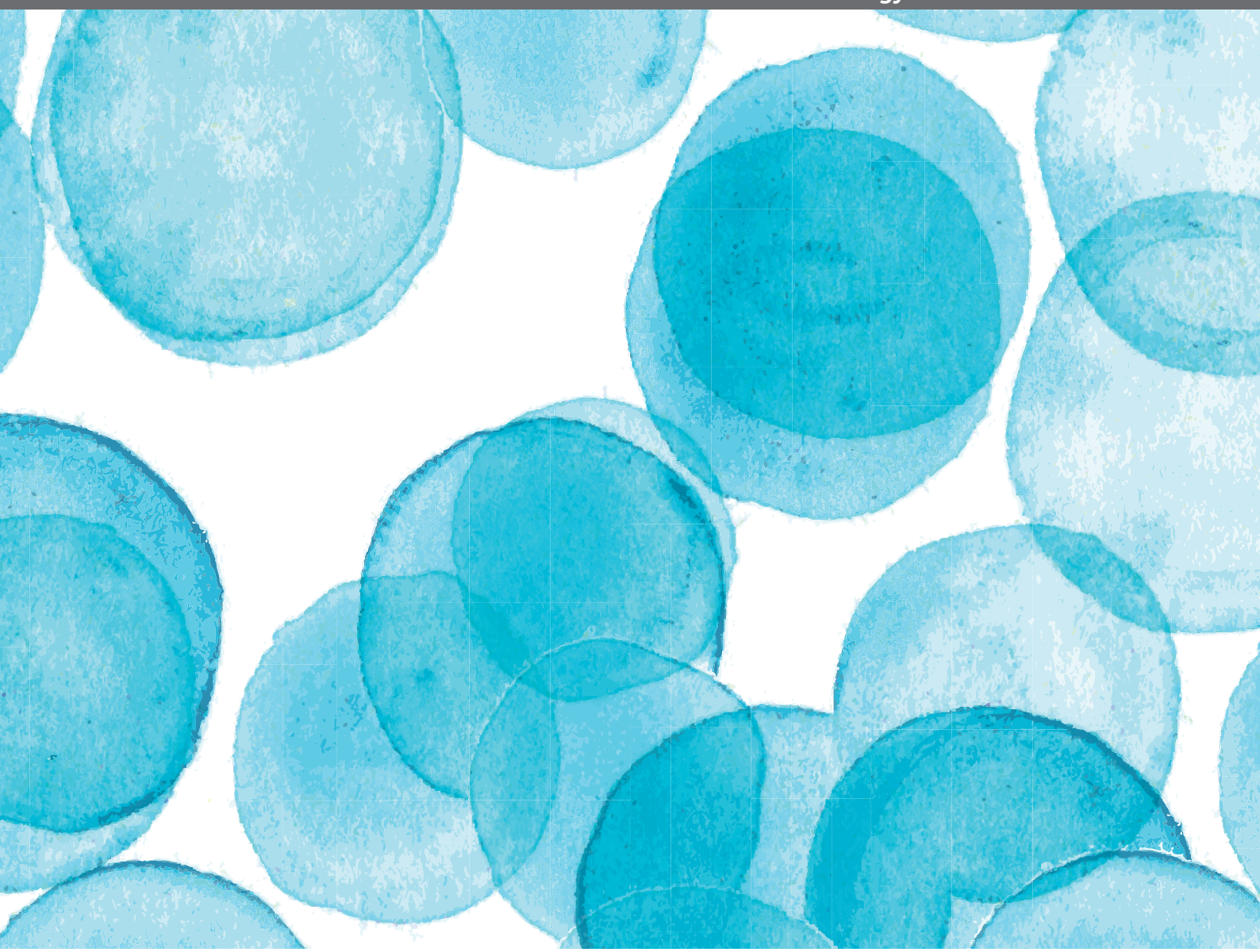
BACTERIAL PATHOGEN GENOMICS: RECENT ACHIEVEMENTS, CURRENT APPLICATIONS AND FUTURE CHALLENGES

EDITED BY: Stephan Fuchs, Guido Werner, Se-Ran Jun, Edward Feil, Kristin Hegstad and Yang Wang

PUBLISHED IN: Frontiers in Microbiology and Frontiers in Cellular and Infection Microbiology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-680-2
DOI 10.3389/978-2-88976-680-2

About Frontiers

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

Frontiers Journal Series

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

Dedication to Quality

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

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

BACTERIAL PATHOGEN GENOMICS: RECENT ACHIEVEMENTS, CURRENT APPLICATIONS AND FUTURE CHALLENGES

Topic Editors:
Stephan Fuchs, Robert Koch Institute (RKI), Germany
Guido Werner, Robert Koch Institute (RKI), Germany
Se-Ran Jun, University of Arkansas for Medical Sciences, United States
Edward Feil, University of Bath, United Kingdom
Kristin Hegstad, University Hospital of North Norway, Norway
Yang Wang, China Agricultural University, China

Citation: Fuchs, S., Werner, G., Jun, S.-R., Feil, E., Hegstad, K., Wang, Y., eds. (2022). Bacterial Pathogen Genomics: Recent Achievements, Current Applications and Future Challenges. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88976-680-2

Table of Contents

frontiers in Microbiology

ORIGINAL RESEARCH
published: 03 December 2020
doi: 10.3389/fmicb.2020.579802

A Biological Inventory of Prophages in *A. baumannii* Genomes Reveal Distinct Distributions in Classes, Length, and Genomic Positions

Belinda Loh[1†], Jiayuan Chen[1†], Prasanth Manohar[1], Yunsong Yu[2], Xiaoting Hua[2,3] and Sebastian Leptihn[1,2,4]**

[1] Zhejiang University-University of Edinburgh (ZJU-UoE) Institute, Zhejiang University, Haining, China, [2] Department of Infectious Diseases, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China, [3] Key Laboratory of Microbial Technology and Bioinformatics of Zhejiang Province, Hangzhou, China, [4] University of Edinburgh Medical School, Biomedical Sciences, College of Medicine & Veterinary Medicine, The University of Edinburgh, Edinburgh, United Kingdom

OPEN ACCESS

Edited by:
Se-Ran Jun,
University of Arkansas for Medical Sciences, United States

Reviewed by:
Li-Kuang Chen,
Tzu Chi University, Taiwan
Rapee Thummeepak,
Naresuan University, Thailand

***Correspondence:**
Xiaoting Hua
xiaotinghua@zju.edu.cn
Sebastian Leptihn
Leptihn@intl.zju.edu.cn;
sebalep@yahoo.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Microbiology

Received: 03 July 2020
Accepted: 10 November 2020
Published: 03 December 2020

Citation:
Loh B, Chen J, Manohar P, Yu Y, Hua X and Leptihn S (2020) A Biological Inventory of Prophages in A. baumannii Genomes Reveal Distinct Distributions in Classes, Length, and Genomic Positions. Front. Microbiol. 11:579802. doi: 10.3389/fmicb.2020.579802

Acinetobacter baumannii is of major clinical importance as the bacterial pathogen often causes hospital acquired infections, further complicated by the high prevalence of antibiotic resistant strains. Aside from natural tolerance to certain antibiotic classes, resistance is often acquired by the exchange of genetic information via conjugation but also by the high natural competence exhibited by *A. baumannii*. In addition, bacteriophages are able to introduce resistance genes but also toxins and virulence factors via phage mediated transduction. In this work, we analyzed the complete genomes of 177 *A. baumannii* strains for the occurrence of prophages, and analyzed their taxonomy, size and positions of insertion. Among all the prophages that were detected, *Siphoviridae* and *Myoviridae* were the two most commonly found families, while the average genome size was determined to be approximately 4 Mbp. Our data shows the wide variation in the number of prophages in *A. baumannii* genomes and the prevalence of certain prophages within strains that are most "successful" or potentially beneficial to the host. Our study also revealed that only two specific sites of insertion within the genome of the host bacterium are being used, with few exceptions only. Lastly, we analyzed the existence of genes that are encoded in the prophages, which may confer antimicrobial resistance (AMR). Several phages carry AMR genes, including OXA-23 and NDM-1, illustrating the importance of lysogenic phages in the acquisition of resistance genes.

Keywords: bacteriophage, prophage, *A. baumannii*, horizontal gene transfer, evolution, viral classification, antimicrobial resistance genes, phage genomes

INTRODUCTION

The opportunistic pathogen *Acinetobacter baumannii* is the causative agent for bloodstream infections, meningitis and urinary tract infections, and is responsible for 2–10% of all Gram-negative hospital-acquired infections (Joly-Guillou, 2005). Such infections include ventilator-associated pneumonia and bacteremia with a mortality rate of 35–52% (Dijkshoorn et al., 2007; Kempf et al., 2013; Antunes et al., 2014). As a multitude of strains cause nosocomial infections,

A. baumannii has become an important pathogen in hospital care and is of global concern. Many clinical isolates have acquired genes coding for virulence factors, such as toxins or efflux pumps, through various genetic uptake mechanisms (Morris et al., 2019). While *A. baumannii* easily acquires genetic material by conjugation, natural transformation is also widespread as many strains are highly naturally competent (Hu et al., 2019). Such mechanisms ultimately give rise to an increasing number of strains that display high levels of antimicrobial resistance (AMR), against which antibiotics show little or no effect. Genetic information for AMR genes is often embedded in genetic elements such as transposons or plasmids (Partridge et al., 2018; Rozwandowicz et al., 2018). In addition, bacteriophages (or phages) are able to transfer non-viral genetic information through a process called transduction, which can include genes coding for toxins or antimicrobial resistance (Wagner and Waldor, 2002; Derbise et al., 2007; Wachino et al., 2019). Therefore, phages play an important role in the development of AMR.

