Degree project

UNIVERSITY OF SKÖVDE

The battle against sepsis: exploring the genotypic diversity of *pseudomonas* and *proteus* clinical isolates

Master Degree Project (120 credits) in Systems Biology

Second Cycle 45 credits
Spring term Year 2023

Student: Suud Ahmed

Email: a21suah@student.his.se

Supervisor: Magnus Fagerlind

Email: magnus.fagerlind@his.se

Examiner: Sanja Jurcevic

Email: sanja.jurcevic@his.se

Abstract

Sepsis is a dangerous and potentially fatal condition that has a mysterious origin, underscoring the significance of prompt and accurate diagnosis and treatment. Bacterial whole-genome sequencing, which is widely used in clinical microbiology, stands at the forefront of sequencing technologies, particularly to combat sepsis. The aim of this thesis is to improve sepsis treatment by examining the genetic characteristics and drug resistance patterns of the common sepsis-causing bacteria *Pseudomonas* and *Proteus spp*., by analyzing the whole-genome sequencing data of bacterial isolates using an in-house-developed pipeline. The result was compared with a commercial cloud-based platform from 1928 Diagnostic (Gothenburg, Sweden), as well as the results from a clinical laboratory. Using Illumina HiSeq X next-generation sequencing technology, whole-genome data from 88 isolates of *Pseudomonas* and *Proteus spp.* was obtained. The isolates were obtained during a prospective observational study of community-onset severe sepsis and septic shock in adults at Skaraborg Hospital in Sweden's western region. The collected isolates were characterized using approved laboratory techniques, such as phenotypic antibiotic susceptibility testing (AST) in accordance with EUCAST guidelines and species identification by MALDI-TOF MS analysis. The species identification result matched the phenotypic method, with the exception of two isolates from Pseudomonas samples and four isolates from Proteus samples. When benchmarking the in-house pipeline and 1928 platform for *Pseudomonas spp*., predicted 97% of the isolates were resistant to at least one class of the tested antibiotics, of which 94% shows multi-drug resistance. In phenotypes, 88% of the isolates had at least one antibiotic resistance future, of which 68% shows multi-drug resistance. The most prevalent sequence types (STs) identified were ST 3285 and ST111 (9.3%) and ST564 and ST17 (6.98%) each, and both pipelines accurately predicted the number of multilocus types. The in-house pipeline reported 9820 *Pseudomonas* virulence genes, with PhzB1, a metabolic factor, being the most common gene. It was discovered that there was a significant correlation between the virulence factor gene count and the multilocus sequence typing (MLST) ($p = 0.00001$). With a Simpson's Diversity Index of 0.98, the urine culture specimens showed the greatest ST diversity. Plasmids were detected in twelve samples (20.93%) in total. In general, this study provided a detailed description of the bacterial future for *Pseudomonas* and *Proteus* organisms using WGS data. This research shows the applicability of the in-house and 1928 pipelines in the identification of sepsis-causing organisms with accuracy. It also showed the need for an organized and easy-to-use international pipeline to implement and analyze WGS bacterial data and to compare it with laboratory results as needed.

Popular Scientific Overview

Infection prevention is considered one of the most common ways to keep humans healthy. As humans, numerous diseases can affect us; one of the most common is sepsis, also referred to as "blood poisoning." This condition is brought on by the body's exaggerated response to infection by its defense systems, which use more than necessary protection, leading to self-destruction. This condition will harm the majority of the vital organs. This will manifest in different symptoms, like fever, lowering blood pressure, and even the possibility of death. This devastating feature of sepsis, which has a widespread impact all over the world, makes it in need of accurate and timely intervention. The ultimate aim of many researchers who study biology is to identify the infection-causing organisms behind sepsis and treat them as quickly as possible, but this have been challenging. Traditional techniques take a long time and frequently miss early-stage organisms. However, a revolutionary technique known as whole genome sequencing (WGS) has recently come to light. Thanks to this modern technology, the genetic code of bacteria like *Pseudomonas* and *Proteus* species can be fully decoded within minutes. These bacteria often coexist with us but can transform into harmful invaders, causing severe infections like sepsis. This research aimed to explore the "secret code" of these bacteria using WGS. By understanding the genes that enhance their different futures or make them resistant to antibiotics, one can essentially have a map that shows where the enemy's fortifications and weapons are located. This knowledge allows us to create specialized defenses against them. This study involved studying the genetic features of 88 samples of *Pseudomonas* and *Proteus spp.* Using both an in-house-developed approach, having many software's and a commercial platform developed with a lot of biological functions, the 1928 combined with a comparison of the result to traditional clinical laboratory results. The results were fascinating. With the exception of a few samples, there was an excellent match between the organisms detected by the in-house organized software, the 1928 and the methods used in traditional labs. The identification of prevalent bacterial epidemiological areas of distribution and accurate prediction of their specific antibiotic resistance gene was an important achievement. The study was able to find a gene that makes them resistant to certain antibiotics among the isolates, with the pipelines somehow predicting different percentages of resistance to various antibiotics. This information is vital as it helps in the treatment of the infected patient with sepsis or other infectious condition. This test also found a lot of harmful genes in the *Pseudomonas* bacteria, with some types being more frequent than others. It was able to find a strong link between the number of these harmful genes and certain genetic features, suggesting that these features might make the bacteria more effective at causing infections. The future implications of this study are vast. By revealing the genetic features and resistance mechanisms of *Pseudomonas* and *Proteus spp.*, it will create a pathway to develop an effective treatment strategy. Tools like 1928 and the in-house approach are essential in this development, using next-generation sequencing technology and driving researchers' efforts towards a future where sepsis becomes a less terrifying and more manageable conditions. But acknowledging the challenges remains at the forefront. While these tools are powerful, there are obstacles to their widespread use. The time and cost involved in sequencing and analyzing the data are significant. In addition, there is not a lot of trained personal to use these mechanisms. And sometimes there will be miss-diagnosis due to certain conditions. However, with further development, training and investment, these barriers can be overcome. In conclusion, the world of medical research is on the edge of a new era where condition like sepsis can be understood and treated with unprecedented accuracy and efficiency. The combination of whole genome sequencing data with powerful tools like 1928 and in-house-developed pipelines has opened doors that were previously unimaginable. The fight against sepsis and other severe infections has been revolutionized, and this study contributes valuable insights to this rapidly advancing field. The path is now set; the journey continues.

Table of Contents

Abbreviations:

AMR	Antimicrobial Resistance
AST	Antibiotic Susceptibility Testing
ANI	Average Nucleotide Identity
CARD	The comprehensive Antibiotic Resistance genes Database
CGE	Center for Genomic Epidemiology
chewBBACA	Comprehensive and Highly Efficient Workflow for BSR-Based Allele Calling Algorithm
cgMLST	Core Genome Multi-Locus Sequence Typing
EUCAST	European Committee on Antimicrobial Susceptibility Testing
ESBL	Extended-Spectrum Beta-lactamase
ESKAPE	*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa,* and *Enterobacter* species
GC content	Guanine – Cytosine content
MLS	Macrolides, Lincosamide, and Streptogramin
ME	Major Error
MALDI-TOF MS	Matrix-assisted laser desorption ionization-time of flight mass spectrometry
MLST	Multi-Locus Sequence Typing
MDR	Multi Drug Resistance
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
QUAST	Quality Assessment Tool for Genome Assembly
ST	Sequence Type
SBS	Sequencing by Synthesis
SOFA	Sequential Organ Failure Assessment score
SNP	Single Nucleotide Polymorphism
SIRS	Systemic Inflammatory Response Syndrome
VF	Virulence Factor
VME	Very Major Error
1928	1928Diagnostic platforms

Introduction

Sepsis

Sepsis is a serious condition resulting from an overactive immune response to infection, which can lead to widespread inflammation, tissue damage, and organ dysfunction (Rhodes et al., 2017; Bone et al., 1992). It can escalate to septic shock, characterized by issues in the circulatory system and cellular metabolism, significantly increasing mortality (Cohen et al., 2015; Dellinger et al., 2013). Diagnostic criteria for sepsis have evolved many times, starting with Sepsis-1 in the 1990s, based on SIRS which relied on clinical signs and symptoms but often led to overdiagnosis and unnecessary treatment (Bone et al., 1992; Dellinger et al., 2013). Sepsis-2, developed in 2001, based on the Systemic Inflammatory Response Syndrome (SIRS), which incorporated laboratory markers of inflammation for better differentiation but had challenges in patients with chronic illnesses or conditions (Levy et al., 2003; Dellinger et al., 2013). To address SIRS limitations, Sepsis-3 was introduced in 2016, defining sepsis as ''a life-threatening organ dysfunction caused by a dysregulated host response to infection'' (Seymour et al., 2016; Singer et al., 2016; Vincent et al., 2015). It included the Sequential Organ Failure Assessment (SOFA) and quick SOFA (qSOFA) scores for improved diagnostic accuracy and early detection (Seymour et al., 2016; Singer et al., 2016). Despite diagnostic advancements, sepsis remains a significant public health concern, with millions of cases globally each year and with one death every 2.8 seconds, especially in developing countries with limited healthcare resources (Kempker & Martin, 2020; Rudd et al., 2020; Vincent et al., 2013). Anyone can develop sepsis, which may be acquired from any infection. Sepsis present as Fever, shortness of breath, feeling very cold, extreme pain, and the like. Sepsis-associated mortality exceeds that of conditions like ST-segment elevation myocardial infarction (Shah et al., 2015).

In Sweden, Sepsis-3 is the current criteria for diagnosing sepsis according to Brink, M. (2018) and where sepsis is one of the commonest reasons of ICU admission and mortality, particularly in critically ill patients (Lengquist et al., 2020; Brink et al., 2018). The Skaraborg area of southern Sweden had a high incidence rate of community-onset sepsis at 838 cases per 100,000 people per year, which is among the highest reported in Sweden (Ljungström et al., 2017; Ljungström et al., 2019). About 11% of hospital admissions met the criteria for septic shock, with a 33% mortality rate while in the hospital (Lengquist et al., 2020). To improve patient outcomes and lessen the impact of sepsis on healthcare systems, it is essential to understand the genotypic characteristics of bacterial pathogens that cause sepsis, such as *Pseudomonas* and *Proteus spp.* By characterizing the genotypes of these pathogens, researchers can better understand the mechanisms of pathogenesis and identify targets for the development of new treatments and interventions. Causative agents of sepsis can be bacteria, viruses, fungi, or parasites (Chun et al., 2015). According to Mayr et al. (2013), bacteremia due to Gram-negative bacteria was linked to a greater mortality rate than Gram-positive bacteria. In the same study, coagulase-negative *staphylococci* and *Escherichia coli* (*E. coli*) were the most frequent causes of bacteremia, which is a bloodstream infection, while *Pseudomonas aeruginosa* (*P. aeruginosa*), a gram-negative bacterium, had the greatest mortality rate of all. *P. aeruginosa*, is one of the sepsis-causing bacteria with numerous virulence genes (Alamu et al., 2022; Talbot et al., 2006). *Proteus mirabilis* (*P. mirabilis*) is an ESBL (extended spectrum beta-lactamase)-forming bacterium, which is the species that is mostly isolated from *Proteus spp*. clinical samples (Schaffer & Pearson, 2015). These two gram-negative bacteria possess features of resistance to different antibiotics (Chen et al., 2015; Shelenkov et al., 2020). For now, there is no known therapeutic treatment that targets the exact mechanisms of sepsis. Fluid resuscitation and oxygen supplementation, combined with timely antibiotic therapy, are critical for reducing the severity of the illnesses (Evans, 2018; Thompson et al., 2019). Broad-spectrum antibiotics having one or more antimicrobials features to cover all possible pathogens are preferable (Thompson et al., 2019). Speaking of sepsis, the most worrying issue is antibiotic resistance, which is widely prevalent throughout the world (Frieri et al., 2017). According to Martnez (2008) and Collignon & McEwen (2019), the definition of antibiotic resistance is the

capacity of a particular bacterium when it develops a way to overcome the effect of antibiotics and limit its effect. Antibiotic resistance, including multidrug resistance (MDR), emerges because of improper or inappropriate antibiotic use (Pradipta et al., 2013). MDR can limit patients' access to the proper antibiotics, which may have an adverse effect and increase mortality (Frieri et al., 2017). In several prevalent infections, new antibiotic resistances are developing and leading to treatment failures (Frieri et al., 2017).

Identifying the sepsis-causing microorganisms

Blood cultures, the current gold standard for diagnosing diseases, are typically used to determine sepsis, a dangerous condition frequently accompanied by bloodstream infections (Opota et al., 2015). A`Tziolos & Giamarellos-Bourboulis (2016) and Vincent et al. (2015) note that this method has drawbacks, such as a lengthy processing time, the potential for false results as a result of ongoing antibiotic treatment or sample contamination, and difficulty in identifying specific pathogen types. Although this method's sensitivity is not very high, it does allow for the evaluation of antibiotic susceptibility. Despite its limitations in simultaneously identifying multiple species, polymerase chain reaction (PCR) is being used more frequently to diagnose infectious diseases (Smith et al., 2009), including sepsis (Li & Yan, 2021; Ruiz-Villalba et al., 2017). For instance, during the COVID-19 pandemic, it was crucial in identifying emerging pathogens (Sule & Oluwayelu, 2020). Matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) is a different technique that is gaining popularity because it is quick and accurate at identifying different bacteria. According to some studies (Kawahara-Matsumizu et al., 2018; Poonawala et al., 2018; Singhal et al., 2015), it may mistakenly identify certain types of microorganisms, particularly in sepsis patients. Antimicrobial susceptibility testing (AST) is vital for identifying bacterial resistance before broad-spectrum antibiotics are administered (Chun et al., 2015; Ljungström, 2017). By identifying the minimum inhibitory concentration (MIC) of antibiotics required to effectively treat the infection. This AST based procedure, which includes tests like disk diffusion, broth dilution, and the E-test, optimizes sepsis treatment and management (Chun et al., 2015). All this approach highlights the need for new dynamic approach for Sepsis diagnosis and treatments.

Pseudomonas aeruginosa

Pseudomonas aeruginosa is an aerobic, chemoheterotrophic, motile, rod-shaped bacterium with a 5.5-7 Mbp genome, a GC content of 65–67% and a variable number of plasmids that was first discovered by botanist Walter Migula in the late 19th century (Zen and Ussery, 2012; Klockgether & Tümmler, 2017; Nicas, T., & Hancock, 1983). Its genome, which is one of the biggest among bacteria and is divided into core and accessory parts, both contribute to the virulence of various strains (Kung et al., 2010; Ozer et al., 2014; Parkins et al., 2018; Özen et al., 2012). All body parts are susceptible to *P. aeruginosa* infections, which can result in acute and chronic infections, the latter of which is seen in diseases like cystic fibrosis (Valentini et al., 2018; Riquelme et al., 2020). This bacterium has become the leading multi-drug resistant bacterium, rendering most antibiotics ineffective (Kerr and Snelling, 2009; Kung et al., 2010; Horcajada et al., 2019; Poole, 2011; Thaden et al., 2017; Rahme et al., 1995; Reynolds, 2021). It belongs to the ESKAPE pathogen group, which is resistant to common antimicrobial treatments (Pendleton et al., 2013). According to Tacconelli et al. (2018), new diagnostic methods may significantly enhance the management of *P. aeruginosa*. Greipel et al. (2016) examined loci for antimicrobial susceptibility and resistance and discovered high-frequency variants in several genes that suggested the presence of MDRs. Additionally, the MultiLocus Sequence Analysis (MLSA) has been suggested for *Pseudomonas* detection, focusing on several housekeeping genes (Girard et al., 2020, 2021). High-resolution typing, like multilocus sequence typing (MLST) or the more advanced core genome MLST (cgMLST), is often necessary for these ubiquitous bacteria (Tönnies et al., 2021). *P. aeruginosa* employs several survival mechanisms against antibiotics, including intrinsic, acquired, and adaptive resistance (Breidenstein et al., 2011). Its biofilm-forming ability contributes significantly to its antibiotic resistance, as does its inherent resistance to several drugs such as carbapenem, ceftazidime,

ciprofloxacin, aminoglycosides, and fosfomycin (Fagerlind et al., 2012; Horcajada et al., 2019; Walters et al., 2019; Wong et al., 2014; Worth et al., 2015). Despite extensive research on *P. aeruginosa*, the current understanding of this bacterium is insufficient for developing novel, effective therapeutic approaches (Qin et al., 2022). Key questions include how invasive strains develop drug resistance, how many unidentified virulence factors exist, and the mechanisms behind the rise in antibiotic resistance (Qin et al., 2022). These areas require further exploration and research to successfully combat this resilient pathogen.

Proteus mirabilis

P. mirabilis is a gram-negative facultative anaerobe bacillus that belongs to the *Morganellaceae* family (Marcon et al., 2019; Shelenkov et al., 2020). It can self-extend and exhibit swarming behavior, which enables it to adhere to and move along surfaces like catheters, intravenous lines, and other medical equipment (Mobley & Belas, 1995; Mathur et al., 2005; Nicolle, 2005; Jacobsen et al., 2008). Bacteremia induced by *P. mirabilis* is most common after a UTI (urinary tract infection) or CAUTI (catheters-associated UTI), and both bacteremia and sepsis caused by *P. mirabilis* have a high fatality rate (Clarke et al., 2019; Hooton et al., 2010). The community-acquired infection by the virulent *P. mirabilis* is the cause of 90% of illnesses that are induced by the *Proteus* genus (Armbruster et al., 2018; Bush, 2010; Nordmann et al., 2011). *P. mirabilis* can also cause skin and respiratory tract infections, as well as infections of several other organs (Mobley & Belas, 1995; Mathur et al., 2005; Nicolle, 2005; Jacobsen et al., 2008). The virulence factors produced by *Proteus spp.* include an S-form lipopolysaccharide (LPS) with a long-chain O-polysaccharide (OPS), termed the O antigen (Yu et al., 2017). The variability of the OPS structure is the foundation for the serotyping of these bacteria. Currently, there are 83 O serogroups included in the *Proteus* serological classification system, which is continuously expanded and updated. Chemical examination of the O77-O81 antigens found both normal and unusual elements in their structures, such as Kdo in the O79-polysaccharide and other elements known from the *Proteus* strain (Zabotni et al., 2018; Arbatsky et al., 2013). It is coated with phosphocholine (ChoP), which is found in O18 serogroup strains, and which protects this bacterium from innate and adaptive immune system responses and also modulates interactions with host proteins involved in human infection, giving it a distinctive appearance to *P. mirabilis* (Czerwonka et al., 2021; Fudala et al., 2003; Zabłotni et al., 2018). The coding sequences and chromosomal locations of previously described virulence factors were identified by genome annotation (Yu et al., 2017). For the extensively researched strain HI4320 (serogroup O28) of *P. mirabilis*, isolated from the urine of a patient with a long-term indwelling urinary catheter, the first complete genome was characterized (Yu et al., 2017). In later studies, it was demonstrated that the *Proteus* serogroups might be genetically distinct based on the sequences of each O antigen biosynthesis cluster; however, data on the other serogroups was lacking (Yu et al., 2017). Numerous studies have been conducted on *P. mirabilis*' antibiotic resistance (Bush, 2010; Nordmann et al., 2011). Polymyxin and tetracycline resistance are innate in *P. mirabilis*, and the future of MDR includes beta-lactams, aminoglycosides, fluoroquinolones, phenicol, streptothricin, and trimethoprim-sulfamethoxazole (Chen et al., 2015).

Next-generation sequencing (NGS)

The "massively parallel sequencing" (MPS) method of DNA sequencing, also known as "next-generation sequencing" (NGS) (Lee et al., 2022; Hu et al., 2021; Behjati & Tarpey, 2013), has revolutionized genomic research. With their high throughput and ability to multiplex samples, which is the addition of numerous bar codes to separate samples before sequencing, NGS technologies have greatly lowered the cost of sequencing (Hu et al., 2021; Mardis, 2017). The main advantage of MPS in microbiology is the replacement of more traditional methods of identification, like morphology, staining characteristics, and metabolic properties, with a genetic description. The genomes of microorganisms provide extensive information about their drug sensitivity, antibiotic resistance, and virulence (Besser et al., 2018; Li, B., & Yan, T., 2021). Prenatal diagnostics, sepsis, organ transplantation, and oncology have all significantly benefited from the

use of NGS (Lee et al., 2022; Schütz et al., 2017; Ulrich & Paweletz, 2018). A number of research studies are currently using NGS-based analysis to investigate the causal microorganisms in patients with respiratory, digestive, and central nervous system infections (Lee et al., 2022; Chiu & Miller, 2019; Joensen et al., 2017; Mizrahi et al., 2017). The main difficulty in NGS-based research is separating contamination from infection (Dargère et al., 2018; Lee et al., 2022). NGS-based diagnostic testing has several advantages over traditional blood culture, including the speed, accuracy and ability to distinguish between bacterial, viral, and fungal pathogens and rationality in the combination of quantitative values and statistical significance calculation (Hu et al., 2021; Chiu & Miller, 2019). This ultimately leads NGS results to have a higher sensitivity and specificity than blood cultures (Lee et al., 2022; Chiu & Miller, 2019).

Illumina technology for sequencing

A prominent technology in NGS is Illumina sequencing, which uses the reversible termination sequencing by synthesis (SBS) technique to generate short reads (Hu, T. et al., 2021; Pereira et al., 2020). The SBS technique involves fragmenting DNA or cDNA, attaching adapters, and adhering each fragment to a flow cell surface to produce dense clusters of double-stranded DNA. Laser excitation and imaging are then used to determine the attached fluorescent dye (Pereira et al., 2020; Dahui, 2019) (Figure 1). Illumina sequencing boasts an accuracy of 99.7% and has advantages such as high throughput, a low base-level error rate, and the capability for paired-end sequencing, which provides more data and larger sequence reads (Goodwin et al., 2016; Pereira et al., 2020; Ambardar et al., 2016; Reuter et al., 2015). Moreover, the Swedish biotech company 1928 seeks to promote infection control globally by optimizing DNA sequencing technology. Their cloud-based platform supports the analysis of 13 bacterial species, until this thesis is published, which are responsible for most hospital-acquired illnesses, offering different analysis processes depending on the pathogen and providing results for quality control, species identification, outbreak tracing, antibiotic resistance profiles, epidemiological typing, and virulence factor identification (1928Diagnostic, 2022; Mahmoud, 2021; Lember, 2021).

