDB) and comprehensive genome analysis service (PubMLST) are used.

Quality control

In this study, a thorough assessment of sequenced data quality was conducted to ensure quality parameters for reliable genomic analysis using the FastQC software tool (version 0.11.9) (Andrews, 2010). The BV-BRC database uses FastQC (version 0.12.1). The tool FastQC in each pipeline was used again after trimming to recheck the quality of trimmed reads after trimming. The quality of the reads was evaluated using various parameters, including basic statistics, per base sequence quality, per sequence quality scores, per base sequence content, per base GC content, per base N content, sequence length distribution, sequence duplication levels and adapter content (Babraham bioinformatics).

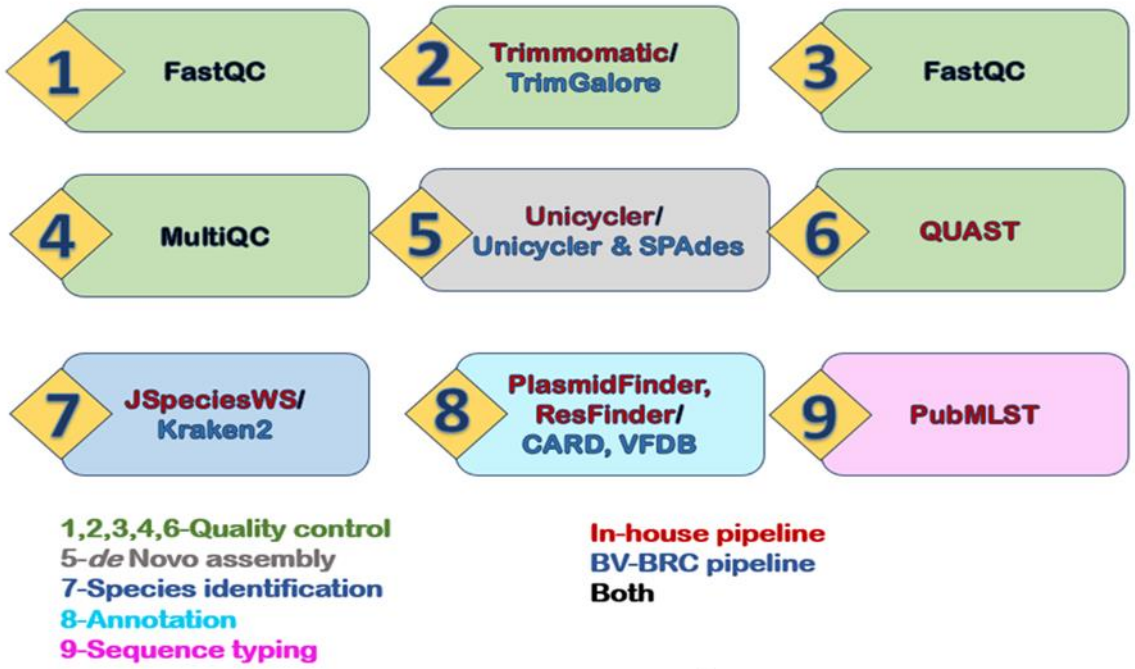

Figure 4 | Summary of the workflow and tools used for *Citrobacter* isolate raw data analysis. This analysis leverages both an in-house pipeline (red) and the BV-BRC (Bacterial and Viral Bioinformatics Resource Center) pipeline (blue) concurrently. Red components signify tools specific to the in-house pipeline, while blue represents those used in the BV-BRC pipeline. Black colored tools are common to both. The analysis begins with initial quality control for both pipelines, followed by a re-check using FastQC. Reads are then trimmed with Trimmomatic (in-house) and TrimGalore (BV-BRC). MultiQC is employed to generate combined FastQC reports. *de novo* assembly in the in-house pipeline utilizes Unicycler, with quality evaluation performed by QUAST. Subsequently, Centre for Genomic Epidemiology (CGE) tools (PlasmidFinder and ResFinder) are used for plasmid identification and antibiotic resistance gene identification, respectively. Species identification is achieved with JSpeciesWS, and multilocus sequence typing is performed using PubMLST. The BV-BRC pipeline utilizes Unicycler and SPAdes for genome assembly. Species classification is done with Kraken2 in the Taxonomic Classification service. Finally, the BV-BRC pipeline identifies antibiotic resistance genes from the Comprehensive Antibiotic Resistance Database CARD database, virulence factor genes from the VFDB database, and performs sequence typing with PubMLST.

Trimming

To improve the accuracy of the analysis, the adapter sequences should be removed and low-quality bases should be trimmed. The in-house pipeline used Trimmomatic (Bolger et al., 2014) for data trimming of adapter sequences and quality filtering of raw reads. It is a flexible and very efficient preprocessing tool specially optimized for handling large sets of paired-end reads. Adapter sequences, low-quality bases at the ends of the reads and unpaired reads are removed during trimming. However, it can lead to false positives during downstream analysis. To guarantee that there are no mismatches, parameters were set to perform trimming. They are (i) the "ILLUMINACLIP" command used to remove Nextera XT adapter sequences from the reads. The seed mismatches, palindrome clip threshold and simple clip threshold are set to 0, 30 and 10 respectively. (ii) The "SLIDING WINDOW" option is used to perform quality-based trimming on reads. It evaluates the quality Phred score corresponding to each base with the quality threshold set to 20 and window size 4. (iii)To specify the number of bases to be removed from the beginning

of reads "HEADCROP" flag is used and set to 12. (iv) To ensure the minimum length of reads to be retained after trimming "MINLEN" parameter is set to 30 (Bolger et al., 2014). The code used for running trimmomatic for FastQ forward and reverse files with all parameters is given in Appendix 1, Figure 1.

TrimGalore (Krueger, 2012) was used for trimming in BV-BRC under FastQ utilities service (Babraham bioinformatics). This tool consistently does quality and adapter trimming by creating a wrapper tool around Cutadapt (Martin, 2011a) and FastQC (Davis et al., 2019). Trim Galore version 0.6.5 dev and Cutadapt version 4.2 were used for paired-end trimming. A quality Phred score cutoff of 20 was applied, and the selected quality encoding type was ASCII+33. Nextera adapter sequences (sequence `'CTGTCTCTTATA'` Nextera Transposase sequence) are auto-detected. A minimum adapter overlap of one base pair was necessary for trimming, and the highest permitted trimming error rate was set at 0.1. For each pair of reads, sequences that were less than 20 base pairs were eliminated (Babraham bioinformatics).

MultiQC

For generating a single report of FastQC results of all isolates, MultiQC version v1.9 was used. MultiQC helps to summarise the results from different samples and different steps of analysis into one consolidated output (Ewels et al., 2016). The report contains information about the FastQC report of raw FastQ files and the FastQC report of files after trimming. The MultiQC report contains Sequence Counts, Sequence Quality Histograms, Per Sequence Quality Scores, Per Base Sequence Content, Per Sequence GC Content, Per Base N Content, Sequence Length Distribution, Sequence Duplication Levels, Overrepresented sequences, Adapter Content and Status Checks. The taxonomic classification service of BV-BRC generates a MultiQC report with the Kraken2 (Wood et al., 2019) and FastQC results. The MultiQC reports were generated for both raw FastQ files and trimmed FastQ files, and all reports were combined to generate an aggregate result using MultiQC version v1.9.

de novo assembly

During de novo assembly for the in house pipeline, Unicycler (Wick et al., 2017) utilizes specialized algorithms to reconstruct genomic sequences obtained from NGS, without relying on a reference genome. This tool excels in handling short reads by optimizing the SPAdes assembler (Bankevich et al., 2012) and systematically examining a variety of k-mer sizes to choose the best ones based on factors such as contig length and graph connectivity. Unicycler enhances assembly quality by effectively scaffolding the assembly graph using SPAdes repetition resolution features, thus improving continuity. Before assembly, SPAdes read error correction is applied to reduce mistakes associated with short reads, thereby enhancing the reliability of the completed assembly. Additionally, Unicycler automatically determines the k-mer range based on the length of the input reads to optimize the assembly process. It refines the assembly graph by creating SPAdes contig bridges and meticulously cleaning the graph, resulting in a bridged assembly graph. Finally, Pilon is employed to systematically identify and rectify flaws in the assembly, further refining it for downstream analysis (Wick et al., 2017). The code used for running Unicycler is given in Appendix 1, Figure 2.

In the BV-BRC pipeline, with the genomic assembly service of BV-BRC, assemblers Unicycler and SPAdes were used to compare the quality of assembly. The assembly strategy parameter was set

to Unicycler with advanced parameters trimming reads before assembly, and pilon iteration was selected as 4 with a minimum contig length and coverage set to 300 and 5, respectively, as default settings. Additionally, SPAdes were employed as an assembly tool, utilizing the BayesHammer module with default settings for error correction to generate a high-quality genome assembly from the FastQ files. SPAdes leverages a multi-kmer de Bruijn graph for short genome assembly, particularly useful for bacterial genomes (Bankevich et al., 2012). The assembly strategy parameter was set to SPAdes, with advanced parameters trimming reads before assembly, and pilon iteration selected as 4, with minimum contig length and coverage set to 300 and 5, respectively, as default settings. The output of Unicycler and SPAdes are FASTA files that contain assembled contigs that represent the reconstructed genome. The BV-BRC genome assembly report generates a visual representation of the assembly process, called Bioinformatics Application for Navigating De novo Assembly Graphs Easily or Bandage plot (Wick et al., 2015). It is similar to a map that shows both nodes and edges and helps to analyse how the smaller pieces of DNA sequence (reads) were connected to form the larger, complete genome.

QUAST

After *de novo* assembly, the quality of assembly like N50, L50, genome fraction, genome coverage and the number of contigs are to be evaluated. For this when compared with a reference genome it enhances the results of the assembly being assessed. The reference genome is selected on the following parameters such as assembly statistics, quality analysis, annotation details etc. Reference genomes were selected from the National Center for Biotechnology Information (NCBI) separately for each isolate. In the BV-BRC pipeline, the search bar in the BV-BRC home page was used to select which reference genomes were then used for analysis. To maintain consistency for both pipelines the same reference genomes were used and the list of reference genomes and details are given in Appendix 1, Table 1. QUAST (quality assessment tool for genome assemblies) helps to select good quality FASTA files by providing completely covered genes, the number of predicted genes, and the number of misassemblies for each assembly (Gurevich et al., 2013). QUAST helps to choose the best pipeline by analyzing, comparing and evaluating assemblies. It also generates statistical results in graphical and PDF formats. The code used for running QUAST is given in Appendix 1, Figure 3. The genome assembly report generated by the BV-BRC genome assembly service includes a QUAST report for evaluating and comparing the genome assemblies.

Species identification

The probability of a group of genome assemblies against reference data belonging to the same species was measured using the Average Nucleotide Identity (ANI) method (Ciufo et al., 2018). JSpecies Web Server (JspeciesWS) version 4.1.1 (Richter et al., 2016) is an online tool used for calculating the average nucleotide identity between two genomes. The ANI calculations are based on the BLAST+ method (ANIb). The server is linked to a genome cart and the reference genome can be selected for comparison against the assembly. The results are validated with the average nucleotide identity (ANI) indicating a more accurate cutoff point of around 95–96% (Richter & Rosselló-Móra, 2009). The Taxonomic classification service by BV-BRC includes the species identification tool Kraken2 (Wood et al., 2019) for the classification of isolates. It uses exact-match databases of k-mers for sequence alignment. A set of lowest common ancestor (LCA) is used to select the proper label for the sequence. This service is managed by Snakemake for species identification (Lu et al., 2022). The output generated by Kraken2 includes a Krona plot and a Sankey plot (Ondov et al., 2011). The parameters selected include WGS and the Kraken2 standard

database. The default confidence interval was 0.1 for all sequence types and analysis types selected as species identification.

Genotypic analysis

The Center for Genomic Epidemiology (CGE) offers sophisticated bacterial analysis tools for various aspects of microbial genomics. PlasmidFinder version 2.1 was utilized for the identification of plasmids. The default settings used are 95% as the threshold for the minimum percentage of identity and 60% as the minimum percentage of coverage. The database selected was *Enterobacteriales* before submitting the assembled contigs. PlasmidFinder detects and characterizes plasmid sequences based on draft genomes (Carattoli & Hasman, 2020). The antimicrobial resistance genes are identified through ResFinder version 4.4.2. The ResFinder identifies antimicrobial resistance genes from assembly FASTA files by setting the default threshold of 90% sequence identity and a minimum length of 60% (Bortolaia et al., 2020). The phenotypic antibiotic resistance results (antibiotic susceptibility test data) obtained through EUCAST (www.eucast.org) criteria using the disc diffusion method are compared with genotypic results from the ResFinder.

In the BV-BRC pipeline, the genome annotation service uses the RASTtool kit (RASTtk) (Brettin et al., 2015) for genome annotation of bacteria from the SEED database(Overbeek et al., 2014). The antimicrobial resistance genes were identified from the Comprehensive Antibiotic Resistance Database (CARD). This database utilizes an ontology system allowing efficient analysis of resistance genes across diverse bacterial strains. A single resistance gene sequence can be linked to multiple *Citrobacter* strains based on their taxonomic relationships (McArthur et al., 2013). The virulence factor genes were identified from the Virulence factor database (VFDB). VFDB uses a BLAST (Basic Local Alignment Search Tool) functionality for identifying potential virulence factors within query genomes by comparing them against known virulence factor sequences in the database. This can analyze the newly sequenced bacterial strains. The antimicrobial resistance genes and virulence factor genes with identity above 90% from CARD and VFDB were selected respectively.

Sequence typing

MLST based on seven housekeeping genes (*arcA, aspC, clpX, dnaG, fadD, lysP and mdh*) was done with the PubMLST tool (Jolley et al., 2018). Different alleles, or sequences for housekeeping genes, are found in every species of bacteria. An isolate's allelic profile, also known as its sequence type (ST), is identified by the combination of alleles at each of its seven loci. It distinguishes closely related strains by assessing whether the allele numbers differ in their nucleotide sequences (Urwin & Maiden, 2003). In the BV-BRC pipeline, the genome annotation service uses PubMLST (www.pubmlst.org) for assigning sequence types.

Statistical analysis

All statistical analyses were done using the RStudio version 4.3.3. A p-value <0.05 was considered statistically significant. The distributions of N50, L50, and number of contigs between *Citrobacter freundii* and *Citrobacter koseri* isolates were evaluated using the Wilcoxon signed-rank test. This non-parametric test is suitable for analyzing paired, non-normally distributed data (Kerby, 2014). It ranks the differences between corresponding samples (isolates from different species) and assesses whether the median of these differences deviates significantly from zero (Li & Johnson,

2014). The Figure 4 was generated using ggplot2. To determine if there is a significant difference between the performance of Unicycler and SPAdes in the BV-BRC pipeline for the parameters (N50, L50 and number of contigs) the Mann-Whitney U test was conducted. Also to compare the distribution of plasmid counts between two species types the Mann-Whitney U test was used. The Kruskal-Walli's rank sum test was conducted to evaluate potential variations in the number of plasmids across different sample types.

Results

Quality control, trimming and assembly

Analysis of raw FastQ files using FastQC revealed high quality for most metrics during the initial assessment, irrespective of the in-house pipeline or BV-BRC pipeline. The basic statistics, per base sequence quality, per sequence quality, per base N content, sequence length distribution and overrepresented sequences for both pipelines were of very good quality with a "green "mark. The per tile sequence quality remained passable quality with an "orange" mark for a single file, while all files for both pipelines showed "good" quality overall. Similarly, both pipelines received passable marks for per base sequence content and the per sequence GC content across all files. Although the quality scores meet acceptable thresholds, the decision to proceed with further genotypic analysis recognizes the paramount importance of data quality for downstream applications and its biological relevance. The sequence duplication levels obtained mixed results for both pipelines. Two files from each pipeline achieved high quality, while four files from each received low quality and the remaining files were "passable". The adapters were a point of concern, with all files for both pipelines receiving "red "marks for "poor" quality.

In-House Pipeline: Trimmomatic improved the overall read quality by removing adapter sequences. Appendix 2, Figure 1, displays the number of reads before and after trimming for sample ID CI11. However, FastQC analysis after trimming revealed some remaining issues. While per-base sequence content was generally good, four files fell into the "passable" category. Similarly, GC content remained a concern for two files, categorized as "poor," while others were "passable." Sequence duplication levels also raised concerns, with one file receiving a "warning" and only two achieving "good" quality. The remaining files were "passable" in this metric. The sequence length distribution remained "passable" for all files after trimming. Trimming with Trimmomatic resulted in a significantly larger reduction in read count (average difference of 1,559,947 reads). Appendix 2, Table 1 shows total read counts before and after trimming for forward and reverse reads of isolates with Trimmomatic.

BV-BRC Pipeline: Subsequent FastQC analysis following trimming with TrimGalore yielded mixed results. Appendix2, Figure2 shows the read counts before and after trimming for sample ID CI619. Although per-tile sequence quality remained a "warning" for one isolate (CI926), some quality metrics for specific isolates fell short of optimal levels. Notably, per-base sequence quality for isolates CI589, CI747, CI828, and CI1035 remained in the 'poor' category. Similarly, GC content remained 'passable' for all isolates except CI1539, which received a 'poor'quality. Sequence length distribution also shifted to 'passable' from the previously observed 'good' for all isolates. Finally, sequence duplication levels exhibited variability, with CI926, CI268, CI1539, and CI1035 displaying 'low' quality. The average reduction in read count after trimming with TrimGalore was 667 reads and the total read counts before and after trimming is given in Appendix 2, Table 1.