Regardless of their morphology or infection mechanism, phages can be divided into two types based on their life cycle: Lytic phages and lysogenic phages (sometimes also called temperate). Both eventually kill the host cell by lysis, employing various enzymes that create holes in the membrane and disintegrate the bacterial cell envelope, to allow the release of phage progeny. Few exceptions exist, such as the filamentous phages that are assembled in the membrane and secreted from the host while the bacterium continues to grow and divide (Loh et al., 2017, 2019; Kuhn and Leptihn, 2019). Nonetheless, phages that destroy the host by lysis upon completion of their life cycle, either start their viral replication immediately after entry (lytic phages) or integrate their genome into that of the host first (lysogenic). Lysogenic phages can remain "dormant" without replicating their genome or initiating phage coat protein synthesis. This way, lysogenic phages are being inherited by daughter cells, and might only replicate to form phage particles after many generations. The trigger for phage replication and synthesis is usually a stress signal produced by the host, such as a SOS response after DNA damage (Howard-Varona et al., 2017). However, if the host resides under favorable conditions, lysogenic phages continue their passive co-existence as so-called prophages embedded inside the DNA of the host.

Prophages are a major source of new genes for bacteria, occupying up to 20% of bacterial chromosomes and therefore may provide new functions to its host (Brüssow et al., 2004; Brüssow, 2007; Cortez et al., 2009; Fortier and Sekulovic, 2013; Wang and Wood, 2016). These functions include virulence factors and drug resistance mechanisms which include extracellular toxins and effector proteins involved in adhesion factors, enzymes, super antigens and invasion (Tinsley and Khan, 2006; Wang and Wood, 2016; Argov et al., 2017; Fortier, 2017). In some cases, the acquisition of virulence genes allows non-virulent bacteria to become a virulent pathogen. The most prominent example is that of the CTXΦ cholera toxin, encoded by a filamentous phage, making *Vibrio cholerae* the clinical pathogen that poses a substantial socio-economic burden on developing countries with poor hygiene due to frequent cholera outbreaks (Davis and Waldor, 2003). Another example is the Shiga toxin-encoding prophages found in highly virulent *Escherichia coli* strains, causing food-borne infections across the world (Gamage et al., 2004; Tozzoli et al., 2014). As part of the bacteriophage life cycle, prophages of lytic phages are a double edged sword; while they provide the advantage of increasing chances of survival in challenging environments, they could also lead to the killing of the host through the release of progeny at the end of the phage life cycle. As our relationship with microorganisms is a complex and vital one, from the important role of gut microbiota to the increase in mortality due to virulent microorganisms, understanding bacterial genomes is crucial. Yet in order to obtain detailed insight, we also need to be able to identify viral genes and to comprehend the impact of these genes on its host. As part of the bacterial genome, prophages are subjected to the general effects of mutation, recombination and deletion events. For some phages it has been clearly established that prophage genes have an influence on the host, such as motility or biofilm formation, both important aspects with regards to virulence. However for many other prophages, it is less well understood. Previous work on *A. baumannii* prophages have identified putative virulence factors and antibiotic resistance genes in host genomes deposited on GenBank (Costa et al., 2018; López-Leal et al., 2020). The work presented here analyses clinical *A. baumannii* strain genomes in search of possible prophages. We describe the identification of active prophages, analyzed their taxonomy, size and detail the regions in the bacterial genomes where these prophages have been found. Our data reveal a wide variation in the number of prophages in *A. baumannii* genomes and the prevalence of certain prophages within strains. From an evolutionary perspective, these might represent the "most successful" phages, or the ones that bring a benefit to the host. Our data analysis also allowed us to identify two major sites of insertion within the genome of the host bacterium, with most phage genomes inserting in these two regions, while only a few exceptions are being observed. In addition, our study indicates two distinct genome size distributions of prophages, as we observe a bimodal distribution when analyzing all prophage genomes. Furthermore, we describe genes coding for virulence factors, in particular for antimicrobial resistance in the prophages.

MATERIALS AND METHODS

A. baumannii genomes: Complete genome sequences of *A. baumannii* only were selected for this study. Detailed information of each *A. baumannii* strain used in this study is disclosed in supplementary material (**Supplementary Table S1**). For the characterization of genome lengths of *A. baumannii*, the following values were obtained: mean, median, mode, the smallest and the largest genome. The distribution of genome lengths was plotted using the geom_density function provided in the ggplot2 package in R (Wickham, 2016).

Alignment of *A. baumannii* genomes: All strains were aligned so that their starting position is identical, with the gene *dnaA*

defined as the start. BLAST Scoring Parameters provided by NCBI (Gertz et al., 2005; available at https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to blast the locations of *dnaA* and the adjustment of genome sequences was achieved by SnapGene software (from Insightful Science; available at snapgene.com).

Identification of Prophage Genes: The tool used to identify prophages in *A. baumannii* genomes was Prophage Hunter (Song et al., 2019; available at https://pro-hunter.bgi.com/). Here, we obtained data on the start, end, length, score, category, and the name of the closest phage. Phaster (Arndt et al., 2016; available at http://phaster.ca/) was used to further confirm some conflicting results. According to the algorithm created by the authors of Prophage Hunter, an "active" prophage is defined by a score close to 1, while the probability decreases the lower the score gets. This means that an active prophage region received a scoring of higher that 0.8 while 0.5–0.8 is defined as "ambiguous," and a score lower than 0.5 as "inactive."

Prophage number analysis: Calculations of the mean, median, and mode on the prophage number (total, only active, and only ambiguous) were performed after removing the overlaps of the same prophage in the same strain. The 10 strains with the fewest and with the largest numbers of total, only active, and only ambiguous prophages were selected to show the 2 extremes, while the density plot achieved through geom_density function provided by ggplot2 package in R were conducted to describe the general distribution. The boxplot produced by R reflected the relationship between total, only active, and only ambiguous prophage number and *A. baumannii* genome length.

Phylogenetic analysis of host strains: Prokka v1.13 (Seemann, 2014) was used to generate the gff files for the genome sequences of 177 *A. baumannii* strains. The core genome alignment was constructed with Roary v3.12.0 (Page et al., 2015). A maximum-likelihood phylogenetic tree was created using FastTree v2.1.10 (Price et al., 2010). The tree was annotated and visualized with ggtree.

Prophage classification and phylogeny analysis: Prophage classification presented was provided by the program Prophage Hunter which was based on the NCBI's database. Different orders and families were taken into consideration. The prophage number in different families and their proportions were revealed by histograms and pie charts generated in Microsoft Excel, respectively. Different bacterial hosts were referred to in the calculation of the number of prophages and the number of phage species they had. Based on the species of prophages, 10 most common ones for total, only active, and only ambiguous were selected, with a heatmap which was completed through geom_tile provided by ggplot2 package in R showing their number with the activity value in different strains. The phylogeny of phages was mapped according to their sequences. The alignment of the phage sequences were performed using Multiple Alignment using Fast Fourier Transform (MAFFT) (Katoh and Standley, 2013) with default options. Maximum-likelihood phylogenetic trees were created using FastTree v2.1.10. The tree was annotated and visualized using ggtree.

Prophage location analysis: The positions of all prophages were first showed in a stacked bar chart in Microsoft Excel. Considering the overlaps of different prophages in the same strain, all the prophage starts and ends were mapped in a density plot created by ggplot2 geom_density and aes functions in R. The stacked bar charts of different prophage species were used to estimate their preference of insertion which were then summarized in tables.

Prophage length analysis: The use of ggplot2 geom_density function in R facilitated the creation of density plots of prophage length. With the help of aes function, the mapping of prophage categories (active and ambiguous) and families were achieved in the density plots.

Identification of virulence factors and antibiotic resistance genes: No program is currently available that allows the search of virulence genes which are embedded in prophage sequences within bacterial genomes. Thus, we identified prophages first followed by manually correlating genomic positions of virulence genes with those that were also identified to belong to prophage genes.

Mapping of prophage-encoded antimicrobial resistance genes (ARG): To search for the specific virulence genes we identified (above), we first downloaded all available 4,128 *A. baumannii* Illumina sequencing reads from the Sequence Read Archive (SRA) with the cut-off date for deposited sequences on 2019/11/17. The raw Illumina sequencing reads were mapped against the ARG prophage sequences employing BWA-MEM v0.7.17 (80% coverage cutoff) (Li, 2013).