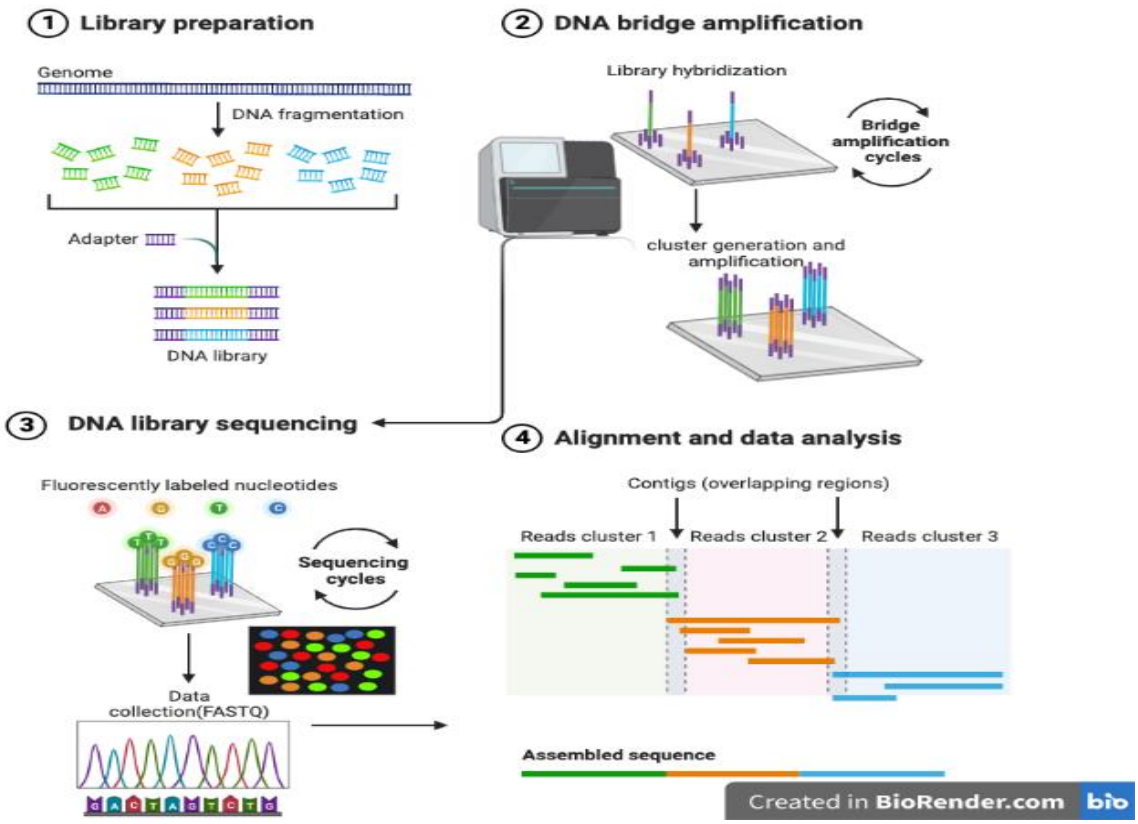

Figure 1 shows the Illumina SBS procedure, which begins with the extraction of genomic DNA from different sources like blood or urine (1). A DNA library is then created by fragmenting this DNA and attaching adapters with barcodes (2). By synthesizing a strand of DNA that is complementary to the template DNA on a flow cell, the DNA sequence of this library is determined. Bridge amplification, a type of PCR, enriches the DNA template and improves the detection signal for sequencing. This procedure, which is essential for cluster generation, guarantees that there will be enough raw material for sequencing. (3). Fluorescently labeled nucleotides are incorporated into the growing DNA strand, and their sequence is recorded. This produces a raw data that is subjected to bioinformatics analysis and alignment, enabling the discovery of AMR genes and other helpful insights like gene expression studies.

Aim

By gaining a thorough understanding of the genetic traits and resistance patterns of *Pseudomonas* and *Proteus spp.*, two commonly occurring sepsis-causing agents, this thesis aims to improve early detection and treatment of sepsis. The accomplishment of a number of specific objectives is required to achieve this aim. Firstly, whole-genome sequencing (WGS) data from *Proteus* and *Pseudomonas* isolates obtained from patients with sepsis was to be analyzed using an in-house-developed bioinformatic pipeline. In order to complete this analysis and benchmark the result, the data had to be uploaded into an autonomous pipeline known as the 1928 Tool. Exploring these bacterial WGS data such as antimicrobial resistance (AMR) genes, virulence factors, and epidemiological relatedness and comparing the predicted AMR genes with the phenotypic AST data was the area of focus. Utilizing statistical software, such as RStudio, the findings were to be interpreted accordingly. The study focused specifically on two types of bacteria: *Pseudomonas* and *Proteus*. An in-depth look to these organisms was planned, which take at account of, their diversity, their disease-causing potential, and their characteristics, irrespective of their drug-resistance profiles. Through these clearly defined objectives, the ability of identifying and characterizing sepsis causing microorganism through genetic insights into causative bacteria in short may be effectively addressed and which will be an input in the way of fighting drug resistance occurrences in the society.

Materials and Methods

Isolates collection

Pseudomonas spp. and *Proteus spp.* clinical isolates used in this study were collected as part of the prospective observational "Sepsis Study Skaraborg" conducted in Skaraborg hospital from September 8, 2011, to June 7, 2012, in collaboration with Unilabs and the systems biology center at Skövde University in the Västra Götaland region in southwest Sweden (Ljungström et al., 2017; Ljungström et al., 2019). All adult patients (≥18 years old) who were admitted with a diagnosis of confirmed or suspected community-onset sepsis or septic shock from the emergency department included in this study. Intravenous antibiotics administration within 48 hours of admission were the other criteria to be able to include in the study (Ljungström et al., 2017; Ljungström et al., 2019). Before administering empirical antibiotics, 1800 pathogenic bacterial isolates were gathered from sampling sites, like blood, wound, urine, and the upper respiratory tract. Colonial material was transferred to MicrobankTM vials (Pro-Lab Diagnostics, Ontario, Canada) and stored at -80 °C, where these isolates were cooled and preserved till they further needed (cryopreserved) at the time of recovery (Ljungström et al., 2017).

Species identification

All isolates were identified as *P. aeruginosa* (n = 45) and *P. mirabilis spp.* (n = 43) by standard microbiological techniques, with culture and MALDI-TOF MS DB-4110 (Bruker Daltonics, Germany) and included in this study. Spectral scores above 2.0 were used as a cut-off for correct species identification. At the time of the study, the Bruker microorganism database MBT Compass Library DB-4110 (Bruker Daltonics, Germany), released in April 2011, was used for species identification (Shemirani et al., 2023). Out of the total *Pseudomonas* samples isolated, 15.6% were

obtained from blood culture, 11.1% were obtained from upper respiratory culture, 51.1% were obtained from urine culture, and 22.2% were obtained from wound culture. Among the 43 samples examined for *Proteus*, the predominant culture type was also urine culture, accounting for approximately 48.83% of the total. Blood cultures (both aerobic and anaerobic) comprised about 27.91% of the total, while wound cultures represented 16.28%, and upper respiratory cultures contributed to 6.98% of the samples.

Antibiotic susceptibility testing

Antibiotic susceptibility testing (AST) for the *Pseudomonas* and *Proteus spp.* Isolates were checked in vitro in the clinical microbiology lab at Unilabs, Skövde. AST determination implement the disc diffusion method on Mueller-Hinton media in according to European Committee on AST (EUCAST) guidelines (www.eucast.org). The phenotypic AST result sample type was used to determine which antibiotics would be good to test for certain bacterial isolates (Shemirani et al., 2023).

DNA extraction

Genomic DNA extraction was performed for *Pseudomonas spp.* and *Proteus spp.* isolates. The extraction of DNA was done from pure cultures by the MagNA Pure 96 DNA and Viral NA Small Volume Kit (Roche Diagnostics, Switzerland) according to the Pathogen Universal 200 procedure on a MagNA Pure 96 instrument (Shemirani et al., 2023). DNA concentrations were measured using Qubit 3.0, purities were determined by a NanoDrop spectrophotometer (Thermo Fisher Scientific, USA), and the DNA samples were stored at -20 °C until sequenced (Saxenborn et al., 2021; Mahmoud, 2021; Ljungström et al., 2017).

NGS data generation

The genomic DNA samples were extracted from *Pseudomonas spp.* and *Proteus spp.* Isolates were sequenced using the Illumina HiSeq X platform at SciLifeLab in Solna, Sweden, using the high-throughput protocol for bacterial genomes. NexteraXT libraries and quality control were performed according to the manufacturer's protocol (Illumina, San Diego, CA). The resulting raw FastQ files were uploaded and stored as zipped files (.gz) at Skövde University for further analysis (Saxenborn et al., 2021; Mahmoud, 2021). The raw data have information about sequencer identifiers, reads, and quality scores. All the *P. aeruginosa* (n = 45) and *P. mirabilis spp.* (n = 43) Isolates that were identified using phenotypic methods underwent WGS analysis. In total, 88 isolates were included in this study.

In-house developed pipeline

WGS data bioinformatics analysis was done using, an in-house-developed pipeline (Figure 2) (Shemirani et al., 2023; Saxenborn et al., 2021). Which involved all standard procedures for analyzing the NGS data. This pipeline was designed to perform quality control (both before and after trimming and assembly), trimming, assembly, and functional annotation of reads (Shemirani et al., 2023; Saxenborn et al., 2021; Mahmoud, 2021). The analysis tools used in the pipeline were selected based on the 1928 tools to enable comparison of the results. Figure 2 presents a schematic representation of the pipeline.

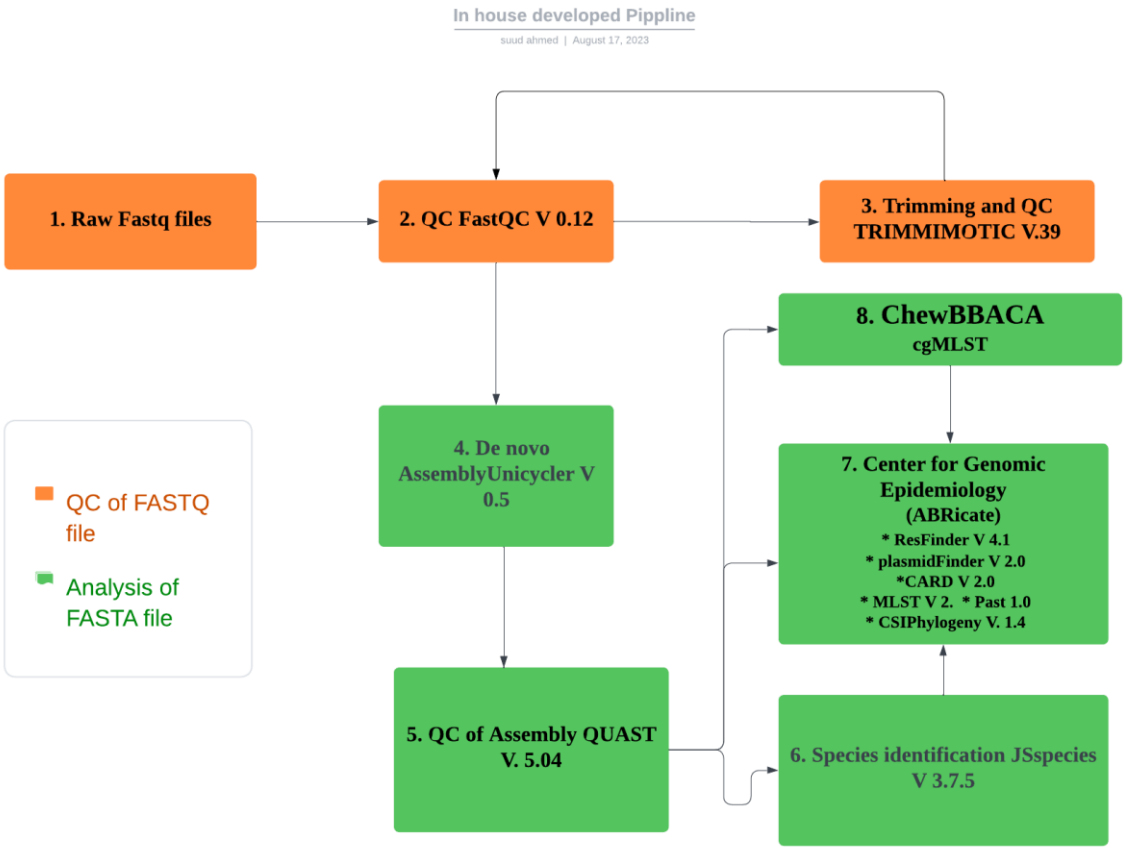

Figure 2: This figure illustrates a step-by-step method developed for analyzing of *Pseudomonas* and *Proteus* bacteria isolates WGS data. Starting with the initial untrimmed FastQ files (1), their quality is first checked using a program called FastQC (2). Next, any unnecessary parts are removed, and the quality is enhanced using Trimmomatic software (3). These refined data pieces are then assembled into larger units (called contigs) with a tool named Unicycler (4). The quality of these contigs is assessed with QUAST (5). For identifying the specific species of the bacteria, the JSpeciesWS online service (6) is employed. Afterward, specific features and functional annotation of reads unique to these two bacteria, *Pseudomonas* and *Proteus spp*. isolates are searched using the tools from the Center for Genomic Epidemiology (CGE) (7). Finally, chewBBACA tools, a multi-function tool (8) used for identifying the total number of loci.

This in-house developed pipeline (Figure 2), which included the steps and software's used for analyzing the sequenced data, such as quality control, trimming, assembly, and functional annotation of reads. The FastQ files were subjected to primary quality control using FastQC v. 0.11.5 (Babraham Bioinformatics, 2023). Trimmomatic v. 0.36 (Bolger et al., 2014) using these codes for trimming (Appendix I, Figures 1–2) was done for adapter removal and quality trimming, with a sliding window of size 4 and a minimum quality of 20. Furthermore, the HEADCROP parameter cut the first 12 bases, and reads shorter than 30 bp were eliminated. Next, the FastQ files were built into contigs using the Unicycler assembly program for Linux v. 0.4.8 (Ubuntu 5.8.0-45 generic; Wick et al., 2017; Mahmoud, 2021). The Quality Assessment Tool for Genome Assemblies, QUAST v. 4.6.0 (Gurevich et al., 2013; Wick et al., 2017), was used to evaluate the assembled contigs' quality. Quality was evaluated in QUAST using the default settings and with their respective reference sequences, the genomes of *P. aeruginosa* PAO1 (NCBI accession number NC_002516.2) and *P. mirabilis* HI4320 (NCBI accession number NC_01055.4), received from the NCBI. Species identification was done on the Fasta files using the free web service JSpeciesWS (http://jspecies.ribohost.com/jspeciesws/), which measured the probability of a group of genomes belonging to the same species (in comparison to a reference genome) using the Blast+ average nucleotide identity (ANI) method, with genomes exhibiting a similarity of ≥ 95% ANI

classified as the same species and those with < 95% as different species (Richter et al., 2015). Based on this genome assembly, six samples from different species were excluded from downstream analysis. The CGE tools (https://www.genomicepidemiology.org/), together with ABRicate (Tseemann, 2023), which is a database with a mass screening of contigs for antimicrobial resistance or virulence genes that gives a control for running the data to avoid sharing it with other databases, were utilized to identify the genetic traits of the *P. aeruginosa* and *P. mirabilis* isolates using the Fasta files. CGE tools: MLST 2.0, VirulenceFinder 2.0, PlasmidFinder 2.1, CSIPhylogeny 1.4, and ResFinder 4.1 were used. Antibiotic resistance genes (AMR) that have been acquired or developed due to mutations in chromosomal genes that can cause antibiotic resistance were located using a threshold of 90% and a minimum length of 60% by ResFinder 4.1 (Zankari et al., 2012). The presence or absence of a resistance gene classified the sample as resistant or susceptible, respectively. VirulenceFinder 2.0 (Panayidou et al., 2020) and VFDB (http://www.mgc.ac.cn/VFs/main.htm) were used with a threshold of 90% and a minimum length of 60% to predict the virulence genes. MLST 2.0 (Larsen et al., 2012) and CSIPhylogeny 1.4 tools were used to detect relatedness between the bacterial strains. CSIPhylogeny 1.4 analysis was performed using the respective reference genome and default settings. MLST, which is the gold standard for typing, was used to determine and describe how the *P. aeruginosa* isolates and the *P. mirabilis* isolates were closely related genetically based on the WGS data (Larsen et al., 2012). The seven housekeeping genes according to the MLST database of *P. aeruginosa* (acsA, aroE, guaA, mutL, nuoD, ppsA, and trpE) (Jolley et al., 2018) served as its foundation. In order to identify the total number of loci found in the genome, a core genome cgMLST scheme was created using chewBBACA (A Comprehensive and Highly Efficient Workflow for BSR-Based Allele Calling Algorithm) (Silva et al., 2018). The *P. aeruginosa* PAO1 and *P. mirabilis* HI4320 reference genomes were utilized just to predict the cgMLST loci and were excluded from further investigation. Each genome's coding sequences (CDSs) were annotated, and using an all-against-all BLASTP search and blast score ratio (BSR) computations, distinct loci were found. Candidate loci were chosen based on their inclusion in all full genomes. *Pseudomonas aeruginosa* serotyper (PAst 1.0), a tool from the Technical University of Denmark that is available in CGE, was employed in this study for in silico serotyping of *P. aeruginosa* isolates using next-generation sequencing data (Thrane et al., 2016). The raw pair end FastQ files, without any processing were uploaded and analyzed by the cloud-based tool, 1928 (1928 Diagnostic platform, Sweden). *P. aeruginosa* and *P. mirabilis* paired-end reads in order to benchmark the performance of 1928 with the "in-house developed pipeline". The platform made up of a specifically designed pipeline to analyze paired-end reads from *P. aeruginosa* and *P. mirabilis*, providing species identification, read quality evaluations, phylogenetic group determination, and antibiotic resistance gene identification. In addition, it offered *P. aeruginosa* MLST typing using a particular pipeline (Shemirani et al., 2023; Saxenborn et al., 2021; Mahmoud, 2021). The platform automatically consolidated most results into an Excel sheet available for download. Additionally, *P. aeruginosa* and *P. mirabilis*' respective reference genomes, *P. aeruginosa* Pa01 and *P. mirabilis* HI3120, were used to conduct single nucleotide polymorphism (SNP) analyses for both bacteria.

Statistical analysis

Using Rstudio (version 3 2022.07.2+576; R studio for Statistical analysis), statistical analysis for *Pseudomonas* and *Proteus* species was carried out. The genotypic results from the internal pipeline and the 1928 pipeline were compared with the clinical laboratory phenotypic results using a descriptive statistical analysis. Mean counts with 95% confidence intervals (CI) were used to show the presence of virulence genes and antibiotic resistance. The counts of resistance genes discovered by 1928 and CGE were compared using the Wilcoxon signed-rank test or the paired t-test (Akeyede et al., 2014). Additionally, the Kruskal-Wallis test or one-way ANOVA (Ostertagová et al., 2014), was used to determine whether there was a statistically significant in the distribution of the virulence genes among the four common STs, ST 111, ST 3285, ST 564, and ST 17. Levene and Shapiro-Wilk tests were used to assess the presence of normality and equal variance, respectively (Schober & Vetter, 2019). ANOVA with Simpson's Diversity Index (Hunter & Gaston,

1988), was used to evaluate the diversity of serotypes in various samples, and McNemar's test was used to ascertain whether any resistant phenotypic traits existed. The precision of predicted antibiotic resistance genes (AMR) in relation to the phenotypic AST results was calculated using major errors (MEs) and very major errors (VMEs). While VMEs happen when the predicted result is susceptible but the phenotypic outcome is resistant, MEs happen when the predicted result is resistant but the phenotypic outcome is susceptible. Previously published research (Gordon et al., 2014; Banerjee et al., 2021; Shemirani et al., 2023) was used to assess these errors. P values of less than 0.05 were considered statistically significant (Wasserstein & Lazar, 2016; Ross, 2014).

Results

Quality control, trimming, and assembly

The quality assessment of raw reads was conducted with FastQC v. 0.11.5 (Babraham Bioinformatics, 2023). Pre- and post-trimming results are shown in Appendix II (Figures 1–3). Using Unicycler (Appendix I, Figure 3), trimmed files were assembled, and Fasta files were generated. After trimming, from *Pseudomonas* samples in most of the samples retained more than 70% of the reads (Table 1, Appendix II). In order to assess the assembly's quality, Quast was used to consider the assembly's overall length, the number of contigs, the genome fraction (%), and the percentage of GC content. Tables 1 and 2 show the QUAST result mean values of these parameters along with their 95% confidence intervals for the respective organisms.

Table 1. *P. aeruginosa* isolates with their mean and 95% CI of the reads, the GC content, and genome fraction with the respective reference genome.

Parameter	Mean (95% CI)
Total length (bp*)	6435738 bp (6275857 bp-6595620 bp)
GC content (%)	65.7% (64.5% - 66.9%)
Genome fraction (%)	97.8% (96.0 % - 99.7%)

*bp = base pair

Table 2. *P. mirabilis* isolates with their mean and 95% CI of reads, the GC content, and genome fraction with the respective reference genome.

Parameter	Mean (95% CI)
Total length (bp*)	4960803 bp (4932575 bp-4989030 bp)
GC content (%)	50.7% (50.5% - 50.8%)
Genome fraction (%)	84.7% (83.7% - 85.8%)

*bp = base pair

Species identification

In this study, 88 isolates that had been identified as *P. aeruginosa* and *P. mirabilis* by MALDI-TOF MS (DB-4110) were genotyped for species identification using Average Nucleotide Identity (ANIb) through JSpeciesWS web services (Appendix II, Table 1). *P. aeruginosa* was identified in 43 out of 45 *Pseudomonas* samples, with an average reference genome similarity of 97.8% CI (96.0–99.7). However, it was discovered that Acinetobacter pittii and *Pseudomonas* putida (Appendix II, Table 2), were present in two isolates (PS947 and PS1217). The results of the Maldi-Tof 2021, 1928, and

JSpeciesWS analyses are consistent for the rest 43 samples that were identified as *P. aeruginosa* (Appendix II, Table 1). For *Proteus spp.* group, 39 out of 43 samples were classified as *P. mirabilis*, with an average similarity to the reference genome of 84.7% CI (83.7–85.8). Four samples initially attributed to Pr. hauseri and Pr. vulgaris in Maldi-Tof 2021, were subsequently identified as *Proteus* genomosp. 4 and *Proteus* columbae, respectively (Appendix II, Table 2). Consistent results were observed from Maldi-Tof 2021, 1928, and JSpeciesWS in all the rest 39 samples. In the analysis, Fasta files that did not meet the internal quality control levels of the 1928 platform which were also identified as other organisms were excluded from downstream analysis. A total of 6000 loci for *Pseudomonas* and 3600 loci for *Proteus* were identified from the tools, chewBBACA, respectively (Appendix V, Figures 4 and 5). The study focused on antibiotic resistance gene identification, virulence gene characterization, and sequence typing, considering the remaining 43 *P. aeruginosa* and 39 *P. mirabilis* isolates.