MultiQC generated summary heatmaps for both pipelines (in-house and BV-BRC) FastQC output, provided in Appendix 2, Figure 3 and Figure 4 respectively.

Following quality control, trimmed FastQ files were assembled using the Unicycler assembler. For the in-house pipeline, QUAST was employed to evaluate the quality of the assembly and its coverage against the reference genome. These results were then compared with the genome assembly results obtained from the BV-BRC pipeline, which also utilized Unicycler. Table 1 summarizes the minimum, maximum, and median values of L50, N50, and the number of contigs for the in-house and BV-BRC pipelines for *Citrobacter freundii* and *Citrobacter koseri* isolates. Due to limited data, *Citrobacter braakii* was excluded from this analysis. The full QUAST report for both pipelines, including the number of contigs, L50, N50, and GC content, is provided in Appendix 2, Table 2. The Wilcoxon signed-rank test was employed to evaluate statistically significant differences (p-value) between the two pipelines for each assembly statistic and bacterial species. Table 1 shows that a statistically significant difference (p-value = 0.02) was observed only in the number of contigs between the in-house pipeline and the BV-BRC pipeline for the *Citrobacter koseri* group. No significant differences ($p > 0.05$) were observed in the number of contigs for the *Citrobacter freundii* group, or in the L50 and N50 values for either isolate.

Table 1 | This table compares the QUAST results for two bioinformatic pipelines (In-house and BV-BRC) of *Citrobacter koseri* and *Citrobacter freundii* isolates WGS data associated with sepsis.

		In-house pipeline			**BV-BRC pipeline**			
		Min	**Max**	**Median**	**Min**	**Max**	**Median**	**p-value**
Citrobacter koseri	L50	1	5	3	1	6	3	0.85
	N50	281203	2637237	683829	281199	2637237	687363	0.34
	No. of contigs	15	40	21	16	46	21	0.02
Citrobacter freundii	L50	4	11	5	4	11	5	1.00
	N50	144033	547363	335764	194315	547363	335771	0.18
	No.of contigs	35	105	38	30	112	41	0.49

The table shows the minimum, maximum, and median values for various metrics (L50, N50 and number of contigs) obtained from the in-house pipeline and BV-BRC pipeline, along with the p-value calculated using the Wilcoxon signed-rank test to analyze statistically significant differences between the pipelines. $p < 0.05$ was considered to be significant.
L50: The minimum number of contigs that cover half (50%) of the assembly.
N50: The length of the smallest contigs that together represent half (50%) of the assembly.

The comparison results of Unicycler and SPAdes done in BV-BRC genome assembly are given in Table 2. The average was used to compare Unicycler and SPAdes because it reflects the consistent performance across all isolates without extreme variations. Unicycler generated a significantly lower average number of contigs (36.28) compared to SPAdes (48.85). Additionally, the average N50 value, indicating contig length, was substantially higher for Unicycler (788,958.09) compared to SPAdes (347,191.33). Notably, GC content remained similar between both assemblers. Statistical analysis using the Mann-Whitney U test revealed significant differences (p-value 0.01) in the number of contigs generated by SPAdes and Unicycler, with SPAdes producing

a considerably higher number. Similarly, both N50 (p-value 0.0001) and L50 (p-value 0.002) values differed significantly between the assemblers, suggesting distinct contig length distributions. These findings indicate that SPAdes tends to favour a more fragmented assembly with a higher number of contigs, while Unicycler might prioritize longer contigs. The Bandage (Bioinformatics Application for Navigating *de novo* Assembly Graphs Easily) plot generated with the genome assembly report for both Unicycler and SPAdes is given in Appendix 2, Figure 4. It visualizes the assembly graph and the difference between the two assemblers in terms of both nodes and edges.

Table 2 | Comparison of Unicycler and SPAdes from BV-BRC genome assembly with average, minimum and maximum for the number of contigs, L50, N50 and GC%.

Assembler	Parameters	Average	Minimum	Maximum
SPAdes	L50	5.66	3	13
	N50	347191.33	136429	499524
	Number of contigs	48.85	27	144
	GC%	53.08	51.48	56.23
Unicycler	L50	3.76	1	11
	N50	788958.09	194315	2637237
	Number of contigs	36.28	16	112
	GC%	53.09	51.48	56.22

L50: The minimum number of contigs that cover half (50%) of the assembly.
N50: The length of the smallest contigs that together represent half (50%) of the assembly.
GC%: The percentage of Guanine (G) and Cytosine (C) nucleotides.

Species identification

The JSpeciesWS for the inhouse pipeline and Kraken 2 output from BV-BRC result was similar to the phenotypic species identification result of MALDI-TOF MS explained in Appendix 2, Table 3. Still, the average nucleotide identity (ANI%) of three of the isolates was below the threshold of 95% (sample ID, CI589-92.13%, CI839-92.11% and CI1869-93.12%). The Krona chart % generated with Taxonomic classification of BV-BRC by Kraken2 output revealed 12 samples with less species similarity shown in Appendix 2, Table 3. The QUAST output for these isolates (sample ID CI 589 and CI 839) with genome fractions of 14.696% and 19.003% and *Citrobacter farmeri* (sample ID CI 1869) with genome fraction of 26.633% was also found to be too low. These were excluded from the rest of the analysis due to less genome similarity with the reference genome. The sample ID CI 1035 initially identified as *Enterobacter bugandensis* showed 93.217% similarity with the *Enterobacter bugandensis* reference genome and was also excluded from further analysis to maintain the consistency of the data.

Genotypic analysis

Antibiotic resistance

This study investigated the prevalence of antibiotic-resistance genes in 17 *Citrobacter* isolates using ResFinder at the Centre for Genomic Epidemiology. All *Citrobacter* isolates (100%) harboured beta-lactam resistance genes as identified by the tool (Appendix 2, Table 4). *Citrobacter braakii* isolates were found to have two distinct antibiotic resistance genes, belonging to the beta-lactam and quinolone classes, respectively. Conversely, *Citrobacter freundii* exhibited a broader spectrum of resistance, encompassing seven different antibiotic classes: aminocyclitol, aminoglycoside, beta-lactam, fosfomycin, sulfonamide, tetracycline, and trimethoprim. The beta-lactam resistance genes comprised half (50%) of the total identified

resistance genes in *Citrobacter freundii.* The isolates of *Citrobacter koseri* exhibited the highest number of resistance genes (15), of which 100% conferred resistance to beta-lactam antibiotics and 27.27% associated with resistance to fosfomycin. The aantibiotic resistance genes profiles of *Citrobacter* isolates are given in Appendix 2, Table 5. AST results explained in Table 3 showed a 3.6% discrepancy with those predicted by ResFinder, primarily in the aminoglycoside and quinolone antibiotic classes. For instance, AST identified *Citrobacter braakii* as susceptible to quinolones and chloramphenicol, but resistant to aminoglycosides also. Furthermore, all four *Citrobacter freundii* isolates and five *Citrobacter koseri* isolates tested resistant to nitrofurantoin in AST. One isolate of *Citrobacter koseri* even exhibited additional macrolide resistance according to AST. The predicted genotypic antibiotic resistance genes in *Citrobacter* spp. by ResFinder were compared with phenotypic AST and the results are presented in Table 3. Since no AST result was available for four of the isolates (sample ID-CI268, CI619, CI628 and CI926), they were not included in the comparison study.

Table 3 | Comparisons of phenotypic AST and genotypic antibiotic susceptibility by ResFinder.

Antibiotics	**Phenotypic AST and predicted antibiotic resistance by ResFinder**				**Discordant across methods**
	RR	**SS**	**RS**	**SR**	**[*n* (%)]**
Aminoglycoside (*n* =7)	0	6	1	0	1(14.2)
β-lactam (*n* = 13)	13	0	0	0	0(0)
Quinolone (*n* = 13)	0	12	0	1	1(7.6)
Sulfonamide (*n* = 12)	1	11	0	0	0(0)
Trimethoprim (*n* = 11)	1	10	0	0	0(0)
Total (% of 56)	15(26.8)	39(69.6)	1(1.8)	1(1.8)	2(3.6)

AST, antibiotic susceptibility testing; RR resistant by both phenotypic AST and ResFinder; SS susceptible by both phenotypic AST and ResFinder; RS resistant by phenotypic AST and susceptible by ResFinder; SR susceptible by phenotypic AST and resistant by ResFinder.

The genome annotation service of BV-BRC predicts antibiotic resistance genes using the CARD database. A total of 705 genes were identified above 90% identity which harbored a complex combination of genes originating from multiple gene families (Appendix 2, Table 6). These genes belong to categories like mutations in resistance genes, and genes encoding target proteins for specific antibiotics like fosfomycin, rifampin, aminocoumarins, aminoglycosides, sulfonamides, and tetracyclines. Target protection proteins, efflux pumps, and genes influencing antibiotic transport revealed the diverse mechanisms employed by *Citrobacter* spp. to resist antibiotics. The distribution of genes among 17 *Citrobacter* isolates is given in Appendix 2, Table 6. For *Citrobacter braakii*, there were 38 genes (5.4%), *Citrobacter freundii* revealed 223 genes (31.6%) and *Citrobacter koseri* represented a variety of 444 genes (63%). Discrepancies were observed in quinolone resistance gene (*QnrB10)* identification between ResFinder and CARD databases. While ResFinder identified quinolone genes only in *Citrobacter brakki* isolates, the CARD database identified quinolone resistance genes named for all isolates of *Citrobacter freundii* and *Citrobacter brakki* isolate. Analysis of beta-lactam resistance genes using both

databases revealed identical results including *CMY-48, CMY-68, CMY-75, CMY-79,* and *CMY-82* for *Citrobacter freundii* and *Citrobacter brakki* isolates.

Virulence factor genes

The analysis of virulence factor genes for *Citrobacter* isolates from the VFDB database identified a total of 300 genes in the BV-BRC pipeline (Appendix 2, Table 7). For *Citrobacter freundii* (n=5), 21 genes, *Citrobacter koseri* (n=11), 267 genes and *Citrobacter braakii* (n=1), 12 genes. *Citrobacter freundii* displayed only 5 types of genes associated with adhesion (*csgG* and *csgE*), iron uptake (*entE* and *fepC*), and serum resistance (*ompA*). *Citrobacter braakii* harboured a unique set of eight genes (*tviE, tviD, tviC, VexB, vexC, vexD, vexA, vexE*) potentially linked to immune evasion along with adherence and iron uptake genes. *Citrobacter koseri* exhibited the broadest virulence potential possessing genes for enterotoxins (*senB*) in all sample types, endotoxins (*grtB*) in urine culture and blood culture, and secretion system (*flhC*) in urine culture only. The genes for iron uptake (*ybtE, fyuA, ybtQ, irp2, irp1, ybtA, ybtP, fyuA, ybtU, ybtS, ybtX, ybtT, ybtE, iucC, iucB, fepC, fepD, entA, iutA, entE* and *entB*) and adherence (*csgG, csgE, csgF* and *fimD*) were distributed in all sample types.

Plasmid replicons

PlasmidFinder at the Centre for Genomic Epidemiology identified a total of 13 distinct plasmids, which include 9 unique plasmid types given in Appendix 2, Table 8. *Citrobacter freundii* hosted 4 different plasmids and *Citrobacter koseri* identified 5 different plasmids. Plasmid *ColRNAI* was repeated in 6 isolates of *Citrobacter*. A total of 13 distinct plasmids were found distributed among these isolates, Figure 4. PlasmidFinder identified plasmids in the *Citrobacter* spp. isolates, with some showing homology to plasmids commonly found in *Klebsiella, Salmonella, Yersinia pestis, Citrobacter rodentium* and *Citrobacter freundii*. Plasmids were identified in isolates from all sample types, including wound (n=2), blood (n=2), upper respiratory tract (n=2), and urine cultures (n=7). For the BV-BRC pipeline, no plasmids were identified for any of the isolates. The Mann-Whitney U test was conducted to compare the distribution of plasmid counts between two species types *Citrobacter freundii* and *Citrobacter koseri*. The test yielded a W-statistic of 27.5 and a corresponding p-value of 0.31. The Kruskal-Walli's rank sum test was conducted to evaluate potential variations in the number of plasmids across different sample types (blood culture and urine culture). Since wound culture and upper respiratory culture are available for only one isolate each, no test was performed. The analysis revealed Kruskal-Walli's chi-squared value of 0.32 with 1 degree of freedom, yielding a p-value of 0.56.

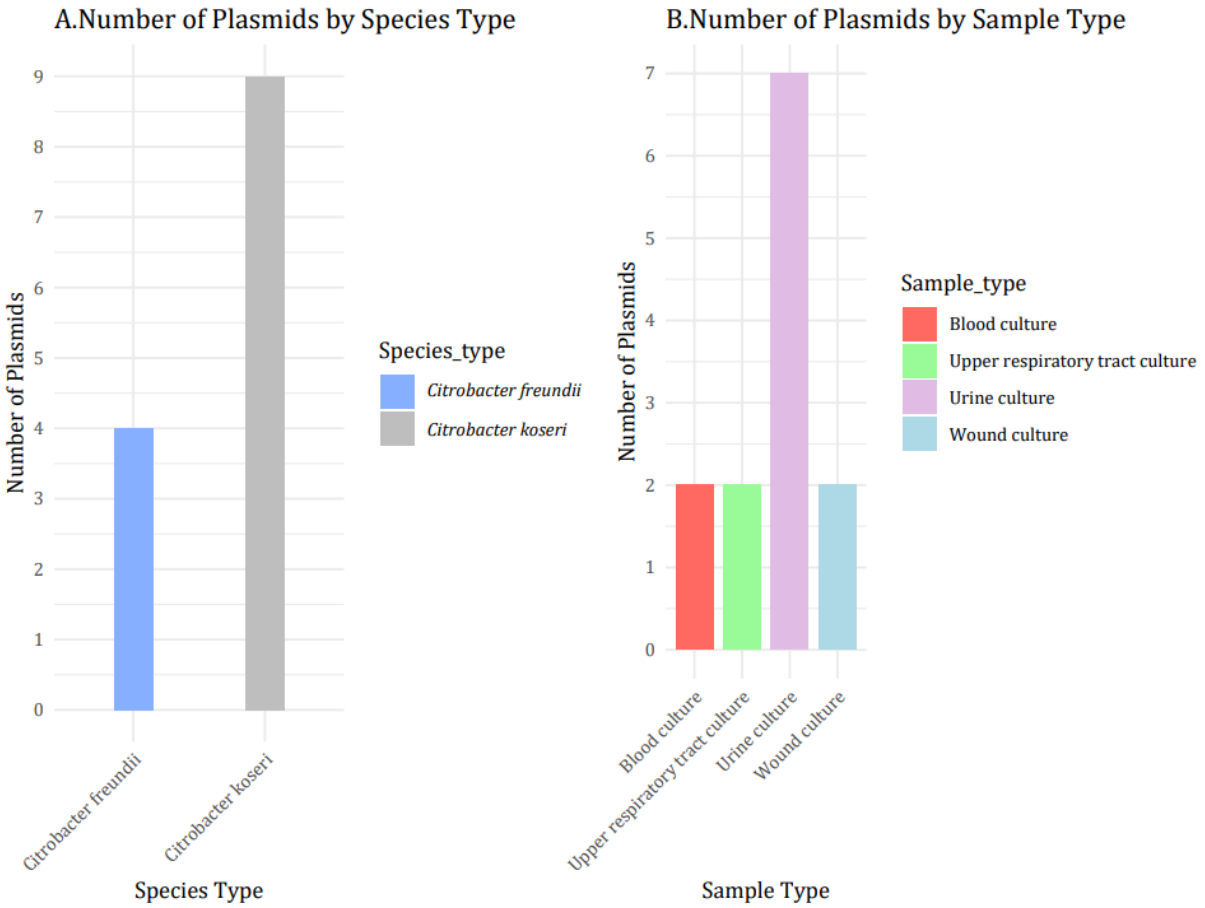

Figure 4 | Distribution of Plasmid Number in *Citrobacter* isolates. This figure depicts the distribution of plasmid numbers within a collection of *Citrobacter* isolates (n=16). Plot A explores the relationship between plasmid content and species type. It compares *Citrobacter freundii* (n=5) and *Citrobacter koseri* (n=9) isolates, with each bar representing the total number of plasmids observed within each species group. Plot B investigates the association between plasmid number and sample origin. The data is categorised by sample type, including urine cultures (n=7), wound cultures (n=2), upper respiratory tract cultures (n=2), and blood cultures (n=2). The figure was generated using RStudio version 4.3.3 with the ggplot2 package.