RESULTS

In silico Discovery of Prophages in 177 *A. baumannii* Genomes Identifies 1,156 Prophage Sequences

Our first aim was to analyze how frequently prophages occur in the genomes of *A. baumannii* strains. We randomly chose 177 genome sequences of *A. baumannii* strains, many of them clinical isolates. For the subsequent analyses, we aligned all sequences so that their starting positions are identical. To this end, we defined the gene *dnaA* as the start, which codes for a replication initiation factor that facilitates DNA replication in bacteria. From the sequence alignments, we observed a large variation in genome sizes. The average length of the genomes was 3,981,579 bp with a median value of 4,001,318 bp; the smallest genome had a length of 3,072,399 bp (22.9% shorter than average), and the largest genome displayed a size of 4,389,990 bp (10.25% larger than average) (**Figure 1A**).

Next, we used the online platform Prophage Hunter (Song et al., 2019) to identify active and ambiguous prophage genomes. The algorithm provides an output value for each prophage identified, which allows the researcher to establish whether a sequence contains an "active" or an ambiguous prophage. While "active" prophages exhibit the complete genomic sequence of a prophage, and are therefore likely to allow the production of phage particles, ambiguous prophage sequences are truncated, mutated or otherwise incomplete, and unlikely to be able to form infectious phages. Among the 177 *A. baumannii* genomes analyzed, we identified 1,156 prophages, with 459 of them being

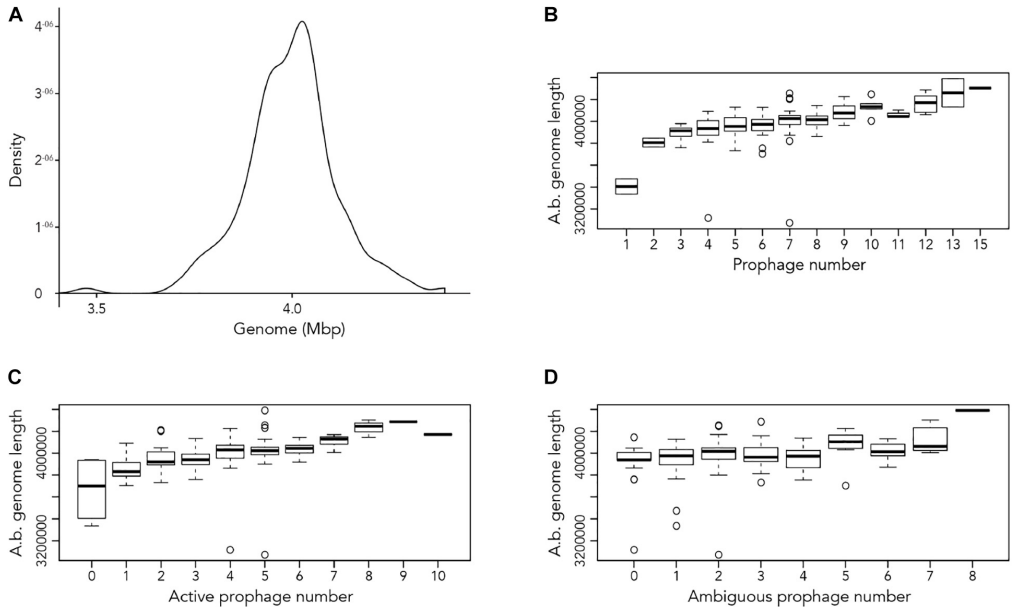


FIGURE 1 | Length of *A. baumannii* genomes and its correlation with number of prophages present. **(A)** Density graph of 177 genome sequences of *A. baumannii* strains, indicating the distribution of the lengths of genome sequences analyzed. **(B–D)** Correlation between *A. baumannii* genome length and number of **(B)** all prophages, **(C)** active prophages, and **(D)** ambiguous prophages identified in genomes.

defined as "ambiguous" while the remaining 697 sequences were labeled as "active" according to the program (**Supplementary Table S2**). To determine the prevalence of prophages in the *A. baumannii* genomes, we analyzed the number of prophages per genome. Using a heatmap to illustrate our results, we found that while some prophages are rarely found, others are quite common in the genomes of *A. baumannii* isolates (**Figure 2**). One phage that was only found once, for example, is a prophage with high sequence similarity to the *Yersinia* Podovirus fHe-Yen3-01, while the *Acinetobacter* phage Bphi-B1251, a Siphovirus, has been found in 79.1% (140 in 177) of all analyzed *A. baumannii* genomes. Such high prevalence observed by prophages such as Bphi-B1251 could indicate high infectivity and wide host range of the active phage particle. **Table 1** shows 10 strains with the fewest and with the largest numbers of prophages identified (**Table 1**). We found that the average prophage number in an *A. baumannii* genome is 6.53, with some bacterial genomes containing only one (n = 2) prophage sequence, such as in case of the *A. baumannii* strains DS002 and VB1190. In contrast to these, other strains have been found to contain as many as 10 prophage sequences, such as in strain 9201 (n = 1), that were labeled "active" by Prophage Hunter; additionally this strain contains two prophage sequences that were defined as ambiguous. The highest prophage number was found in the strain AF-401 which contains 15 prophages, however, only 8 were defined as active. Our results show that prophage sequences are relatively common and that most *A. baumannii* strains show a median of seven and a mode of eight prophages per genome. Additional genome analysis of the clinical isolates illustrates a possible relationship between prophages and host strains (**Supplementary Figure S1**). Prophages, such as *Bacillus* phage PfEFR 4 and *Enterobacteria* phage CUS 3 were observed more frequently in strains whose genomes are in the same clade or are closely related, indicating a narrow host range.

We next correlated host genome length with number of prophages identified to determine if there is a relationship between the two variables. Perhaps unsurprisingly, the length of *A. baumannii* genomes increased as more prophages are identified, disregarding whether the prophage genomes are "active" or "ambiguous" (**Figure 1B**). However, when prophages are classified, the correlation between host genome length and number of prophage genomes identified were less distinct (**Figures 1C,D**).

Siphoviridae and *Myoviridae* Are the Two Most Commonly Found Classes of Prophages in *A. baumannii* Genomes

We analyzed the relationship of all prophages we identified and created a phylogenetic tree (**Supplementary Figure S2**). Phage phylogeny is very complex. While bacteria share many common genes, microbial viruses are less related to each other creating large phylogenetic distances. The phylogenetic analysis shows that in some instances the phylogenetic clustering does not necessarily result in grouping of the phages according to their classes. This is, however, not surprising as phages often display diversity by "mosaicity of their genomes" (Dion et al., 2020). After identifying all prophage sequences in the genomes of the 177 *A. baumannii* isolates, we set out to analyze the most prevalent classes of phages present. Our analysis of all prophage sequences (n = 1,156) revealed that the majority

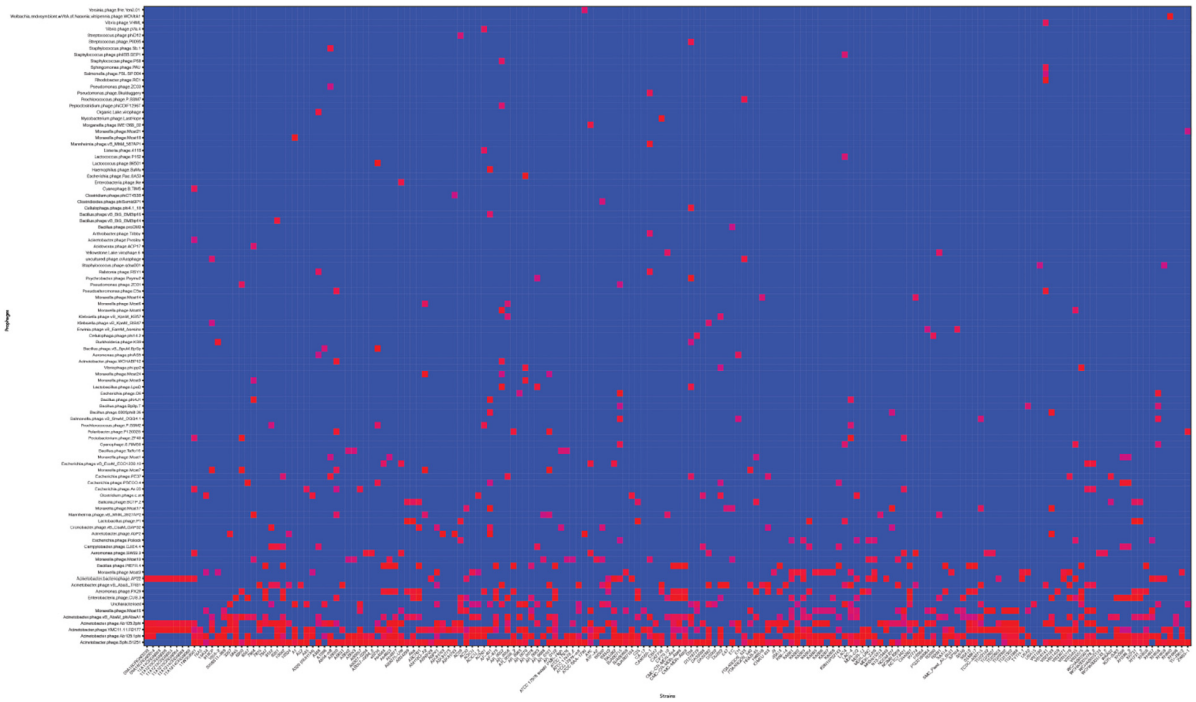


FIGURE 2 | The prevalence of prophages analyzed. Heat map of prophages found in all *A. baumannii* strains analyzed. Prophages (*y*-axis) are plotted against each *A. baumannii* strain (*x*-axis). Red squares indicate the presence of the indicated prophage. Blue squares indicate the lack thereof. Please refer to the PDF of the figure and use the zoom function to identify names of strains and phages.