Antimicrobial Resistance Profile

Clinical isolates of the bacteria *Pseudomonas* and *Proteus* were successfully genotyped using the 1928 feature of, the *Pseudomonas* pipeline and the Others bacteria pipeline in the platform, together with ResFinder/CARD from CGE. In predicting the quantity of resistant isolates, both approaches showed a high degree of agreement. Specifically, 1928 predicted that approximately 98% of the isolates were resistant to at least one class of the investigated antibiotics, while CGE predicted 97% resistance to at least one antibiotic class (Appendix III, Tables 1, 2 and 3). In addition to specific antibiotic resistance genes, the 1928 approach found that all isolates (n = 82) encoded multidrug efflux pumps (Appendix III, Table 2). These efflux pump genes were unique to each method. Overall, the comparison between CGE and 1928 across different antibiotic classes showed small variability (Appendix III, Tables 5 and 6), with some classes having more resistance genes identified by CGE and others by 1928. In a study analyzing the antibiotic resistance of *Pseudomonas* aeruginosa, out of 43 samples, approximately 97.6% exhibited Beta-lactamase resistance, and 95.2% showed resistance to Chloramphenicol. Every sample (100%) demonstrated resistance to both Fosfomycin and the Multidrug efflux pump mechanism. Quinolone resistance was detected in 19% of the samples. Other notable genetic resistance markers include the crpP gene, present in 35.7% of samples, and specific gyrA mutations in about 9.5%. These findings indicate a significant resistance profile across most antibiotics tested, highlighting the challenges in treating infections caused by this bacterium. A paired t-test comparing gene resistance detection by CGE and 1928 across eight antibiotic classes found no statistically significant difference (p-value = 0.76) (Appendix III, Table 6). Both methods showed comparable effectiveness, with CGE detecting 228 instances of resistance genes and 1928 identifying 213 instances across all classes. Despite variations in individual classes, the overall performance of both methods appears equivalent in detecting antibiotic resistance genes in clinical *P. aeruginosa* isolates. In the analysis of clinical isolates of *Proteus* bacteria (total samples: 39), tetracycline exhibited the highest resistance prevalence, which was unanimous across all samples at 100% (39 out of 39) carrying the tet(J) gene (Appendix III Table 1). Chloramphenicol resistance was found in 87.2% of the samples, with the detected gene being cat_1. Only one sample demonstrated resistance to Quinolone via the qnrD1 gene. Additionally, approximately 20.5% of the samples had the dfrA1 gene, conferring resistance to Trimethoprim. No samples showed resistance to Aminoglycosides, Beta-lactamases (excluding catA), MLS, or Sulfonamide. The comparison of resistance gene identification between CGE and 1928 across different antibiotic classes showed variable results (Appendix III, Table 7). For some classes of antibiotics, CGE identified more resistance genes, whereas for others, 1928 identified more. When performing a paired t-test to compare the overall performance of the two methods across all antibiotic classes (Appendix III, Table 8), it shows no significant difference (p = 0.23).

Comparing the phenotypic AST with genotypic predictions from 1928 and CGE across various classes of antibiotics in the *Pseudomonas* and *Proteus* study was done next (Appendix III, Table 4). About 88% of the isolates had at least one antibiotic resistance, of which 68% shows multi-drug

resistance from their phenotype findings. For *Pseudomonas*, there was a noticeable level of discordance characterized by major errors (ME) and very major errors (VME) in this case. The aminoglycoside class recorded the highest level of VME at 100% for both prediction methods, with no ME observed. Beta-lactam and Cephalosporin classes also showed total discordance (100%) between phenotypic and genotypic methods but varied in the distribution of VME and ME. Beta-lactam had a VME rate of 39.5% and a ME rate of 60.5%, whereas Cephalosporin had a VME rate of 26.8% and a ME rate of 73.2%. The fluoroquinolone class had the lowest discordance (61%), with a VME rate of 46.3% and an ME rate of 14.6%. Across the bioinformatic workflows for *Proteus* isolates in CGE and 1928, there were a total of 37 and 41 ME and 37 and 41 VME, respectively (AppendixIII, Table 4). The aminoglycosides class showed the greatest discordance, with 10 VMEs for CGE and 13 VMEs for 1928. Chloramphenicol was the next most discordant, with 11 VMEs for CGE and 10 VMEs for 1928. These discrepancies were also reflected in the ME counts, where aminoglycosides showed the highest number of MEs with 10 for CGE and 13 for 1928, closely followed by chloramphenicol with 11 MEs for CGE and 10 MEs for 1928. It was easy to examine all discrepancies from the bioinformatic workflows and reveals that the outputs, like identified resistance genes, from the CGE or 1928 workflows show certain detection future.

Virulence Factors and MLST results

The CGE Virulence Finder analysis conducted on the WGS data of *Pseudomonas* yielded a total of 9840 virulence genes across all 43 isolates. Out of these, there were 240 different virulence genes found among the isolates, with an average of 228, in all the isolates (n = 43), with a 95% CI of 214–238, per isolate. Different virulence genes per isolate were detected, and PhzB1, a metabolic factor, was the most prevalent virulence gene, found in 73.2% of all samples. Among the next most prevalent virulence factor genes are fliQ, alg44, alg8, algA, algB, algC, algD, algE, algF, algL, algQ, algR, algU, algW, algX, and algZ, each accounting for roughly 0.45% of the total virulence factor genes. Table 3 presents the proportion of the most frequent virulence factors and their genes. No virulence factor was included in the 1928 *Pseudomonas* results. For *Proteus*, there is no tool available to obtain the VF gene in CGE, VFDB, or 1928 until performing this thesis.

Table 3 Most common virulence factors and their genes (in percentage) of *P. aeruginosa* clinical isolates (n = 43)

Virulence genes (n=9840)	**Counts of genes (%)**
Adherence (*air, xcp, iroN*)	1497 (15.2%)
Effector delivery system (*lip, tse1, exs and S fimbriae*)	215 (2%)
Motility (*fliQ, flg, fle, mot*)	1834 (18.6%)
Exotoxin (*toxA, plcH*)	84 (5%)
Biofilm (algA, algB *algR, mucA, rhl*)	1419 (14.4%)
Nutritional/Metabolic factor (*pch, PhzB1, pvd*)	1334 (13.4%)

The MLST analysis of the 43 *Pseudomonas* isolates using MLST 2.0, CGE tools, and 1928 resulted in 26 matching sequence types. The MLST analysis results are shown (Appendix IV, Table 1) together with the serotype and other futures. From the analysis of *P. aeruginosa* samples, four most common Sequence type (ST) were identified. Four samples (9.30%) of ST 3285, four samples (9.30%) of ST 111, three samples (6.98%) of ST 17, and three samples (6.98%) of ST 564 were

analyzed. The number of virulence factor genes was found to be significantly correlated with the MLSTs (p = 0.00001), showing that the STs are a significant predictor of the number of virulence factor genes. The number of virulence genes differed significantly between ST 3285 and ST 111, according to post hoc comparisons using Tukey's HSD test (CI [1.40, 11.10], p = 0.005). On average, ST 3285 possesses 6.25 more virulence genes than ST 111. Additionally, there were significant differences in the mean number of virulence genes between these STs, between ST 111 and ST 17 (mean difference = 5.67, p = 0.03), as well as between ST 3285 and ST 564 (mean difference = 7.25, p = 0.002).

The CGE and 1928 epidemiological studies show similarities in ST distribution (Figure 3). Two isolates (PS1521 and PS1528) could not have their (ST) determined using both the 1928 approach and MLST 2.0. Despite this, both workflows identified the same allelic profile for these isolates. Phylogenetic trees for *Pseudomonas* and *Proteus spp.* were constructed using SNP-based analysis and MLST features using CGE and 1928 epidemiological studies (Appendix V, Figures 1, 2, and 3). The phylogenetic trees constructed using SNP-based analysis showed similar groupings of the sequence types (Appendix V, Figures 1, 2). While the trees do not appear identical, they both show similarities in grouping the same STs together and distinguishing them from other STs.

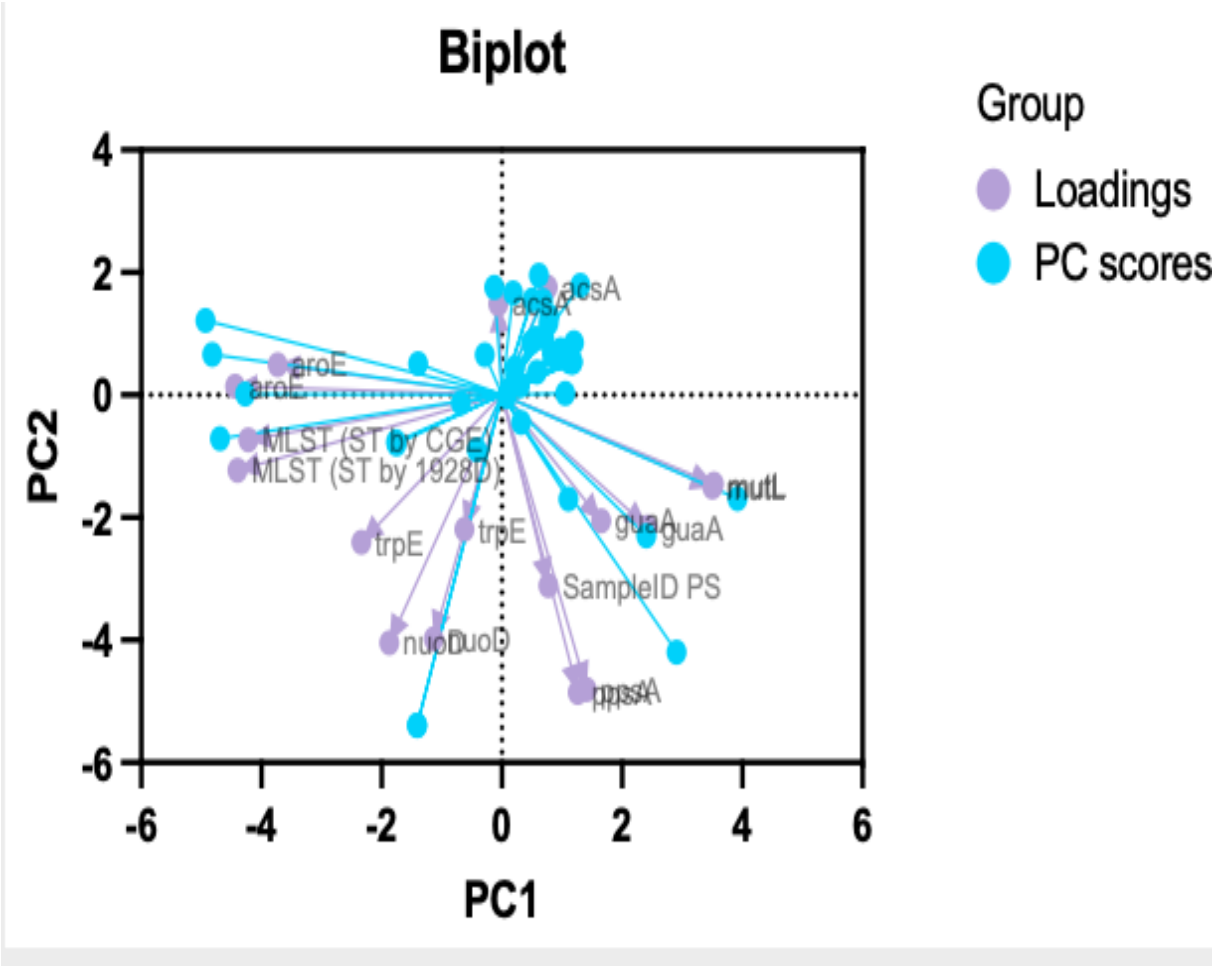

Figure 3. Principal Component Analysis of MLST Genes from CGE and 1928 *Pseudomonas* Samples, this figure presents a PCA plot of the MLST genes identified by CGE and 1928, including acsA, aroE, guaA, mutL, nuoD, ppsA, and trpE. Each gene is represented by two data points: a blue point indicating the PC score and a purple point indicating the loading. The position of the points is determined by the multilocus sequence typing (MLST) and sequence type (ST) as determined by CGE or 1928. Similarities between STs are indicated by proximity in the PCA plot. High degrees of similarity between the STs identified by CGE and 1928 for each gene are indicated by close proximity between the corresponding PC score and loading points.

The distribution of the isolates, in relation for each ST is shown in Table 4 below. Simpson's Diversity Index (Table 4) was used for *Pseudomonas* samples to assess the diversity of *Pseudomonas* isolates across various clinical specimens. Accordingly, the urine culture extended and urine culture routine specimens exhibited the highest diversity, while the wound culture extended specimen showed no variation in the identified STs. Furthermore, the most frequent ST was identified for each specimen type to determine the prevailing *Pseudomonas* type. For *Proteus*, because the MLST result was not available, this statistical analysis could not be performed.

Table 4. Diversity and Prevalence of *Pseudomonas spp.* in Clinical Specimens

Specimen Type	N	Simpson's Diversity Index	Most frequent ST
Blood culture	7	0.952	270
Upper respiratory	5	0.90	111
Urine culture extended	2	1	17
Urine culture routine	20	0.98	564
Wound culture	9	0.97	3285
Wound culture extended	1	NaN	1068
Total	43	0.997	564

Serotype

For *Pseudomonas* samples, the serotype group was identified from CGE only using *Pseudomonas aeruginosa* serotyper (PAst 1.0) tools (Figures 4 and Appendix IV, Figure 1). A total of nine different serotypes were identified. Serotype O6, was the commonest, with urine culture being the specimen with the most variety of serotypes. In this study of Pseudomonas serotype data, a total of nine serotypes, O1, O2, O3, O4, O5, O6, O7, O9, and O11 were discovered (Figure 4). Serotypes O6, O1, and O9, were the commonest serotypes to be identified (Appendix IV, Figure 1), which were detected in about 54.8 % of the isolates (Appendix IV, Table 1). The serotype significantly influenced coverage, as evidenced by a statistically significant result ($p = 0.001$). The influence of both serotype and specimen type on coverage was statistically significant for the serotypes ($p = 0.001$), but the specimen type alone did not yield any significant effect in the serotype ($p = 0.601$). A comprehensive breakdown of the data revealed the count and percentage of each serotype within the specimen types, and another future (Appendix IV, Table 1), which provides a detailed perspective of the distribution. There were no tools available to identify the possible serotypes in the case of P. mirabilis in the CGE or 1928 pipeline for both organisms, until this research was published.

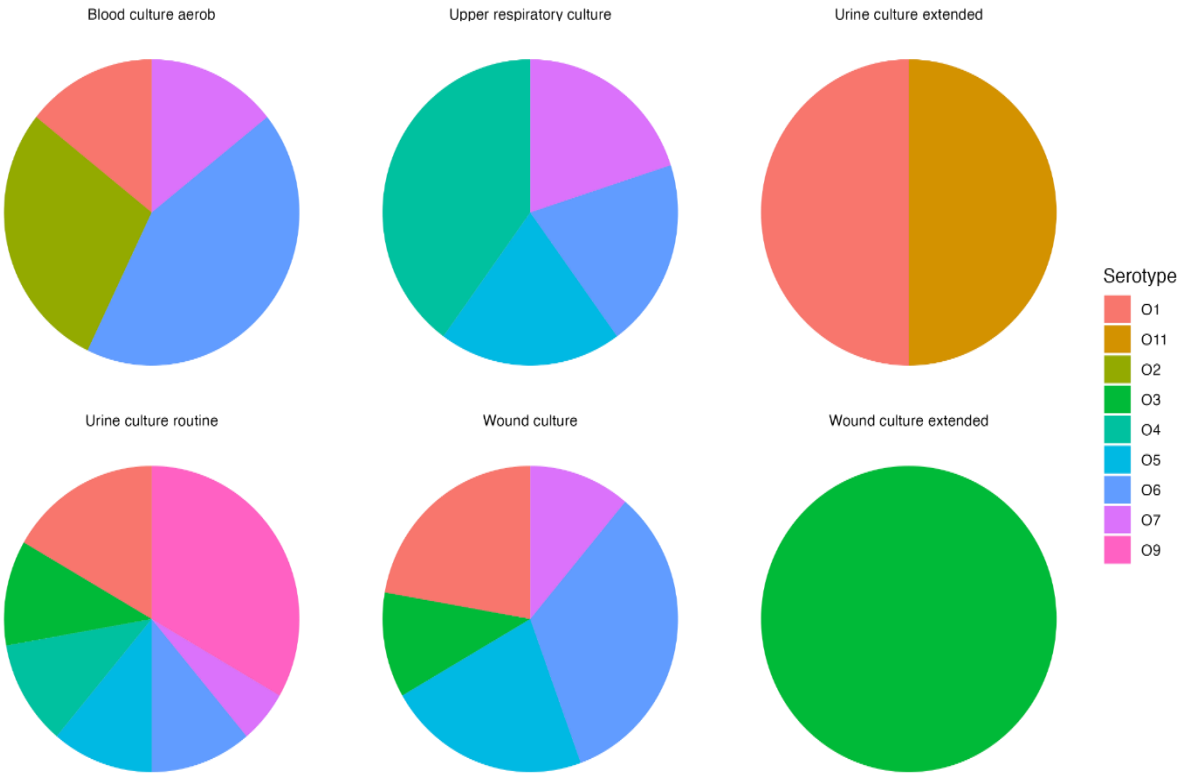

Figure 4: The distribution and variety of serotypes in different samples of *Pseudomonas aeruginosa*

Plasmids

For a comprehensive understanding of the plasmids and their functions, the next step involved utilizing ResFinder 2.0 in CGE to successfully identify plasmids in both *Proteus spp.* and *Pseudomonas spp.* Isolates (Appendix IV, Table 2). Out of the 39 analyzed samples of *Proteus spp.*, plasmids were detected in nine samples (20.93%). Among these samples, the most commonly identified plasmid was Col3M_1, found in five samples (11.63%) obtained from various sources, including wound cultures, blood cultures, and urine cultures. The ColE10_1 plasmid was detected in two samples (4.65%), specifically from urine cultures. Additionally, the IncN_1 and ColRNAI_1 plasmids were each identified in one sample (2.33%). Notably, the majority of the *Proteus spp.* samples, 30 out of 39 (79.07%), did not exhibit an identifiable plasmid. Similarly, in the case of *Pseudomonas spp.* plasmids were found in four out of the 43 analyzed samples (9.30%). Among these, the IncQ2_1 plasmid was the most prevalent, being present in three samples (6.98%) obtained from wound cultures, blood cultures, and urine cultures. Furthermore, the IncFII (pRSB107) plasmid was identified in one sample (2.33%), specifically from wound cultures. The majority of the *Pseudomonas spp.* samples, 39 out of 43 (90.70%), did not exhibit an identifiable plasmid. No plasmid finder was included in the 1928 tools for both bacteria.

Discussion

WGS methods have emerged as effective tools for identifying pathogens quickly, detecting resistance genes precisely, and tracking outbreaks (Váradi et al., 2017). In contrast to traditional techniques that depend on the ongoing development of species-specific probes and primers, WGS offers a thorough and quick approach and provides extensive data without the need for target-specific reagents (Váradi et al., 2017). Recently, there has been an increase in the use of WGS analysis in clinical settings, especially for the diagnosis and treatment of sepsis and related infections, such as those brought on by *Pseudomonas* and *Proteus spp.* (Lee et al., 2022; Schütz et al., 2017). WGS analysis has been shown in studies to be effective at identifying antibiotic-resistant strains of *P. aeruginosa* and *P. mirabilis*, and directing appropriate antibiotic therapy (Horcajada et al., 2019; Walters et al., 2019). Furthermore, Roach et al. (2018) showed how effective WGS is at determining the origin and spread of hospital-acquired infections, particularly in sepsis cases. This thesis examines the genotypic identification of *Pseudomonas* and *Proteus spp.* and benchmark the performance of the 1928 tool against an in-house-developed pipeline using data from *P. aeruginosa* and *Proteus spp.* obtained from suspected sepsis patients. Various freely available bioinformatic tools, including JSpeciesWS, chewBBACA, and the CGE service, which have all been validated in numerous studies (Saxenborn et al., 2021; Zankari et al., 2013; Joensen et al., 2014; Silva et al., 2018), are integrated into this in-house-developed pipeline. This evaluation aims to provide a rough estimate of the precision and efficiency of the relatively recent 1928 tool for automated WGS analysis. To increase the accuracy and dependability of sequencing analysis, WGS data must be preprocessed. In this process, trimming raw reads is crucial and forefront, because it must preserve as many bases as possible while maximizing read quality. According to Del Fabbro et al. (2013), trimming thresholds of 20–30 are ideal for keeping a lot of high-quality reads. The in-house-developed pipeline in this study employed a quality threshold of 20 for trimming reads, consistent with these recommendations (Andrews, S. et al., 2010; Shemirani et al., 2023). Low-quality regions, very short reads, and adapter sequences was removed. Trimming was also performed to improve de novo assembly of the reads (Bolger et al., 2014). Most of the *Pseudomonas* and *Proteus* samples kept more than half of the reads (Table 1 and 2), after trimming, proving the efficiency of the trimming technique (Klockgether & Tümmler, 2017). These results support the importance of careful selection and application of quality control and trimming procedures for high-quality sequencing data, consistent with Del Fabbro et al.'s (2013) and Bolger et al. (2014) recommendations. These findings also suggest the trimming approach effectively retained a high percentage of good-quality reads within the expected range (Del Fabbro et al., 2013; Bolger et al., 2014), which assures that careful pre-processing and quality control procedures are critical for WGS data analysis. The results of this study, along with previous research, highlight the importance of careful selection and application of quality control and trimming procedures for high-quality sequencing data suitable for downstream analysis.

The use of a reference genome in analyzing WGS data for *Pseudomonas* and *Proteus spp.* varies based on research questions and availability. Several high-quality reference genomes have been utilized in *Pseudomonas*, while there are fewer for *Proteus*, resulting in a significant challenge for performing an assembly (Bacci et al., 2017; Markussen et al., 2014). The choice of reference genome can improve analysis accuracy and efficiency, particularly for AMR gene detection, comparative genomics, and phylogenetic analysis (Bacci et al., 2017; Markussen et al., 2014; Yang et al., 2019). In this study, the well-known complete reference genomes of *Pseudomonas* PAO1 (Juhas, 2015) and *Proteus* HI4320 (Yu et al., 2017) were used, respectively. According to Luo et al. (2012), using a reference genome in QUAST analysis is critical for accurately evaluating the quality of genome assemblies in WGS. Key evaluation metrics might not be available prior to the launch of QUAST without a reference genome, which would limit the evaluation's accuracy (Gurevich et al., 2013). JSpeciesWS calculates the ANI-based approach for specious identification using the reference genomes of the two organisms. The results were comparable; the JSpeciesWS-based method and 1928 tools matched almost 99% of the time (Appendix II, Table 1). Numerous studies (Shemirani et al., 2023; Bonnelly et al., 2023) have supported this. Two of the total samples were

determined to be *P. vulgaris*, a common habitant of the human gastrointestinal tract (Hamilton et al., 2018), and two of the samples were determined to be *P. columbae* in *Proteus* samples, the most prevalent microorganisms that can be diagnosed alongside *P. mirabilis* (Hamilton et al., 2018). Similar to the findings from 1928, two *Pseudomonas* strains were found to be Acinetobacter pittii and *Pseudomonas* stutzeri, respectively. All the results agreed with the 2021 updated clinical microbiology result from Unilabs, Skövde, using MBT Compass Library DB-7854 (Bruker Daltonics, Germany).