Multi-locus sequence typing

The Multi-Locus Sequence Typing (MLST) scheme available through the PubMLST database was employed to characterize the genetic diversity of the 17 *Citrobacter* isolates obtained via the in-house pipeline. Analysis of the sequenced housekeeping genes using the PubMLST platform revealed the presence of eight distinct sequence types (STs) among the *Citrobacter* isolates. Table 4 illustrates the distribution of sequence types among the *Citrobacter* isolates and the presence of plasmids and associated resistance genes and virulence genes. Among *Citrobacter freundii* isolates, four isolates exhibited diverse STs (ST22, ST112, ST114, ST725), while *Citrobacter brakki* had ST567. Four STs (ST854, ST862, ST114, and ST1149) were found in blood, urine, and wound cultures of *Citrobacter koseri*. Interestingly, ST114 was identified in both C*itrobacter koseri* and *Citrobacter freundii* isolates. In contrast, ST854 and ST862 were culture type-specific, detected across various clinical samples such as wound, urine, and blood cultures. No sequence types were observed in upper respiratory tract culture. The BV-BRC pipeline found one ST

(ST22), for isolate with sample 1D CI935, which is consistent with the findings of the in-house pipeline.

Table 4 | Summary of MLST data, plasmids, resistance genes and virulence genes of *Citrobacter* isolates in the study.

Sample ID	Type of Species	Type of sample culture	MLST #1	No. of Plasmids ≥ 95%	No. of resistance genes ≥90%	No. of virulence genes≥90%
CI 1674	*Citrobacter braakii*	Blood	ST 567	0	2	12
CI 11	*Citrobacter freundii*	Urine	ST114	1	1	5
CI 926	*Citrobacter freundii*	Urine	ST725	0	1	5
CI 935	*Citrobacter freundii*	Urine	ST22	0	5	4
CI 1480	*Citrobacter freundii*	Urine	ST112	3	1	3
CI 1578	*Citrobacter freundii*	Urine	UK	0	3	4
CI 619	*Citrobacter koseri*	Blood	ST854	1	2	23
CI 1631	*Citrobacter koseri*	Blood	ST862	1	1	27
CI 1539	*Citrobacter koseri*	Upper respiratory	UK	2	1	27
CI 268	*Citrobacter koseri*	Urine	ST1149	0	1	22
CI 507	*Citrobacter koseri*	Urine	ST114	0	1	23
CI 628	*Citrobacter koseri*	Urine	ST854	1	2	23
CI 828	*Citrobacter koseri*	Urine	ST854	0	2	23
CI 1475	*Citrobacter koseri*	Urine	UK	0	1	25
CI 1626	*Citrobacter koseri*	Urine	ST862	1	1	27
CI 1936	*Citrobacter koseri*	Urine	ST1149	1	1	24
CI 747	*Citrobacter koseri*	Wound	ST854	2	2	23

The table details 17 *Citrobacter* isolates investigated for the presence of plasmids and associated resistance and virulence genes. Plasmid identification was performed using PlasmidFinder, while resistance genes were detected using ResFinder, both tools from the Center for Genomic Epidemiology (CGE). Additionally, virulence genes were determined through the Virulence Factor Database (VFDB). Multilocus sequence typing (MLST) was employed to characterize the isolates based on the *Citrobacter* MLST scheme available at http://pubmlst.org/cfreundii/.

Threshold for sequence identity for plasmids ≥ 95%, for resistance genes and virulence genes ≥ 90%

CI- *Citrobacter* isolate; MLST- multi-locus sequence typing; ST- Sequence type; UK- unknown

Discussion

Sequencing technologies have revolutionized the field of genomics, enabling researchers to generate vast amounts of data for analysis. These massive datasets offer unprecedented opportunities for scientific discoveries and advancements in our understanding of biological function (Satam et al., 2023). This study employed WGS to genotypically characterise *Citrobacter* isolates obtained from patients with suspected community-onset sepsis. To elucidate the diversity and attributes of these isolates, both an in-house developed bioinformatics pipeline and the BV-BRC pipeline were utilized. The analysis of raw FastQ files using FastQC (Andrews, 2010b; Guo et al., 2014) is an essential step in assessing the quality and integrity of sequencing data before downstream analysis. It allows researchers to identify potential issues or biases in the data, and make informed decisions on data filtering, trimming, and quality control (Guo et al., 2014). Evaluating various metrics such as per base sequence quality, per base sequence content, GC content, and adapter presence is crucial for identifying potential issues. Initial quality assessment of raw FastQ files in this study revealed generally high-quality metrics for both the in-house pipeline and the BV-BRC pipeline. However, sequence duplication levels and adapter contamination emerged as the primary areas of concern. The in-house pipeline employed Trimmomatic (Bolger et al., 2014), which effectively addressed the presence of the adapter as evidenced by improved FastQC scores after adapter trimming. It employs a sliding window approach for quality trimming, enabling researchers to remove low-quality bases within a defined window size (Bolger et al., 2014). When comparing Trimmomatic with TrimGalore, the key difference lies in read trimming. Trimmomatic removes more read bases leading to improved data quality. In contrast, TrimGalore is gentler, resulting in fewer trimmed reads but potentially retaining more biological information.

Analysis of sequence duplication levels revealed discrepancies between the in-house pipeline and the BV-BRC pipeline. Interestingly, the sequence length distribution improved for all samples, transitioning to passable quality after trimming. Conversely, the BV-BRC pipeline, which used TrimGalore (Krueger, 2012), failed to improve sequence duplication levels. TrimGalore prioritizes user-friendliness for Illumina data analysis. It uses Cutadapt (Martin, 2011) for efficient adapter removal, particularly for Nextera adapters, with automatic detection. While offering basic quality trimming with simplified parameters, it lacks flexibility in handling adapters beyond Nextera. However, TrimGalore efficiently utilizes multiple cores for faster processing of large datasets (Krueger, 2012). The observed variability in sequence duplication level could stem from several factors. In terms of per sequence GC content, both Trimmomatic and TrimGalore were able to maintain passable quality, except for isolate (CI1631) which showed a drop to poor quality after trimming with Timmomatic. This decline may be attributed to the sliding window method employed by Trimmomatic, which can inadvertently trim high-quality reads containing low-quality GC-rich regions, leading to a potential bias in GC content (Williams et al., 2016). Trimmomatic is known for its comprehensive suite of functionalities, including advanced quality trimming options like leading/trailing base removal and customizable sliding window parameters, allowing for precise quality control tailored to the dataset (Bolger et al., 2014). In contrast to TrimGalore, Trimmomatic's adaptability stands out. DNA-Seq data trimming requires careful selection of trimmers, adjustment of parameters, and consideration of the dataset's overall quality distribution to ensure optimal data integrity and analytical outcomes (Del Fabbro et al., 2013).

Genome assembly, a critical step in genomic research, involves reconstructing the complete genome sequence from a collection of sequenced DNA fragments (Nagarajan & Pop, 2013). Computational algorithms and tools are employed that align and merge these overlapping fragments, ultimately generating a contiguous sequence representing the entire genome (Wajid & Serpedin, 2016). De Bruijn graph-based assembly is a popular and efficient method, which simplifies the process by breaking down reads into shorter sequences called k-mers. These k-mers form nodes in a graph, with edges connecting them based on single-base overlaps, thus avoiding the computationally intensive task of aligning long reads directly to each other (Medvedev et al., 2011). However, the choice of assembly method can significantly impact the accuracy and quality of the final assembly, as various tools are available. Encouragingly, the application of Unicycler (Wick et al., 2017) within both the in-house developed pipeline and the BV-BRC pipeline yielded highly comparable genome assembly results for the *Citrobacter* isolates. Reinforcing the findings from the study (Chen et al., 2020), using both the in-house and BV-BRC pipelines demonstrates the consistent performance of Unicycler in assembling *Citrobacter* genomes. This consistency across diverse bioinformatic workflows suggests Unicycler's robustness as a reliable assembly tool for this bacterial genus.

To evaluate the performance of different assembly algorithms within the BV-BRC pipeline, Unicycler and SPAdes were employed for the *de novo* assembly of the *Citrobacter* isolates. The QUAST results of assembly done by Unicycler, comparing the in-house pipeline and the BV-BRC pipeline for *Citrobacter koseri* and *Citrobacter freundii* isolates provide insights into the quality of genome assembly achieved by each pipeline. For *Citrobacter koseri,* the in-house pipeline had a minimum L50 of 1, a maximum of 5, and a median of 3. The BV-BRC pipeline had a minimum L50 of 1, a maximum of 6, and a median of 3. The p-value for this comparison was 0.85, indicating no significant difference between the two pipelines in terms of contig lengths. The N50 values for both pipelines were similar, with no significant difference observed (p-value of 0.34). The number of contigs ranged from 15 to 40 for the in-house pipeline and from 16 to 46 for the BV-BRC pipeline. The median number of contigs was 21 for both pipelines. The p-value for this comparison was 0.02, suggesting a significant difference in the number of contigs between the two assemblers. For *Citrobacter freundii*, both pipelines showed similar results for L50, with no significant difference observed (p-value of 1.00). The N50 values for both pipelines were also comparable, with no significant difference (p-value of 0.18). The number of contigs ranged from 35 to 105 for the in-house pipeline and 30 to 112 for the BV-BRC pipeline. The median number of contigs was 38 for the in-house pipeline and 41 for the BV-BRC pipeline. The p-value for this comparison was 0.49, indicating no significant difference in the number of contigs between the two assemblers. In summary, the QUAST results suggest that Unicycler produced assemblies with a lower number of contigs and a higher N50, indicating a more contiguous assembly with fewer fragmented sequences. This finding aligns with observations from a previous study comparing these assemblers (Dida & Yi, 2021). Unicycler outperforms SPAdes in assembling bacterial genomes when compared to the BV-BRC pipeline. This is evident by its significantly higher N50 value, indicating longer and more contiguous assemblies. SPAdes, on the other hand, produces a more fragmented assembly with a higher number of shorter contigs. The study utilized short-read sequencing data for the assembly of bacterial genomes. Despite this, Unicycler demonstrated superior performance over SPAdes, which can be attributed to its hybrid assembly algorithm that is optimized for short-read data. Unicycler's algorithm is designed to create more contiguous assemblies even when only short-read data is available, as it employs a combination of de Bruijn graph and overlap-layout-consensus methods. This allows for better handling of repetitive

regions and results in longer contigs with higher N50 values (Wick et al., 2017). These findings align with prior research emphasizing Unicycler's advantages in contiguity and completeness (Chen et al., 2020b). Consequently, Unicycler emerges as the preferred approach for analyzing bacterial pathogens using combined sequencing data (Wick et al., 2017). Its ability to generate contiguous assemblies and handle even low-quality long reads makes it a valuable tool for bioinformatics research.

Accurate taxonomic classification and evaluation of genomic similarity are crucial for *Citrobacter* isolates. To achieve this, assembled genomes were aligned with reference genomes. The QUAST (Gurevich et al., 2013) genome fraction metric can be effectively integrated with results from other species identification pipelines for a more robust approach. MALDI-TOF MS served as the initial reference for the species designation of all isolates. While JSpeciesWS, Kraken2, and MALDI-TOF MS generally showed concordance in species identification for most isolates, some discrepancies emerged. Three isolates (CI589, CI839, CI1869) exhibited an average nucleotide identity (ANIb) below 95% compared to the reference genome using JSpeciesWS. Additionally, these isolates had low QUAST genome fractions suggesting a poor fit with the reference. Furthermore, Kraken2 Krona chart data for some isolates hinted at lower species similarity. Due to these inconsistencies, these three isolates were excluded from further analysis. This ensures data quality and allows to focus on confidently identified *Citrobacter* species. Similarly, the exclusion of CI1035 (identified as *Enterobacter bugandensis*) was justified because it was not a *Citrobacter* species according to all three identification methods (MALDI-TOF MS, JSpeciesWS, and Kraken2).

The genotypic analysis using ResFinder revealed resistance genes associated with various antibiotic classes, including beta-lactams, aminoglycosides, quinolones, fosfomycin, sulphonamides, tetracycline, and trimethoprim, in *Citrobacter* isolates. The AST result by MALDI-TOF MS resulted in beta-lactamase, lincosamide, macrolide, nitrofurantoin, sulphonamide, and trimethoprim resistance, also susceptible to quinolone, aminoglycoside classes. This study further employed the CARD database to identify a diverse array of antibiotic-resistance genes (ARGs). A comparison of these methods for *Citrobacter freundii* isolates showed high concordance for resistance determinants related to beta-lactam, tetracycline, aminoglycoside, sulphanilamide, aminocyclitol, and trimethoprim. However, a notable discrepancy emerged regarding the *qnrB10* gene, a quinolone resistance determinant. ResFinder only identified *qnrB10* in *Citrobacter braakii* isolate, whereas CARD predicted its presence along with genes for aminocoumarin, fluoroquinolone, fosfomycin, rifampin, nitrofurantoin, and streptothricin resistance in *Citrobacter freundii.* This highlights potential limitations in either detection method. Consistent with previous findings (Delgado et al., 2013; Liu et al., 2020; Yuan et al., 2019c; Zhou et al., 2019), the prevalence of quinolone resistance and associated mutations varied among *Citrobacter* species. Notably, *Citrobacter braakii* exhibited the highest proportion of quinolone-resistant isolates, followed by *Citrobacter freundii*. Conversely, isolates across all species displayed resistance to beta-lactam antibiotics, underlining the prevalence of beta-lactam resistance genes (*blaCMY-48, blaCMY-68, blaCMY-75, blaCMY-79, blaCMY-82, blaMAL-1, blaCKO-1*). Building on prior in silico analyses suggesting reduced susceptibility of *Citrobacter freundii* to antibiotics (Ramsamy et al., 2020; Yuan et al., 2019) this study employed whole-genome sequencing to identify a greater diversity of resistance genes (*blaCMY, fosA sul2, dfrA1, tetD, aadA* and *aac(6)*) in *Citrobacter freundii* isolates compared to *Citrobacter koseri.* This finding underscores the potential for *Citrobacter freundii* to pose a more significant clinical challenge in the face of antibiotic therapy.

CARD provides detailed annotation results for each resistance gene based on their molecular mechanisms and consolidates information from various sources and genome databases. The protein homolog models detect antimicrobial resistance protein sequences present in CARD based on BLASTP hits to curated protein sequences. This study aligns with previous findings that reported that multidrug efflux pumps, mutations in DNA gyrase subunits A and B, contribute to high-level fluoroquinolone resistance in *Enterobacterales,* including *Citrobacter freundii* (Frenk et al., 2021). The identification of these diverse ARGs in *Citrobacter* isolates highlights the growing challenge of antibiotic resistance in this pathogen. When comparing ResFinder-predicted genes to genes identified by CARD, this discrepancy could be due to several factors. The emergence of new resistance genes and the spread of existing ones through horizontal gene transfer are ongoing processes. Additionally, the distribution of resistance genes varied among the *Citrobacter* species, with *Citrobacter freundii* exhibiting a broader spectrum of resistance genes. This study characterized several critical ARGs in *Citrobacter freundii* isolates, including *sul2, aac(6), dfrA1, tetD*, and *aadA1* along with beta-lactam and fosfomycin resistance genes. These genes confer resistance to specific antibiotic classes, potentially compromising therapeutic options for *Citrobacter freundii* infections (Dolejska et al., 2013). The *aadA1* gene encodes an enzyme that inactivates both streptomycin and spectinomycin (*aadA1* gene). The *dfrA1* gene is the most common cause of trimethoprim resistance and is often found alongside genes for resistance to other important antibiotics (Jabeen et al., 2023). This co-resistance with carbapenems and extended-spectrum beta-lactamases in CRE strains creates a critical need for new antibiotics that can target these multidrug-resistant bacteria (Lombardo et al., 2016). The prevalence of sul2 in other Gram-negative bacteria such as *Salmonella* and *Escherichia coli* (Kerrn, 2002) suggests it is a common resistance mechanism. Horizontal gene transfer through plasmids allows bacteria like *Citrobacter freundii* to acquire *sul2* from other bacteria in the environment (Antunes et al., 2005).

A particularly concerning finding involved *Citrobacter koseri* isolates. These isolates displayed not only the highest overall number of resistance genes but also harboured genes associated with resistance to fosfomycin, a last-resort antibiotic for treating serious infections (Sheu et al., 2019). The presence of *ColRNAI, IncCX4, Col44011,* and *IncX10* plasmids in these *Citrobacter koseri* strains aligns with growing concerns about the evolution and spread of multidrug resistance. *Citrobacter rodentium* exhibits a restricted host range, primarily infecting mice. This murine pathogen serves as a valuable in vivo model for attaching and effacing A/E lesion-forming enteric pathogens, including human and animal-associated strains of *Escherichia coli* (Mundy et al., 2005 ; Jacoby et al., 2002). This study investigates *Citrobacter koseri* isolates harbouring a plasmid (*InCX4*) potentially related to *Citrobacter rodentium.* The presence of this plasmid in *Citrobacter koseri,* a species occasionally associated with human infections, warrants further investigation into the potential for zoonotic transmission of A/E lesion-forming mechanisms. Previous studies, such as the one by (Galindo et al., 2022), have not necessarily reported widespread fosfomycin resistance in *Citrobacter koseri*, this study's findings suggest a potential shift in resistance profiles. This underscores the importance of continuous monitoring of antibiotic resistance patterns to inform effective treatment strategies. Another concern emerged from a *Citrobacter freundii* isolate carrying multiple *IncF* plasmids (*IncFIBK* and *IncHIB*(*pNDM-CIT*)). *IncF* plasmids are major contributors to antibiotic resistance in *Enterobacteriaceae,* but current typing methods for these plasmids need improvement (Villa et al., 2010). These methods struggle to differentiate between various *IncF* plasmids and handle situations where a single cell harbour multiple *IncF* plasmids. This limitation makes it challenging to fully characterize the antibiotic resistance potential

conferred by these plasmids in *Citrobacter freundii.* This study also identified the *Yersinia pestis* plasmid *(IncFIIYp)* in *Citrobacter freundii* isolates. This finding contrasts with observations in another study by (Yuan et al., 2019d), where *Citrobacter koseri* exhibited factors associated with flagellar movement and iron uptake, including a High Pathogenicity Island (HPI) like those found in highly pathogenic *Yersinia* strains. These contrasting findings highlight the potential for significant heterogeneity in virulence mechanisms within *Citrobacter* species. Further research is necessary to elucidate the specific virulence factors encoded by the *Yersinia pestis* plasmid in *Citrobacter freundii* and assess their contribution to its pathogenicity.