of them, ∼57% (n = 660) of the prophages, belong to the *Siphoviridae* group (**Figure 3A**). The *Siphoviridae* is a class of head-and-tail phages, with the best known representative being phage lambda, that exhibit long, non-contractile but comparably flexible tail structures (Nobrega et al., 2018). The second most commonly found prophages are *Myoviridae*, with a percentage of ∼33% (n = 385) (**Figure 3A**). With the best-known Myovirus, the *E. coli* phage T4, these phages have a stiff, contractible tail that allows the active penetration of the bacterial host envelope (Hu et al., 2015). Together, these two phage classes make up 90% of all prophage genomes. The third most common class, albeit only 4.7% (n = 55) of all prophage genomes, belongs to *Podoviridae*. The best known Podovirus is probably T7, which has a short, stubby tail and internal core proteins that get ejected for the formation of a DNA-translocating channel across the bacterial cell envelope (Guo et al., 2014; Lupo et al., 2015; Leptihn et al., 2016). Prophages that could not be conclusively classified to a viral group accounted for 3.3% (n = 38). *Siphoviridae*, *Myoviridae* and *Podoviridae* all belong to the order *Caudovirales*, phages that exhibit a head-and-tail structure. Within the two phage classes we found several phages that were most successful, i.e., most common. Examples are the *A. baumannii* phages Bphi-B1251 and YMC11/11/R3177, which both belong to the *Siphoviridae* (**Table 2A**). The most common Myovirus was Ab105-1phi. Not only are these the most common prophages found, they are also the most common active prophages identified (**Table 2B**). In addition, the distribution of the classes was similar if only active prophages were analyzed.

TABLE 1 | *A. baumannii* stains with the highest and the fewest number of prophages identified.

Highest number of prophages identified		Fewest number of prophages identified	
Strain	**Prophage number**	**Strain**	**Prophage number**
AbPK1	11	DS002	1
AR_0056	11	VB1190	1
9201	12	CA-17	2
10042	12	E47	2
AR_0101	12	11A14CRGN003	3
DU202	12	11A1213CRGN008	3
VB35435	12	11A1213CRGN055	3
11W359501	13	11A1213CRGN064	3
AB030	13	11A1314CRGN088	3
AF-401	15	11A1314CRGN089	3

Here, 62% (432/697) belonged to the *Siphoviridae* and 32% (223/697) to the *Myoviridae* (**Figure 3B**).

The Genomic Position of Prophages Shows Two Main Locations for Genome Integration

To determine where all prophages -regardless of their class- are found in the bacterial genome, or if they are possibly

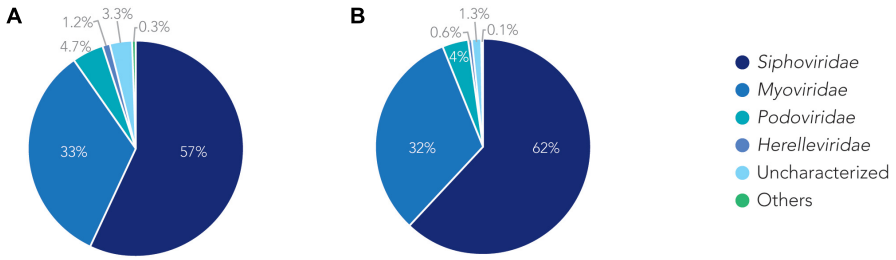


FIGURE 3 | The families of prophages found. Pie charts of prophages identified showing the percentage make up of each family. **(A)** Classification for all prophages. **(B)** Classification of active prophages only.

TABLE 2 | The most common prophages identified in *A. baumannii* stains.

A	10 the most common prophages (total)		B	10 the most common prophages (active)	
	Phages	**Number found**		**Phages**	**Number found**
	Acinetobacter phage Bphi-B1251	228		*Acinetobacter* phage Bphi-B1251	177
	Acinetobacter phage Ab105-1phi	143		*Acinetobacter* phage Ab105-1phi	111
	Acinetobacter phage YMC11/11/R3177	118		*Acinetobacter* phage YMC11/11/R3177	78
	Acinetobacter phage Ab105-2phi	95		*Acinetobacter* phage Ab105-2phi	65
	Acinetobacter phage vB_AbaM_phiAbaA1	56		*Aeromonas* phage PX29	31
	Moraxella phage Mcat16	42		*Enterobacteria* phage CUS-3	25
	Uncharacterized	38		*Acinetobacter* bacteriophage AP22	23
	Enterobacteria phage CUS-3	35		*Bacillus* phage PfEFR-4	20
	Aeromonas phage PX29	34		*Acinetobacter* phage vB_AbaS_TRS1	19
	Acinetobacter phage vB_AbaS_TRS1	33		*Moraxella* phage Mcat3	16

distributed at random within the host DNA, we visualized the position of the prophages in all 177 genomes (**Figure 4A**). We then plotted all prophage positions found in all genomes against the position in the bacterial host genome sequence. Surprisingly, we observed a bimodal distribution, with two clear peaks in the position of prophages (**Figure 4B**), indicating that there are two main sites of attachment for prophages and their genomic insertion. While this reflects the situation for all prophages, we then analyzed the position of several individual prophages within the bacterial genome. First, we assessed one of the most commonly found phages YMC11/11/R3177. The position of this phage reflects the overall distribution of all phages in the analyzed genomes, with two main areas of insertion. However, in some cases the position is outside the main area of insertion, possibly due to recombination of the bacterial genome (**Figure 5A**). The second phage we analyzed was phage vB_AbauS_TRS1. Here, the distribution of the phage within the bacterial genome seems to be more random as compared to the overall distribution (**Figure 5B**). In case of the *Aeromonas* phage PX29, insertion seems to be very "strict," i.e., only observed in one location within the genome (**Figure 5C**). The observations made when analyzing the prophage positions show that the insertion of phages could be described as "directed," and less random, indicating that attachment sites, if they are required, are found more commonly in certain positions of the bacterial genome.

The Sequence Length of Prophages Reveals Distinct Groups

Using the data provided by the program Prophage Hunter, we were interested in evaluating the size distribution of prophage genomes. Therefore, we plotted sizes against the frequency of prophages present in the bacterial genomes and calculated average prophage genome sizes. In the case of ambiguous prophages, a main population at 15 kb became visible followed by a minor peak of substantial size at approximately 60 kb, leaving the average and median length of ambiguous prophage genomes at 29.2 and 25.8 kb, respectively (**Figure 6A** and **Table 3**). In contrast to this, two main peaks were observed when analyzing only active prophages. Here, one peak is observed at around 17 kb while the other at around 36 kb was observed (**Figure 6A**). The average genome length of active prophages is 34 kb (**Table 3**). As these peaks include all phage categories, we re- analyzed the genomic length of the active prophages according to their classes: *Siphoviridae*, *Myoviridae* and *Podoviridae*, which together constitutes almost 95% of all prophages (see **Figure 3A**). When analyzing the length of all *Siphoviridae* sequences, we observed two main populations, one sharp peak at around 20 kb and one broad peak with a shoulder containing larger sequences from 18 to 56 kb (**Figure 6B**). The average prophage length of *Siphoviridae* is 36.7 kb (**Table 3**). *Myoviridae* sequences similarly exhibited two sharp peaks (17 and 36 kb), with a third minor one of around 60 kb (**Figure 6C**). The average prophage genome size of active *Myoviridae* is 32.4 kb (**Table 3**). The *Podoviridae* showed

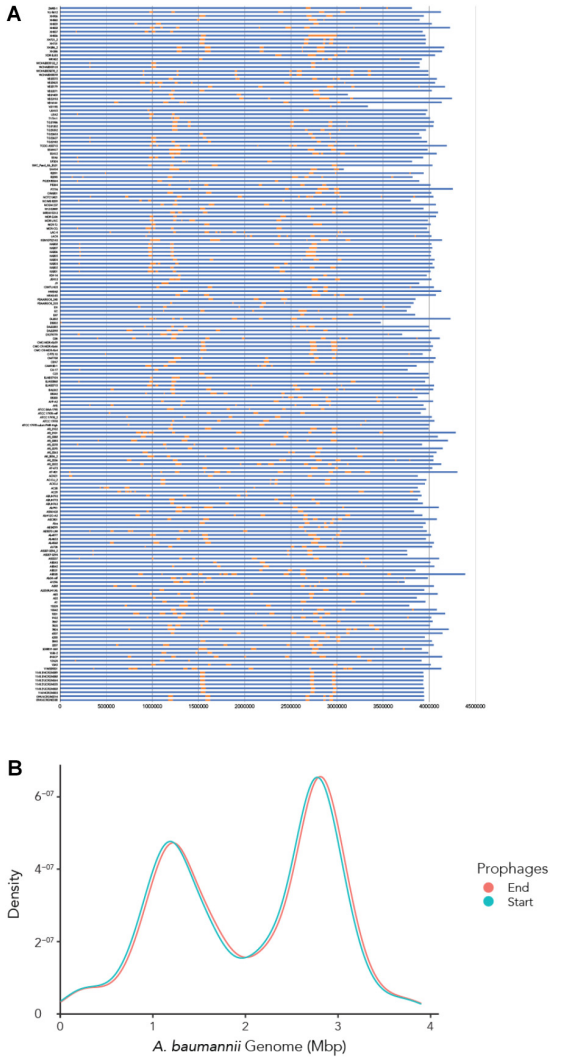


FIGURE 4 | The location of prophages found in each *A. baumannii* genome. **(A)** Stacked bar chart of each bacterial strain (*y*-axis). Yellow segments indicate prophage sequences identified on the genome (in blue). **(B)** Density graph compiled from the stacked bar chart. Please refer to the PDF of the figure and use the zoom function to identify names of strains and phages.