Studies have shown that *P. aeruginosa* and *P. mirabilis* can develop antibiotic resistance through a number of mechanisms, including efflux pumps, target modification, toxic production, and enzymatic inactivation (Girlich et al., 2020; Chambers et al., 2017; Martinez et al., 2015; Sabnis et al., 2021; Maldonado et al., 2016). *P. aeruginosa* and the host interact in complex ways that involve a number of different host components and signaling pathways. These involve triggering immune reactions, triggering inflammation, and modifying host cell death and survival pathways (Chambers et al., 2017). Studies utilizing WGS to identify resistance genes demonstrated substantial agreement between genotypic and phenotypic resistance profiles (Jeukens et al., 2020; Shemirani et al., 2023), though some discrepancies have been noted. The dataset used for this study shows a mixed pattern of agreement and disagreement between the phenotypic and genotypic resistance profiles. Despite some differences in specific classes, the genotypic approaches of CGE and 1928 both showed comparable efficiency in identifying antibiotic resistance genes in clinical *P. aeruginosa* isolates. For beta-lactam and fluoroquinolone resistance, the majority of the samples demonstrate agreement between the phenotypic and genotypic results (Appendix III, Tables 1, 2, 3, and 5). This result agrees with previous similar studies (Dégi et al., 2021; Dötsch et al., 2009; Saxenborn et al., 2021). These findings provide valuable insights into the prevalence and distribution of antibiotic resistance genes in clinical *Proteus* isolates, which can inform better antibiotic treatment strategies for this bacterial species. However, there are also a number of samples where the phenotypic and genotypic outcomes were different. Both the in-house and 1928 pipelines encountered challenges in accurately predicting AST results that agreed with the phenotypic AST for aminoglycoside and cephalosporin antibiotics. This issue was observed in both the *P. aeruginosa* study and, specifically for aminoglycoside, in the *P. mirabilis* study (Appendix III, Table 3 and 4). These discrepancies can be observed across various specimen types and resistance categories, emphasizing the importance of considering both phenotypic and genotypic data when assessing antibiotic resistance in these organisms (Jia et al., 2017; Boero & Bernardi, 2014). Numerous factors may contribute to the differences between phenotypic AST results and genotypic predictions of AMR (Boero & Bernardi, 2014). One factor that makes predictions difficult is that some AMR genes are linked to particular antibiotics, while others have complex mechanisms linked to different antibiotic classes (Mahfouz et al., 2020; Jia et al., 2017). Another consideration is that the existence of an AMR gene does not ensure that it will be expressed or that it will be resistant to certain antibiotics (Zankari et al., 2013). Furthermore, it is important to compare predicted AMR gene results from WGS analysis with phenotypic AST results using a variety of WGS datasets (Mahfouz et al., 2020). Unidentified resistance mechanisms, unexplored resistance genes, gene expression regulation, epigenetics, bacterial growth conditions, antibiotics used for grouping in both sides, and limitations in prediction algorithms, databases, or experimental techniques are additional potential causes for these discrepancies (Larsen et al., 2012; Jia et al., 2017; Pachori et al., 2019). In addition to highlighting the ongoing problem of antibiotic resistance in *P. aeruginosa* and *P. mirabilis*, the results from the samples support previous researches (Lupo et al., 2019; Chen et al., 2015). Additionally, the discovery of the mexA, mexE, and mexX genes in the samples is consistent with the study's analysis of the significance of efflux pumps in developing resistance to fluoroquinolones, tetracyclines, and other antibiotics (Lupo et al., 2019).

The analysis of 39 isolated clinical samples of *Proteus spp.* revealed the presence of antibiotic resistance genes such as aph(3')-IIb (aminoglycosides), catB7 (chloramphenicol), fosA (fosfomycin), blaPDC-374 (beta-lactams), and mexA, mexE, and mexX (efflux pump-related

resistance in fluoroquinolones, tetracyclines, and other antibiotics), showing the possibility of MDR in *Proteus spp.* (Girlich et al., 2020). Isolations, similar to the finding in this research (Chen et al., 2015; Pachori et al., 2019). The analysis of antibiotic resistance gene prevalence in clinical *Proteus* isolates revealed significant resistance to tetracycline, with the tetJ gene being highly prevalent. This finding underscores the importance of vigilant monitoring and judicious use of tetracycline antibiotics to mitigate further resistance development (Girlich et al., 2020). Chloramphenicol resistance was also significant, indicating the potential challenges in treating infections caused by *Proteus* with this class of antibiotics (Bush, 2010). This data underlines the prevailing resistance challenges in *Proteus* mirabilis against specific antibiotics, especially Tetracycline and Chloramphenicol. The detection of aminoglycoside, broad-spectrum beta-lactam, sulfonamide, and trimethoprim resistance genes warrants attention as well, as these are important antibiotics commonly used in clinical settings (Girlich et al., 2020; Bush, 2010). The absence of quinolone resistance is a positive outcome, but continued surveillance remains crucial to prevent its emergence (Girlich et al., 2020). Overall, this study shows the resistance ability of *Proteus* isolates and emphasizes the need for ongoing efforts in antimicrobial management to combat antibiotic resistance effectively. The crpP gene, present in the samples, is associated with resistance to ciprofloxacin, which aligns with previous research on *P. mirabilis* resistance patterns, where mirabilis' antibiotic resistance is widely available (Girlich et al., 2020; Bush, 2010; Nordmann et al., 2011). The resistance patterns observed from the two genotypic analysis methods, CGE and 1928, revealed similarities and differences in both antibiotic resistance genes and their associated antibiotics. Both methods detected the presence of antibiotic resistance genes, such as aph(3')-IIb (aminoglycosides), catB7 (chloramphenicol), fosA (fosfomycin), and blaOXA variants (beta-lactams). However, CGE identified additional genes, including qnrVC1 (quinolones) and tet(42) (tetracyclines), whereas 1928 detected the presence of mexA, mexE, mexX (efflux pumps related to multiple antibiotics), and crpP (cyclic peptide antibiotics). The similarities in antibiotic resistance genes identified by both methods are related to resistance against aminoglycosides, chloramphenicol, fosfomycin, and beta-lactams. In contrast, the differences in the detected genes highlight the potential variability in resistance patterns depending on the chosen analysis method, with CGE detecting quinolone and tetracycline resistance genes and 1928 identifying genes related to efflux pumps and cyclic peptide antibiotic resistance.

Comparing phenotypic and genotypic antibiotic susceptibility testing (AST) reveals the subtle complexities and differences (Appendix 3, Tables 3 and 4). While some antibiotics show notable consistency across methodologies, others show startling differences, particularly in terms of Very Major Errors (VMEs) and Major Errors (MEs). Tetracycline emerges as having an unsettling 100% discordance across all bioinformatic workflows, starting with *Proteus* species (Appendix3, Table 3). This significant inconsistency results in 36 VMEs for both the 1928 and CGE platforms, suggesting that there may be gaps in the workflows or bioinformatic databases used (Shemirani et al., 2023; Saxenborn et al., 2021; Mahmoud, 2021). Tetracycline's multifactorial resistance mechanism may imply to genetic determinants that have not yet been identified and included in prediction systems. Additionally, nearly half of the tested samples showed discrepancies for both Aminoglycosides and Chloramphenicol, highlighting the dangers of solely relying on genotypic predictions without concurrent phenotypic validations (Krause et al., 2016). By focusing on *Pseudomonas* samples (Appendix 3, Table 4), a similar pattern of discrepancy is visible, particularly for aminoglycoside and beta-lactam. The 100% discordance for Aminoglycoside, with 38 VMEs across both platforms, echoes a conceivable gap for understanding or representation of this antibiotic's resistance in *Pseudomonas* species (Krause et al., 2016). Beta-lactam, on the other hand, offers a blend of errors. The predominant MEs suggest a potential over-prediction of resistance, pointing to the need for refining the genotypic prediction tools. It's crucial to acknowledge that for both bacterial species, antibiotics like MLS, Quinolone, and Sulfonamide consistently demonstrated congruence between phenotypic and genotypic methods. This synchrony could reflect well-characterized genetic markers adeptly recognized by both the 1928 and CGE algorithms used (Saxenborn et al., 2021; Mahmoud, 2021). Further complicating the

perspective, as described in this sepsis study, is the considerable level of discrepancies for *Pseudomonas* samples, casting doubts over the efficacy of genotypic methods as standalone predictive tools (Collignon & McEwen, 2019). Distinct error patterns emerge across the antibiotic classes. For instance, the aminoglycoside class, despite phenotypic sensitivities, consistently showed genotypic resistance, highlighting the probable shortcomings in current genotypic databases or algorithms (Collignon & McEwen, 2019; Krause et al., 2016). Drawing upon these multifaceted observations, it becomes evident that a balanced integration of both phenotypic and genotypic testing is essential for reliable and effective antibiotic resistance determination (Shemirani et al., 2023; Mahmoud, 2021). The insights presented align with previous research, emphasizing the urgency of enhancing genotypic prediction platforms, ensuring they are routinely validated and updated in conjunction with phenotypic tests, and striving for a more holistic approach to resistance detection (Collignon, P., & McEwen, S. 2019).

To effectively address infections caused by *different organisms,* it is also crucial to have a comprehensive understanding of the wide variety of virulence genes and their relationship with antibiotic resistance (Sónia et al., 2015). This knowledge is also vital in developing effective strategies for the diagnosis, prevention, and treatment of *P. aeruginosa* and *P. mirabilis* infections (Chambers et al., 2017; Sabnis et al., 2021; Sónia et al., 2015; Ozer et al., 2021; Maldonado et al., 2016). Healthcare professionals can learn crucial information that will help in the development of targeted therapies and enhance patient outcomes by looking at the complex relationship between virulence factors and antibiotic resistance (Girija et al., 2019). A total of 9840 virulence genes were found in 43 *P. aeruginosa* isolates using the CGE Virulence Finder analysis of WGS data. The top virulence genes found in more than 40% of the isolates (Table 3) were phzB1, fliQ, and algA, with various functions, such as pyocyanin production, flagellar motility, and alginate biosynthesis being the commonest (Winsor et al., 2015; Dasgupta et al., 2002). These were the same as previous research findings (Ertuğrul et al., 2017; Beasley et al., 2020). Overall, these findings provide valuable insights into the distribution of virulence genes among *Pseudomonas* isolates, which can be useful in understanding their pathogenicity and designing targeted treatments. Their relation with ST values and their variation from specimen to specimen were able to be identified (Beasley et al., 2020; Pachori et al., 2019). All identified virulence genes in this study were validated against the Virulence Factor of Pathogenic Bacteria database (VFDB, 2023), confirming their presence and reliability. There was no VF in 1928 for *Pseudomonas* or *Proteus* spp., and which limit to perform any benchmarking between the pipelines when it came to VF. Phylogenetic analysis, utilizing

WGS data is also a crucial tool for determining epidemiological relationships between bacterial samples and detecting outbreaks (Sawa et al., 2020; Safarirad et al., 2021; Besser et al., 2018). This approach is widely adopted in clinical microbiology and infection control laboratories (Quick et al., 2014; Ellington et al., 2017). The high concordance observed between the 1928 pipeline and the CGE CSI Phylogeny analysis indicates the reliability of these methods. Notably, samples clustered according to their ST (Appendix VI). Complementing WGS-based phylogenetic analysis, Multi-Locus Sequence Typing (MLST) assigns bacteria to distinct STs based on allelic variations in seven housekeeping genes (Safarirad et al., 2021; Fischer et al., 2020; Feng et al., 2021). The diversity of isolates for each ST is shown in Table 4. Using Simpson's Diversity Index revealed varying frequent STs for different specimen types, some with multiple frequent STs. These findings highlight the diverse patterns of *Pseudomonas* diversity and prevalence across clinical specimens, with the findings agreeing with previous research (Gužvinec et al.,2014). Which potentially influence the diagnosis and treatment of *Pseudomonas* infections. Combining these techniques provides a comprehensive view of bacterial relationships and aids in detecting outbreaks. When it comes to serotyping, it was only able to generate a serotype for *Pseudomonas* samples. Interestingly, the serotype distribution (Appendix IV, Figure 1) mirrored findings from earlier studies (Nasrin et al., 2022; Pirnay et al., 2009). In particular, Serotype O6, commonly found in the urinary tract and in the upper airway (Nasrin et al., 2022), was mostly detected (Appendix IV, Table 1). This understanding of serotype distributions holds implications for future vaccine development and disease prevention, among others bacterial futures (Nasrin et al., 2022;

Pirnay et al., 2009). In the study, plasmids were detected within both *Proteus spp*. and *Pseudomonas spp*. Isolates. COL3M-1 plasmid, was the most frequent in *Proteus spp*. samples (Appendix IV, Table 2), and has been previously associated with fluoroquinolone resistance through the qnrD1 gene (Bitar et al., 2020). Conversely, for *Pseudomonas spp*., the most identified plasmid (Appendix IV, Table 2), IncQ2, is known to harbor multiple resistance genes (Rozwandowicz et al., 2018). Notably, the 1928 toolset lacked a plasmid finder for these bacteria.

In conclusion, WGS proves itself as a critical tool for swiftly identifying pathogens and detecting antibiotic resistance, providing essential insights into managing infectious diseases. Careful pre-processing and selection of a suitable reference genome are key to accurate analysis. While the phenotype-genotype relationship isn't always straightforward, it's crucial for understanding antibiotic resistance. Method selection can impact results, as observed in the different resistance patterns detected by CGE and 1928. Ultimately, effective use of WGS could revolutionize the approach to diagnosing and managing diseases, making it an invaluable tool in Sepsis prevention mechanisms.

Ethical aspects and its impacts on Society

The foundational principles of ethics and ethical behavior are the tenets of a civilized society, and it has been argued that ethics motivates students to learn more and keeps research up-to-date (Hudek, 2009; Sivasubramaniam et al., 2021). Ethics can be referred to as any philosophical theory of what is ethically right or wrong, as well as any group of moral norms, mandates, or objectives (Hudek, 2009, Understanding Ethics and Types, 2022). Obtaining informed consent is the main ethical concern while performing research on humans or animals, according to studies by Sivasubramaniam et al. (2021). According to Mantzorou (2011), informed consent is permission given to researchers voluntarily, knowingly, and openly by completing a form with comprehensive information about the research (Manti and Licari, 2018). Ethical committees are necessary to regulate every part of research work, including informed consent. The Skaraborg Sepsis Study (Ljungström et al., 2017), which was approved by the Gothenburg Regional Ethical Review Board (376-11), was the impetus for the research that led to this project. In this project, clinical isolates of *P. aeruginosa* and *P. mirabilis* were able to have their genomes sequenced without the need to provide informed consent. Moreover, each patient who participated in the "Skaraborg sepsis study" provided signed informed consent (Ljungström et al., 2017). Personal data such as names or social security numbers will not be published in this research. While collecting isolates, no patient's gender was purposefully neglected or overrepresented; the samples were taken from both male and female patients. In the long run, sepsis, a critical health emergency often associated with multidrug resistance, must be addressed. This project aims to benefit the community by pioneering methods for early sepsis identification and countering antibiotic resistance. This thesis will present potential strategies for infection detection and mitigation. Collaborating with various stakeholders, which aspire to broaden the research's horizons and contribute meaningfully to the society in developing an outstanding mechanism to prevent infection specially, in the health care systems, to which where it will be implemented.

Future perspectives

NGS has indeed revolutionized biomedical research over time-consuming and expensive wet lab-based procedures (Hu, T et al., 2021). Many clinical laboratories have already adopted NGS technology to identify causal variants for the diagnosis of constitutional disorders, genomic profiling for precision oncology, and pathogen detection for infectious diseases (Cohen et al., 2015). NGS also has widespread clinical application in precision medicine, such as understanding sepsis-causing organisms like *Pseudomonas* and *Proteus spp.*, which are responsible for millions of infections worldwide (Zhong et al., 2021). This thesis's findings highlight *Pseudomonas* and *Proteus spp.*, shedding light on their roles in a serious disease, known as sepsis. However, in clinical settings, the transition from traditional clinical microbiological methods to NGS faces significant challenges. Despite the existence of free bioinformatics tools, the cost of NGS methodology to the lack of skilled professionals, make it unpracticable in many places. Tools, such as CGE and 1928, can manage raw data within an hour with the identification of resistance genes, virulence genes, sequence types (STs), and many other futures without the need for a lot of intervention, potentially accelerating the process. These tools could improve the applicability and relevance of NGS data for research. *P. mirabilis* and *P. aeruginosa* are the commonest pathogens from the *Proteus* and *Pseudomonas spp.*, respectively, in causing sepsis, especially in healthcare setups in their groups. WGS study on *Proteus spp.*, are not too many, especially because there was a big challenge in identifying the virulence future and the MLST because of the lack of organized sources, and due to the small number of genes in the MLST used for classification, the resolution could be limited (Chen et al., 2023; Silva et al., 2018). Although *Proteus* and *Pseudomonas* have been identified as susceptible organisms in this study, little is known about the pathogens that they use to cause a wide range of serious illnesses. Future similar research that incorporates the core genome MLST result will greatly benefit from the genetic analysis of these bacterial strains and other bacterial species. ChewBBACA is a highly customizable platform for the cgMLST scheme. Scalability and flexibility in handling genomic data offer a standardized and reproducible workflow, facilitating data sharing and large-scale multi-center studies (Silva et al., 2018; Chen et al., 2023). Nowadays, most research uses the cgMLST approach to study and characterize organisms (Chen et al., 2023; De Been et al., 2015). The chewBBACA algorithm, which is open source and circumvents those restrictions from MLST, offers a straightforward bioinformatics pipeline for creating target strains' cgMLST schemes. It is possible to type *P. aeruginosa* and *P. mirabilis* globally, keep track of the clonal groups (CGs) of *P. aeruginosa* and *P. mirabilis*, and verify the cgMLST scheme of them (Chenn et al., 2023). It is necessary to consider ethical standards, data privacy, and socioeconomic barriers as researchers move towards this new frontier in medicine (Sivasubramaniam et al., 2021). The ultimate objective must be the incorporation of NGS into routine clinical practice, either in place of or in addition to current practices. Collaboration across disciplines, method standardization, and ongoing innovation will be necessary for this. In conclusion, NGS holds promise for the future of sepsis research and treatment, but there are still obstacles to overcome (Cohen et al., 2015; Hu, T. et al., 2021; Li, B., & Yan, T., 2021). It is important to carefully strike a balance between efficiency, accuracy, cost, and clinical applicability (Li, B., & Yan, T., 2021).

Acknowledgments

I would like to begin by extending my deepest gratitude to my principal thesis advisor, Professor Magnus Fagerlind, of the University of Skövde. His unwavering support, faith in my abilities, and courage to let me navigate my own challenges have proven to be invaluable life lessons. Magnus' inspiring ideas, coupled with the freedom he granted me to develop my own views within my research, have imparted key life skills I will always cherish. His continuous advice for refining the project workflow was essential to the thesis's success. Once again, thank you, Magnus, for everything. I am eternally grateful to you. I also wish to acknowledge the Swedish Institute, which provided me with a scholarship opportunity, enabling me to immerse myself in and enrich my knowledge together of experiencing cultures of Swedish society. My heartfelt thanks go out to Sweden. Next, my sincere appreciation goes to the faculty members of the Systems Biology team at the University of Skövde. Their consistent guidance and insightful suggestions greatly enhanced my interpretation and understanding of the project's results at each stage. I owe the successful completion of my thesis to Professors, Diana Tilevik, Andreas Tilevik, Anna-Karin Pernestig, and also Kristoffer Nilsson, and John Baxter. In addition, I am thankful to Dimitrios Arnellos from 1928 for his patient responses to my queries, his assistance with the platform's utilization, and his help in understanding the obtained results. I also extend my gratitude to my colleagues Sihaam and Josna for them continues support, encouragement and teamwork mentality throughout my thesis journey. Lastly, my family, and especially my wife Hanan, deserve a special mention. Hanan, the light of my life, was always there when I needed her. My family's emotional support throughout my study period was truly a boon. Thank you; I am forever grateful. In conclusion, I extend my heartfelt thanks to everyone who has contributed to my journey. Your help and support will always be remembered.

References

Akeyede, I., Usman, M., & Chiawa, M. A. (2014). On Consistency and Limitation of paired t-test, Sign and Wilcoxon Sign Rank Test. IOSR Journal of Mathematics, 10(1), 01–06. https://doi.org/10.9790/5728-10140106

Alamu, J., Kakithakara Vajravelu, L., Venkatesan, B., & Thulukanam, J. (2022). Correlation of phenotypic and genotypic virulence markers, antimicrobial susceptibility patterns, and outcomes of *Pseudomonas aeruginosa* sepsis infection Microbial Pathogenesis, 170, 105716. https://doi.org/10.1016/j.micpath.2022.105716

Ambardar, S., Gupta, R., Trakroo, D., Lal, R., & Vakhlu, J. (2016). High-Throughput Sequencing: An Overview of Sequencing Chemistry Indian Journal of Microbiology, 56(4), 394–404. https://doi.org/10.1007/s12088-016-0606-4

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

Arbatsky, N. P., Drzewiecka, D., Palusiak, A., Shashkov, A. S., Zabotni, A., Siwiska, M., & Knirel, Y. A. (2013). structure of a Kdo-containing O polysaccharide representing *Proteus* O79, a newly described serogroup for some clinical *Proteus* genomospecies isolates from Poland. Carbohydrate Research, 379, 100–105. https://doi.org/10.1016/j.carres.2013.07.001

Armbruster, C. E., Mobley, H. L. T., & Pearson, M. M. (2018). Pathogenesis of *Proteus mirabilis* Infection. EcoSal Plus, 8(1).

Banerjee, R., Cunningham, S. A., Beisken, S., Posch, A. E., Johnston, B., Johnson, J. R., & Patel, R. (2021). Core Genome Multilocus Sequence Typing and Prediction of Antimicrobial Susceptibility Using Whole-Genome Sequences of Escherichia coli Bloodstream Infection Isolates Antimicrobial Agents and Chemotherapy, 65(11). https://doi.org/10.1128/aac.01139-21

Bitar, I., Marchetti, V. M., Mercato, A., Nucleo, E., Anesi, A., Bracco, S., Rognoni, V., Hrabak, J., & Migliavacca, R. (2020). Complete Genome and Plasmids Sequences of a Clinical Proteus mirabilis Isolate Producing Plasmid Mediated NDM-1 From Italy. Microorganisms, 8(3), 339. https://doi.org/10.3390/microorganisms8030339

Behjati, S., & Tarpey, P. S. (2013). What is next-generation sequencing? Archives of Disease in Childhood: Education and Practice Edition, 98(6), 236–238. https://doi.org/10.1136/archdischild-2013-304340

Beasley, K. L., Cristy, S. A., Elmassry, M. M., Dzvova, N., Colmer-Hamood, J. A., and Hamood, A. N. (2020). During bacteremia, *Pseudomonas aeruginosa* PAO1 adapts by altering the expression of numerous virulence genes, including those involved in quorum sensing. PLoS One 15:e0240351. doi: 10.1371/journal.pone.0240351

Besser, J., Carleton, H., Gerner-Smidt, P., Lindsey, R., & Trees, E. (2018). Next-generation sequencing technologies and their application to the study and control of bacterial infections Clinical Microbiology and Infection, 24(4), 335–341. https://doi.org/10.1016/j.cmi.2017.10.013

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

Bonnelly, R., Calderón, V. V., Ortiz, I., Ovando, A., Pinales, C., Lara, W., Mateo-Perez, S. E., Cardenas-Alegria, O. V., Ramos, R. T. J., Rodriguez-Rodriguez, Y., Martín, L. O. M., & Franco, E. F. (2023). Comparison of Two Bacterial Characterization Techniques for the Genomic Analysis of River Microbiome. preprints.org. https://doi.org/10.20944/preprints202307.0696.v1

Bone, R. C., Balk, R. A., Cerra, F. B., Dellinger, R. P., Fein, A. M., Knaus, W. A.,... & 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

Breidenstein, E. B., de la Fuente-Núñez, C., & Hancock, R. E. (2011). *Pseudomonas aeruginosa*: all roads lead to resistance. Trends in Microbiology, 19(8), 419–426. https://doi.org/10.1016/j.tim.2011.04.005

Brink, M. (2018, March 26). *Nu gäller Sepsis-3 för definitioner och diagnostiska kriterier*. Läkartidningen. https://lakartidningen.se/klinik-och-vetenskap-1/artiklar-1/klinisk-oversikt/2018/03/nu-galler-sepsis-3-for-definitioner-och-diagnostiska-kriterier/

Boero, F., & Bernardi, G. (2014). Phenotypic vs genotypic approaches to biodiversity, from conflict to alliance. Marine Genomics, 17, 63–64. https://doi.org/10.1016/j.margen.2014.03.005

Bitar, I., Marchetti, V. M., Mercato, A., Nucleo, E., Anesi, A., Bracco, S., Rognoni, V., Hrabak, J., & Migliavacca, R. (2020). Complete Genome and Plasmids Sequences of a Clinical Proteus mirabilis Isolate Producing Plasmid Mediated NDM-1 From Italy. Microorganisms, 8(3), 339. https://doi.org/10.3390/microorganisms8030339

Bush, K. (2010). Alarming β-lactamase-mediated resistance in multidrug-resistant Enterobacteriaceae. *Current Opinion in Microbiology*, *13*(5), 558–564.https://doi.org/10.1016/j.mib.2010.09.006

Chambers, J. R.; Cherny, K. E. & Sauer, K. The susceptibility of *Pseudomonas aeruginosa* dispersed cells to antimicrobial agents is dependent on the dispersion cue and class of the antimicrobial agent used. Antimicrob. Agents Chemother. 61, 1–18 (2017).