The potential interplay between virulence and antibiotic resistance is not well-understood for *Citrobacter* species. Some theories suggest that increased virulence can sometimes lead to enhanced bacterial survival and persistence within a host, potentially creating a selective pressure that favours the emergence of antibiotic-resistance mutations (Bakkeren et al., 2020). The detection of specific virulence factor genes (*senB, gtrB,* and *flhC*) in *Citrobacter koseri* isolates, and invasion serum resistance genes (*ompA*) in *Citrobacter braakii* suggests their potential role in the pathogenesis of *Citrobacter* infections. *Citrobacter* species with low virulence can persist in hosts for a long time, potentially acquiring and accumulating resistance genes. These low-virulence bacteria act as a potential reservoir for resistance genes. Their persistence within immunocompromised patients allows them to accumulate and share these resistance genes with more virulent bacteria, potentially leading to outbreaks of highly resistant pathogens (Pepperell et al., 2002). To summarize, discrepancies between genotypic predictions and phenotypic AST results in terms of resistance/susceptibility calls can arise from the complex interplay of genetic, regulatory, methodological, and environmental factors that influence antibiotic resistance in bacterial populations. Integrating both genotypic and phenotypic data, along with clinical context, is essential for accurate antibiotic susceptibility testing and informed treatment decisions.

The PubMLST analysis of 17 *Citrobacter* isolates revealed eight distinct sequence types (STs), including three novel ones. While this study offers valuable insights into ST diversity within this limited sample set, a recent large-scale PubMLST analysis (Osei Sekyere & Reta, 2021) reported a significantly higher number of STs (84) within *Citrobacter freundii* alone, highlighting the extensive genetic diversity within this species. Interestingly, some prevalent STs observed in their study, such as ST114 and ST22 reported in Australia, were also identified in *Citrobacter freundii* isolates. This suggests potential overlap in circulating STs of *Citrobacter freundii*, even across geographically distinct locations. According to global PubMLST submissions, ST567, previously found in Sweden, was also identified in a *Citrobacter braakii* isolate of this study. Similarly, *Citrobacter koseri* with ST1149, previously identified in the USA (Jolley et al., 2018c), was present in two of the *Citrobacter koseri* isolates. *Citrobacter koseri* isolates with diverse STs (ST854 and ST862) may represent novel STs due to the lack of available references. The same applies to ST725 and ST112 of *Citrobacter freundii* isolates. Identifying these potentially novel STs would contribute to a more comprehensive understanding of their global distribution and potential virulence diversity. Among the isolates, *Citrobacter koseri* and *Citrobacter freundii* displayed the most diverse ST profile with four different STs each, while *Citrobacter braakii* possessed a single unique ST. Interestingly, ST114 appeared in both *Citrobacter koseri* and *Citrobacter freundii* isolates. However, no ST was identified in upper respiratory tract cultures which warrants further investigation with a larger sample size to confirm this observation. Future studies comparing these STs with virulence profiles and antibiotic resistance patterns could provide deeper insights into the pathogenic potential and epidemiology of *Citrobacter* isolates.

In conclusion, existing literature often portrays *Citrobacter* bloodstream infections and *Citrobacter freundii*-induced sepsis in adults as infrequent occurrences (Chen et al., 2023). However, this study challenges the prevailing view of *Citrobacter* as a rare cause of sepsis by analyzing 17 *Citrobacter* isolates from adult sepsis patients. It aimed to comprehensively evaluate *Citrobacter's* role in sepsis, focusing on both clinical aspects and bioinformatic analysis. This study employed WGS analysis to achieve these. By comparing the performance of different assemblers within the BV-BRC pipeline, the research assessed their effectiveness in reconstructing *Citrobacter* genomes. This comparison not only aids in evaluating the strengths and weaknesses of different bioinformatic pipelines but also directly contributes to the identification of unique genomic signatures to differentiate *Citrobacter* species accurately**.** This differentiation is crucial for preventing misdiagnosis and unnecessary antibiotic use, as outlined in the study's goals. Ultimately, this research contributes valuable knowledge regarding the epidemiology, virulence factors, and clinical consequences of *Citrobacter*-mediated sepsis. By elucidating the pathogenic mechanisms and patient outcomes associated with *Citrobacter* infections, this research emphasizes the critical importance of early diagnosis and intervention strategies to reduce morbidity and mortality associated with this increasingly recognized pathogen.

Ethical aspects, gender perspectives, and impact on the society

Ethical principles are paramount for ensuring the integrity, credibility, and responsible conduct of research, especially within thesis work (National Institutes of Health, 2018). These principles encompass a wide range of crucial considerations, including obtaining informed consent from participants, upholding participant confidentiality, and maintaining research integrity through honesty and transparency. Researchers are obligated to minimize potential harm while maximizing the benefits for participants and society. Treating participants with respect and sensitivity is equally important. Disclosure of any potential conflicts of interest is essential, alongside adherence to established ethical guidelines and regulations governing research practices (National Institutes of Health, 2018). By meticulously addressing these ethical aspects, researchers not only uphold the ethical standards of their fields but also demonstrate their commitment to responsible and ethical research conduct. The European Union's research framework program, Horizon Europe, exemplifies this commitment. It emphasizes the inclusion of gender perspectives within research activities, fostering progress towards gender equality. Integrating gender perspectives into research is essential for promoting inclusivity and addressing potential gender bias (Horstmann et al., 2022). This necessitates a gender analysis that considers how gender norms, roles, and identities may influence research outcomes. Disaggregating data by gender and examining its intersection with other social factors like race and socioeconomic status are crucial steps in this process. A gender-sensitive approach helps uncover nuances and disparities, leading to more comprehensive and equitable research findings. By centring gender perspectives, research not only gains quality and relevance but also contributes to advancements in gender equality and social justice (Rai et al., 2020). The "Sepsis Skaraborg study" adhered to these ethical principles, receiving approval from The Regional Ethical Review Board in Gothenburg (376-11) (Ljungström et al., 2017). Since bacterial isolates, unlike human subjects, do not require informed consent. However, the patients involved in the original "Skaraborg sepsis study" did provide written informed consent, as detailed by Ljungström (Ljungström et al., 2017). Patient confidentiality was ensured by excluding personal data like names and identification numbers from this thesis. This study offers valuable insights into the genotypic characteristics of *Citrobacter* isolates. This information can significantly benefit society by strengthening surveillance of pathogenic strains, combating antibiotic resistance through the development of targeted therapies, and guiding clinical decision-making for *Citrobacter*-mediated sepsis infections. By providing valuable information for these areas, this study has the potential to improve public health outcomes and contribute to a more effective healthcare system.

Future perspectives

WGS presents a paradigm shift in the diagnosis and management of infectious diseases, with the potential to revolutionize sepsis treatment. As we overcome challenges and embrace continuous innovation, WGS has the power to transform the future of infectious disease control, leading to more effective therapies, improved patient outcomes, and a world better equipped to combat the ever-evolving threat of infectious agents. A survey study found a rapid increase in the use of WGS for tracking infectious diseases across Europe by 2016 (Revez et al., 2017). While challenges remain in data analysis and standardization, WGS holds promise for future pan-EU disease surveillance. Sequencing the entire genome of bacteria isolated from patients could be the best way to quickly track the spread of multi-drug resistant (MDR) strains within hospitals. Implementing WGS as a regular practice in clinical labs for MDR pathogens seems achievable and has the potential to guide specific actions for preventing and controlling hospital-acquired infections (Forde et al., 2023). Recent studies utilizing genome-wide association studies (GWAS) have identified novel genetic variants associated with decreased 28-day survival in adult sepsis patients. The study pinpointed three previously unknown variants, one of which alters protein structure. However, the collective effect of these variants on survival outcomes remains modest, highlighting the need for larger studies to comprehensively understand the genetic landscape of sepsis (Hernandez et al., 2022). This study delves into the WGS analysis of *Citrobacter* isolates obtained from sepsis patients and aimed to identify and characterize the antibiotic-resistance genes harboured by these bacteria, unveiling potential chinks in our current treatment armor. Additionally, exploring the virulence factors encoded within their genomes, and the malicious tools they wield to cause havoc in the human body. Furthermore, by determining the sequence types (STs) and plasmid replicons present, the *Citrobacter* isolates are differentiated and gain insights into their potential for spread within healthcare settings. By identifying plasmids potentially harboring virulence factors from *Citrobacter rodentium* and *Yersinia pestis* in human-associated *Citrobacter* species, this study opens avenues for exploring the possibility of *Citrobacter rodentium*-like pathogenesis in humans. By integrating these findings with existing knowledge on *Citrobacter* and sepsis, this study aspires to illuminate the intricate mechanisms of *Citrobacter*-mediated sepsis, paving the way for more effective diagnostic tools, targeted antibiotic therapies, and robust infection control strategies.

Acknowledgments

I extend my deepest gratitude to my supervisor, Magnus Fagerlind. His guidance, expertise, and unwavering support have been instrumental throughout my thesis journey. His mentorship and encouragement inspired me to push the boundaries of my research and strive for excellence. He fostered my independence and patiently offered suggestions that significantly improved my work.

I am grateful to my examiner, Björn Olsson, for his valuable suggestions and feedback during my study, which significantly refined this thesis. I also appreciate the expertise and attention to detail provided by Andreas Tilevik, Diana Tilevik, and Anna-Karin Pernestig. Their contributions undoubtedly shaped the quality of my work. I would like to express my sincere gratitude to Kristoffer Nilsson, for sharing his knowledge and experience. He provided me with the necessary tools and resources to undertake this research endeavour. Furthermore, I am thankful to John Baxter and Nada Mahmood for their continuous support and inspiration throughout my thesis and course.

A special thanks goes to Sanja Jurcevic, whose advice and support were instrumental in my academic success. Her encouragement empowered me to persevere through challenges and pursue my aspirations.

I extend heartfelt thanks to my colleagues, Trang and Tissy, for their continuous support, encouragement, and teamwork mentality. A special thank you to Udari for her invaluable support, love, and care throughout our entire program. Finally, I extend my love to Tabitha, Labiba, Jahir, Manuel, and Afzal for their encouragement and companionship, which enriched my life in countless ways. I am truly grateful to everyone who played a part, big or small, in the realisation of this thesis. Your contributions have made this journey both meaningful and rewarding.

I am profoundly grateful to my parents (Achaye and Ammachi) and in-laws (Daddy and Mummy) for their unwavering love, encouragement, and unwavering support throughout this journey.

Finally, I thank God and the boys of my life Ishan, Rishan, and Binish. Their unwavering belief in my abilities and constant encouragement were my rock. Their understanding and support provided the mental strength I needed to achieve this goal.

References

Andrews, S. (2010a). *FastQC: a quality control tool for high throughput sequence data. Available at. http://www.bioinformatics.babraham.ac.uk/projects/fastqc*.

Andrews, S. . (2010b). *FastQC: a quality control tool for high throughput sequence data. Available at. http://www.bioinformatics.babraham.ac.uk/projects/fastqc (Accessed February 27, 2018)*.

Antunes, P., Machado, J., Sousa, J. C., & Peixe, L. (2005). Dissemination of Sulfonamide Resistance Genes (sul1, sul2 , and sul3) in Portuguese Salmonella enterica Strains and Relation with Integrons. *Antimicrobial Agents and Chemotherapy*, *49*(2), 836–839. https://doi.org/10.1128/AAC.49.2.836-839.2005

Arora, J., Mendelson, A. A., & Fox-Robichaud, A. (2023). Sepsis: network pathophysiology and implications for early diagnosis. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *324*(5), R613–R624. https://doi.org/10.1152/ajpregu.00003.2023

Avatsingh, A. U., Sharma, S., Kour, S., Arora, Y., Sharma, S., Joshi, D., Chaudhary, P. P., Perveen, K., Kamal, M. A., & Singh, N. (2023). Prevalence of antibiotic-resistant Gram-negative bacteria having extended-spectrum β-lactamase phenotypes in polluted irrigation-purpose wastewaters from Indian agro-ecosystems. *Frontiers in Microbiology*, *14*. https://doi.org/10.3389/fmicb.2023.1227132

Bakkeren, E., Diard, M., & Hardt, W.-D. (2020). Evolutionary causes and consequences of bacterial antibiotic persistence. *Nature Reviews Microbiology*, *18*(9), 479–490. https://doi.org/10.1038/s41579-020-0378-z

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Bassetti, M., Peghin, M., & Pecori, D. (2016). The management of multidrug-resistant Enterobacteriaceae. *Current Opinion in Infectious Diseases*, *29*(6), 583–594. https://doi.org/10.1097/QCO.0000000000000314

Baur, D., Gladstone, B. P., Burkert, F., Carrara, E., Foschi, F., Döbele, S., & Tacconelli, E. (2017). Effect of antibiotic stewardship on the incidence of infection and colonisation with antibiotic-resistant bacteria and Clostridium difficile infection: a systematic review and meta-analysis. *The Lancet Infectious Diseases*, *17*(9), 990–1001. https://doi.org/10.1016/S1473-3099(17)30325-0

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bone, R. C., Balk, R. A., Cerra, F. B., Dellinger, R. P., Fein, A. M., Knaus, W. A., Schein, R. M. H., & Sibbald, W. J. (1992). Definitions for Sepsis and Organ Failure and Guidelines for the Use

of Innovative Therapies in Sepsis. *Chest*, *101*(6), 1644–1655. https://doi.org/10.1378/chest.101.6.1644

Borenshtein, D., & Schauer, D. B. (2006). The Genus Citrobacter. In *The Prokaryotes* (pp. 90–98). Springer New York. https://doi.org/10.1007/0-387-30746-X_5

Bortolaia, V., Kaas, R. S., Ruppe, E., Roberts, M. C., Schwarz, S., Cattoir, V., Philippon, A., Allesoe, R. L., Rebelo, A. R., Florensa, A. F., Fagelhauer, L., Chakraborty, T., Neumann, B., Werner, G., Bender, J. K., Stingl, K., Nguyen, M., Coppens, J., Xavier, B. B., … Aarestrup, F. M. (2020). ResFinder 4.0 for predictions of phenotypes from genotypes. *Journal of Antimicrobial Chemotherapy*, *75*(12), 3491–3500. https://doi.org/10.1093/jac/dkaa345

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., Olson, R., Overbeek, R., Parrello, B., Pusch, G. D., Shukla, M., Thomason, J. A., Stevens, R., Vonstein, V., Wattam, A. R., & Xia, F. (2015). RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports*, *5*(1), 8365. https://doi.org/10.1038/srep08365

Carattoli, A., & Hasman, H. (2020). *PlasmidFinder and In Silico pMLST: Identification and Typing of Plasmid Replicons in Whole-Genome Sequencing (WGS)* (pp. 285–294). https://doi.org/10.1007/978-1-4939-9877-7_20

Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., Møller Aarestrup, F., & Hasman, H. (2014). In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. *Antimicrobial Agents and Chemotherapy*, *58*(7), 3895–3903. https://doi.org/10.1128/AAC.02412-14

Carniel, E. (1999). The Yersinia high-pathogenicity island. *International Microbiology : The Official Journal of the Spanish Society for Microbiology*, *2*(3), 161–167.