several minor peaks with a large sharp peak at about 12 kb and the average prophage genome length is calculated to be 17.4 kb (**Figure 6D** and **Table 3**). The results of these analyses show that there are distinct distributions of bacteriophage genome sizes. Two clearly separated groups of prophages can be observed just based on size, in the case of *Myoviridae*. In the case of *Siphoviridae*, we saw a less defined area with possibly multiple species within the broad distribution.

Prophage Encoded Antibiotic Resistance Genes

As prophages are able to encode genes that might allow its host to become more virulent and therefore more evolutionary successful, we aimed to analyze prophage-encoded virulence factors. However, in contrast to e.g., *E. coli*, a databank for *A. baumannii* virulence factors currently does not exist. We therefore searched for prophage sequences that contain genes that contribute to antibiotic resistance. **Table 4** lists the start and end of the genes that are encoded within a respective prophage. Among others, we found AMR genes for OXA-23 and NDM-1. OXA-23 is the most widespread carbapenem resistance gene globally (Hamidian and Nigro, 2019). NDM-1 encodes a carbapenemase, a beta-lactamase enzyme with a broad substrate specificity capable of hydrolyzing penicillins, carbapenems, cephalosporins, and monobactams. Other beta-lactamase genes were bla_{ADC-5}, bla_{OXA-67}, $bla_{OXA-115}$, and bla_{TEM-12}. In addition, we were able to identify genes coding for N-Acetyltransferases(aac(3)-I, *aac(3)-Id*, *aacA16)*, Aminoglycoside phosphotransferases (*aph(3′)-Ia*, *aph(3′)-VI*, *aph(6)-Id*, *aph(3″)-Ib*), both groups mediating aminoglycoside resistance. Other genes that contribute to antibiotic resistance were sulfonamide resistance gene (*sul2*), and the macrolide-resistance conferring genes *msr(E)*, encoding an efflux pump, and *mph(E)*, coding for a macrolide-inactivating phosphotransferase.

Interestingly, one *A. baumannii* strain, ACN21, contains a prophage which encodes three antibiotic resistance genes [*aph(3')-VI*, bla_{NDM-1}, ble_{MBL}]. The phage is most closely related to *Vibrio* phage pVa-4, a Myovirus that infects *V. alginolyticus* (Kim et al., 2019). In contrast to its relative, phage pVa-4 is a lytic phage was grouped to be part of the phiKZ-like phages (Phikzviruses), which are considered as "jumbo" phages. A second *A. baumannii* strain, AR_0078, contains a prophage sequence that shares a high degree of similarity to the *Bacillus* phage PfEFR-4, a Siphovirus with a prolate head, in contrast to its *E. coli* relative lambda (Geng et al., 2017). The prophage encodes three antimicrobial resistance genes that confer macrolide-resistance, *msr(E)*, encoding an efflux pump, and *mph(E)*, coding for a macrolide-inactivating phosphotransferase. The third gene encodes the Aminoglycoside phosphotransferases [*aph(3')-Ia*], inactivating aminoglycoside antibiotics.

A. baumannii strain DU202 contains a prophage sequence that is related to the lytic *E. coli* myovirus PBCO 4 (Kim et al., 2013). Within the genome sequence, two antimicrobial resistance genes are encoded: *aac(3)-Id* codes for an N-Acetyltransferase mediating aminoglycoside resistance and OXA-23, which is encodes the most widespread resistance mechanism toward the β-lactamase inhibitor sulbactam. The gene was also embedded in the prophage sequences of two phages in the *A. baumannii* strain XH859; here, the *A. baumannii* phage Bphi-B1251 was found to be the most closely related phage, a lytic Podovirus, that was previously shown to be able to infect and lyse an OXA-23- harboring *A. baumannii* isolate from a septic patient (Jeon et al., 2012).

To determine how prevalent the prophage encoded antimicrobial resistance genes are in other genomes, we mapped all available 4,128 *A. baumannii* Illumina sequence reads that were accessible by 2019/11/17 on the Sequence Read Archive (SRA) to the AMR prophage sequences using an 80% cutoff for the coverage. We identified 174 *A. baumannii* genomes that contain prophage sequences, or about 4.2% of all available

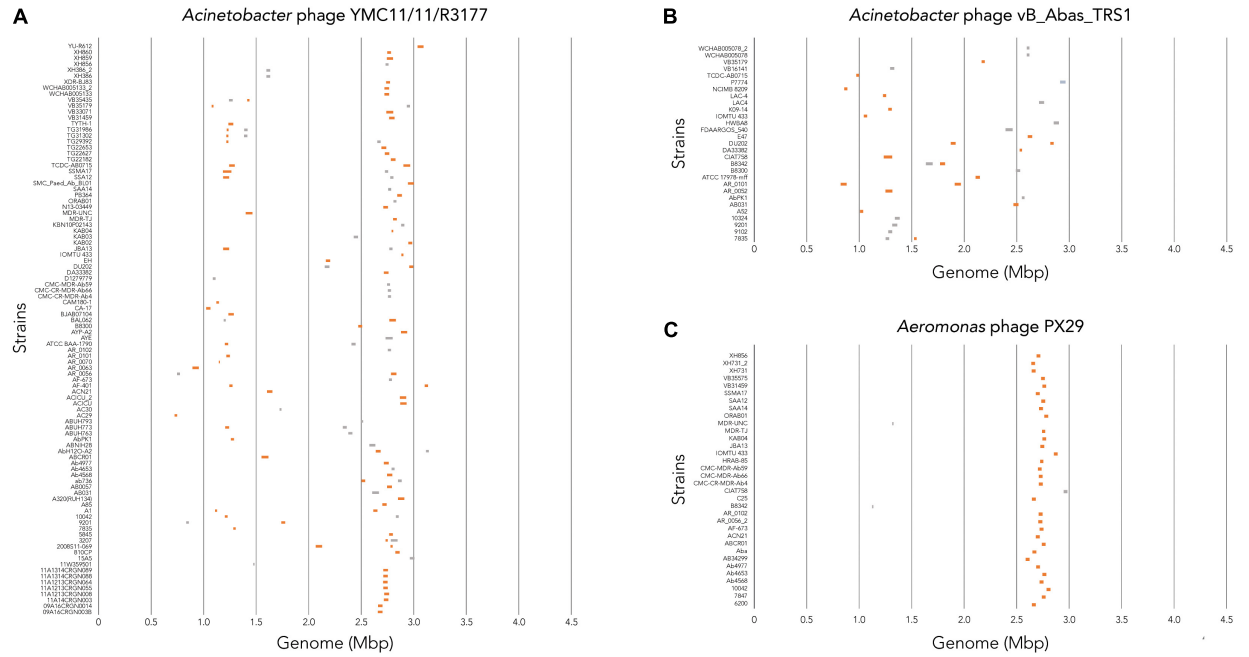


FIGURE 5 | Location of prophage genome insertion differs between phages. **(A)** Comparison of phage location for *Acinetobacter* phage YMC11/11/R3177. **(B)** Comparison of insertion locations for *Acinetobacter* phage vB_Abas_TRS1. **(C)** Comparison of prophage insertion sites for *Aeromonas* phage PX29. Boxes in orange indicate active prophages identified. Gray boxes indicate ambiguous prophage sequences. Please refer to the PDF of the figure and use the zoom function to identify names of strains and phages.