Chen, L., Al Laham, N., Chavda, K. D., Mediavilla, J. R., Jacobs, M. R., Bonomo, R. A., & Kreiswirth, B. N. (2015). First Report of an OXA-48-Producing Multidrug-Resistant *Proteus mirabilis* Strain from Gaza, Palestine. *Antimicrobial Agents and Chemotherapy*, *59*(7), 4305–4307. https://doi.org/10.1128/aac.00565-15

Chen, S. L., Kang, Y. T., Liang, Y. H., Qiu, X. T., & Li, Z. J. (2023). A Core Genome Multilocus Sequence Typing Scheme for *Proteus mirabilis* PubMed, 36(4), 343–352. https://doi.org/10.3967/bes2023.040

Chiu, C. Y., & Miller, S. A. (2019). Clinical metagenomics. *Nature Reviews Genetics*, *20*(6), 341–355. https://doi.org/10.1038/s41576-019-0113-7

Chun, K., Syndergaard, C., Damas, C., Trubey, R., Mukindaraj, A., Qian, S., Jin, X., Breslow, S., & Niemz, A. (2015). Sepsis Pathogen Identification. SLAS Technology, 20(5), 539–561. https://doi.org/10.1177/2211068214567345

Clarke, K., Hall, C. L., Wiley, Z., Tejedor, S. C., Kim, J. S., Reif, L., Witt, L., & Jacob, J. T. (2019). Catheter-Associated Urinary Tract Infections in Adults: Diagnosis, Treatment, and Prevention *Journal of Hospital Medicine*, *15*(9), 552–556. https://doi.org/10.12788/jhm.3292

Cohen, J., Vincent, J. L., Adhikari, N. K. J., Machado, F. R., Angus, D. C., Calandra, T., Jaton, K., Giulieri, S., Delaloye, J., Opal, S., Tracey, K., van der Poll, T., & Pelfrene, E. (2015). Sepsis: a roadmap for future research. *The Lancet Infectious Diseases*, *15*(5), 581–614. https://doi.org/10.1016/s1473-3099(15)70112-x

Collignon, P., & McEwen, S. (2019). One Health—Its Importance in Helping to Better Control Antimicrobial Resistance. *Tropical Medicine and Infectious Disease*, *4*(1), 22. https://doi.org/10.3390/tropicalmed4010022

Czerwonka, G., Gmiter, D., & Durlik-Popińska, K. (2021). Draft Genome of *Proteus mirabilis* Serogroup O18 Elaborating Phosphocholine-Decorated O Antigen. *Frontiers in Cellular and Infection Microbiology*, *11*. https://doi.org/10.3389/fcimb.2021.620010

Dahui, Q. (2019). Next-generation sequencing and its clinical application Cancer Biology and Medicine, 16(1), 4–10. https://doi.org/10.20892/j.issn.2095-3941.2018.0055

Dargère, S., Cormier, H., & Verdon, R. (2018). Contaminants in blood cultures: importance, implications, interpretation, and prevention *Clinical Microbiology and Infection*, *24*(9), 964–969. https://doi.org/10.1016/j.cmi.2018.03.030

Dasgupta, N., Ferrell, E. P., Kanack, K. J., West, S. E. H., & Ramphal, R. (2002). fleQ , the Gene Encoding the Major Flagellar Regulator of Pseudomonas aeruginosa , Is σ 70 Dependent and Is Downregulated by Vfr, a Homolog of Escherichia coli Cyclic AMP Receptor Protein. Journal of Bacteriology, 184(19), 5240–5250. https://doi.org/10.1128/jb.184.19.5240-5250.2002

De Been, M., Pinholt, M., Top, J., Bletz, S., Mellmann, A., Van Schaik, W., Brouwer, E., Rogers, M. R. C., Kraat, Y., Bonten, M. J. M., Corander, J., Westh, H., Harmsen, D., & Willems, R. J. L. (2015). Core Genome Multilocus Sequence Typing Scheme for High-Resolution Typing of Enterococcus faecium. Journal of Clinical Microbiology, 53(12), 3788–3797. https://doi.org/10.1128/jcm.01946-15

Dellinger, R. P., Levy, M. M., Rhodes, A., Annane, D., Gerlach, H., Opal, S. M., Sevransky, J. E., Sprung, C. L., Douglas, I. S., Jaeschke, R., Osborn, T. M., Nunnally, M. E., Townsend, S. R., Reinhart, K., Kleinpell, R. M., Angus, D. C., Deutschman, C. S., Machado, F. R., Rubenfeld, G. D., . . . Moreno, R. (2013). Surviving Sepsis Campaign. *Critical Care Medicine*, *41*(2), 580–637. https://doi.org/10.1097/ccm.0b013e31827e83af

Dégi, J., Moţco, O. A., Dégi, D. M., Suici, T., Mareş, M., Imre, K., & Cristina, R. T. (2021). Antibiotic Susceptibility Profile of *Pseudomonas aeruginosa* Canine Isolates from a Multicentric Study in Romania. *Antibiotics*, *10*(7), 846. https://doi.org/10.3390/antibiotics10070846

Dötsch, A., Becker, T., Pommerenke, C., Magnowska, Z., Jänsch, L., & Häussler, S. (2009). Genomewide Identification of Genetic Determinants of Antimicrobial Drug Resistance in *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*, *53*(6), 2522–2531. https://doi.org/10.1128/aac.00035-09

Ertuğrul, B., Oryaşin, E., Lipsky, B. A., Willke, A., & Bozdogan, B. (2017). Virulence genes *fliC*, *toxA* and *phzS* are common among *Pseudomonas aeruginosa* isolates from diabetic foot infections. *Infectious Diseases*, *50*(4), 273–279. https://doi.org/10.1080/23744235.2017.1393839

Evans, T. (2018). Diagnosis and management of sepsis. *Clinical Medicine*, *18*(2), 146–149. https://doi.org/10.7861/clinmedicine.18-2-146

Fagerlind, M. G., Webb, J. S., Barraud, N., McDougald, D., Jansson, A., Nilsson, P., Harlén, M., Kjelleberg, S., & Rice, S. A. (2012). Dynamic modelling of cell death during biofilm development. *Journal of Theoretical Biology*, *295*, 23–36. https://doi.org/10.1016/j.jtbi.2011.10.007

Frieri, M., Kumar, K., & Boutin, A. (2017). Antibiotic resistance. *Journal of Infection and Public Health*, *10*(4), 369–378. https://doi.org/10.1016/j.jiph.2016.08.007

Girard, L., Lood, C., Rokni-Zadeh, H., van Noort, V., Lavigne, R., & De Mot, R. (2020). Reliable Identification of Environmental *Pseudomonas*Isolates Using the rpoD Gene. *Microorganisms*, *8*(8), 1166. https://doi.org/10.3390/microorganisms8081166

Girard, L., Lood, C., Höfte, M., Vandamme, P., Rokni-Zadeh, H., van Noort, V., Lavigne, R., & De Mot, R. (2021). The Ever-Expanding Pseudomonas Genus: Description of 43 New Species and Partition of the *Pseudomonas*putida Group *Microorganisms*, *9*(8), 1766. https://doi.org/10.3390/microorganisms9081766

Girlich, D., Bonnin, R. A., & Dortet, L. (2020). Genetics of Acquired Antibiotic Resistance Genes in Proteus spp. Frontiers in Microbiology, 11. https://doi.org/10.3389/fmicb.2020.00256

Girija, S.A.S., Priyadharsini, J.V., Madhavan, T., & Karutha Pandian, S. (2019). Virulence factors as predictive tools for drug resistance in *Pseudomonas aeruginosa*. Journal of Applied Microbiology, 127(1), 38-54. doi: 10.1111/jam.14255.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics, 17(6), 333–351. https://doi.org/10.1038/nrg.2016.49

Greipel, L., Fischer, S., Klockgether, J., Dorda, M., Mielke, S., Wiehlmann, L., Cramer, N., & Tümmler, B. (2016). Molecular Epidemiology of Mutations in Antimicrobial Resistance Loci of *Pseudomonas aeruginosa* Isolates from Airways of Cystic Fibrosis Patients. *Antimicrobial Agents and Chemotherapy*, *60*(11), 6726–6734. https://doi.org/10.1128/aac.00724-16

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

Gužvinec, M., Izdebski, R., Butić, I., Jelić, M., Abram, M., Koscak, I., Baraniak, A., Hryniewicz, W., Gniadkowski, M., & Andrašević, A. T. (2014). Sequence Types 235, 111, and 132 Predominate among Multidrug-Resistant *Pseudomonas aeruginosa* Clinical Isolates in Croatia. Antimicrobial Agents and Chemotherapy, 58(10), 6277–6283. https://doi.org/10.1128/aac.03116-14

Hamilton, A. L., Kamm, M. A., Ng, S. C., & Morrison, M. (2018). Proteus spp. as Putative Gastrointestinal Pathogens. Clinical Microbiology Reviews, 31(3). https://doi.org/10.1128/cmr.00085-17

Hooton, T. M., Bradley, S. F., Cardenas, D. D., Colgan, R., Geerlings, S. E., Rice, J. C., Saint, S., Schaeffer, A. J., Tambayh, P. A., Tenke, P., & Nicolle, L. E. (2010). Diagnosis, Prevention, and Treatment of Catheter-Associated Urinary Tract Infection in Adults: 2009 International Clinical Practice Guidelines from the Infectious Diseases Society of America *Clinical Infectious Diseases*, *50*(5), 625–663. https://doi.org/10.1086/650482

Hunter, P., & Gaston, M. A. (1988). Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. Journal of Clinical Microbiology, 26(11), 2465–2466. https://doi.org/10.1128/jcm.26.11.2465-2466.1988

Horcajada, J. P., Montero, M., Oliver, A., Sorlí, L., Luque, S., Gómez-Zorrilla, S., Benito, N., & Grau, S. (2019). Epidemiology and Treatment of Multidrug-Resistant and Extensively Drug-Resistant *Pseudomonas aeruginosa* Infections. *Clinical Microbiology Reviews*, *32*(4). https://doi.org/10.1128/cmr.00031-19

Hudek, K. (2009). Emergence Delirium: A Nursing Perspective. *AORN Journal*, *89*(3), 509–520. https://doi.org/10.1016/j.aorn.2008.12.026

Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology*, *82*(11), 801–811. https://doi.org/10.1016/j.humimm.2021.02.012

Jacobsen, S. M., Stickler, D. J., Mobley, H. L. T., & Shirtliff, M. E. (2008). Complicated Catheter-Associated Urinary Tract Infections Due to *Escherichia coli* and *Proteus mirabilis*. *Clinical Microbiology Reviews*, *21*(1), 26–59. https://doi.org/10.1128/cmr.00019-07

Jeukens, J., Kukavica-Ibrulj, I., Emond-Rheault, J. G., Freschi, L., Levesque, R. C., & Bourret, G. (2020). Genotypic and phenotypic characterization of *Pseudomonas aeruginosa* isolates from urinary tract infections. International Journal of Medical Microbiology, 310(3), 151417.

Jia, B., Raphenya, A. R., Alcock, B., Waglechner, N., Guo, P., Tsang, K. K.,... & McArthur, A. G. (2017). CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database Nucleic Acids Research, 45(D1), D566-D573.

Joensen, K. G., Engsbro, A. L., Lukjancenko, O., Kaas, R. S., Lund, O., Westh, H., & Aarestrup, F. M. (2017). Evaluating next-generation sequencing for direct clinical diagnostics in diarrhoeal disease. *European Journal of Clinical Microbiology &Amp; Infectious Diseases*, *36*(7), 1325–1338. https://doi.org/10.1007/s10096-017-2947-2

Juhas, M. (2015). *Pseudomonas aeruginosa* essentials: an update on investigation of essential genes. *Microbiology*, *161*(11), 2053–2060. https://doi.org/10.1099/mic.0.000161

JSpeciesWS. Retrieved 2023 November 16 from http://jspecies.ribohost.com/jspeciesws/

Kawahara-Matsumizu, M., Yamagishi, Y., & Mikamo, H. (2018). Misidentification of Neisseria cinerea as Neisseria meningitidis by matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS). Japanese journal of infectious diseases, 71(1), 85-87.

Kempker, J. A., & Martin, G. S. (2020). A global accounting of sepsis *The Lancet*, *395*(10219), 168–170. https://doi.org/10.1016/s0140-6736(19)33065-x

Kerr, K. G., and Snelling, A. M. (2009). *Pseudomonas aeruginosa*: a formidable and ever-present adversary The Journal of Hospital Infection, 73, 338-44.

Klockgether, J., & Tümmler, B. (2017). Recent advances in understanding *Pseudomonas aeruginosa* as a pathogen. *F1000Research*, *6*, 1261. https://doi.org/10.12688/f1000research.10506.1

Krause, K. M., Serio, A. W., Kane, T. R., & Connolly, L. (2016). Aminoglycosides: An overview. Cold Spring Harbor Perspectives in Medicine, 6(6), a027029. https://doi.org/10.1101/cshperspect.a027029

Kung, V. L., Ozer, E. A., & Hauser, A. R. (2010). The Accessory Genome of *Pseudomonas aeruginosa*. *Microbiology and Molecular Biology Reviews*, *74*(4), 621–641. https://doi.org/10.1128/mmbr.00027-10

Larsen, M. V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R. L., ... & Lund, O. (2012). Multilocus sequence typing of total genome sequenced bacteria. Journal of Clinical Microbiology, 50(4), 1355-1361.

Lee, I. K., Chang, J. P., Huang, W. C., Tai, C. H., Wu, H. T., & Chi, C. H. (2022). Comparative of clinical performance between next-generation sequencing and standard blood culture diagnostic method in patients suffering from sepsis. *Journal of Microbiology, Immunology and Infection*, *55*(5), 845–852. https://doi.org/10.1016/j.jmii.2022.07.011

Lember, G. (2021). Sepsis-associated Escherichia coli whole-genome sequencing analysis using in-house developed pipeline and 1928 tool.

Lengquist, M., Lundberg, O. H. M., Spångfors, M., Annborn, M., Levin, H., Friberg, H., & Frigyesi, A. (2020). Sepsis is underreported in Swedish intensive care units: A retrospective observational multicentre study. *Acta Anaesthesiologica Scandinavica*, *64*(8), 1167–1176. https://doi.org/10.1111/aas.13647

Levy, M. M., Fink, M. P., Marshall, J. C., Abraham, E., Angus, D., Cook, D., ... & Ramsay, G. (2003). 2001 sccm/esicm/accp/ats/sis international sepsis definitions conference. Intensive care medicine, 29(4), 530-538.

Li, B., & Yan, T. (2021). Next-generation sequencing reveals the limitations of qPCR methods in quantifying emerging antibiotic resistance genes (ARGs) in the environment. *Applied Microbiology and Biotechnology*, *105*(7), 2925–2936. https://doi.org/10.1007/s00253-021-11202-4

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

Ljungström, L., Pernestig, A. K., Jacobsson, G., Andersson, R., Usener, B., & Tilevik, D. (2017). Diagnostic accuracy of procalcitonin, neutrophil-lymphocyte count ratio, C-reactive protein, and lactate in patients with suspected bacterial sepsis. PloS one, 12(7)

Magnus Brink, Jonas Cronqvist, Anneli Fagerberg, Pär Lindgren, Miklos Lipcsey, Mantas Okas, & Lisa Kurland. (2018). [New definition of and diagnostic criteria for sepsis - Swedish use of Sepsis-3]. *Läkartidningen*, *115*.

Mahmoud, N. (2021). Next-generation diagnostics of E. coli from community-onset sepsis patients in Sweden: Studying the biodiversity of E. coli genomes

Mantzorou Marianna. (2011). What are the major ethical issues in conducting research? Is there a conflict between research ethics and the nature of nursing? *Health Science Journal*, *5*(1).

Mahfouz, N., Ferreira, I., Beisken, S., von Haeseler, A., & Posch, A. E. (2020). Critical assessment of the automated MPA Portable library for antibiotic resistance gene prediction Bioinformatics, 36(16), 4501-4506.

Martinez, J. L., Coque, T. M., & Baquero, F. (2015). What is a resistance gene? Ranking risk in resistomes. Nature Reviews Microbiology, 13(2), 116-123.

Mardis, E. R. (2017). DNA sequencing technologies: 2006–2016. Nature Protocols, 12(2), 213–218. https://doi.org/10.1038/nprot.2016.182

Marcon, J., Schubert, S., Stief, C. G., & Magistro, G. (2019). In vitro efficacy of phytotherapeutics suggested for prevention and therapy of urinary tract infections *Infection*, *47*(6), 937–944. https://doi.org/10.1007/s15010-019-01315-4

Mathur, S., Sabbuba, N. A., Suller, M. T. E., Stickler, D. J., & Feneley, R. C. L. (2005). Genotyping of urinary and fecal *Proteus mirabilis* isolates from individuals with long-term urinary catheters. *European Journal of Clinical Microbiology &Amp; Infectious Diseases*, *24*(9), 643–644. https://doi.org/10.1007/s10096-005-0003-0

Mayr, F. B., Yende, S., & Angus, D. C. (2013). Epidemiology of severe sepsis. *Virulence*, *5*(1), 4–11. https://doi.org/10.4161/viru.27372

Mizrahi, H., Peretz, A., Lesnik, R., Aizenberg-Gershtein, Y., Rodríguez-Martínez, S., Sharaby, Y., Pastukh, N., Brettar, I., Höfle, M. G., & Halpern, M. (2017). Comparison of sputum microbiome of legionellosis-associated patients and other pneumonia patients: indications for polybacterial infections. *Scientific Reports*, *7*(1). https://doi.org/10.1038/srep40114

Mobley, H. L., & Belas, R. (1995). Swarming and pathogenicity of *Proteus mirabilis* in the urinary tract *Trends in Microbiology*, *3*(7), 280–284. https://doi.org/10.1016/s0966-842x(00)88945-3

Nasrin, S., Hegerle, N., Sen, S., Nkeze, J., Sen, S., Permala-Booth, J., Choi, M., Sinclair, J. L., Tapia, M. D., Johnson, J. K., Sow, S. O., Thaden, J. T., Fowler, V. G., Krogfelt, K. A., Brauner, A., Protonotariou, E., Christaki, E., Shindo, Y., Kwa, A. L., . . . Tennant, S. M. (2022). Distribution of serotypes and antibiotic resistance of invasive Pseudomonas aeruginosa in a multi-country collection. BMC Microbiology, 22(1). https://doi.org/10.1186/s12866-021-02427-4

Nicas, T. I., & Hancock, R. E. (1983). *Pseudomonas aeruginosa* outer membrane permeability: isolation of a porin protein F-deficient mutant. *Journal of Bacteriology*, *153*(1), 281–285. https://doi.org/10.1128/jb.153.1.281-285.1983

Nicolle, L. E. (2005). Catheter-Related Urinary Tract Infection. *Drugs and Aging*, *22*(8), 627–639. https://doi.org/10.2165/00002512-200522080-00001

Nordmann, P., Naas, T., & Poirel, L. (2011). Global Spread of Carbapenemase-producing*Enterobacteriaceae*. *Emerging Infectious Diseases*, *17*(10), 1791–1798. https://doi.org/10.3201/eid1710.110655

Opota, O., Jaton, K., & Greub, G. (2015). Microbial diagnosis of bloodstream infection: towards molecular diagnosis directly from blood. *Clinical Microbiology and Infection*, *21*(4), 323–331. https://doi.org/10.1016/j.cmi.2015.02.005

Ostertagová, E., Ostertag, O., & Kováč, J. (2014). Methodology and application of the Kruskal-Wallis Test. Applied Mechanics and Materials, 611, 115–120. https://doi.org/10.4028/www.scientific.net/amm.611.115

Ozer, E. A., Allen, J. P., & Hauser, A. R. (2014). Characterization of the core and accessory genomes of *Pseudomonas aeruginosa* using bioinformatic tools Spine and AGEnt *BMC Genomics*, *15*(1), 737. https://doi.org/10.1186/1471-2164-15-737

Ozer, E. et al. An inside look at a biofilm: *Pseudomonas aeruginosa* flagella biotracking. Sci. Adv. 7, 1–15 (2021).

Pachori, P., Gothalwal, R., & Gandhi, P. (2019). Emergence of antibiotic resistance *Pseudomonas aeruginosa* in intensive care unit; a critical review. *Genes and Diseases*, *6*(2), 109–119. https://doi.org/10.1016/j.gendis.2019.04.001

Panayidou, S., Georgiades, K., Christofi, T., Tamana, S., Promponas, V. J., & Apidianakis, Y. (2020). *Pseudomonas aeruginosa* core metabolism exerts a widespread growth-independent control on virulence. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-66194-4

Parkins, M. D., Somayaji, R., & Waters, V. J. (2018). Epidemiology, Biology, and Impact of Clonal *Pseudomonas aeruginosa* Infections in Cystic Fibrosis. Clinical Microbiology Reviews, 31(4). https://doi.org/10.1128/cmr.00019-18

Pendleton, J. N., Gorman, S. P., & Gilmore, B. F. (2013). Clinical relevance of the ESKAPE pathogens *Expert Review of Anti-Infective Therapy*, *11*(3), 297–308. https://doi.org/10.1586/eri.13.12

Pereira, R., Oliveira, J., & Sousa, M. (2020). Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. *Journal of Clinical Medicine*, *9*(1), 132. https://doi.org/10.3390/jcm9010132

Pirnay, J., Bilocq, F., Pot, B., Cornelis, P., Zizi, M., Van Eldere, J., Deschaght, P., Vaneechoutte, M., Jennes, S., Pitt, T. L., & De Vos, D. (2009). Pseudomonas aeruginosa Population Structure Revisited. PLOS ONE, 4(11), e7740. https://doi.org/10.1371/journal.pone.0007740

Poonawala, H., Conner, T. M., & Peaper, D. R. (2018). The brief case: misidentification of Brucella melitensis as Ochrobactrum anthropi by matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS).