Chen, X., Zhou, Y., & Feng, Y. (2023). Citrobacter freundii induces sepsis with new-onset status seizure in an adult: A case report and literature review. *Medicine*, *102*(1), e32549. https://doi.org/10.1097/MD.0000000000032549

Chen, Z., Erickson, D. L., & Meng, J. (2020a). Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanopore sequencing. *BMC Genomics*, *21*(1), 631. https://doi.org/10.1186/s12864-020-07041-8

Chen, Z., Erickson, D. L., & Meng, J. (2020b). Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanopore sequencing. *BMC Genomics*, *21*(1), 631. https://doi.org/10.1186/s12864-020-07041-8

Cillóniz, C., Torres, A., & Niederman, M. S. (2021). Management of pneumonia in critically ill patients. *BMJ*, e065871. https://doi.org/10.1136/bmj-2021-065871

Ciufo, S., Kannan, S., Sharma, S., Badretdin, A., Clark, K., Turner, S., Brover, S., Schoch, C. L., Kimchi, A., & DiCuccio, M. (2018). Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *International Journal of Systematic and Evolutionary Microbiology*, *68*(7), 2386–2392. https://doi.org/10.1099/ijsem.0.002809

Davis, J. J., Wattam, A. R., Aziz, R. K., Brettin, T., Butler, R., Butler, R. M., Chlenski, P., Conrad, N., Dickerman, A., Dietrich, E. M., Gabbard, J. L., Gerdes, S., Guard, A., Kenyon, R. W., Machi, D., Mao, C., Murphy-Olson, D., Nguyen, M., Nordberg, E. K., … Stevens, R. (2019). The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz943

Del Fabbro, C., Scalabrin, S., Morgante, M., & Giorgi, F. M. (2013). An Extensive Evaluation of Read Trimming Effects on Illumina NGS Data Analysis. *PLoS ONE*, *8*(12), e85024. https://doi.org/10.1371/journal.pone.0085024

Delgado, G., Souza, V., Morales, R., Cerritos, R., González-González, A., Méndez, J. L., Vázquez, V., & Cravioto, A. (2013). Genetic characterization of atypical Citrobacter freundii. *PloS One*, *8*(9), e74120. https://doi.org/10.1371/journal.pone.0074120

Dida, F., & Yi, G. (2021). Empirical evaluation of methods for de novo genome assembly. *PeerJ Computer Science*, *7*, e636. https://doi.org/10.7717/peerj-cs.636

Dolejska, M., Villa, L., Poirel, L., Nordmann, P., & Carattoli, A. (2013). Complete sequencing of an IncHI1 plasmid encoding the carbapenemase NDM-1, the ArmA 16S RNA methylase and a resistance-nodulation-cell division/multidrug efflux pump. *Journal of Antimicrobial Chemotherapy*, *68*(1), 34–39. https://doi.org/10.1093/jac/dks357

Emerson, D., Agulto, L., Liu, H., & Liu, L. (2008). Identifying and Characterizing Bacteria in an Era of Genomics and Proteomics. *BioScience*, *58*(10), 925–936. https://doi.org/10.1641/B581006

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Forde, B. M., Bergh, H., Cuddihy, T., Hajkowicz, K., Hurst, T., Playford, E. G., Henderson, B. C., Runnegar, N., Clark, J., Jennison, A. V, Moss, S., Hume, A., Leroux, H., Beatson, S. A., Paterson, D. L., & Harris, P. N. A. (2023). Clinical Implementation of Routine Whole-genome Sequencing for Hospital Infection Control of Multi-drug Resistant Pathogens. *Clinical Infectious Diseases*, *76*(3), e1277–e1284. https://doi.org/10.1093/cid/ciac726

Frenk, S., Rakovitsky, N., Kon, H., Rov, R., Abramov, S., Lurie-Weinberger, M. N., Schwartz, D., Pinco, E., Lellouche, J., & Carmeli, Y. (2021). OXA-900, a Novel OXA Sub-Family Carbapenemase Identified in Citrobacter freundii, Evades Detection by Commercial Molecular Diagnostics Tests. *Microorganisms*, *9*(9), 1898. https://doi.org/10.3390/microorganisms9091898

Galindo-Méndez, M., Navarrete-Salazar, H., Baltazar-Jiménez, F., Muñoz-de la Paz, E., Sánchez-Mawcinitt, M. F., Gómez-Pardo, A., Garza-González, E., Ponce-de-León-Garduño, L. A., Franco-Cendejas, R., Morfín-Otero, R., Rojas-Larios, F., Mena-Ramírez, J. P., Morales-de-la-Peña, C. T., García-Mendoza, L., Choy-Chang, E. V., Avilés-Benítez, L. K., López-Gutiérrez, E., Canizales-Oviedo, J. L., Barlandas-Rendón, N. E., … Ostos-Cantú, H. L. (2022). Emergence of Fosfomycin Resistance by Plasmid-Mediated fos Genes in Uropathogenic ESBL-Producing E. coli Isolates in Mexico. *Antibiotics*, *11*(10), 1383. https://doi.org/10.3390/antibiotics11101383

Gatica, S., Fuentes, B., Rivera-Asín, E., Ramírez-Céspedes, P., Sepúlveda-Alfaro, J., Catalán, E. A., Bueno, S. M., Kalergis, A. M., Simon, F., Riedel, C. A., & Melo-Gonzalez, F. (2023). Novel evidence on sepsis-inducing pathogens: from laboratory to bedside. *Frontiers in Microbiology*, *14*. https://doi.org/10.3389/fmicb.2023.1198200

Gauer, R., Forbes, D., & Boyer, N. (2020). Sepsis: Diagnosis and Management. *American Family Physician*, *101*(7), 409–418.

Geroulanos, S., & Douka, E. T. (2006). Historical perspective of the word "sepsis." *Intensive Care Medicine*, *32*(12), 2077–2077. https://doi.org/10.1007/s00134-006-0392-2

Gül, F., Arslantaş, M. K., Cinel, İ., & Kumar, A. (2017). Changing Definitions of Sepsis. *Turkish Journal of Anaesthesiology and Reanimation*, *45*(3), 129–138. https://doi.org/10.5152/TJAR.2017.93753

Gul, F., Arslantas, M. K., Cinel, I., & Kumar, A. (2017). Changing Definitions of Sepsis. *Turkish Journal of Anesthesia and Reanimation*, *45*(3), 129–138. https://doi.org/10.5152/TJAR.2017.93753

Guo, L., Ye, L., Zhao, Q., Ma, Y., Yang, J., & Luo, Y. (2014). Comparative study of MALDI-TOF MS and VITEK 2 in bacteria identification. *Journal of Thoracic Disease*, *6*(5), 534–538. https://doi.org/10.3978/j.issn.2072-1439.2014.02.18

Guo, Y., Ye, F., Sheng, Q., Clark, T., & Samuels, D. C. (2014). Three-stage quality control strategies for DNA re-sequencing data. *Briefings in Bioinformatics*, *15*(6), 879–889. https://doi.org/10.1093/bib/bbt069

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Gyawali, B., Ramakrishna, K., & Dhamoon, A. S. (2019). Sepsis: The evolution in definition, pathophysiology, and management. *SAGE Open Medicine*, *7*, 205031211983504. https://doi.org/10.1177/2050312119835043

Hernandez-Beeftink, T., Guillen-Guio, B., Lorenzo-Salazar, J. M., Corrales, A., Suarez-Pajes, E., Feng, R., Rubio-Rodríguez, L. A., Paynton, M. L., Cruz, R., García-Laorden, M. I., Prieto-González, M., Rodríguez-Pérez, A., Carriedo, D., Blanco, J., Ambrós, A., González-Higueras, E., Espinosa, E., Muriel, A., Tamayo, E., … Flores, C. (2022). A genome-wide association study of survival in patients with sepsis. *Critical Care*, *26*(1), 341. https://doi.org/10.1186/s13054-022-04208-5

Hirakawa, H., Bordi, C., & Tomita, H. (2021). Editorial: Gram-Negative Pathogenesis. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.813062

Horstmann, S., Schmechel, C., Palm, K., Oertelt-Prigione, S., & Bolte, G. (2022). The Operationalisation of Sex and Gender in Quantitative Health–Related Research: A Scoping Review. *International Journal of Environmental Research and Public Health*, *19*(12), 7493. https://doi.org/10.3390/ijerph19127493

Jabeen, I., Islam, S., Hassan, A. K. M. I., Tasnim, Z., & Shuvo, S. R. (2023). A brief insight into Citrobacter species - a growing threat to public health. *Frontiers in Antibiotics*, *2*. https://doi.org/10.3389/frabi.2023.1276982

Jacoby, R. O., Fox, J. G., & Davisson, M. (2002). Biology and Diseases of Mice. In *Laboratory Animal Medicine* (pp. 35–120). Elsevier. https://doi.org/10.1016/B978-012263951-7/50006-5

Jarczak, D., Kluge, S., & Nierhaus, A. (2021). Sepsis—Pathophysiology and Therapeutic Concepts. *Frontiers in Medicine*, *8*. https://doi.org/10.3389/fmed.2021.628302

Jazayeri, A., Yang, C. C., & Capan, M. (2023). Frequent temporal patterns of physiological and biological biomarkers and their evolution in sepsis. *Artificial Intelligence in Medicine*, *143*, 102576. https://doi.org/10.1016/j.artmed.2023.102576

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

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

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

Kerby, D. S. (2014). The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation. *Comprehensive Psychology*, *3*, 11.IT.3.1. https://doi.org/10.2466/11.IT.3.1

Kerrn, M. B. (2002). Susceptibility of Danish Escherichia coli strains isolated from urinary tract infections and bacteraemia, and distribution of sul genes conferring sulphonamide resistance. *Journal of Antimicrobial Chemotherapy*, *50*(4), 513–516. https://doi.org/10.1093/jac/dkf164

Khezri, A., Avershina, E., & Ahmad, R. (2021). Hybrid Assembly Provides Improved Resolution of Plasmids, Antimicrobial Resistance Genes, and Virulence Factors in Escherichia coli and Klebsiella pneumoniae Clinical Isolates. *Microorganisms*, *9*(12), 2560. https://doi.org/10.3390/microorganisms9122560

Kolínská, R., Španělová, P., Dřevínek, M., Hrabák, J., & Žemličková, H. (2015). Species identification of strains belonging to genus Citrobacter using the biochemical method and MALDI-TOF mass spectrometry. *Folia Microbiologica*, *60*(1), 53–59. https://doi.org/10.1007/s12223-014-0340-4

Kozarewa, I., Ning, Z., Quail, M. A., Sanders, M. J., Berriman, M., & Turner, D. J. (2009). Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nature Methods*, *6*(4), 291–295. https://doi.org/10.1038/nmeth.1311

Krueger, F. ,. (2012). *Trim Galore: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisufite-Seq) libraries.* .

Kumar, P., Meghvansi, M. K., & Kamboj, D. V. (2021). Phenotypic Characterization and Whole-Genome Analysis of a Novel Bacteriophage HCF1 Infecting Citrobacter amalonaticus and C. freundii. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.644013

Lalaoui, R., Djukovic, A., Bakour, S., Hadjadj, L., Sanz, J., Salavert, M., López-Hontangas, J. L., Sanz, M. A., Ubeda, C., & Rolain, J.-M. (2019). Genomic characterization of Citrobacter freundii strains coproducing OXA-48 and VIM-1 carbapenemase enzymes isolated in leukemic patient in Spain. *Antimicrobial Resistance & Infection Control*, *8*(1), 167. https://doi.org/10.1186/s13756-019-0630-3

Leone, M., Lakbar, I., & Vincent, J.-L. (2023). Sepsis : Actual numbers and uncertainties. *Revue d'Épidémiologie et de Santé Publique*, *71*(6), 102176. https://doi.org/10.1016/j.respe.2023.102176

Levy, M. M., Fink, M. P., Marshall, J. C., Abraham, E., Angus, D., Cook, D., Cohen, J., Opal, S. M., Vincent, J.-L., & Ramsay, G. (2003a). 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Critical Care Medicine*, *31*(4), 1250–1256. https://doi.org/10.1097/01.CCM.0000050454.01978.3B

Levy, M. M., Fink, M. P., Marshall, J. C., Abraham, E., Angus, D., Cook, D., Cohen, J., Opal, S. M., Vincent, J.-L., & Ramsay, G. (2003b). 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Intensive Care Medicine*, *29*(4), 530–538. https://doi.org/10.1007/s00134-003-1662-x

Li, H., & Johnson, T. (2014). Wilcoxon's signed-rank statistic: what null hypothesis and why it matters. *Pharmaceutical Statistics*, *13*(5), 281–285. https://doi.org/10.1002/pst.1628

Lin, G.-L., McGinley, J. P., Drysdale, S. B., & Pollard, A. J. (2018). Epidemiology and Immune Pathogenesis of Viral Sepsis. *Frontiers in Immunology*, *9*. https://doi.org/10.3389/fimmu.2018.02147

Liu, L., Qin, L., Hao, S., Lan, R., Xu, B., Guo, Y., Jiang, R., Sun, H., Chen, X., LV, X., Xu, J., & Zhao, C. (2020). Lineage, Antimicrobial Resistance and Virulence of Citrobacter spp. *Pathogens*, *9*(3), 195. https://doi.org/10.3390/pathogens9030195

Liu, L., Zhang, L., Zhou, H., Yuan, M., Hu, D., Wang, Y., Sun, H., Xu, J., & Lan, R. (2021a). Antimicrobial Resistance and Molecular Characterization of Citrobacter spp. Causing Extraintestinal Infections. *Frontiers in Cellular and Infection Microbiology*, *11*. https://doi.org/10.3389/fcimb.2021.737636

Liu, L., Zhang, L., Zhou, H., Yuan, M., Hu, D., Wang, Y., Sun, H., Xu, J., & Lan, R. (2021b). Antimicrobial Resistance and Molecular Characterization of Citrobacter spp. Causing Extraintestinal Infections. *Frontiers in Cellular and Infection Microbiology*, *11*. https://doi.org/10.3389/fcimb.2021.737636

Liu, L., Zhang, L., Zhou, H., Yuan, M., Hu, D., Wang, Y., Sun, H., Xu, J., & Lan, R. (2021c). Antimicrobial Resistance and Molecular Characterization of Citrobacter spp. Causing Extraintestinal Infections. *Frontiers in Cellular and Infection Microbiology*, *11*. https://doi.org/10.3389/fcimb.2021.737636

Ljungström L. (2017). *Community onset sepsis in Sweden: a population based study (Doctoral dissertation), Sahlgrenska academy at University of Gothenburg, Sweden.*

Ljungström, L., Andersson, R., & Jacobsson, G. (2019). Incidences of community onset severe sepsis, Sepsis-3 sepsis, and bacteremia in Sweden – A prospective population-based study. *PLOS ONE*, *14*(12), e0225700. https://doi.org/10.1371/journal.pone.0225700

Lombardo, M. N., G-Dayanandan, N., Wright, D. L., & Anderson, A. C. (2016). Crystal Structures of Trimethoprim-Resistant DfrA1 Rationalize Potent Inhibition by Propargyl-Linked Antifolates. *ACS Infectious Diseases*, *2*(2), 149–156. https://doi.org/10.1021/acsinfecdis.5b00129

Lu, J., Rincon, N., Wood, D. E., Breitwieser, F. P., Pockrandt, C., Langmead, B., Salzberg, S. L., & Steinegger, M. (2022). Metagenome analysis using the Kraken software suite. *Nature Protocols*, *17*(12), 2815–2839. https://doi.org/10.1038/s41596-022-00738-y

Luo, Y., Wang, H., Liang, J., Qian, H., Ye, J., Chen, L., Yang, X., Chen, Z., Wang, F., Octavia, S., Payne, M., Song, X., Jiang, J., Jin, D., & Lan, R. (2021). Population Structure and Multidrug Resistance of Non-O1/Non-O139 Vibrio cholerae in Freshwater Rivers in Zhejiang, China. *Microbial Ecology*, *82*(2), 319–333. https://doi.org/10.1007/s00248-020-01645-z

Mahapatra, S., Heffner, A. C., & Atarthi-Dugan, J. M. (2023). *Septic Shock (Nursing)*.