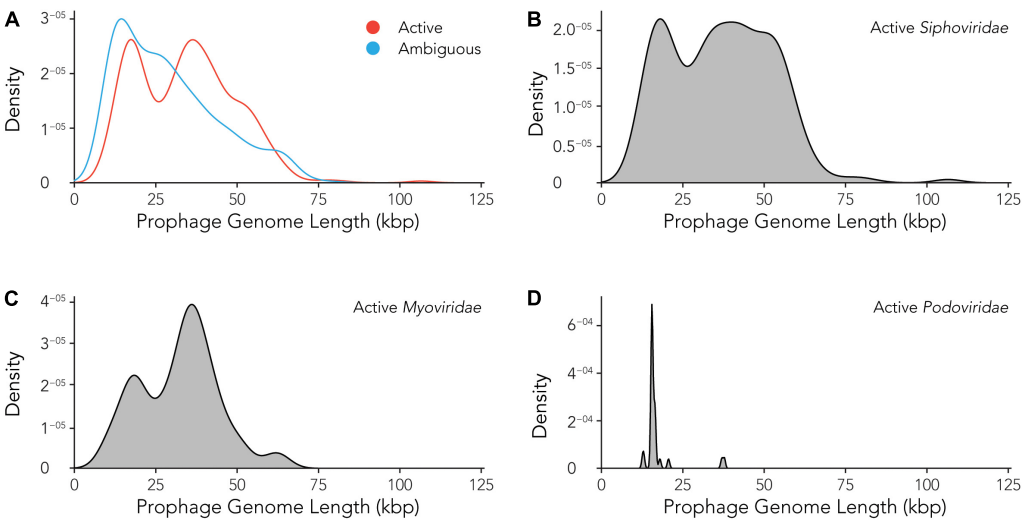


FIGURE 6 | Active prophages categorized by prophage lengths. **(A)** Comparison of prophage lengths between active and ambiguous prophages. **(B–D)** Distribution of prophage length for active *Siphoviridae* **(B)**, *Myoviridae* **(C)**, and *Podoviridae* **(D)**.

reads, not including the genomes we used for the initial analysis. **Supplementary Figure S3** illustrates the prevalence of the prophage-encoded AMR genes (ARGs) and their respective prophages. Fairly "successful," i.e., widely distributed, were two prophages: ABUH793 and AR_0056. ABUH793 is a close relative of the *Clostridium* phage phiCT453B, containing the resistance gene $bla_{OXA-115}$. The second prophage sequence that was found often in *A. baumannii* genomes in comparison to other prophages encoding ARGs, was AR_0056, a relative of the *Moraxella* phage Mcat6, encoding the ARG *sul2*. While the prophage does not necessarily render the host antibiotic resistant as genetic regulators might be missing, prophages containing ARGs can present an evolutionary advantage for the host (Wendling et al., 2020).

TABLE 3 | Prophage genome lengths for active and ambiguous prophages, among the three major families in the order of *Caudovirales*.

		All prophages	Active prophages	Ambiguous prophages
All prophages	Average	32211.62	34182.66	29218.56
	Median	31,296	35,074	25,759
Siphoviridae	Average	34586.75	36668.1	30625.76
	Median	33,018	36562.5	27,679
Myoviridae	Average	32148.62	32351.98	31870.4
	Median	34,249	35,075	30,792
Podoviridae	Average	19422.25	17357.86	21563.11
	Median	15,498	15,498	14,963

Our finding demonstrates the importance of phages in the acquisition of antimicrobial resistance; the above described genes may confer the ability to grow in the presence of antibiotics when the bacterial host is infected by a phage that encodes not only the information for its own replication but also genes that inactivate or remove antibiotic compounds.

DISCUSSION

Our search for prophages in the genomes of *A. baumannii* strains revealed several interesting findings. One surprising observation was the positions of the prophages within the genome of the bacterial host. When analyzing the prophage positions one might expect that the insertion of phages would be less directed, and more random. However, we found that the majority of phages inserted into two locations as seen by a bimodal density plot with a sharp separation between the two peaks. Prophage genome integration can either be a site-specific recombination event at so-called *att* sites or occurs in a non-directed manner by transposition into random sites (Ramisetty and Sudhakari, 2019). Our data could indicate that the two areas in the genome contain most of the attachment sites for the majority of phages. A previously published analysis of *Salmonella* and *E. coli* genomes found a large number of distinctive phage integration loci; in the case of *Salmonella*, 24 loci were shared among 102 *Salmonella* phages, amounting to four phages statistically sharing one integration site. In case of *E. coli*, 58 distinctive integration loci were identified for 369 phages, with statistically 6.6 phages per site (Bobay et al., 2013). It might be reasonable to assume that *A. baumannii* contains similar numbers of attachment sites, although we have not analyzed potential sites in the genomes we investigated. However, regardless of whether a phage inserts via one of the various attachment sites or randomly via transposition, only two "hot spots" were observed in our study. In addition to the explanation that attachment sites might be more frequent in these two sections of the bacterial genome, prophage insertion

TABLE 4 | Antimicrobial resistance genes found in prophages embedded in *A. baumannii* stains.

Strain	AMR gene	Start of resistance gene	End of resistance gene	Phage name	Start of prophage sequence	End of prophage sequence
AB030	blaADC-5	3131953	3133104	*Acinetobacter* phage Bphi-B1251	3061121	3133273
AB5075-UW	blaOXA-23	562998	563819	*Escherichia* phage vB_EcoM_ECO1230-10	545886	581029
AbPK1	aac(3)-I	1359578	1360042	*Acinetobacter* phage Ab105-1phi	1357668	1404464
ABUH793	blaOXA-115	2017307	2018131	*Clostridium* phage phiCT453B	2016119	2031153
AC29	blaTEM-12	728807	729667	*Acinetobacter* phage YMC11/11/R3177	723894	746651
AC29	aph(3')-Ia	732720	733535	*Acinetobacter* phage YMC11/11/R3177	723894	746651
ACN21	aph(3')-VI	110687	111466	*Vibrio* phage pVa-4	98299	123387
ACN21	blaNDM-1	112744	113556	*Vibrio* phage pVa-4	98299	123387
ACN21	ble-MBL	113560	113925	*Vibrio* phage pVa-4	98299	123387
ACN21	ble-MBL	113560	113925	*Listeria* phage A118	113004	131863
AR_0056	sul2	3643229	3644044	*Moraxella* phage Mcat6	3631231	3645626
AR_0078	aph(3')-Ia	1454389	1455204	*Bacillus* phage PfEFR-4	1448981	1463460
AR_0078	msr(E)	1456488	1457963	*Bacillus* phage PfEFR-4	1448981	1463460
AR_0078	mph(E)	1458019	1458903	*Bacillus* phage PfEFR-4	1448981	1463460
BJAB0715	blaOXA-23	1040633	1041454	*Pseudomonas* phage ZC01	1035624	1064722
DU202	blaOXA-23	1304225	1305046	*Escherichia* phage PBECO 4	1300481	1331113
DU202	aac(3)-Id	1307973	1308424	*Escherichia* phage PBECO 4	1300481	1331113
EC	blaOXA-67	2088567	2089391	*Bacillus* phage proCM3	2082124	2102020
LAC4	aph(6)-Id	3852495	3853331	*Lactococcus* phage P162	3843825	3855365
LAC4	aph(3″)-Ib	3853331	3854133	*Lactococcus* phage P162	3843825	3855365
MDR-UNC	aacA16	1317505	1318056	*Aeromonas* phage PX29	1315787	1325918
MDR-UNC	aac(3)-I	1321744	1322208	*Aeromonas* phage PX29	1315787	1325918
TCDC-AB0715	sul2	2557327	2558142	*Acinetobacter* phage vB_AbaM_phiAbaA1	2530565	2567304
XH858	blaOXA-23	1093489	1094310	N/A	1088870	1114410
XH859	blaOXA-23	1070735	1071556	*Acinetobacter* phage Bphi-B1251	1038026	1092046

into segments crucial for e.g., over-all gene regulation, or into household genes, would be an evolutionary disadvantage and might therefore be less commonly found.

Bacteriophages are classified into 12 families (ICTV, 2019). Pioneering work in taxonomy divided tailed phages into three classes based on the morphology of the phages; *Myoviridae* have long contractile tails, *Siphoviridae* have long non contractile tails, and *Podoviridae* have short tails. Recently, the International Committee on Taxonomy of Viruses (ICTV) expanded the order *Caudovirales*, describing tailed bacteriophages to include six additional families, i.e.,: *Ackermannviridae*, *Autographiviridae*, *Chaseviridae*, *Demerecviridae*, *Drexlerviridae*, and *Herelleviridae*, taking additional characteristics into consideration such as genome sequence, gene content, protein homology and the host (Adriaenssens et al., 2020). When analyzing the families of prophages in this population of *A. baumannii* strains, we observed a prevalence of *Siphoviridae* which constituted 57% of all identified prophages. Together with the next family of phages, *Myoviridae*, which consists 1/3 of all prophages, the two groups make up 90% of all prophages identified. Among the remaining 10%, the largest group belongs to *Podoviridae*. These ratios are very similar to the ones that have been reported in other studies, and also the ratio of the most commonly found phages in nature (Costa et al., 2018).

Interestingly, the ratio between ambiguous to active prophages in case of the ones that have been identified as *Siphoviridae*, 0.654, markedly differs from the ratio calculated for the prophages that belong to *Myoviridae*, 0.579. It is unlikely that *Siphoviridae* prophage sequences are less prone to mutations, as they should be occurring at random. However, a mechanism that would specifically "protect" *Siphoviridae* prophages might be the case if daughter cells, where mutations in the prophages occur, would have an evolutionary disadvantage. This would imply that prophages influence the host behavior positively, which has previously been shown in some cases (Bondy-Denomy and Davidson, 2014; Nanda et al., 2015; Loh et al., 2019). Could the most likely scenario be that the genomes of *Myoviridae* are possibly larger than those of the *Siphoviridae*, making them more prone to random mutations and deletions? However, the size comparisons of the genomic sequences of the prophages that we identified, does not support this possible explanation: The *Siphoviridae* sequences display distribution with one clear peak at around 20 kb followed by a fairly broad peak with a plateau and a shoulder toward larger genome sizes, ranging from 28 to 65 kb (**Figure 6B**). In contrast to this, the genomic size distributions of *Myoviridae* showed two peaks, one around 20 kb, the second around 42 kb (**Figure 6C**). The size estimations are corroborated by the findings of a previous study which estimated the genome sizes of *Siphoviridae* to be approximately 50 kb in average, with a broad distribution between 24 and 101 kb, while *Myoviridae* display smaller genomes or around 34 kb (Costa et al., 2018).