Poole, K. (2011). *Pseudomonas aeruginosa*: Resistance to the Max *Frontiers in Microbiology*, *2*. https://doi.org/10.3389/fmicb.2011.00065

Pradipta, I., Sodik, D., Parwati, I., Lestari, K., Halimah, E., Diantini, A., & Abdulah, R. (2013). Antibiotic resistance in sepsis patients: Evaluation and recommendation of antibiotic use. *North American Journal of Medical Sciences*, *5*(6), 344. https://doi.org/10.4103/1947-2714.114165

Qin, S., Xiao, W., Zhou, C., Pu, Q., Deng, X., Lan, L., Liang, H., Song, X., & Wu, M. (2022). *Pseudomonas aeruginosa*: pathogenesis, virulence factors, antibiotic resistance, interaction with the host,

technological advances, and emerging therapeutics *Signal Transduction and Targeted Therapy*, *7*(1). https://doi.org/10.1038/s41392-022-01056-1

Rahme L. G., Stevens E. J., Wolfort S. F., Shao J., Tompkins R. G., Ausubel F. M. (1955). Common virulence factors for bacterial pathogenicity in plants and animals. Science 268, 1899-902.

Quick, J., Cumley, N., Wearn, C. M., Niebel, M., Constantinidou, C., Simpson, T. J., Pallen, M. J., Moiemen, N., Bamford, A., Oppenheim, B., & Loman, N. J. (2014). Seeking the source of*Pseudomonas aeruginosa*infections in a recently opened hospital: an observational study using whole-genome sequencing. BMJ Open, 4(11), e006278. https://doi.org/10.1136/bmjopen-2014-006278

Riquelme, S. A., Liimatta, K., Wong Fok Lung, T., Fields, B., Ahn, D., Chen, D., Lozano, C., Sáenz, Y., Uhlemann, A. C., Kahl, B. C., Britto, C. J., DiMango, E., & Prince, A. (2020). *Pseudomonas aeruginosa* Utilizes Host-Derived Itaconate to Redirect Its Metabolism to Promote Biofilm Formation. *Cell Metabolism*, *31*(6), 1091-1106.e6. https://doi.org/10.1016/j.cmet.2020.04.017

Reynolds, D. (2021, November 7). *The Epidemiology, Pathogenesis, and Treatment of Pseudomonas aeruginosa Infections: An Update* SpringerLink. https://link.springer.com/article/10.1007/s40265-021-01635-6?error=cookies_not_supported&code=0451117a-943e-4d8c-acf3-0769fc4e67b4

Rhodes, A., Evans, L. E., Alhazzani, W., Levy, M. M., Antonelli, M., Ferrer, R., Kumar, A., Sevransky, J. E., Sprung, C. L., Nunnally, M. E., Rochwerg, B., Rubenfeld, G. D., Angus, D. C., Annane, D., Beale, R. J., Bellinghan, G. J., Bernard, G. R., Chiche, J. D., Coopersmith, C., . . . Dellinger, R. P. (2017). Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. *Intensive Care Medicine*, *43*(3), 304–377. https://doi.org/10.1007/s00134-017-4683-6

Richter, M., Rosselló-Móra, R., Glöckner, F. O., & Peplies, J. (2015). 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., Johnson, S. C., Agesa, K. M., Shackelford, K. A., Tsoi, D., Kievlan, D. R., Colombara, D. V., Ikuta, K. S., Kissoon, N., Finfer, S., Fleischmann-Struzek, C., Machado, F. R., Reinhart, K. K., Rowan, K., Seymour, C. W., Watson, R. S., West, T. E., Marinho, Hay, S. I. . . Naghavi, M. (2020). Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study The Lancet, 395(10219), 200–211. https://doi.org/10.1016/s0140-6736(19)32989-7

Ross, S. M. (2014). Introduction to probability and statistics for engineers and scientists Academic Press.

Rozwandowicz, M., Brouwer, M. S., Fischer, J., Wagenaar, J. A., Gonzalez-Zorn, B., Guerra, B., Mevius, D., & Hordijk, J. (2018). Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. Journal of Antimicrobial Chemotherapy, 73(5), 1121–1137. https://doi.org/10.1093/jac/dkx488

Ruiz-Villalba, A., van Pelt-Verkuil, E., Gunst, Q. D., Ruijter, J. M., & van den Hoff, M. J. (2017). Amplification of nonspecific products in quantitative polymerase chain reactions (qPCR). Biomolecular Detection and Quantification, 14, 7–18. https://doi.org/10.1016/j.bdq.2017.10.001

Safarirad, S., Arzanlou, M., Mohammadshahi, J., Vaez, H., Sahebkar, A., & Khademi, F. (2021). Prevalence and characteristics of metallo-beta-lactamase-positive and high-risk clone ST235 *Pseudomonas aeruginosa* at Ardabil hospitals. Jundishapur Journal of Microbiology, 14(3).

Saxenborn, P., Baxter, J., Tilevik, A., Fagerlind, M., Dyrkell, F., Pernestig, A. K., Enroth, H., & Tilevik, D. (2021). Genotypic Characterization of Clinical Klebsiella *spp.* Isolates Collected From Patients With Suspected Community-Onset Sepsis, Sweden. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.640408

Schaffer, J. N., & Pearson, M. M. (2015). *Proteus mirabilis* and Urinary Tract Infections. *Microbiology Spectrum*, *3*(5). https://doi.org/10.1128/microbiolspec.uti-0017-2013

Schütz, E., Fischer, A., Beck, J., Harden, M., Koch, M., Wuensch, T., Stockmann, M., Nashan, B., Kollmar, O., Matthaei, J., Kanzow, P., Walson, P. D., Brockmöller, J., & Oellerich, M. (2017). Graft-derived cell-free DNA, a noninvasive early rejection and graft damage marker in liver transplantation: A prospective, observational, multicenter cohort study. *PLOS Medicine*, *14*(4), e1002286. https://doi.org/10.1371/journal.pmed.1002286

Seymour, C. W., Liu, V. X., Iwashyna, T. J., Brunkhorst, F. M., Rea, T. D., Scherag, A., Rubenfeld, G., Kahn, J. M., Shankar-Hari, M., Singer, M., Deutschman, C. S., Escobar, G. J., & Angus, D. C. (2016). Assessment of Clinical Criteria for Sepsis. *JAMA*, *315*(8), 762. https://doi.org/10.1001/jama.2016.0288

Shah, R. U., Henry, T. D., Rutten-Ramos, S., Garberich, R. F., Tighiouart, M., & Bairey Merz, C. N. (2015). Increasing Percutaneous Coronary Interventions for ST-Segment Elevation Myocardial Infarction in the United States *JACC: Cardiovascular Interventions*, *8*(1), 139–146. https://doi.org/10.1016/j.jcin.2014.07.017

Shelenkov, A., Petrova, L., Fomina, V., Zamyatin, M., Mikhaylova, Y., & Akimkin, V. (2020). Multidrug-Resistant *Proteus mirabilis* Strain with Cointegrate Plasmid. *Microorganisms*, *8*(11), 1775. https://doi.org/10.3390/microorganisms8111775

Shemirani, M. I., Tilevik, D., Tilevik, A., Jurcevic, S., Arnellos, D., Enroth, H., & Pernestig, A. (2023). Benchmarking of two bioinformatic workflows for the analysis of whole-genome sequenced Staphylococcus aureus collected from patients with suspected sepsis. BMC Infectious Diseases, 23(1). https://doi.org/10.1186/s12879-022-07977-0

Schober, P., & Vetter, T. R. (2019). Two-Sample Unpaired t Tests in Medical Research. Anesthesia & Analgesia, 129(4), 911. https://doi.org/10.1213/ane.0000000000004373

Silva M, Machado MP, Silva DN, Rossi M, Moran-Gilad J, Santos S, Ramirez M, Carriço JA. 2018. chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. Microb Genom 4:000166. doi:10.1099/mgen.0.000166

Singer, A. C., Shaw, H., Rhodes, V., & Hart, A. (2016). Review of Antimicrobial Resistance in the Environment and Its Relevance to Environmental Regulators *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.01728

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, Rubenfeld, G. D., van der Poll, 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

Silva, M., Machado, M. P., Silva, D. a. L., Rossi, M., Moran-Gilad, J., Santos, S., Ramirez, M., & Carriço, J. A. (2018). chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. Microbial Genomics, 4(3). https://doi.org/10.1099/mgen.0.000166

Sivasubramaniam, S., Dlabolová, D. H., Kralikova, V., & Khan, Z. R. (2021). Assisting you to advance with ethics in research: an introduction to ethical governance and application procedures *International Journal for Educational Integrity*, *17*(1). https://doi.org/10.1007/s40979-021-00078-6

Smith, C. J., & Osborn, A. M. (2009). Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. FEMS Microbiology Ecology, 67(1), 6–20. https://doi.org/10.1111/j.1574-6941.2008.00629.x

Sónia Gonçalves Pereira, Ana Cristina Rosa & Olga Cardoso (2015) Virulence factors as predictive tools for drug resistance in *Pseudomonas aeruginosa*, Virulence, 6:7, 679-683, DOI: 10.1080/21505594.2015.1048958

Tacconelli, E., Carrara, E., Savoldi, A., Harbarth, S., Mendelson, M., Monnet, D. L., Pulcini, C., Kahlmeter, G., Kluytmans, J., Carmeli, Y., Ouellette, M., Outterson, K., Patel, J., Cavaleri, M., Cox, E. M., Houchens, C. R., Grayson, M. L., Hansen, P., Singh, N., . . . Zorzet, A. (2018). Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *The Lancet Infectious Diseases*, *18*(3), 318–327. https://doi.org/10.1016/s1473-3099(17)30753-3

Talbot, G. H., Bradley, J., Edwards, J. E., Gilbert, D., Scheld, M., & Bartlett, J. G. (2006). Bad Bugs Need Drugs: An Update on the Development Pipeline from the Antimicrobial Availability Task Force of the Infectious Diseases Society of America *Clinical Infectious Diseases*, *42*(5), 657–668. https://doi.org/10.1086/499819

Thaden, J. T., Park, L. P., Maskarinec, S. A., Ruffin, F., Fowler, V. G., & van Duin, D. (2017). Results from a 13-Year Prospective Cohort Study Show Increased Mortality Associated with Bloodstream Infections Caused by *Pseudomonas aeruginosa* Compared to Other Bacteria *Antimicrobial Agents and Chemotherapy*, *61*(6). https://doi.org/10.1128/aac.02671-16

Thrane, S. W., Taylor, V. L., Lund, O., Lam, J. S., & Jelsbak, L. (2016). Application of Whole-Genome Sequencing Data for O-Specific Antigen Analysis and In Silico Serotyping of Pseudomonas aeruginosa Isolates. Journal of Clinical Microbiology, 54(7), 1782–1788. https://doi.org/10.1128/jcm.00349-16

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

Tseemann. (2023). GitHub - tseemann/abricate: :mag_right: Mass screening of contigs for antimicrobial and virulence genes. GitHub. https://github.com/tseemann/abricate

Ulrich, B. C., & Paweletz, C. P. (2018). Cell-Free DNA in Oncology: Gearing up for Clinic. *Annals of Laboratory Medicine*, *38*(1), 1–8. https://doi.org/10.3343/alm.2018.38.1.1

Understanding ethics and types. (n.d.). Encyclopedia Britannica. https://www.britannica.com/summary/ethics-philosophy

Valentini, M., Gonzalez, D., Mavridou, D. A., & Filloux, A. (2018). Lifestyle transitions and adaptive pathogenesis of *Pseudomonas aeruginosa*. *Current Opinion in Microbiology*, *41*, 15–20. https://doi.org/10.1016/j.mib.2017.11.006

VFDB: Pathogenesis of *Pseudomonas* (2023) http://www.mgc.ac.cn/cgi-bin/VFs/genus.cgi?Genus=*Pseudomonas*

Vincent, J. L., Brealey, D., Libert, N., Abidi, N. E., O'Dwyer, M., Zacharowski, K.,... & Picard-Maureau, M. (2015). Rapid diagnosis of infection in the critically ill: a multicenter study of molecular detection in bloodstream infections, pneumonia, and sterile site infections Critical care medicine, 43(11), 2283

Vincent, J. L., Opal, S. M., Marshall, J. C., & Tracey, K. J. (2013). Sepsis definitions: time for change. Lancet (London, England), 381(9868), 774.

Váradi, L., Luo, J., Hibbs, D. E., Perry, J. R. B., Anderson, R. J., Orenga, S., & Groundwater, P. W. (2017). Methods for the detection and identification of pathogenic bacteria: past, present, and future Chemical Society Reviews, 46(16), 4818–4832. https://doi.org/10.1039/c6cs00693k

Walters, M. S., Grass, J. E., Bulens, S. N., Hancock, E. B., Phipps, E. C., Muleta, D., Mounsey, J., Kainer, M. A., Concannon, C., Dumyati, G., Bower, C., Jacob, J., Cassidy, P. M., Beldavs, Z., Culbreath, K.,

Phillips, W. E., Hardy, D. J., Vargas, R. L., Oethinger, M., . . . Kallen, A. (2019). Carbapenem-Resistant *Pseudomonas aeruginosa* at US Emerging Infections Program Sites, 2015. *Emerging Infectious Diseases*, *25*(7), 1281–1288. https://doi.org/10.3201/eid2507.181200

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p-Values: Context, Process, and Purpose. The American Statistician, 70(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

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

Winsor, G. L., Griffiths, E. J., Lo, R., Dhillon, B. K., Shay, J. A., & Brinkman, F. (2015). Enhanced annotations and features for comparing thousands of Pseudomonas genomes in the Pseudomonas genome database *Nucleic Acids Research*, *44*(D1), D646–D653. https://doi.org/10.1093/nar/gkv1227

Wong, P. H., von Krosigk, M., Roscoe, D. L., Lau, T. T., Yousefi, M., & Bowie, W. R. (2014). Antimicrobial co-resistance patterns of gram-negative bacilli isolated from bloodstream infections: a longitudinal epidemiological study from 2002–2011. *BMC Infectious Diseases*, *14*(1). https://doi.org/10.1186/1471-2334-14-393

Worth, L. J., Bull, A. L., Spelman, T., Brett, J., & Richards, M. J. (2015). Diminishing Surgical Site Infections in Australia: Time Trends in Infection Rates, Pathogens and Antimicrobial Resistance Using a Comprehensive Victorian Surveillance Program, 2002–2013. *Infection Control &Amp; Hospital Epidemiology*, *36*(4), 409–416. https://doi.org/10.1017/ice.2014.70

Yu, X., Torzewska, A., Zhang, X., Yin, Z., Drzewiecka, D., Cao, H., Liu, B., Knirel, Y. A., Rozalski, A., & Wang, L. (2017). Genetic diversity of the O antigens of *Proteus* species and the development of a suspension array for molecular serotyping PLOS ONE, 12(8), e0183267. https://doi.org/10.1371/journal.pone.0183267

Zabłotni, A., Arbatsky, N. P., Drzewiecka, D., Shashkov, A. S., & Knirel, Y. A. (2018). Chemical characterization and serological properties of a unique O-polysaccharide of the *Proteus mirabilis* Sm 99 clinical strain: – Identification of a new O81 serotype *International Journal of Biological Macromolecules*, *118*, 1131–1135. https://doi.org/10.1016/j.ijbiomac.2018.06.174

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

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., ... & Aarestrup, F. M. (2013). Identification of acquired antimicrobial resistance genes. Journal of Antimicrobial Chemotherapy, 67(11), 2640-2644.

Zhen, X., Stålsby Lundborg, C., Sun, X., Gu, S., & Dong, H. (2020). Clinical and Economic Burden of Carbapenem-Resistant Infection or Colonization Caused by Klebsiella pneumoniae, *Pseudomonas aeruginosa*, Acinetobacter baumannii: A Multicenter Study in China. *Antibiotics*, *9*(8), 514. https://doi.org/10.3390/antibiotics9080514

Zhong, Y., Xu, F., Wu, J., Schubert, J., & Li, M. M. (2021). Application of next-generation sequencing in laboratory medicine Annals of Laboratory Medicine, 41(1), 25–43. https://doi.org/10.3343/alm.2021.41.1.25

Özen A.I., Ussery D.W. (2012). Defining the *Pseudomonas* genus: where do we draw the line with Azotobacter? Microbial Ecology 63, 239-48

Appendix

Appendix I: Coding Guide Used for In-House-Developed Pipeline

```
#### The modived trimming code ###

chmod +x <unix_Quast.bash>.

bash unix_Quast.bash
cd /Users/suudahmed/Desktop/Raw_WGS_data/Pseudomonas/PS66
java -jar $Trim

#!/bin/bash
Trim='/Users/suudahmed/Desktop/NGS_Folder/NGS_Tools/Trimmomatic-0.39/-
trimmomatic-0.39.jar' #path to trimmomatic
Files1='/Users/suudahmed/Desktop/Raw_WGS_data/Pseudomonas/PS66'
mkdir out4_Trim #the output file
all_forw=($(find $Files1/*_1.fastq.gz)) #path to all forwad files
all_rev=($(find $Files1/*_2.fastq.gz)) #path to all reverse files
for file in ${!all_forw[@]}
do
 echo $file
 names1=$(basename ${all_forw[file]} _1.fastq.gz)
 main=$(echo ${all_forw[file]})
 mate=$(echo ${all_rev[file]})
 java -jar "$Trim" PE -phred40 $main $mate -baseout out4_Trim/${names1}.fastq.gz
ILLUMINACLIP:NexteraPE-PE.fa:0:30:10 HEADCROP:12 SLIDINGWINDOW: 4:20 MINLEN:30
done


rm out4_Trim/ *U.fastq.gz # Delete all unpaired files
```

Figure 1: Trimmomatic Execution Code This figure displays the code utilized to run Trimmomatic as per the bash script in Python.

```
#!/bin/bash

for file in *__1P.fastq.gz ; do
    id=$(basename "${file}" __1P.fastq.gz)
    echo "running unicycler assembly on ${id}"
    unicycler --spades_path SPAdes-3.13.0-Linux/bin/spades.py --short1 "${id}
__1P.fastq.gz" --short2 "${id}__2P.fastq.gz" --out OUTloop/${id}_FASTA
    echo "unicycler assembly on ${id} finished"
done
```

```
#!/bin/bash
mkdir $File OUTPUT_plasmidfinder
for file in *_assembly.fasta ; do
    id=$(basename "${file}" _assembly.fasta)
    echo "running abdicate on ${id}"
    abricate "${id}_assembly.fasta" --db plasmidfinder --csv > OUTPUT_plasmidfinder/${id}_plasmidfinder.csv
    echo "abricate on ${id} finished"
done
```

```
#!/bin/bash
mkdir $File OUTPUT_vfdb
for file in PS* ; do
    id=$(basename "${file}" PS)
    echo "running abricate on ${id}"
    abricate "${id}.fasta" --db vfdb --csv > OUTPUT_ vfdb/${id}_VFDB.csv
    echo "abricate on ${id} finished"
done
```

Figure 2: Unicycler, QUAST, and Abricate Execution Codes This figure shows the code used to run Unicycler and QUAST in the loop and in code form using Abricate.

```
#!/bin/bash

for file in *_1P.fastq.gz ; do
    id=$(basename "${file}" _1P.fastq.gz)
    echo "running unicycler assembly on ${id}"
    unicycler --spades_path SPAdes-3.13.0-Linux/bin/spades.py --short1 "${id}
_1P.fastq.gz" --short2 "${id}_2P.fastq.gz" --out Out_josna_unicycler/${id}_FASTA
    echo "unicycler assembly on ${id} finished"
done
cd tools


### My favorite species has no training file. What can I do? ###
prodigal -i myGoldStandardGenome.fna -t myTrainedFile.trn -p single


~/tools/OUTloop
chewBBACA.py CreateSchema -i ~/tools/OUTloop/*.fasta -o OutputSchemaFolder --ptf
ProdigalTrainingFile


chewBBACA.py CreateSchema -i ~/tools/OUTloop -o OutputSchemaFolder6 --ptf '/home/-
linux4/Downloads/Listeria_monocytogenes.trn'

chewBBACA.py CreateSchema -i ~/tools/OUTloop -o OutputSchemaPROTEUS --ptf ~/tools/-
OUTloop/OUTloop1/myTrainedFile.trn
###Allelecall
chewBBACA.py AlleleCall -i ~/tools/OUTloop -g /home/linux4/OutputSchemaFolder6/-
schema_seed  -o OutputFolderName




chewBBACA.py AlleleCall -i ~/tools/OUTloop -g OutputSchemaFolder/schema_seed -o
ALLELEPR_Out




### SchemaEvaluator ###

chewBBACA.py SchemaEvaluator -i /path/to/SchemaName -o /path/to/OutputFolderName --
cpu 4

chewBBACA.py SchemaEvaluator -i /home/linux4/OutputSchemaFolder6/schema_seed -o
SchemaEvaluatorProteus --cpu 4




###Determine the loci that constitute the cgMLST ####

chewBBACA.py ExtractCgMLST -i ~/OutputFolderName/results_20230321T133059/-
results_alleles.tsv -o ExtractCgMLSTPROTEUS


```

Figure 3: The figure depicts the code workflow for cgMLST determination using chewBBACA software. It involves training file preparation based on the reference genome, followed by code execution to generate the schema and cgMLST data. This enables comprehensive genetic analysis and comparison of organisms using allelic profiles. It also explains the code that will be used for Unicycler.

Appendix II: MultiQC Quality Assessment of FastQ Files with Species Analysis

The analysis carried out by the MultiQC tool on ten random *Proteus* samples is shown in the figures below. The isolate's forward file, both before (Figure 1) and after (Figures 2-3) trimming, is the subject of this analysis. A summary of the pre-trimming analysis is provided in Figure 1. Per-base sequence content, adapter content, and levels of sequence duplication were the three warnings that this initial analysis raised. The quality of the forward file significantly improved after trimming Figure 3, with the exception of sequence duplication levels, which have faint warning signs. Figure 2 shows a summary of the quality assessment after trimming.