Martin, M. (2011a). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Martin, M. (2011b). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Massilamany, C., Mohammed, A., Loy, J. D., Purvis, T., Krishnan, B., Basavalingappa, R. H., Kelley, C. M., Guda, C., Barletta, R. G., Moriyama, E. N., Smith, T. P. L., & Reddy, J. (2016). Whole genomic sequence analysis of Bacillus infantis: defining the genetic blueprint of strain NRRL B-14911, an emerging cardiopathogenic microbe. *BMC Genomics*, *17*(S7), 511. https://doi.org/10.1186/s12864-016-2900-2

McArthur, A. G., Waglechner, N., Nizam, F., Yan, A., Azad, M. A., Baylay, A. J., Bhullar, K., Canova, M. J., De Pascale, G., Ejim, L., Kalan, L., King, A. M., Koteva, K., Morar, M., Mulvey, M. R., O'Brien, J. S., Pawlowski, A. C., Piddock, L. J. V., Spanogiannopoulos, P., … Wright, G. D. (2013). The Comprehensive Antibiotic Resistance Database. *Antimicrobial Agents and Chemotherapy*, *57*(7), 3348–3357. https://doi.org/10.1128/AAC.00419-13

McLeod, M., Ahmad, R., Shebl, N. A., Micallef, C., Sim, F., & Holmes, A. (2019). A whole-health–economy approach to antimicrobial stewardship: Analysis of current models and future direction. *PLOS Medicine*, *16*(3), e1002774. https://doi.org/10.1371/journal.pmed.1002774

Meckawy, R., Stuckler, D., Mehta, A., Al-Ahdal, T., & Doebbeling, B. N. (2022). Effectiveness of early warning systems in the detection of infectious diseases outbreaks: a systematic review. *BMC Public Health*, *22*(1), 2216. https://doi.org/10.1186/s12889-022-14625-4

Medvedev, P., Pham, S., Chaisson, M., Tesler, G., & Pevzner, P. (2011). Paired de Bruijn Graphs: A Novel Approach for Incorporating Mate Pair Information into Genome Assemblers. *Journal of Computational Biology*, *18*(11), 1625–1634. https://doi.org/10.1089/cmb.2011.0151

Mellhammar, L., Wullt, S., Lindberg, Å., Lanbeck, P., Christensson, B., & Linder, A. (2016). Sepsis Incidence: A Population-Based Study. *Open Forum Infectious Diseases*, *3*(4). https://doi.org/10.1093/ofid/ofw207

Mitchell, E., Pearce, M. S., & Roberts, A. (2019). Gram-negative bloodstream infections and sepsis: risk factors, screening tools and surveillance. *British Medical Bulletin*, *132*(1), 5–15. https://doi.org/10.1093/bmb/ldz033

Mundy, R., MacDonald, T. T., Dougan, G., Frankel, G., & Wiles, S. (2005). Citrobacter rodentium of mice and man. *Cellular Microbiology*, *7*(12), 1697–1706. https://doi.org/10.1111/j.1462-5822.2005.00625.x

Nagarajan, N., & Pop, M. (2013). Sequence assembly demystified. *Nature Reviews Genetics*, *14*(3), 157–167. https://doi.org/10.1038/nrg3367

Olson, R. D., Assaf, R., Brettin, T., Conrad, N., Cucinell, C., Davis, J. J., Dempsey, D. M., Dickerman, A., Dietrich, E. M., Kenyon, R. W., Kuscuoglu, M., Lefkowitz, E. J., Lu, J., Machi, D., Macken, C., Mao, C., Niewiadomska, A., Nguyen, M., Olsen, G. J., … Stevens, R. L. (2023). Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Research*, *51*(D1), D678–D689. https://doi.org/10.1093/nar/gkac1003

Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, *12*(1), 385. https://doi.org/10.1186/1471-2105-12-385

Osei Sekyere, J., & Reta, M. A. (2021). Global evolutionary epidemiology and resistome dynamics of Citrobacter species , Enterobacter hormaechei, Klebsiella variicola, and Proteeae clones. *Environmental Microbiology*, *23*(12), 7412–7431. https://doi.org/10.1111/1462-2920.15387

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., & Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, *42*(D1), D206–D214. https://doi.org/10.1093/nar/gkt1226

Pant, A., Mackraj, I., & Govender, T. (2021). Advances in sepsis diagnosis and management: a paradigm shift towards nanotechnology. *Journal of Biomedical Science*, *28*(1), 6. https://doi.org/10.1186/s12929-020-00702-6

Patel, R. (2015). MALDI-TOF MS for the Diagnosis of Infectious Diseases. *Clinical Chemistry*, *61*(1), 100–111. https://doi.org/10.1373/clinchem.2014.221770

Pepperell, C., Kus, J. V., Gardam, M. A., Humar, A., & Burrows, L. L. (2002). Low-Virulence Citrobacter Species Encode Resistance to Multiple Antimicrobials. *Antimicrobial Agents and Chemotherapy*, *46*(11), 3555–3560. https://doi.org/10.1128/AAC.46.11.3555-3560.2002

Pervez, M. T., Hasnain, M. J. ul, Abbas, S. H., Moustafa, M. F., Aslam, N., & Shah, S. S. M. (2022). A Comprehensive Review of Performance of Next-Generation Sequencing Platforms. *BioMed Research International*, *2022*, 1–12. https://doi.org/10.1155/2022/3457806

Quail, M., Smith, M. E., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., & Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. *BMC Genomics*, *13*(1), 341. https://doi.org/10.1186/1471-2164-13-341

Rai, S. S., Peters, R. M. H., Syurina, E. V., Irwanto, I., Naniche, D., & Zweekhorst, M. B. M. (2020). Intersectionality and health-related stigma: insights from experiences of people living with stigmatized health conditions in Indonesia. *International Journal for Equity in Health*, *19*(1), 206. https://doi.org/10.1186/s12939-020-01318-w

Ramsamy, Y., Mlisana, K. P., Amoako, D. G., Allam, M., Ismail, A., Singh, R., Abia, A. L. K., & Essack, S. Y. (2020). Pathogenomic Analysis of a Novel Extensively Drug-Resistant Citrobacter freundii Isolate Carrying a blaNDM-1 Carbapenemase in South Africa. *Pathogens*, *9*(2), 89. https://doi.org/10.3390/pathogens9020089

Rayamajhi, N., Cheng, C.-H. C., & Catchen, J. M. (2022). Evaluating Illumina-, Nanopore-, and PacBio-based genome assembly strategies with the bald notothen, *Trematomus borchgrevinki*. *G3 Genes|Genomes|Genetics*, *12*(11). https://doi.org/10.1093/g3journal/jkac192

Revez, J., Espinosa, L., Albiger, B., Leitmeyer, K. C., & Struelens, M. J. (2017). Survey on the Use of Whole-Genome Sequencing for Infectious Diseases Surveillance: Rapid Expansion of European National Capacities, 2015–2016. *Frontiers in Public Health*, *5*. https://doi.org/10.3389/fpubh.2017.00347

Rhee, C., Dantes, R., Epstein, L., Murphy, D. J., Seymour, C. W., Iwashyna, T. J., Kadri, S. S., Angus, D. C., Danner, R. L., Fiore, A. E., Jernigan, J. A., Martin, G. S., Septimus, E., Warren, D. K., Karcz, A., Chan, C., Menchaca, J. T., Wang, R., Gruber, S., & Klompas, M. (2017). Incidence and Trends of Sepsis in US Hospitals Using Clinical vs Claims Data, 2009-2014. *JAMA*, *318*(13), 1241. https://doi.org/10.1001/jama.2017.13836

Richter, M., & Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences*, *106*(45), 19126–19131. https://doi.org/10.1073/pnas.0906412106

Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., & Peplies, J. (2016a). JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics*, *32*(6), 929–931. https://doi.org/10.1093/bioinformatics/btv681

Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., & Peplies, J. (2016b). JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics*, *32*(6), 929–931. https://doi.org/10.1093/bioinformatics/btv681

Rudd, K. E., Kissoon, N., Limmathurotsakul, D., Bory, S., Mutahunga, B., Seymour, C. W., Angus, D. C., & West, T. E. (2018a). The global burden of sepsis: barriers and potential solutions. *Critical Care*, *22*(1), 232. https://doi.org/10.1186/s13054-018-2157-z

Rudd, K. E., Kissoon, N., Limmathurotsakul, D., Bory, S., Mutahunga, B., Seymour, C. W., Angus, D. C., & West, T. E. (2018b). The global burden of sepsis: barriers and potential solutions. *Critical Care*, *22*(1), 232. https://doi.org/10.1186/s13054-018-2157-z

Samonis, G., Karageorgopoulos, D. E., Kofteridis, D. P., Matthaiou, D. K., Sidiropoulou, V., Maraki, S., & Falagas, M. E. (2009). Citrobacter infections in a general hospital: characteristics and outcomes. *European Journal of Clinical Microbiology & Infectious Diseases*, *28*(1), 61–68. https://doi.org/10.1007/s10096-008-0598-z

Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R. P., Banday, S., Mishra, A. K., Das, G., & Malonia, S. K. (2023). Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology*, *12*(7), 997. https://doi.org/10.3390/biology12070997

Schlapbach, L. J., Kissoon, N., Alhawsawi, A., Aljuaid, M. H., Daniels, R., Gorordo-Delsol, L. A., Machado, F., Malik, I., Nsutebu, E. F., Finfer, S., & Reinhart, K. (2020). World Sepsis Day: a global agenda to target a leading cause of morbidity and mortality. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, *319*(3), L518–L522. https://doi.org/10.1152/ajplung.00369.2020

Schwartz, D. A. (2021). Prioritizing the Continuing Global Challenges to Emerging and Reemerging Viral Infections. *Frontiers in Virology*, *1*. https://doi.org/10.3389/fviro.2021.701054

Shankar-Hari, M., Phillips, G. S., Levy, M. L., Seymour, C. W., Liu, V. X., Deutschman, C. S., Angus, D. C., Rubenfeld, G. D., & Singer, M. (2016). Developing a New Definition and Assessing New Clinical Criteria for Septic Shock. *JAMA*, *315*(8), 775. https://doi.org/10.1001/jama.2016.0289

Sheu, C.-C., Chang, Y.-T., Lin, S.-Y., Chen, Y.-H., & Hsueh, P.-R. (2019). Infections Caused by Carbapenem-Resistant Enterobacteriaceae: An Update on Therapeutic Options. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.00080

Shih, C.-C., Chen, Y.-C., Chang, S.-C., Luh, K.-T., & Hsieh, W.-C. (1996). Bacteremia Due to Citrobacter Species: Significance of Primary Intraabdominal Infection. *Clinical Infectious Diseases*, *23*(3), 543–549. https://doi.org/10.1093/clinids/23.3.543

Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., Bellomo, R., Bernard, G. R., Chiche, J.-D., Coopersmith, C. M., Hotchkiss, R. S., Levy, M. M., Marshall, J. C., Martin, G. S., Opal, S. M., Rubenfeld, G. D., van der Poll, T., Vincent, J.-L., & Angus, D. C. (2016). The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, *315*(8), 801. https://doi.org/10.1001/jama.2016.0287

Tang, K. W. K., Millar, B. C., & Moore, J. E. (2023). Antimicrobial Resistance (AMR). *British Journal of Biomedical Science*, *80*. https://doi.org/10.3389/bjbs.2023.11387

Thompson, K., Venkatesh, B., & Finfer, S. (2019). Sepsis and septic shock: current approaches to management. *Internal Medicine Journal*, *49*(2), 160–170. https://doi.org/10.1111/imj.14199

Tuna, M., & Amos, C. I. (2013). Genomic sequencing in cancer. *Cancer Letters*, *340*(2), 161–170. https://doi.org/10.1016/j.canlet.2012.11.004

Urwin, R., & Maiden, M. C. J. (2003). Multi-locus sequence typing: a tool for global epidemiology. *Trends in Microbiology*, *11*(10), 479–487. https://doi.org/10.1016/j.tim.2003.08.006

van Engelen, T. S. R., Joost Wiersinga, W., & van der Poll, T. (2018). Pathogenesis of Sepsis. In *Handbook of Sepsis* (pp. 31–43). Springer International Publishing. https://doi.org/10.1007/978-3-319-73506-1_3

Villa, L., García-Fernández, A., Fortini, D., & Carattoli, A. (2010). Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. *Journal of Antimicrobial Chemotherapy*, *65*(12), 2518–2529. https://doi.org/10.1093/jac/dkq347

Vincent, J.-L., Moreno, R., Takala, J., Willatts, S., De Mendonça, A., Bruining, H., Reinhart, C. K., Suter, P. M., & Thijs, L. G. (1996). The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Medicine*, *22*(7), 707–710. https://doi.org/10.1007/BF01709751

Wajid, B., & Serpedin, E. (2016). Do it yourself guide to genome assembly. *Briefings in Functional Genomics*, *15*(1), 1–9. https://doi.org/10.1093/bfgp/elu042

Wardi, G., Pearce, A., DeMaria, A., & Malhotra, A. (2023). Describing Sepsis as a Risk Factor for Cardiovascular Disease. *Journal of the American Heart Association*, *12*(3). https://doi.org/10.1161/JAHA.122.028882

Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Computational Biology*, *13*(6), e1005595. https://doi.org/10.1371/journal.pcbi.1005595

Wick, R. R., Judd, L. M., & Holt, K. E. (2023). Assembling the perfect bacterial genome using Oxford Nanopore and Illumina sequencing. *PLOS Computational Biology*, *19*(3), e1010905. https://doi.org/10.1371/journal.pcbi.1010905

Wick, R. R., Schultz, M. B., Zobel, J., & Holt, K. E. (2015). Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*, *31*(20), 3350–3352. https://doi.org/10.1093/bioinformatics/btv383

Williams, C. R., Baccarella, A., Parrish, J. Z., & Kim, C. C. (2016). Trimming of sequence reads alters RNA-Seq gene expression estimates. *BMC Bioinformatics*, *17*(1), 103. https://doi.org/10.1186/s12859-016-0956-2

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1), 257. https://doi.org/10.1186/s13059-019-1891-0

Yao, Y., Falgenhauer, L., Falgenhauer, J., Hauri, A. M., Heinmüller, P., Domann, E., Chakraborty, T., & Imirzalioglu, C. (2021). Carbapenem-Resistant Citrobacter spp. as an Emerging Concern in the Hospital-Setting: Results From a Genome-Based Regional Surveillance

Study. *Frontiers in Cellular and Infection Microbiology*, *11*. https://doi.org/10.3389/fcimb.2021.744431

Yatabe, T., Egi, M., & Ogura, H. (2022). New avenues of sepsis research: obtaining perspective by analyzing and comparing SSCG 2021 and J-SSCG 2020. *Journal of Intensive Care*, *10*(1), 11. https://doi.org/10.1186/s40560-022-00606-7

Yuan, C., Yin, Z., Wang, J., Qian, C., Wei, Y., Zhang, S., Jiang, L., & Liu, B. (2019a). Comparative Genomic Analysis of Citrobacter and Key Genes Essential for the Pathogenicity of Citrobacter koseri. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02774

Yuan, C., Yin, Z., Wang, J., Qian, C., Wei, Y., Zhang, S., Jiang, L., & Liu, B. (2019b). Comparative Genomic Analysis of Citrobacter and Key Genes Essential for the Pathogenicity of Citrobacter koseri. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02774

Yuan, C., Yin, Z., Wang, J., Qian, C., Wei, Y., Zhang, S., Jiang, L., & Liu, B. (2019c). Comparative Genomic Analysis of Citrobacter and Key Genes Essential for the Pathogenicity of Citrobacter koseri. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02774

Yuan, C., Yin, Z., Wang, J., Qian, C., Wei, Y., Zhang, S., Jiang, L., & Liu, B. (2019d). Comparative Genomic Analysis of Citrobacter and Key Genes Essential for the Pathogenicity of Citrobacter koseri. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02774

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F. M., & Larsen, M. V. (2012). Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy*, *67*(11), 2640–2644. https://doi.org/10.1093/jac/dks261

Zhang, G., Zhao, Q., Ye, K., Ye, L., Ma, Y., & Yang, J. (2023). Molecular analysis of clinical Citrobacter spp. isolates: Acquisition of the Yersinia high-pathogenicity island mediated by ICEkp in C. freundii. *Frontiers in Microbiology*, *14*. https://doi.org/10.3389/fmicb.2023.1056790

Zhou, W., Chen, Q., Qian, C., Shen, K., Zhu, X., Zhou, D., Lu, W., Sun, Z., Liu, H., Li, K., Xu, T., Bao, Q., & Lu, J. (2019). In Vitro Susceptibility and Florfenicol Resistance in Citrobacter Isolates and Whole-Genome Analysis of Multidrug-Resistance in Citrobacter freundii. *International Journal of Genomics*, *2019*, 1–15. https://doi.org/10.1155/2019/7191935

Appendix 1

```
java -jar ./Trimmomatic-0.39. jar PE 220322_CI11_1.fastq.gz
220322_CI11_2.fastq.qz -baseout Trimmed_out/outCI11.fastq
ILLUMINACLIP:Trimmomatic/adapters/NexteraPE-PE.fa:0:30:10
SLIDINGWINDOW:4:20 HEADCROP:12 MINLEN:30
```

Figure 1 | demonstrates the trimmomatic code of Trimmomatic v0.39 to perform quality trimming of paired-end sequencing reads. The command operates on two input FASTQ files, 220322_CI11_1.fastq.gz (forward read) and 220322_CI11_2.fastq.gz (reverse read) and output trimmed reads to the directory Trimmed_out/ with the filename outCI11.fastq. of *Citrobacter* isolate (CI11).

```
unicycler --short1 outCI11_1P.fastq  --short2 outCI11_2P.
fastq  -o  unicycler_outCI11 --spades_path  SPAdes-3.13.0 -
Linux/bin/spades.py
```

Figure 2 | illustrates the Unicycler command for assembling the *Citrobacter* isolate (CI11) genome from quality trimmed paired-end sequencing reads. The command specifies the paths to the trimmed forward (--short1 outCI11_1P.fastq) and reverse (--short2 outCI11_2P.fastq) reads generated after quality trimming.

```
quast-5.0.2/quast.py assembly.fasta
-r GCF_003812345.1_ASM381234v1_genomic.fna o quast_outCI11
```

Figure 3 | depicts QUAST v5.0.2 command for evaluating the quality of a genome assembly of *Citrobacter* isolate (CI11). The command instructs QUAST to analyse the assembly stored in the assembly. fasta. The reference genome given is (specified by -r), and results are generated in the output directory quast_outCI11.