One question we could pose is why there is no broad distribution of prophage sizes, and why do we observe "peaks"? Can we conclude from this data that certain genome sizes are advantageous from an evolutionary standpoint? The arch-Myovirus might have had a certain size that proved to be sufficient for the successful persistence during the course of evolution. Only smaller increases or decreases of the genome allowed evolutionary success, and no gradual increase or decrease in genome size occurred. However, an evolutionary leap or jump might have happened at some point, which might have led to a major increase of genomic size, creating a new, second type of a Myovirus class which is represented in the second, larger peak. Starting from this size, again only smaller changes, decreases or increases with regards to the genomic size, may have occurred, preserving the sharp separation of each peak. It would be interesting to investigate if the smaller Myovirus display a prolate head as does T4. The increase volume of this geometry allows the packing of a larger genome, which might explain the possible separation in two sizes. To test this hypothesis, smaller Myoviruses should have non-prolate heads. Viral classification is a complex topic. Possibly the genome sizes might help to contribute to classifying of microbial viruses in the future.

While Prophage Hunter extracts prophage genomes from bacterial genomes, the platform is a web-based tool that also distinguishes between "active" and "ambiguous" prophage genomes (Song et al., 2019). The developers of Prophage Hunter have used experimental data and conducted induction experiments with mitomycin C, to validate the program's output, showing its ability to hunt for "active," inducible prophages. Yet, conclusions should not be hastily drawn to assume that all "active" prophages can definitively excise from the host genome to commence the bacteriophage lytic life cycle; false positives may still exist. In this regard, induction experiments should be conducted to confirm that "active" prophages can indeed produce active particles.

Prophages are an important source for acquiring new genetic information, including antibiotic resistance genes, for their bacterial host. Phage-mediated transfer of genes from donor to recipient cells, also called transduction, has been shown to be instrumental in the spread of AMR genes both *in vitro* and *in vivo* (Haaber et al., 2016). In our study, we also investigated AMR genes that are embedded in prophage sequences. Previous studies on prophage diversity in *A. baumannii* had found AMR genes (also called: ARGs) in many prophages that were analyzed (Costa et al., 2018; López-Leal et al., 2020). Yet despite this, it remains to be shown whether prophages confer antimicrobial resistance to its host *A. baumannii*. Our observation illustrates that phages might represent important contributors in the process of AMR acquisition. However, it remains to be said that we found less than 5% of a publicly available, deposited sequence reads to contain the prophage-encoded ARGs we initially identified, arguing that phage transduction is possibly not the prevalent mode of AMR acquisition but is second to other mechanisms such as plasmid uptake via conjugation. Interestingly, despite viruses in general showing highly condensed genomes trying to pack essential information in small volumes, bacterial viruses seem to have co-evolved with their hosts and carry genes that are not directly required

for the virus but are beneficial to the host and thus also to the prophage.

CONCLUSION

Our study attempts to take an inventory of prophages in the important nosocomial pathogen *A. baumannii*. We have analyzed the phylogeny of the prophages, their position in the host genome and characterized their lengths, identifying "successful," i.e., widely distributed phages, and the dominant families, *Myoviridae* and *Siphoviridae*.

Several prophage sequences contained genes coding for antimicrobial resistance genes. By mapping these genes in all deposited illumina *A. baummannii* sequence reads, we found that less than 5% of all available host sequences contain such prophage-embedded genes, indicating that transduction may not be the major contributor to the emergence of antimicrobial resistance.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

SL and XH devised this study. JC, BL, XH, and SL performed the work. BL created the figures. BL, PM, XH, YY, and SL wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

We thank the National Science Foundation of China for supporting this work (NSFC 32011530116). BL was supported by the China Postdoctoral Science Foundation (530000-X91902).

ACKNOWLEDGMENTS

We would like to thank the reviewers who made very valuable suggestions that helped to improve the work. A preprint of this article has been published on BioRxiv (Loh et al., 2020).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.579802/full#supplementary-material

Supplementary Figure 1 | *In silico* detection and distribution of prophage sequences in *A. baumannii* strains. Unrooted core phylogenetic tree of 177 *A. baumannii* genomes from clinical isolates. Presence of prophage sequences are indicated by red squares. Green squares indicate the lack thereof. Please refer to the PDF of the figure and use the zoom function to identify names of strains and phages.

Supplementary Figure 2 | Phylogeny of prophages found in *A. baumannii* genomes.

Supplementary Figure 3 | Prevalence of prophages carrying AMR genes in *A. baumannii* strains.

Supplementary Table 1 | Metadata of all *A. baumannii* genomes used in this study.

Supplementary Table 2 | Details of detected active and ambiguous prophage in all 177 *A. baumannii* strains analyzed in this study.

REFERENCES

Adriaenssens, E. M., Sullivan, M. B., Knezevic, P., van Zyl, L. J., Sarkar, B. L., Dutilh, B. E., et al. (2020). Taxonomy of prokaryotic viruses: 2018-2019 update from the ICTV bacterial and archaeal viruses subcommittee. *Arch. Virol.* 165, 1253–1260. doi: 10.1007/s00705-020-04577-8

Antunes, L. C., Visca, P., and Towner, K. J. (2014). *Acinetobacter baumannii*: evolution of a global pathogen. *Pathog. Dis.* 71, 292–301. doi: 10.1111/2049-632X.12125

Argov, T., Rabinovich, L., Sigal, N., and Herskovits, A. A. (2017). An effective counterselection system for *Listeria monocytogenes* and its use to characterize the monocin genomic region of strain 10403S. *Appl. Environ. Microbiol.* 83:e02927-16. doi: 10.1128/AEM.02927-16

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387

Bobay, L. M., Rocha, E. P., and Touchon, M. (2013). The adaptation of temperate bacteriophages to their host genomes. *Mol. Biol. Evol.* 30, 737–751. doi: 10.1093/molbev/mss279

Bondy-Denomy, J., and Davidson, A. R. (2014). When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness. *J. Microbiol.* 52, 235–242. doi: 10.1007/s12275-014-4083-3

Brüssow, H. (2007). Bacteria between protists and phages: from antipredation strategies to the evolution of pathogenicity. *Mol. Microbiol.* 65, 583–589. doi: 10.1111/j.1365-2958.2007.05826.x

Brüssow, H., Canchaya, C., and Hardt, W. D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol. Mol. Biol. Rev.* 68, 560–602. doi: 10.1128/MMBR.68.3.560-602.2004

Cortez, D., Forterre, P., and Gribaldo, S. (2009). A hidden reservoir of integrative elements is the major source of recently acquired foreign genes and ORFans in archaeal and bacterial genomes. *Genome Biol.* 10:R65. doi: 10.1186/gb-2009-10-6-r65

Costa, A. R., Monteiro, R., and Azeredo, J. (2018). Genomic analysis of *Acinetobacter baumannii* prophages reveals remarkable diversity and suggests profound impact on bacterial virulence and fitness. *Sci. Rep.* 8:15346. doi: 10.1038/s41598-018-33800-5

Davis, B. M., and Waldor, M. K. (2003). Filamentous phages linked to virulence of *Vibrio cholerae*. *Curr. Opin. Microbiol.* 6, 35–42. doi: 10.1016/s1369-5274(02)00005-x

Derbise, A., Chenal-Francisque, V., Pouillot, F., Fayolle, C., Prévost, M. C., Médigue, C., et al. (2007). A horizontally acquired filamentous phage contributes to the pathogenicity of the plague bacillus. *Mol. Microbiol.* 63, 1145–1157. doi: 10.1111/j.1365-2958.2006.05570.x

Dijkshoorn, L., Nemec, A., and Seifert, H. (2007). An increasing threat in hospitals: multidrug-resistant *Acinetobacter baumannii*. *Nat. Rev. Microbiol.* 5, 939–951. doi: 10.1038/nrmicro1789

Dion, M. B., Oechslin, F., and Moineau, S. (2020). Phage diversity, genomics and phylogeny. *Nat. Rev. Microbiol.* 18, 125–138. doi: 10.1038/s41579-019-0311-5

Fortier, L. C. (2017). The Contribution of Bacteriophages to the biology and virulence of pathogenic clostridia. *Adv. Appl. Microbiol.* 101, 169–200. doi: 10.1016/bs.aambs.2017.05.002

Fortier, L. C., and Sekulovic, O. (2013). Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* 4, 354–365. doi: 10.4161/viru.24498

Gamage, S. D., Patton, A. K., Hanson, J. F., and Weiss, A. A. (2004). Diversity and host range of Shiga toxin-encoding phage. *Infect. Immun.* 72, 7131–7139. doi: 10.1128/IAI.72.12.7131-7139.2004