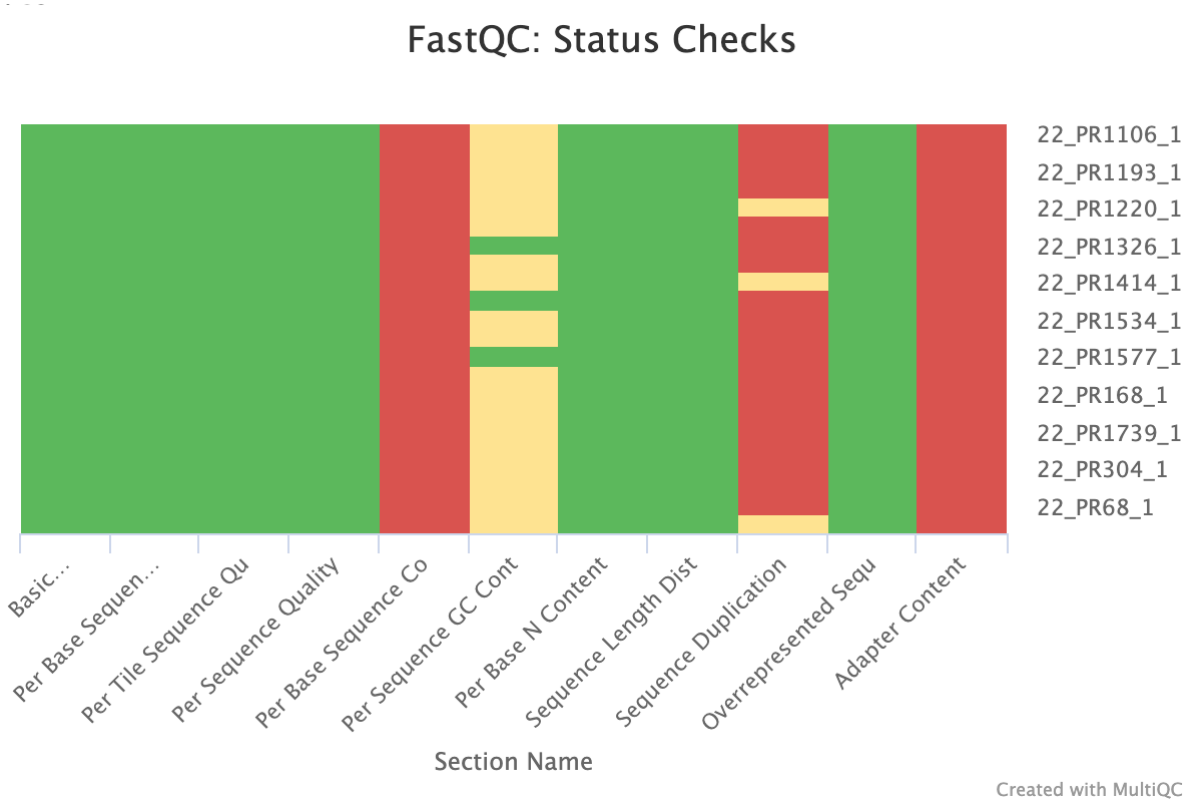

Figure 1: FastQC Status Checks Summary from MultiQC Output: A visual representation of the quality assessment for each FastQC section of ten randomly selected pre-trim files, categorized by normal (green), slightly abnormal (orange), or highly unusual (red) results

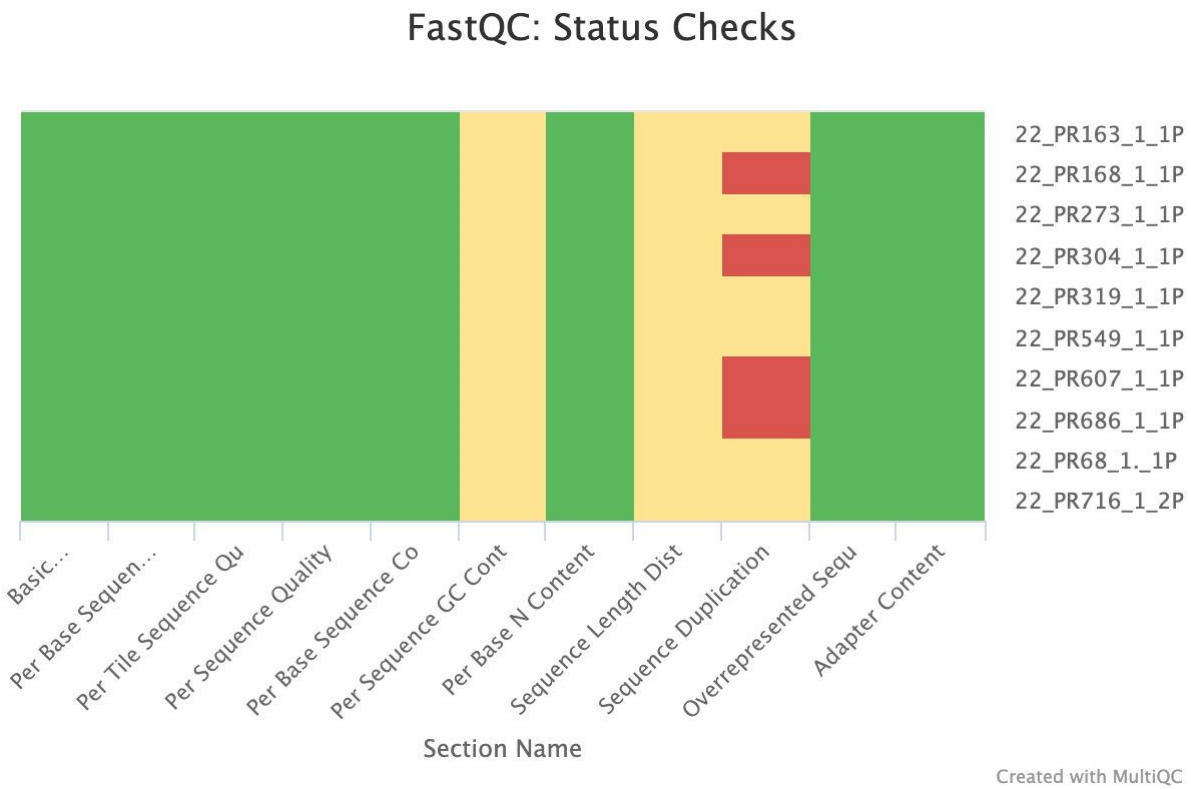

Figure 2. FastQC Status Checks Summary from MultiQC Output: Output: A visual representation of the quality assessment for each FastQC section of ten randomly selected post-trim files, categorized by normal (green), slightly abnormal (orange), or highly unusual (red) results.

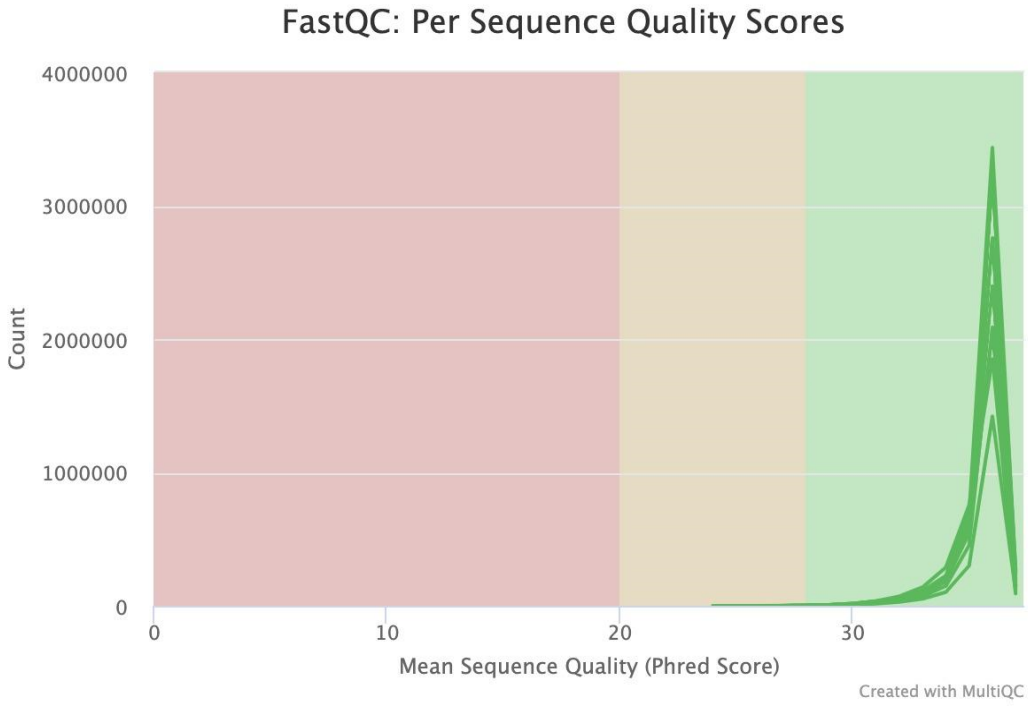

Figure 3: FastQC Per-Sequence Quality Scores Summary from MultiQC Output: A visual representation of the quality assessment for each FastQC section of ten randomly selected post-trim files, categorized by normal (green), slightly abnormal (orange), or highly unusual (red) results, and all grouped in the normal high-quality group (green).

Table 1: Pairwise Genome Sequence Alignment for *Pseudomonas spp.* Using the JSpeciesWS Online Service The table presents the results obtained from the JSpeciesWS online service for pairwise genome sequence alignment of *Pseudomonas spp.* The following parameters are included: file size, GC percent, number of contigs, average nucleotide identity (ANI) calculated using BLAST, percentage of aligned nucleotides, and length of aligned nucleotides.

Sample ID	reads before trim	reads after trim	#Contigs	L50	N50	Similiarity ref genome
PS 66	3769850	2420381	97	8	343902	96%
PS 105	5575250	3674425	91	9	265449	95.18%
PS 131	6905305	3569806	74	7	350291	95.84%
PS 239	5261338	3396056	79	8	268438	94.87%
PS 292	5577039	3576690	52	7	385409	95.78%
PS 325	6148210	3749963	52	5	450546	94.665 %
PS 350	3949642	2707485	58	5	428305	96.818 %
PS 581	4056012	2389816	74	8	304162	94.937 %
PS 644	7310637	4323656	54	7	359245	95.761 %
PS 857	6041739	3437264	73	7	362584	94.445 %
PS 864	5034074	2879197	55	6	398526	95.94%
PS 876	6185860	3655746	54	6	459032	95.94%
PS 886	4414184	2670256	93	9	241464	95.26%
PS 887	5433131	3410727	61	8	257358	95.408 %
PS 900	4628265	3085236	93	9	265448	95.24 %
PS 920	5258792	2889911	55	8	325211	95.761 %
PS 947	4245870	2646366	94	13	92376	0.004 %
PS 977	5057912	2993902	83	11	197613	94.801 %
PS 1024	2804553	1795382	61	6	367803	95.991 %
PS 1033	5116255	3038458	92	5	454882	94.83%
PS 1120	5014457	3219869	83	7	281612	96.35 %
PS 1194	14705044	8373984	53	6	386484	95.45%
PS 1198	10298697	6781025	54	5	428305	96.83%
PS 1217	5183389	2825174	15	3	720765	0.04%
PS 1314	11098502	6524474	127	6	371829	95.33%
PS 1344	9793159	5414976	37	5	426252	93.56%
PS 1358	9509651	6427801	84	9	226087	94.90%
PS 1405	11151815	6928504	83	7	391909	96.97%
PS 1512	4243998	2437923	55	5	360567	96.817
PS 1521	5354679	3300572	61	6	381604	96.31%
PS 1528	8478123	4632436	64	6	440574	96.061 %
PS 1538	7046813	4541664	59	6	349686	93.21%
PS 1620	7016177	4463788	65	7	321590	93.20%
PS 1654	5835176	3581731	109	9	286630	96.89%
PS 1670	10213305	6794473	73	7	363240	94.53%
PS 1744	22284993	4472437	102	10	223519	95.22%
PS 1781	4859633	2951885	53	6	409753	94.81%
PS 1782	6376214	4257861	57	6	430606	95.222
PS 1901	4909458	3087021	52	7	426255	94.397
PS 1912	7288910	4783023	83	7	314975	95.697
PS 1934	7801422	4864239	60	6	409879	95.411
PS 1943	6157516	3853402	64	7	349391	95.401
PS 1972	2782855	1951941	57	6	374982	95.31
PS 1984	7642434	5021138	48	6	425830	95.344
PS 2017	7702238	4626670	84	8	278770	96.174

Table 2: Samples exhibiting variations in species detection in different pipelines

Sample	JSpeciesWS	1928
PS947	Acinetobacter pittii, 96.76%	Acinetobacter pittii, 80.7 %
PS1217	*Pseudomonas* stutzeri, 98.1%	*Pseudomonas* stutzeri, 98.2 %
PR549	*Proteus* genomosp. 4	*Proteus* vulgaris, 91.02%
PR978	*Proteus* cloumbae, 99.79%	*Proteus* cloumbae, 50.6%
PR1486	*Proteus* cloumbae, 93.5%	*Proteus* cloumbae, 93.5%
PR1950	*Proteus* vulgaris, 91.02%	*Proteus* vulgaris, 94.9%

Appendix III: AMR result

Table 1. The frequency of detected antibiotic resistance genes in all isolates (n = 82) by Rasfinder of CGE and 1928 *Pseudomonas* pipeline for *P. aeruginosa*, and other pipelines for *P. mirabilis* samples.

Antibiotic_Class	Prevalence	Count	Total_Sample	Detected_Gene
Tetracycline	79.07	34	43	tet(J)
Chloramphenicol	32.56	14	43	cat_1
Aminoglycosides	18.60	8	43	ant(3'')-Ia, aph(3'')-Ib, aph(3')-Ia,
Beta-lactams	6.98	3	43	blaTEM-1B, blaCARB-2, blaCTX...
Sulfonamide	4.65	2	43	sul1
Trimethoprim	9.30	4	43	dfrA1
Quinolone	0	0	43	None

Table 2. The frequently detected antibiotic resistance genes with their description in all isolates (n = 82) from CGE and 1928 *Pseudomonas* pipeline for *P. aeruginosa*, and other pipelines for *Proteus* and CGE.

Description	1928 (Isolates genes)	CGE (Isolates genes)
Multi-drug efflux pumps ***Major facilitator superfamily***	mexA mexE mexX	None
Aminoglycoside 3'-phosphotransferase *(APHs)*	APH(3')-la APH(3'')-llb APH(6)-ld sat2	APH(3'')-llb
Broad-spectrum beta lactams oxacillin-hydrolysing-Beta lactamase family	OXA-905, 395, 50, 906, 853, 494, 847 BRO-1 ACT-74	BlaOXA-905, 906, 50, 494, 395
Chloramphenicol chloramphenicol acetyltransferase (CAT)	catB7	catB7

FLUOROQUINOLONE			
DNA mutations	gyrA (T83I) crpP parC (S87L) gyrA (D87N)	gyrA(S83L) parE(I529L)	parC(A56T)
Fosfomycin	fosA	fosA	
MLS **erythromycin ribosome methylase (erm)**	ermC		
Sulfonamides			
Sulfonamides resistant sul	Sul1 Sul2 Sul3	Sul1 Sul2 Sul3	
Tetracycline			
Major facilitator super family	Tet(A) Tet(B)	Tet(A) Tet(B)	
Trimethoprim			
Trimethoprim resistant dihydrofolate reductase dfr	dfrG	dfrA1 dfrA5	

Table 3. Comparative Analysis of Phenotypic and Genotypic Antibiotic Susceptibility Testing in *Proteus* Species

Antibiotic Class	Phenotypic AST (n)	Predicted genotypic AST from 1928 (n)	Predicted genotypic AST from CGE (n)	Discordant across methods (n [%])	Very major errors (n [%])	Major errors (n [%])
Aminoglycosides	27	27	27	13 (48.1%)	10 (37%)	3 (11.1)
Beta-lactams	32	32	32	4 (12.5%)	2 (6.3%)	2 (6.3%)
Chloramphenicol	29	29	29	14 (48.3%)	10 (34.5%)	4 (13.8)
Trimethoprim	32	32	32	10 (31.3%)	8 (25%)	2 (6.3%)
MLS	36	36	36	0 (0%)	0 (0%)	0 (0%)
Quinolone	27	27	27	0 (0%)	0 (0%)	0 (0%)
Sulfanamide	18	18	18	0 (0%)	0 (0%)	0 (0%)
Tetracycline	36	36	36	36 (100%)	36 (100%)	0 (0%)

This table presents a comparative analysis of phenotypic and genotypic antibiotic susceptibility testing (AST) in *Proteus* species across eight antibiotic classes. It outlines the number of phenotypic ASTs conducted, the corresponding genotypic ASTs predicted by the 1928 and CGE algorithms, and the extent of

discordance across these methods, expressed as a percentage. The table further highlights the number and percentage of Very Major Errors (VMEs) and Major Errors (MEs), providing an assessment of the accuracy of genotypic AST predictions in comparison to phenotypic tests.

Table 4. Comparative Analysis of Phenotypic and Genotypic Antibiotic Susceptibility Testing in *Pseudomonas* Species

Antibiotic	Phenotypic AST (n)	Predicted genotypic AST from 1928 (n)	Predicted genotypic AST from CGE (n)	Discordant across methods (n [%])	Very major errors (n [%])	Major errors (n [%])
Aminoglycoside	38	38	38	38 (100%)	38 (100%)	0 (0%)
Beta-lactam	38	38	38	38 (100%)	15 (39.5%)	23 (60.5%)
CEPHALOSPORIN	41	40	40	41 (100%)	11 (26.8%)	30 (73.2%)
FLUOROQUINOLONE	41	41	41	25 (61%)	19 (46.3%)	6 (14.6%)

This table offers a comparative analysis of phenotypic and genotypic antibiotic susceptibility testing (AST) in *Pseudomonas* species across four antibiotic classes. Similar to Table 1, it provides a detailed breakdown of the conducted phenotypic ASTs and the predicted genotypic ASTs from the 1928 and CGE algorithms. The level of discordance between these methods is indicated as a percentage. The occurrences and percentages of Very Major Errors (VMEs) and Major Errors (MEs) are also listed, shedding light on the challenges and limitations of genotypic AST predictions in reflecting true antibiotic resistance as determined by phenotypic tests.

Table 5. A comparison between the predicted genotypic antibiotic resistance results from 1928 and CGE for the *P. aeruginosa* isolates (n = 43)

Antibiotic Class	RR	SS	RS	SR	Discordant methods
Aminoglycosides	41	1	0	1	1
Broad Beta-lactams	42	1	0	0	0
FOSFOMYCIN	42	1	0	0	0
Chloramphenicol	42	1	0	0	0
CEPHALOSPORIN	42	1	0	0	0
FLUOROQUINOLONE	25	18	0	0	0
MLS	0	43	0	0	0

Table 5 presents an analysis of antibiotic resistance among bacterial isolates. The "Antibiotic Class" column represents the specific antibiotic class being evaluated. The "RR" and "SS" columns denote the count of isolates identified as resistant and susceptible, respectively, by both CGE and 1928. "RS" indicates the count of isolates deemed resistant by 1928 but susceptible by CGE, while "SR" represents those deemed susceptible by 1928 but resistant by CGE. Finally, the "Discordant Methods" column shows the total count of isolates where CGE and 1928' predictions disagreed.

Table 6. The counts of resistance genes detected by 1928 and CGE CGE for each antibiotic class from the clinical *P. aeruginosa* isolates (n = 43)

Antibiotics class	CGE	1928	P-value*
Aminoglycosides	43	46	-
Broad Beta-lactams	43	45	-
Chloramphenicol	32	45	-
MLS	0	1	-
Cephalosporin	42	0	-
Fosfomycin	43	43	-
Fluoroquinolone	25	32	-
Trimethoprim	0	1	-
Total	228	213	0.76

*The p-values are derived from paired t-tests; the choice of the test was based on the result of the normality test. The comparison was performed for all antibiotic classes.

Table 7. A comparison between the predicted genotypic antibiotic resistance results from 1928 and ResFinder for the *P. mirabilis* isolates (n = 39)

Antibiotic Class	RR	SS	RS	SR	Discordant across methods
Aminoglycosides	0	24	12	3	15
Broad Beta-lactams	0	26	3	10	13
Trimethoprim	0	28	6	5	11
Chloramphenicol	39	0	0	0	0
Sulfonamide	3	36	0	0	0
Quinolone	4	35	0	0	0
MLS	0	34	0	5	5
Tetracycline	29	0	10	0	10

Table 7 presents an analysis of antibiotic resistance among bacterial isolates. The "Antibiotic Class" column represents the specific antibiotic class being evaluated. The "RR" and "SS" columns denote the count of isolates identified as resistant and susceptible, respectively, by both CGE and 1928. "RS" indicates the count of isolates deemed resistant by 1928 but susceptible by CGE, while "SR" represents those deemed susceptible by 1928 but resistant by CGE. Finally, the "Discordant Methods" column shows the total count of isolates where CGE and 1928' predictions disagreed.

Table 8. The counts of resistance genes detected by 1928 and CGE for each antibiotic class from the clinical *P. mirabilis* isolates (n = 39)

Antibiotics class	CGE	1928	P-value*
Aminoglycosides	10	13	-
Broad Beta-lactams	6	3	-
Chloramphenicol	32	35	-
MLS	NA	1	-
Quinolones	1	1	-
Sulfonamides	3	3	-
Tetracyclines	36	38	-
Trimethoprim	7	10	-
Total	95	104	0.2308

*The p-values are derived from paired t-tests; the choice of the test was based on the result of the normality test. The comparison was performed for all antibiotic classes.

Appendix IV: MLST and serotype

Table 1. Comparative Analysis of *Pseudomonas* Strains Based on MLST, Serotype, Plasmids, Virulence Factor (VF) Genes, and Resistance Genes

SampleIDPS...	Specimen_type	MLST	Serotype	No. Plasmids	No. VF gene	No. of resistance gene
22_PS1024	Urine culture routine	968	O1	NA	227	6
22_PS1033	Wound culture	17	O1	NA	225	6
22_PS105	Upper respiratory culture	111	O4	NA	230	6
22_PS1120	Urine culture routine	245	O5	NA	231	6
22_PS1194	Wound culture extended	1068	O3	NA	223	5
22_PS1198	Wound culture	3285	O5	NA	238	5
22_PS131	Urine culture routine	1244	O9	NA	225	6
22_PS1314	Urine culture routine	3449	O6	NA	28	5
22_PS1344	Urine culture routine	671	O9	NA	214	5
22_PS1358	Blood culture aerob	270	O7	NA	231	5
22_PS1405	Blood culture aerob	3285	O2	NA	231	5
22_PS1512	Urine culture routine	3285	O5	NA	238	5
22_PS1521	Wound culture	0	O7	NA	230	6
22_PS1528	Urine culture routine	0	O9	NA	228	6
22_PS1538	Upper respiratory culture	1207	O6	NA	226	6
22_PS1620	Wound culture	1207	O6	NA	226	5
22_PS1654	Blood culture aerob	270	O2	NA	230	6
22_PS1670	Upper respiratory culture	560	O7	NA	230	6
22_PS1744	Urine culture routine	111	O4	NA	230	6
22_PS1781	Blood culture aerob	17	O1	NA	225	5
22_PS1782	Urine culture routine	27	O1	1	228	5
22_PS1901	Blood culture aerob	1062	O6	NA	226	6
22_PS1912	Wound culture	27	O1	1	229	6
22_PS1934	Wound culture	274	O3	NA	223	6

22_PS1943	Urine culture routine	274	O3	NA	223	6
22_PS1972	Urine culture routine	1090	O6	NA	225	6
22_PS1984	Blood culture aerob	1090	O6	NA	226	5
22_PS2017	Urine culture routine	1197		1	224	8
22_PS239	Urine culture extended	532	O11	NA	216	5
22_PS292	Urine culture routine	564	O9	NA	229	6
22_PS325	Urine culture extended	17	O1	NA	223	5
22_PS350	Wound culture	3285	O5	NA	238	5
22_PS581	Urine culture routine	1480	O1	NA	224	5
22_PS644	Urine culture routine	564	O9	NA	229	6
22_PS66	Wound culture	395	O6	4	229	6
22_PS857	Upper respiratory culture	244	O5	NA	235	5
22_PS864	Blood culture aerob	1485	O6	NA	226	5
22_PS876	Wound culture	1485	O6	NA	226	6
22_PS886	Urine culture routine	111	O4	NA	230	6
22_PS887	Urine culture routine	562	O3	NA	223	6
22_PS900	Upper respiratory culture	111	O4	NA	230	5
22_PS920	Urine culture routine	564	O9	NA	229	6
22_PS977	Urine culture routine	377	O7	NA	222	

Table 2. Analysis of Plasmid Types and Their Occurrences in *Proteus* Samples, Including Sample Types and Roles

Plasmid Type	No. of Occurrences	Sample Types	Role
Col3M_1	5	Wound culture, Urine culture routine, Blood culture anaerob	This plasmid is part of the ColE1-like plasmid family, often involved in resistance to antibiotics like tetracyclines, aminoglycosides, and beta-lactams.
ColE10_1	2	Urine culture routine	ColE10_1 is a plasmid from the ColE1-like plasmid family. It's frequently associated with the production of Colicin E10, a bactericidal protein, and antibiotic resistance.
IncN_1	1	Wound culture	IncN plasmids, including IncN_1, are known to carry multiple antibiotic resistance genes and are responsible for the spread of resistance to several classes of antibiotics. They are often associated with resistance to aminoglycosides, beta-lactams, and quinolones.
No identifiable plasmid	35(*Proteus*), 38(*Pseudomonas*)	Various specimens	The absence of identifiable plasmids in these samples suggests that they may either carry non-typical or novel plasmids not included in the database, or that the bacteria may not be relying on plasmid-mediated resistance mechanisms.