Table 1 | List of reference genomes selected for 5 different species types from BV-BRC and NCBI

	BV-BRC	**NCBI**		
Species	**Taxon ID**	**Accession No.**	**Strain**	**Genome size**
Citrobacter farmeri	1114922	GCA_000764735.1	*GTC1319*	4.9Mb
Citrobacter braakii	57706	GCA_004331545.1	*HH7*	4.8Mb
Citrobacter freundii	546	GCA_002871775.1	*UMB1094*	5.2Mb
Citrobacter koseri	290338	GCF_000018045.1	*ATCC BAA-895*	4.7Mb
Enterobacter bugandensis	881260	GCF_003964645.1	*WCHEBO9OO31*	4.6Mb

BV-BRC- (Bacterial and Viral Bioinformatics Resource Center)
NCBI- (National Center for Biotechnology Information)

Appendix 2

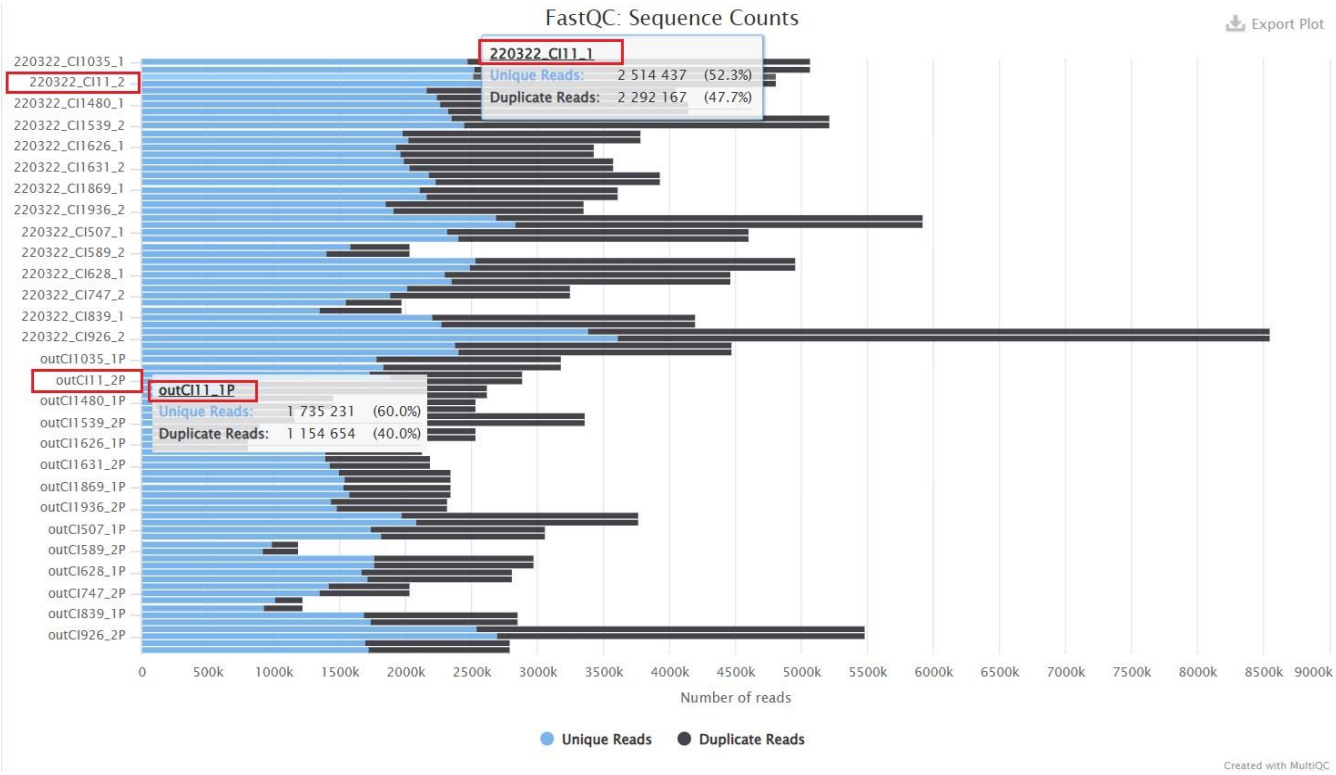

Figure 1 | Displays the number of reads before and after trimming with Trimmomatic for various sample IDs in the in-house pipeline. The red box highlights the read counts for sample ID CI11. On the Y-axis, '220322_CI11_2' represents the initial read count (in millions) for both read 1 and read 2 (with an enlarged portion in box '220322_CI11_1' specifically showing read 1). Additionally, 'outCI11_2P' on the Y-axis corresponds to the final read count (in million reads) after trimming, while 'outCI11_1P' within the graph represents read 1.

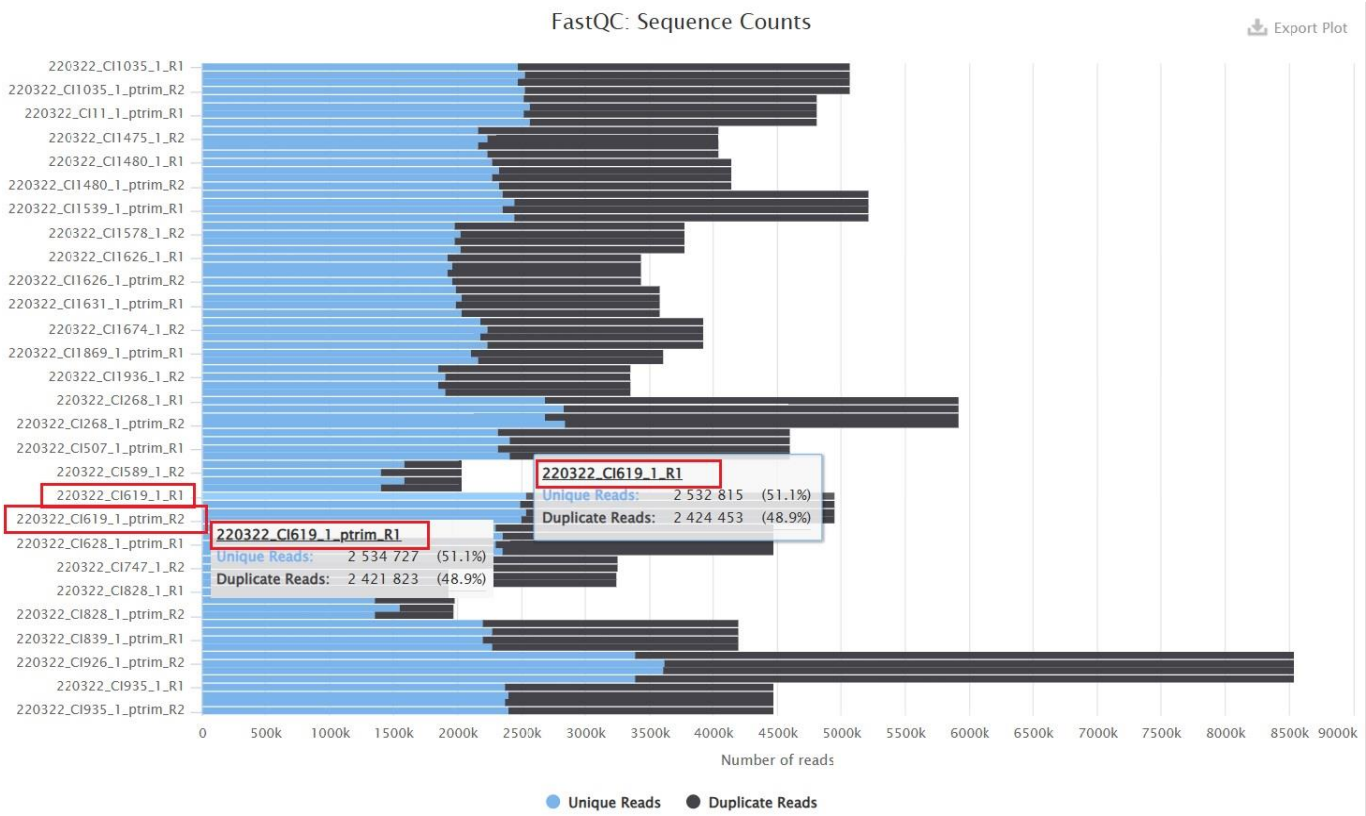

Figure 2| Illustrates the read counts before and after trimming using TrimGalore for various sample IDs in the BV-BRC pipeline. The red box highlights the read counts for sample ID CI619. On the Y-axis, '220322_CI619_1_R1' represents the initial read count (in millions) for both read 1 and read 2 (with an enlarged portion in box '220322_CI619_1_R1' specifically showing read 1). Additionally, 'OutCI619_1_ptrim_R2' on the Y-axis corresponds to the final read count (in million reads) after trimming, while '220322_CI619_1_ptrim_R1' within the graph represents read 1.

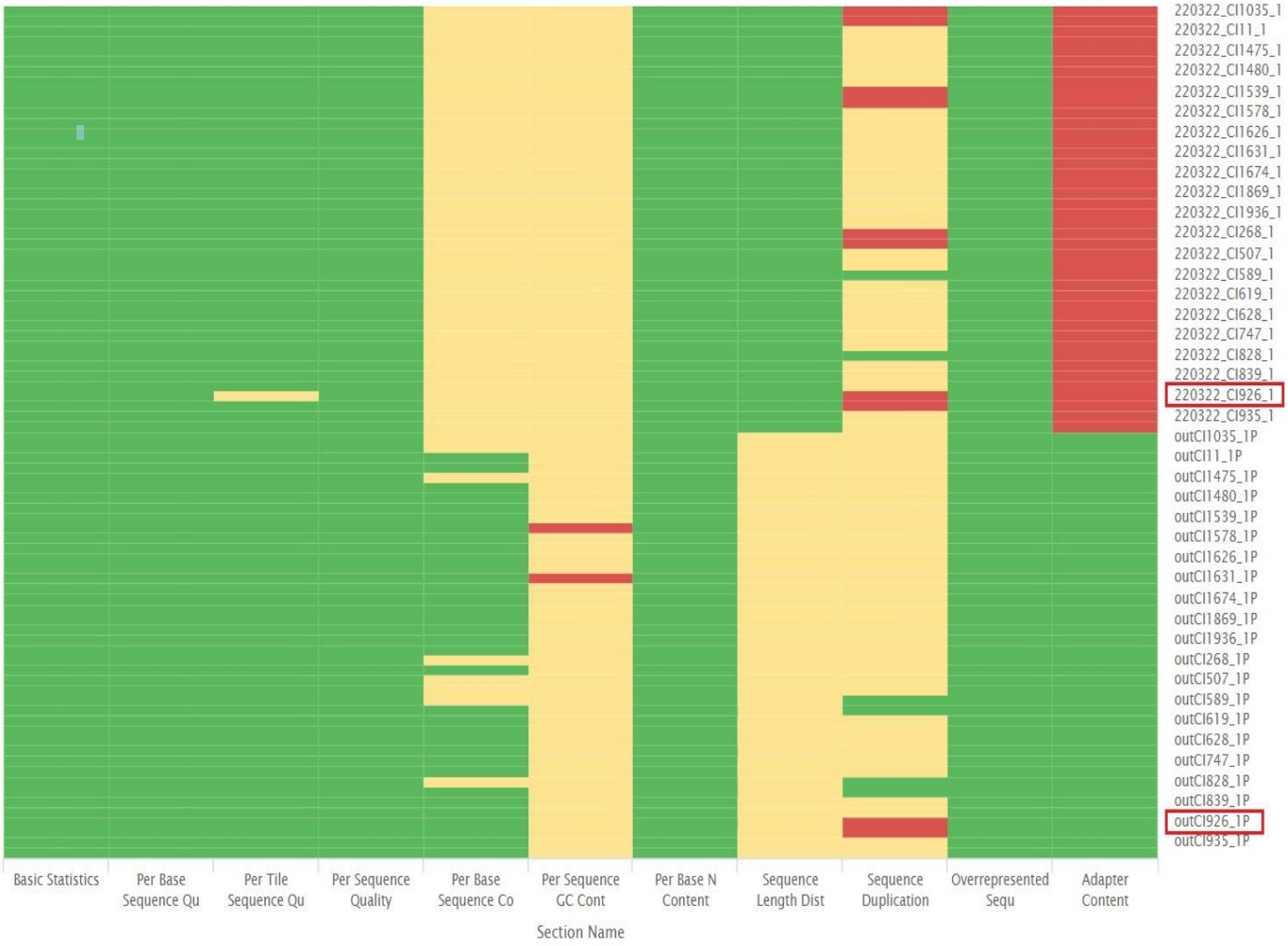

Figure 3 | MultiQC report for in-house pipeline summarizes quality control metrics for DNA sequencing data. Key sections include, Basic Statistics, Per Base Sequence Quality, Per Tile Sequence Quality, Per Sequence GC Content, Per Base N Content, Sequence Length Distribution, Sequence Duplication Levels, Overrepresented Sequences and Adapter Content. Green indicates good quality, while yellow warning and red poor quality. Results of 21 *Citrobacter* isolates with sample ID before trimming (eg:220322_CI1926_1) and after trimming (eg: outCI926_1P) are given.

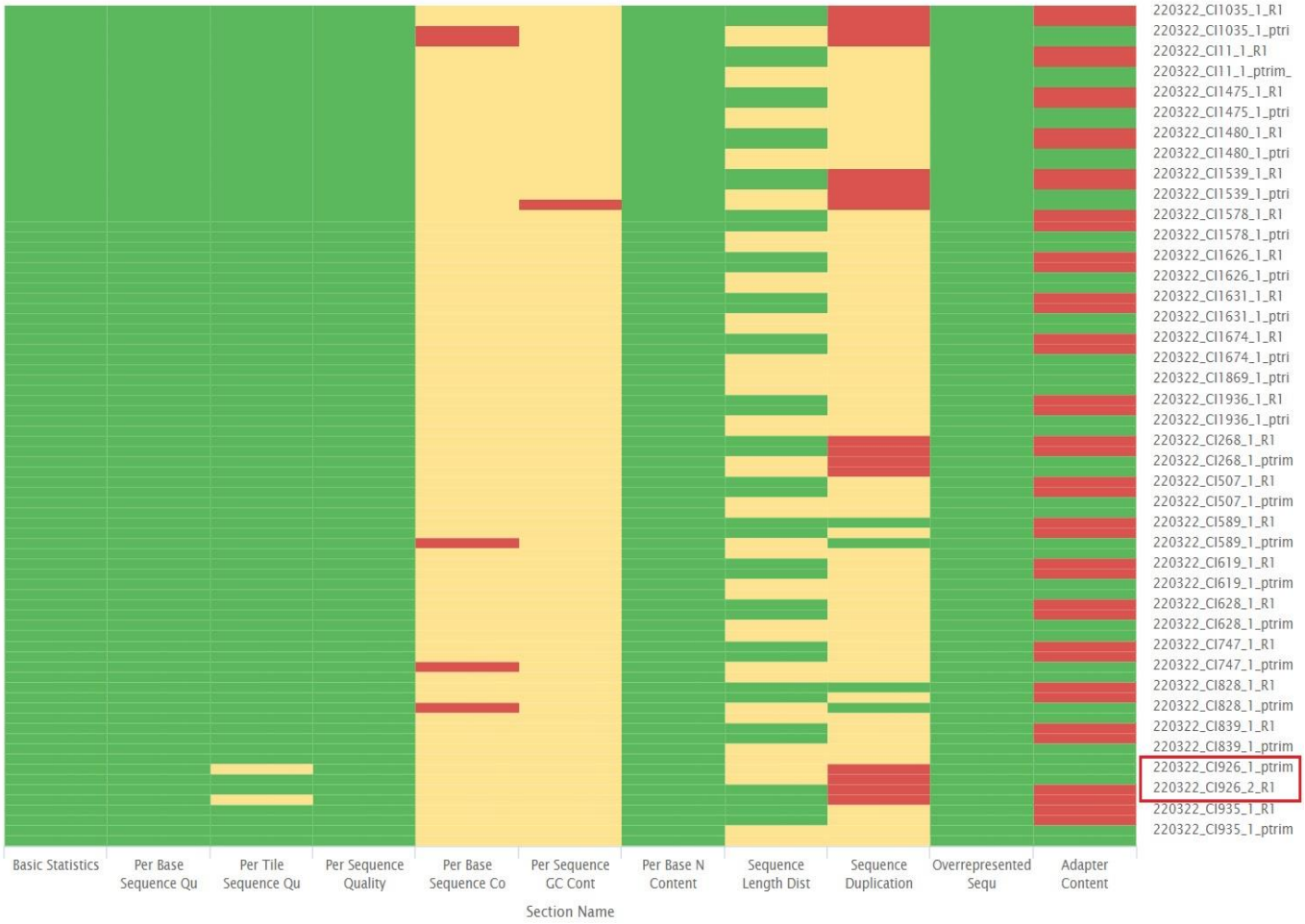

Figure 4 | MultiQC report for BV-BRC pipeline summarizes quality control metrics for DNA sequencing data. Key sections include Basic Statistics, Per Base Sequence Quality, Per Tile Sequence Quality, Per Sequence GC Content, Per Base N Content, Sequence Length Distribution, Sequence Duplication Levels, Overrepresented Sequences and Adapter Content. Green indicates good quality, while yellow warning and red poor quality. Results of 21 *Citrobacter* isolates with sample ID before trimming (eg:220322_CI1926_1) and after trimming (eg: 220322_CI926_1_ptrim) are given.

Table 1 | Compares the impact of TrimGalore (BV-BRC pipeline) and Trimmomatic (in-house pipeline) on read counts by showing the difference in total read counts before and after trimming for both forward and reverse reads of isolates.

	BV-BRC pipeline					In-house pipeline				
Sample ID	**Forward reads before trimming**	**Forward reads after trimming**	**Reverse reads before trimming**	**Reverse reads after trimming**	**Difference in number of reads**	**Forward reads before trimming**	**Forward reads after trimming**	**Reverse reads before trimming**	**Reverse reads after trimming**	**Difference in number of reads**
CI 11	4806604	4805783	4806604	4805783	**821**	4806604	2889885	4806604	2889885	**1916719**
CI 268	5918987	5918090	5918987	5918090	**897**	5918987	3771103	5918987	3771103	**2147884**
CI 507	4601651	4600845	4601651	4600845	**806**	4601651	3061742	4601651	3061742	**1539909**
CI 589	2034946	2034518	2034946	2034518	**428**	2034946	1190737	2034946	1190737	**844209**
CI 619	4957268	4956550	4957268	4956550	**718**	4957268	2974245	4957268	2974245	**1983023**
CI 628	4469007	4468329	4469007	4468329	**678**	4469007	2811164	4469007	2811164	**1657843**
CI 747	3250282	3249803	3250282	3249803	**479**	3250282	2038252	3250282	2038252	**1212030**
CI 828	1977327	1977046	1977327	1977046	**281**	1977327	1226203	1977327	1226203	**751124**
CI 839	4197790	4196963	4197790	4196963	**827**	4197790	2852065	4197790	2852065	**1345725**
CI 926	8548969	8547772	8548969	8547772	**1197**	8548969	5481321	8548969	5481321	**3067648**
CI 935	4470267	4469623	4470267	4469623	**644**	4470267	2792924	4470267	2792924	**1677343**
CI 1035	5070680	5070086	5070680	5070086	**594**	5070680	3181175	5070680	3181175	**1889505**
CI 1475	4049372	4048607	4049372	4048607	**765**	4049372	2617376	4049372	2617376	**1431996**
CI 1480	4149827	4149181	4149827	4149181	**646**	4149827	2536075	4149827	2536075	**1613752**
CI 1539	5218335	5217519	5218335	5217519	**816**	5218335	3365231	5218335	3365231	**1853104**
CI 1578	3784292	3783740	3784292	3783740	**552**	3784292	2532562	3784292	2532562	**1251730**
CI 1626	3434265	3433705	3434265	3433705	**560**	3434265	2130101	3434265	2130101	**1304164**
CI 1631	3579218	3578654	3579218	3578654	**564**	3579218	2187885	3579218	2187885	**1391333**
CI 1674	3930277	3929644	3930277	3929644	**633**	3930277	2348147	3930277	2348147	**1582130**
CI 1869	3610109	3609580	3610109	3609580	**529**	3610109	2344912	3610109	2344912	**1265197**
CI 1936	3353540	3352962	3353540	3352962	**578**	3353540	2321016	3353540	2321016	**1032524**

Table 2 | QUAST results for the in-house pipeline and BV-BRC pipeline with the number of contigs, L50, N50 and GC%.

	In-house pipeline				BV-BRC pipeline			
Sample ID	**number of contigs**	**L50**	**N50**	**GC%**	**number of contigs**	**L50**	**N50**	**GC%**
CI 11	35	5	440206	51.75	30	4	519577	51.75
CI 268	21	4	687363	53.78	20	4	687363	53.78
CI 507	18	2	753182	53.62	21	2	749785	53.63
CI 589	24	3	559107	51.98	28	2	915701	51.97
CI 619	16	1	2637237	53.81	19	2	1132453	53.81
CI 628	15	1	2637219	53.81	17	1	2637237	53.81
CI 747	17	2	681657	53.74	19	2	586194	53.74
CI 828	16	4	670628	53.81	16	1	2637219	53.81
CI 839	24	2	816950	51.48	31	2	816977	51.48
CI 926	42	5	335764	51.81	41	5	335771	51.81
CI 935	38	6	325351	51.71	39	6	325351	51.71
CI 1035	28	2	667709	56.23	36	2	667711	56.22
CI 1475	40	3	683829	53.88	46	3	683829	53.88
CI 1480	105	11	144033	51.92	112	11	194315	51.92
CI 1539	37	5	281203	53.73	38	6	281199	53.73
CI 1578	38	4	547363	51.79	45	4	547363	51.79
CI 1626	23	4	675014	53.59	26	4	675057	53.59
CI 1631	24	4	675246	53.59	25	4	675058	53.59
CI 1674	27	3	777958	52.02	33	4	468993	52.02
CI 1869	101	8	207722	53.18	99	7	295387	53.18
CI 1936	21	2	735625	53.7	21	3	735580	53.71

L50: The minimum number of contigs that cover half (50%) of the assembly.
N50: The length of the smallest contigs that together represent half (50%) of the assembly.
GC%: The percentage of Guanine (G) and Cytosine (C) nucleotides.

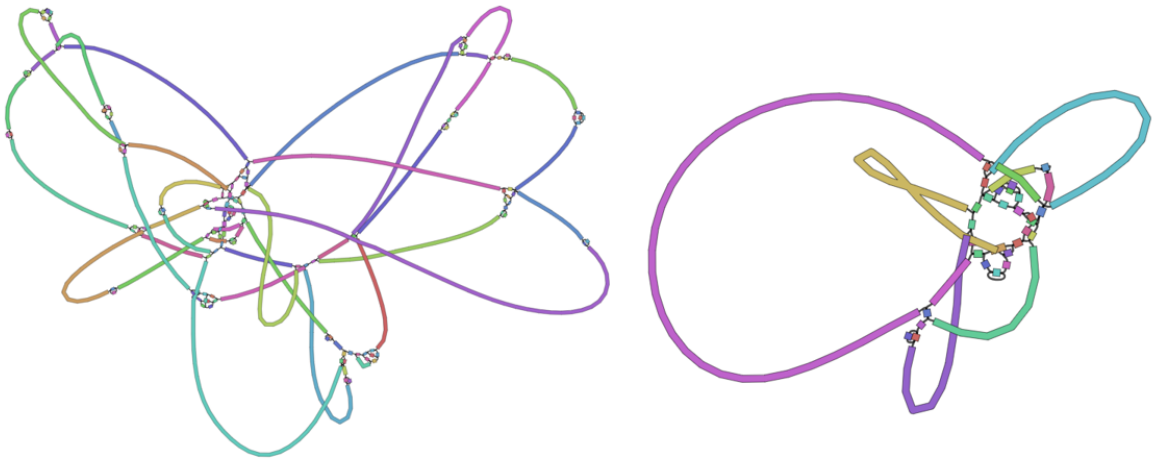

Figure 5 | Shows the Bandage (Bioinformatics Application for Navigating De novo Assembly Graphs Easily) plot generated with the genome assembly report for SPAdes and Unicycler for isolate sample ID CI507. (A) SPAdes show a denser network of smaller nodes and more edges, indicating a higher number of shorter contigs and potential gaps. (B) Unicycler shows fewer, larger nodes (contigs) and fewer edges, reflecting a more contiguous assembly.

Table 3 | Species identification for 21 *Citrobacter* isolates were identified using MALDI-TOF MS, whole genome alignment with JSpecies WS (ANIb% set at ~95-96%), and Kraken2 krona chart visualization. Lower identity percentages and genome fractions are highlighted in red to indicate potential limitations in the identification.

Sample ID	**MALDI-TOF MS**	**JSpeciesWS (in-house pipeline)**	**Kraken2 (BV-BRC)**	**QUAST Genome fraction%**	**JSpeciesWS (ANIb%)**	**Kraken 2 Krona chart %**
CI 11	*Citrobacter freundii*	*Citrobacter freundii*	*Citrobacter freundii*	83.625	98.38	97
CI 268	*Citrobacter koseri*	*Citrobacter koseri*	*Citrobacter koseri*	91.317	98.78	92
CI 507	*Citrobacter koseri*	*Citrobacter koseri*	*Citrobacter koseri*	90.840	98.70	35
CI 589	*Citrobacter freundii*	*Citrobacter freundii*	*Citrobacter freundii*	14.696	92.13	64
CI 619	*Citrobacter koseri*	*Citrobacter koseri*	*Citrobacter koseri*	92.746	98.71	45
CI 628	*Citrobacter koseri*	*Citrobacter koseri*	*Citrobacter koseri*	92.760	98.71	46
CI 747	*Citrobacter koseri*	*Citrobacter koseri*	*Citrobacter koseri*	92.843	98.72	46
CI 828	*Citrobacter koseri*	*Citrobacter koseri*	*Citrobacter koseri*	92.320	98.69	45
CI 839	*Citrobacter freundii*	*Citrobacter freundii*	*Citrobacter freundii*	19.003	92.11	5
CI 926	*Citrobacter freundii*	*Citrobacter freundii*	*Citrobacter freundii*	84.605	98.55	5
CI 1035	*Enterobacter bugandensis*	*Enterobacter bugandensis*	*Enterobacter bugandensis*	97.388	98.74	16
CI 935	*Citrobacter freundii*	*Citrobacter freundii*	*Citrobacter freundii*	85.582	98.49	3
CI 1475	*Citrobacter koseri*	*Citrobacter koseri*	*Citrobacter koseri*	93.148	98.99	35
CI 1480	*Citrobacter freundii*	*Citrobacter freundii*	*Citrobacter freundii*	84.522	98.46	10
CI 1539	*Citrobacter koseri*	*Citrobacter koseri*	*Citrobacter koseri*	93.543	99.02	30
CI 1578	*Citrobacter freundii*	*Citrobacter freundii*	*Citrobacter freundii*	85.140	98.63	12
CI 1626	*Citrobacter koseri*	*Citrobacter koseri*	*Citrobacter koseri*	98.503	99.99	42
CI 1631	*Citrobacter koseri*	*Citrobacter koseri*	*Citrobacter koseri*	98.498	99.98	43
CI 1674	*Citrobacter braakii*	*Citrobacter braakii*	*Citrobacter braakii*	92.375	98.51	3
CI 1869	*Citrobacter farmeri*	*Citrobacter farmeri*	*Citrobacter farmeri*	29.633	93.12	18
CI 1936	*Citrobacter koseri*	*Citrobacter koseri*	*Citrobacter koseri*	91.227	98.82	3

Table 4 | Number of antibiotic resistance genes identified using ResFinder (Centre for Genomic Epidemiology) for *Citrobacter* isolates.

Antibiotic resistance	***Citrobacter braakii* (n=1)**	***Citrobacter freundii* (n=5)**	***Citrobacter koseri* (n=11)**
Aminocyclitol	0	1	0
Aminoglycoside	0	1	0
Beta-lactam	1	5	11
Fosfomycin	0	1	4
Quinolone	1	0	0
Sulphonamide	0	1	0
Tetracycline	0	1	0
Trimethoprim	0	1	0

n= () number of isolates.

Table 5 | The table presents antibiotic resistance genes identified in *Citrobacter* isolates (n=17) using ResFinder v 4.4.2 from the Centre for Genomic Epidemiology.

Sample ID	**Species type**	**Sample culture type**	**Antibiotic class**	**Resistance gene**
CI 1674	*Citrobacter braakii*	Blood culture	Beta -lactam Quinolone	*blaCMY-82* *QnrB10*
CI 11	*Citrobacter freundii*	Urine	Beta -lactam	*blaCMY -48*
CI 926	*Citrobacter freundii*	Urine	Beta -lactam	*blaCMY -68*
CI 935	*Citrobacter freundii*	Urine	Beta -lactam Aminocyclitol Sulphonamide Tetracycline Trimethoprim	*blaCMY -48* *aadA1* *sul 2* *tetD* *dfrA1*
CI 1480	*Citrobacter freundii*	Urine	Beta -lactam	*blaCMY -75*
CI 1578	*Citrobacter freundii*	Urine	Beta -lactam Aminoglycoside Fosfomycin	*blaCMY -79* aaC(6') *fosA7*
CI 268	*Citrobacter koseri*	Urine	Beta -lactam	*blaMAL-1*
CI 507	*Citrobacter koseri*	Urine	Beta -lactam	*blaCKO -1*
CI 619	*Citrobacter koseri*	Urine	Beta -lactam Fosfomycin	*blaMAL-1* *fosA7*
CI 628	*Citrobacter koseri*	Urine	Beta -lactam Fosfomycin	*blaMAL-1* *fosA7*
CI 747	*Citrobacter koseri*	Wound	Beta-lactam Fosfomycin	*blaMAL-1* *fosA7*
CI 828	*Citrobacter koseri*	Urine	Beta -lactam Fosfomycin	*blaMAL-1* *fosA7*
CI 1475	*Citrobacter koseri*	Urine	Beta -lactam	*blaMAL-1*

CI 1539	*Citrobacter koseri*	Upper respiratory tract	Beta -lactam	*blaMAL-1*
CI 1626	*Citrobacter koseri*	Urine	Beta -lactam	*blaCKO -1*
CI 1631	*Citrobacter koseri*	Blood	Beta -lactam	*blaCKO -1*
CI 1936	*Citrobacter koseri*	Urine	Beta -lactam	*blaMAL-1*

Genes with a sequence identity of ≥90% to a reference gene in the database are reported. The table details the distribution of these genes across different antibiotic classes, sample types, and specific *Citrobacter* species identified by sample ID.

Table 6 | List of antibiotic resistance genes identified by CARD (BV-BRC) for *Citrobacter* isolates (n=17)

Classification	**Genes**	***Citrobacter braakii* (n=1)**	***Citrobacter freundii* (n=5)**	***Citrobacter koseri* (n=11)**
aminocoumarin resistance gene	*alaS* *cysB* *gyrB*	2	15	22
aminoglycoside resistance gene	*kdpE* *aaC(6')-If* *aadA*	1	7	0
beta-lactam resistance gene	*CMY-48* *CMY-68* *CMY-75* *CMY-79* *CMY-82*	1	5	0
Efflux pump conferring antibiotic resistance gene	*PhoP* *marR* *soxR* *soxS* *baeR* *baeS* *cpxA* *cpxR* *CRP* *EmrR* *H -NS* *acrB* *acrD* *embed* *mdtB* *mdtC* *mdtG* *mdtH* *msbA* *patA*	23	142	296

	YojI *marA* *ramA* *sdiA* *robA* *rosA* *mdtM* *mdtN* *mdtP*			
fluoroquinolone resistance gene	*gyrA* *gyrB* *parC* *parE* *mfd* *QnrB10*	6	25	55
fosfomycin resistance gene	*GlpT* *murA* *UhpT*	3	15	33
rifampin resistance gene	*rpoB*	1	5	11
sulfonamide resistance gene	*folP*	1	5	11
nitrofuratoin resistance gene	*nfsA*	0	2	0
trimethoprim resistance gene	*dfrA1*	0	1	0
streptothricin resistance gene	*sat-1*	0	1	0
aminoglycoside resistance gene+ tetracycline resistance gene	*fyuA*	0	0	16
Total	705	223	38	444

Table 7 | Distribution of virulence factor genes across species type and classification of gene group identified by VFDB (BV-BRC) for *Citrobacter* isolates (n=17).

Species	**Adherence**	**Endo toxin**	**Entero toxin**	**Immune evasion**	**Iron uptake**	**Secretion system**	**Serum resistance**	**Total genes/ Species**
Citro bacter freundii **(n= 5)**	8	0	0	0	10	0	3	21
Citro bacter koseri **(n= 11)**	38	3	8	0	217	1	0	267
Citro bacter braakii **(n=1)**	2	0	0	8	2	0	0	12
Total genes/ group	48	3	8	8	229	1	3	300

Table 8 | This table presents the results of plasmid identification in nine *Citrobacter* isolates using PlasmidFinder v2.0, at the Centre for Genomic Epidemiology.

Sample ID	**Species**	**Plasmids identified**	**Number of plasmids**	**Identity ≥ 95%**	**Accession number**	**Source organism**
CI 11	*Citrobacter freundii*	*IncFII (SARC14)*	1	95.07	JQ418540	*Salmonella enterica*
CI 619	*Citrobacter koseri*	*ColRNAI*	1	100	DQ298019	*Klebsiella pneumoniae*
CI 628	*Citrobacter koseri*	*ColRNAI*	1	100	DQ298019	*Klebsiella pneumoniae*
CI 747	*Citrobacter koseri*	*ColRNAI*	2	100	DQ298019	*Klebsiella pneumoniae*
		IncX10		99.81	OQ821188.1	*Citrobacter freundii*
CI 1480	*Citrobacter freundii*	*IncFIB(K)*	3	98.75	JN233704	*Klebsiella pneumoniae*
		IncHI1B(pNDM-CIT)		100	JX182975	*Citrobacter freundii*
CI 1539	*Citrobacter koseri*	*Col440II*	2	97.52	CP023921	*Klebsiella pneumoniae*
		ColRNAI		100	DQ298019	*Klebsiella pneumoniae*

CI 1626	*Citrobacter koseri*	*ColRNAI*	1	100	DQ298019	*Klebsiella pneumoniae*
CI 1631	*Citrobacter koseri*	*ColRNAI*	1	100	DQ298019	*Klebsiella pneumoniae*
CI 1936	*Citrobacter koseri*	*IncX4*	1	98.88	FN543504	*Citrobacter rodentium*
		IncFII (Yp)		99.57	CP000670	*Yersinia pestis*

For each isolate, the table details the sample ID, bacterial species, number of identified plasmids, sequence identity (≥95%), accession number, and source organism of each detected plasmid.