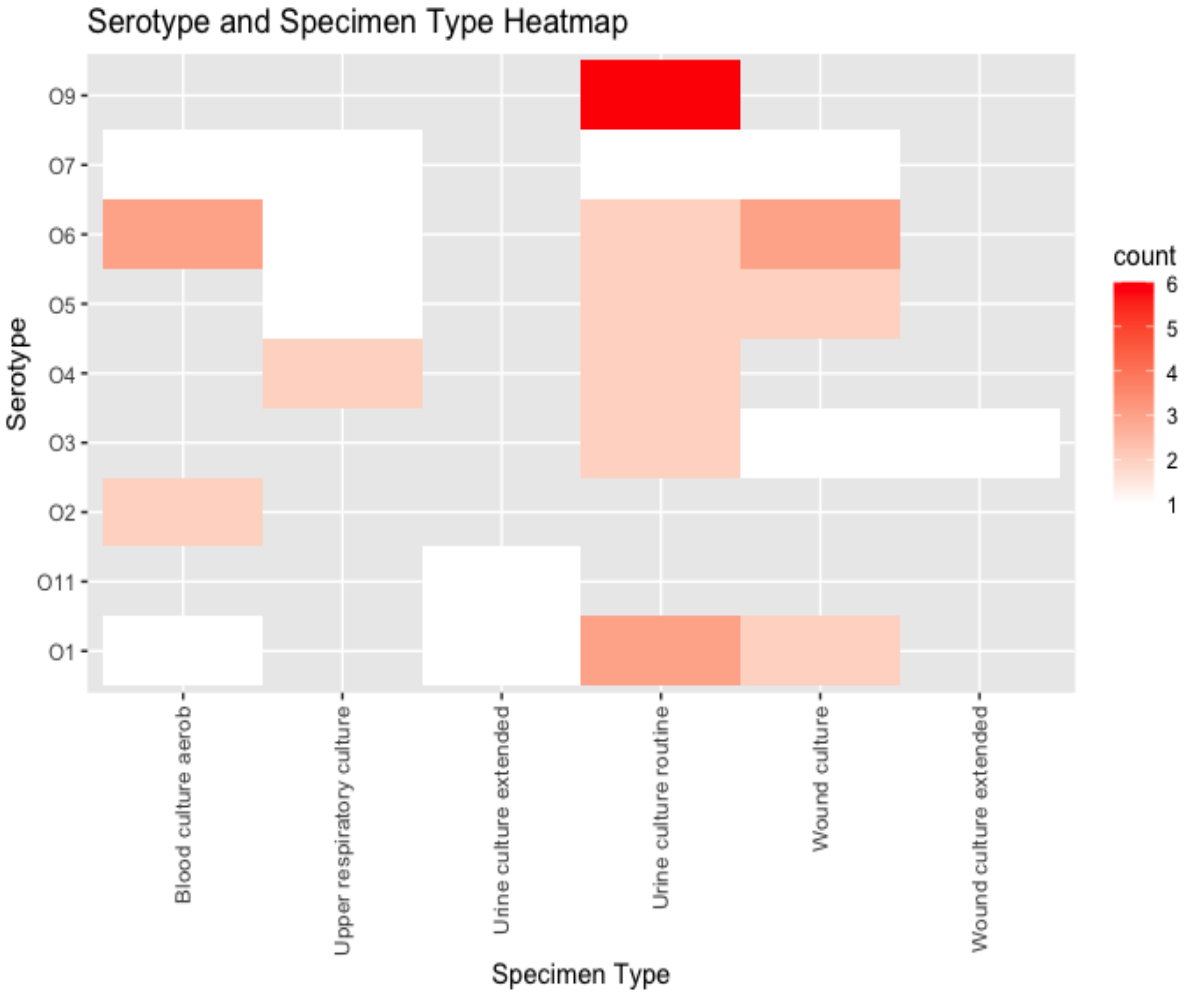

Figure 1: The Heatmap of serotypes in different samples of *P. aeruginosa*.

Appendix V: Phylogenic analysis

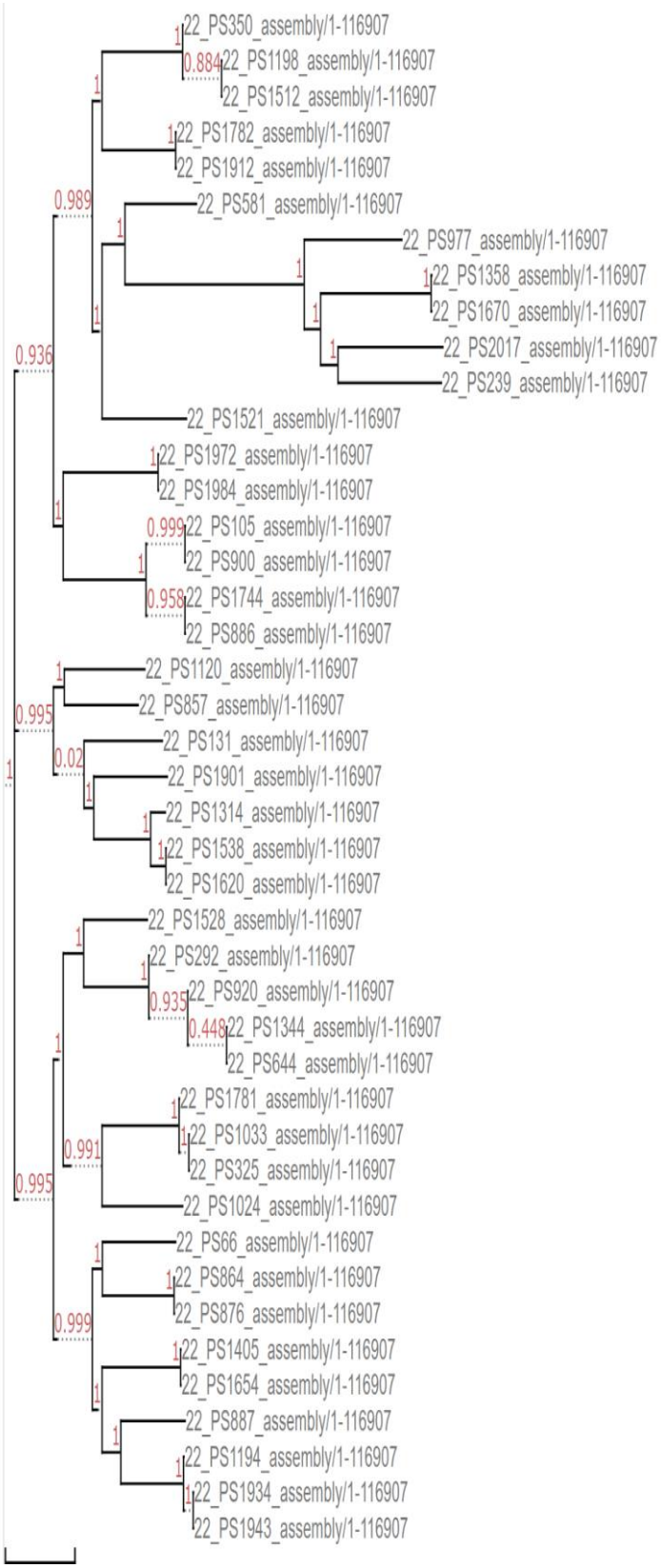

Figure 1. SNP-based phylogeny tree from the CGE CSIPhylogeny analysis.

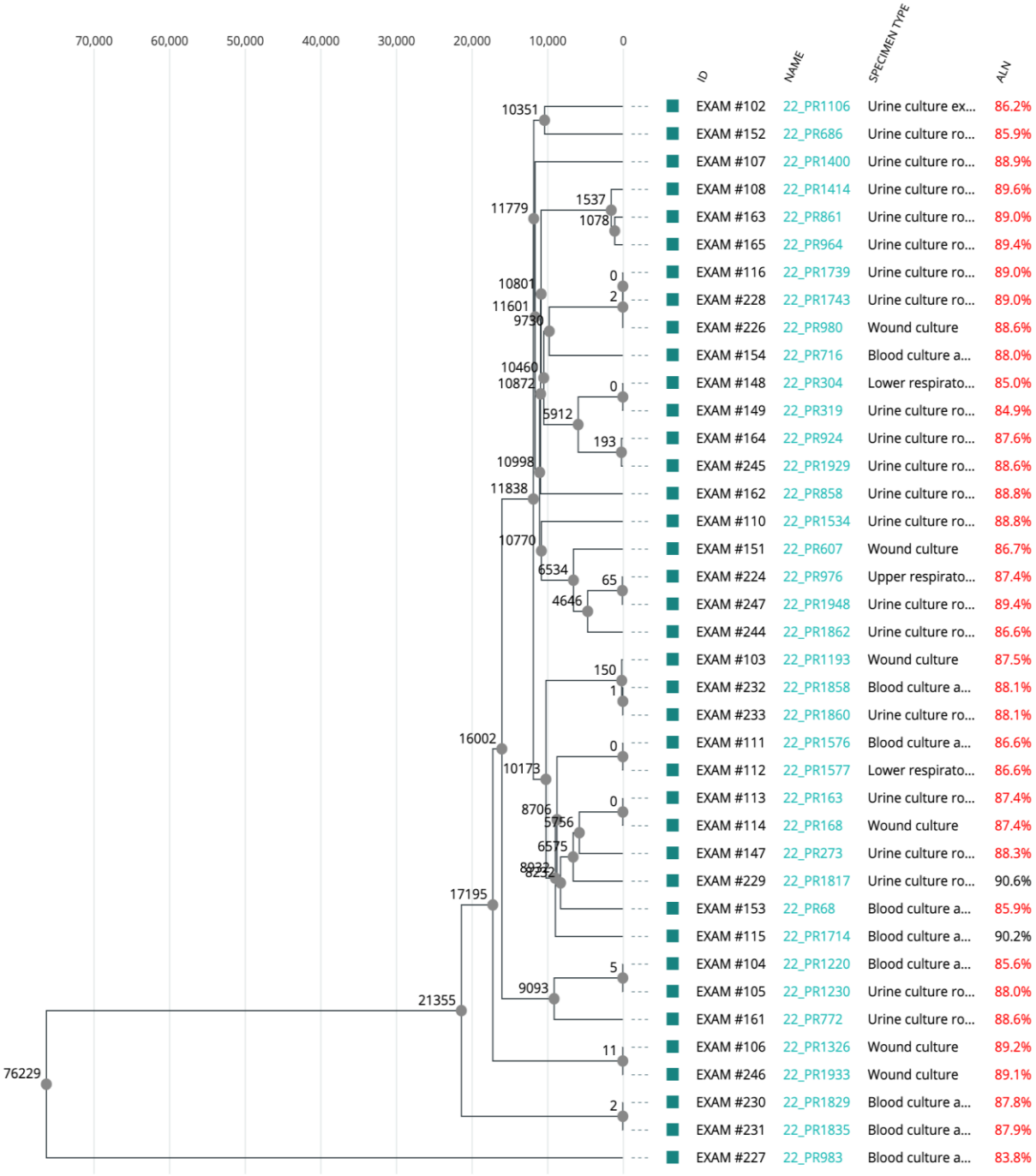

Figure 2: SNP Analysis on the 1928 Resulted in a Clustering Tree. The tree, generated using the reference genome *Proteus mirabilis* HI1230, effectively grouped isolates collected from the same patient at different sample locations into a single cluster

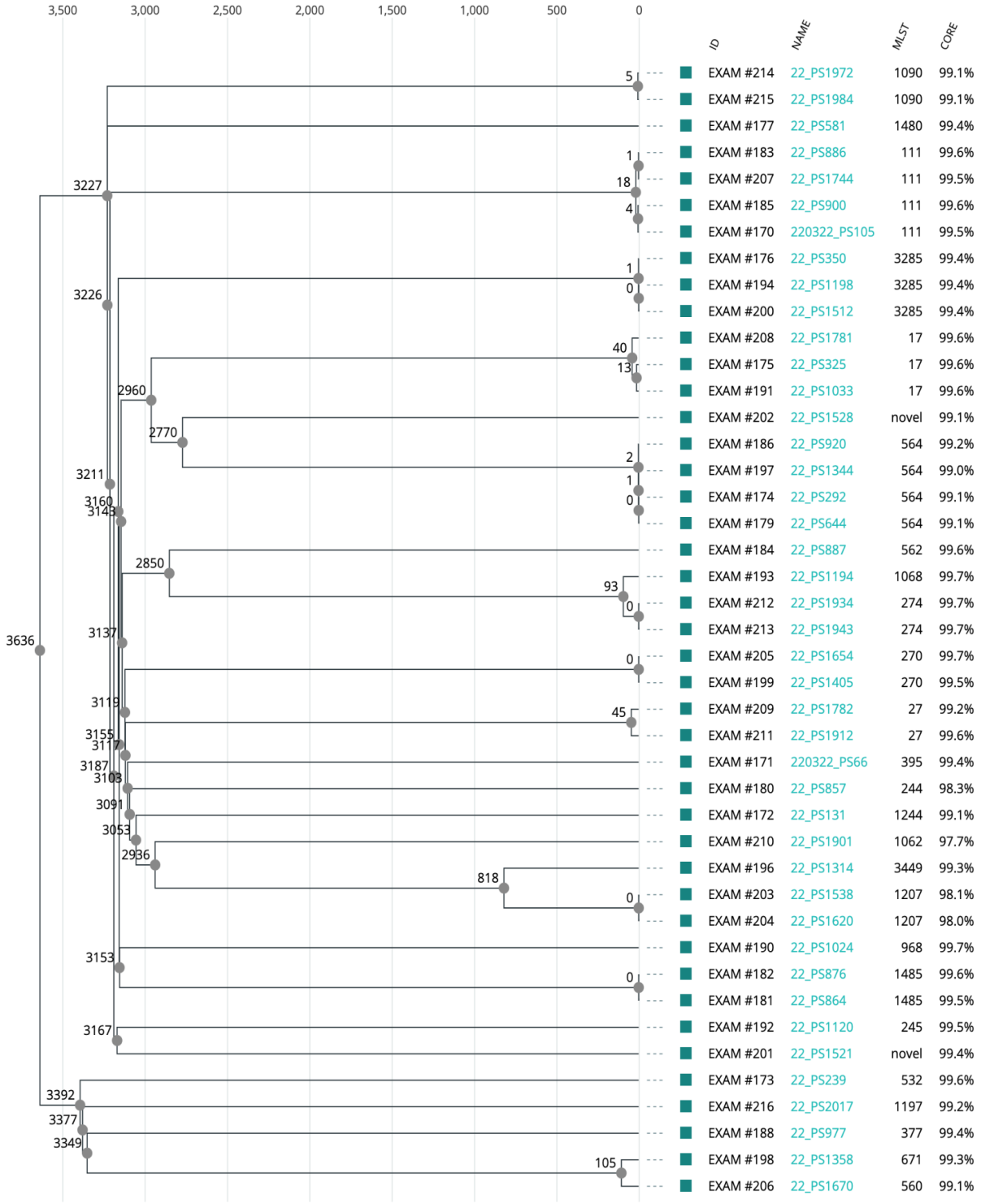

Figure 3: SNP Analysis on the 1928 Resulted in a Clustering Tree. The tree, generated using the reference genome *P. aeruginosa* PA01, effectively grouped isolates collected from the same patient at different sample locations into a single cluster

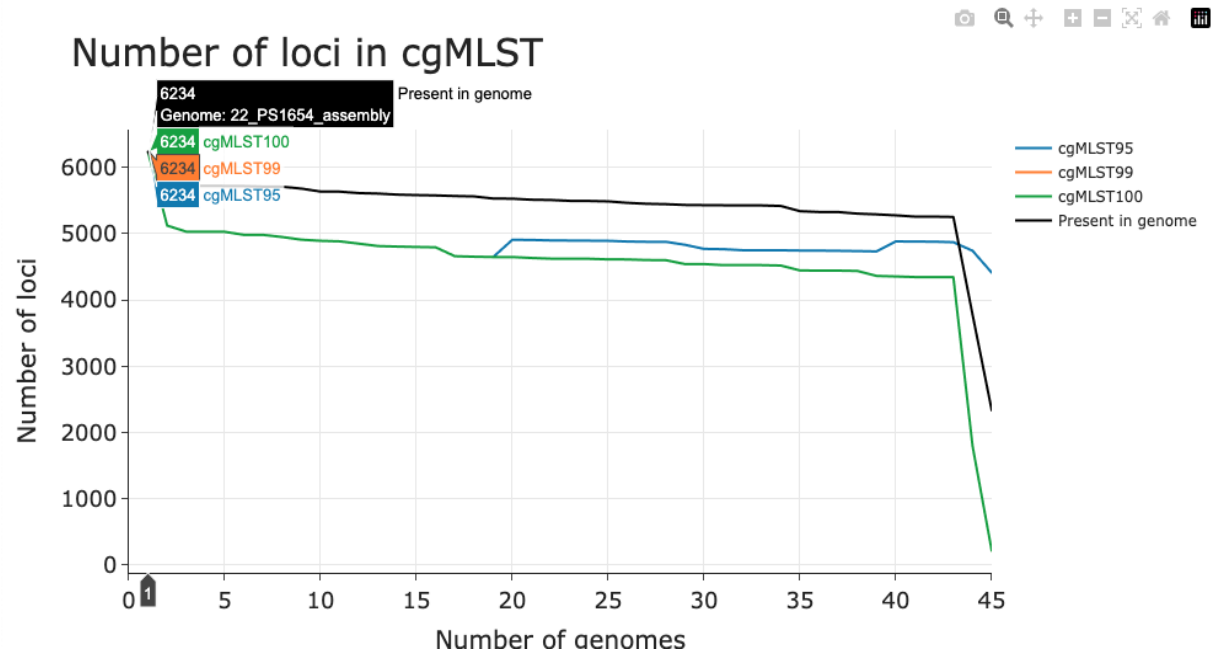

Figure 4: Comparison of cgMLST Loci Counts in *Pseudomonas* Genomes. This figure presents a line graph comparing the number of loci identified in 43 *Pseudomonas* genome samples using three different core genome multilocus sequence typing (cgMLST) thresholds: cgMLST 95 (95% sequence identity), cgMLST 99 (99% sequence identity), and cgMLST 100 (100% sequence identity). Each threshold is represented by a different colored line, with the height at each point indicating the number of loci identified in a sample's genome. The dark-colored points on each line represent loci present in the genome. This comparative visualization highlights the genetic diversity and complexity within these *Proteus* genomes as revealed by cgMLST at varying sequence identity thresholds.

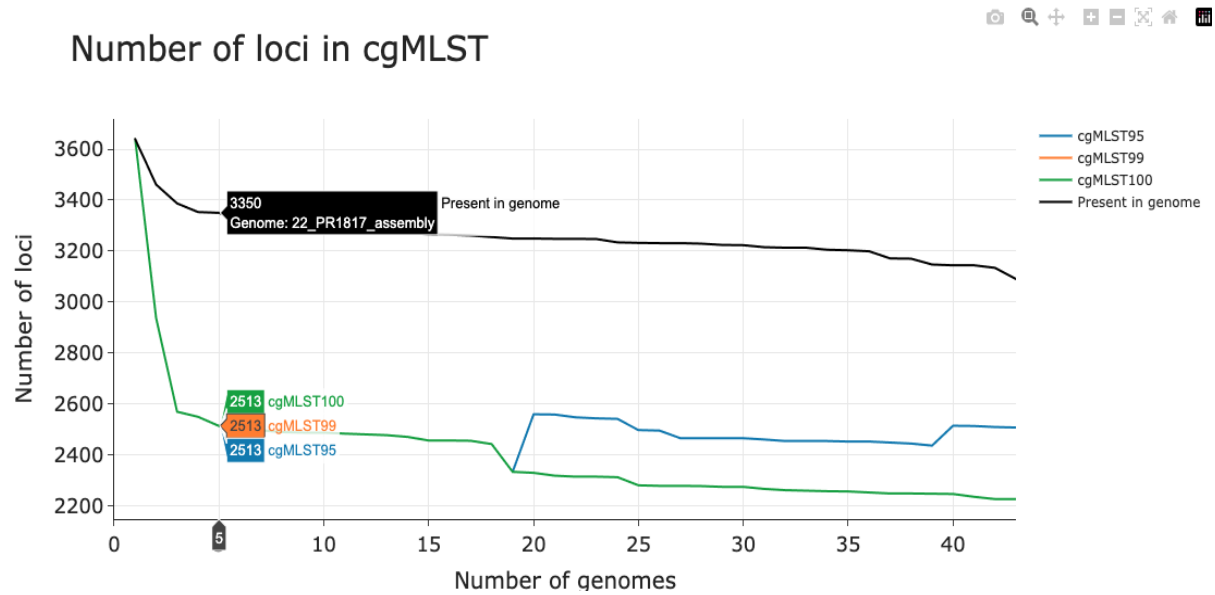

Figure 5: Comparison of cgMLST Loci Counts in *Proteus* Genomes. This figure presents a line graph comparing the number of loci identified in 39 *Proteus* genome samples using three different core genome multilocus sequence typing (cgMLST) thresholds: cgMLST 95 (95% sequence identity), cgMLST 99 (99% sequence identity), and cgMLST 100 (100% sequence identity). Each threshold is represented by a different colored line, with the height at each point indicating the number of loci identified in a sample's genome. The dark-colored points on each line represent loci present in the genome. This comparative visualization highlights the genetic diversity and complexity within these *Proteus* genomes as revealed by cgMLST at varying sequence identity thresholds.