Geng, P., Tian, S., Yuan, Z., and Hu, X. (2017). Identification and genomic comparison of temperate bacteriophages derived from emetic *Bacillus cereus*. *PLoS One* 12:e0184572. doi: 10.1371/journal.pone.0184572

Gertz, J., Riles, L., Turnbaugh, P., Ho, S. W., and Cohen, B. A. (2005). Discovery, validation, and genetic dissection of transcription factor binding sites by comparative and functional genomics. *Genome Res.* 15, 1145–1152. doi: 10.1101/gr.3859605

Guo, F., Liu, Z., Fang, P. A., Zhang, Q., Wright, E. T., Wu, W., et al. (2014). Capsid expansion mechanism of bacteriophage T7 revealed by multistate atomic models derived from cryo-EM reconstructions. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4606–E4614. doi: 10.1073/pnas.1407020111

Haaber, J., Leisner, J. J., Cohn, M. T., Catalan-Moreno, A., Nielsen, J. B., Westh, H., et al. (2016). Bacterial viruses enable their host to acquire antibiotic resistance genes from neighbouring cells. *Nat. Commun.* 7:13333. doi: 10.1038/ncomms13333

Hamidian, M., and Nigro, S. J. (2019). Emergence, molecular mechanisms and global spread of carbapenem-resistant *Acinetobacter baumannii*. *Microb. Genom.* 5:e000306. doi: 10.1099/mgen.0.000306

Howard-Varona, C., Hargreaves, K. R., Abedon, S. T., and Sullivan, M. B. (2017). Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J.* 11, 1511–1520. doi: 10.1038/ismej.2017.16

Hu, B., Margolin, W., Molineux, I. J., and Liu, J. (2015). Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *Proc. Natl. Acad. Sci. U.S.A.* 112, E4919–E4928. doi: 10.1073/pnas.1501064112

Hu, Y., He, L., Tao, X., Meng, F., and Zhang, J. (2019). High DNA uptake capacity of international clone II *Acinetobacter baumannii* detected by a novel planktonic natural transformation assay. *Front. Microbiol.* 10:2165. doi: 10.3389/fmicb.2019.02165

ICTV (2019). *International Committee on Taxonomy of Viruses: New Taxonomy Release*. Available online at: https://talk.ictvonline.org/taxonomy/ (accessed June 03, 2020).

Jeon, J., Kim, J. W., Yong, D., Lee, K., and Chong, Y. (2012). Complete genome sequence of the podoviral bacteriophage YMC/09/02/B1251 ABA BP, which causes the lysis of an OXA-23-producing carbapenem-resistant *Acinetobacter baumannii* isolate from a septic patient. *J. Virol.* 86, 12437–12438. doi: 10.1128/JVI.02132-12

Joly-Guillou, M. L. (2005). Clinical impact and pathogenicity of acinetobacter. *Clin. Microbiol. Infect.* 11, 868–873. doi: 10.1111/j.1469-0691.2005.01227.x

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kempf, M., Rolain, J. M., Azza, S., Diene, S., Joly-Guillou, M. L., Dubourg, G., et al. (2013). Investigation of *Acinetobacter baumannii* resistance to carbapenems in Marseille hospitals, south of France: a transition from an epidemic to an endemic situation. *APMIS* 121, 64–71. doi: 10.1111/j.1600-0463.2012.02935.x

Kim, M. S., Hong, S. S., Park, K., and Myung, H. (2013). Genomic analysis of bacteriophage PBECO4 infecting *Escherichia coli* O157:H7. *Arch. Virol.* 158, 2399–2403. doi: 10.1007/s00705-013-1718-3

Kim, S. G., Jun, J. W., Giri, S. S., Yun, S., Kim, H. J., Kim, S. W., et al. (2019). Isolation and characterisation of pVa-21, a giant bacteriophage with anti-biofilm potential against Vibrio alginolyticus. *Sci. Rep.* 9:6284. doi: 10.1038/s41598-019-42681-1

Kuhn, A., and Leptihn, S. (2019). "Helical and filamentous phages," in *Reference Module in Life Sciences*, ed. M. G. Feiss (Amsterdam: Elsevier), doi: 10.1016/B978-0-12-809633-8.20986-2

Leptihn, S., Gottschalk, J., and Kuhn, A. (2016). T7 ejectosome assembly: a story unfolds. *Bacteriophage* 6:e1128513. doi: 10.1080/21597081.2015.1128513

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* [Preprint]. Available online at: https://arxiv.org/abs/1303.3997 (accessed November 20, 2020).

Loh, B., Haase, M., Mueller, L., Kuhn, A., and Leptihn, S. (2017). The transmembrane morphogenesis protein gp1 of filamentous phages contains Walker A and Walker B motifs essential for phage assembly. *Viruses* 9:73. doi: 10.3390/v9040073

Loh, B., Kuhn, A., and Leptihn, S. (2019). The fascinating biology behind phage display: filamentous phage assembly. *Mol. Microbiol.* 111, 1132–1138. doi: 10.1111/mmi.14187

Loh, B., Chen, J., Manohar, P., Yu, Y., Hua, X., Leptihn, S. A. et al. (2020). Biological inventory of prophages in *A. baumannii* genomes reveal distinct distributions in classes, length and genomic positions. *bioRxiv*. doi: 10.1101/2020.10.26.355222

López-Leal, G., Santamaria, R. I., Cevallos, M. Á, Gonzalez, V., and Castillo-Ramírez, S. (2020). Prophages encode antibiotic resistance genes in *Acinetobacter baumannii*. *Microb. Drug Resist.* 26, 1275–1277. doi: 10.1089/mdr.2019.0362

Lupo, D., Leptihn, S., Nagler, G., Haase, M., Molineux, J. I., and Kuhn, A. (2015). The T7 ejection nanomachine components gp15-gp16 form a spiral ring complex that binds DNA and a lipid membrane. *Virology* 486, 263–271. doi: 10.1016/j.virol.2015.09.022

Morris, F. C., Dexter, C., Kostoulias, X., Uddin, M. I., and Peleg, A. Y. (2019). The mechanisms of disease caused by *Acinetobacter baumannii*. *Front. Microbiol.* 10:1601. doi: 10.3389/fmicb.2019.01601

Nanda, A. M., Thormann, K., and Frunzke, J. (2015). Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. *J. Bacteriol.* 197, 410–419. doi: 10.1128/JB.02230-14

Nobrega, F. L., Vlot, M., de Jonge, P. A., Dreesens, L. L., Beaumont, H. J. E., Lavigne, R., et al. (2018). Targeting mechanisms of tailed bacteriophages. *Nat. Rev. Microbiol.* 16, 760–773. doi: 10.1038/s41579-018-0070-8

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

Partridge, S. R., Kwong, S. M., Firth, N., and Jensen, S. O. (2018). Mobile genetic elements associated with antimicrobial resistance. *Clin. Microbiol. Rev.* 31:e0088-17. doi: 10.1128/CMR.00088-17

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

Ramisetty, B. C. M., and Sudhakari, P. A. (2019). Bacterial 'Grounded' prophages: hotspots for genetic renovation and innovation. *Front. Genet.* 10:65. doi: 10.3389/fgene.2019.00065

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

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

Song, W., Sun, H. X., Zhang, C., Cheng, L., Peng, Y., Deng, Z., et al. (2019). Prophage Hunter: an integrative hunting tool for active prophages. *Nucleic Acids Res.* 47, W74–W80. doi: 10.1093/nar/gkz380

Tinsley, E., and Khan, S. A. (2006). A novel FtsZ-like protein is involved in replication of the anthrax toxin-encoding pXO1 plasmid in Bacillus anthracis. *J. Bacteriol.* 188, 2829–2835. doi: 10.1128/JB.188.8.2829-2835.2006

Tozzoli, R., Grande, L., Michelacci, V., Ranieri, P., Maugliani, A., Caprioli, A., et al. (2014). Shiga toxin-converting phages and the emergence of new pathogenic *Escherichia coli*: a world in motion. *Front. Cell Infect. Microbiol.* 4:80. doi: 10.3389/fcimb.2014.00080

Wachino, J. I., Jin, W., Kimura, K., and Arakawa, Y. (2019). Intercellular transfer of chromosomal antimicrobial resistance genes between *Acinetobacter baumannii* strains mediated by prophages. *Antimicrob. Agents Chemother.* 63:e00334-19. doi: 10.1128/AAC.00334-19

Wagner, P. L., and Waldor, M. K. (2002). Bacteriophage control of bacterial virulence. *Infect. Immun.* 70, 3985–3993. doi: 10.1128/iai.70.8.3985-3993.2002

Wang, X., and Wood, T. K. (2016). Cryptic prophages as targets for drug development. *Drug Resist. Updat.* 27, 30–38. doi: 10.1016/j.drup.2016.06.001

Wendling, C. C., Refardt, D., and Hall, A. R. (2020). Fitness benefits to bacteria of carrying prophages and prophage-encoded antibiotic-resistance genes peak in different environments. *bioRxiv* [Preprint]. doi: 10.1101/2020.03.13.990044

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

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

frontiers in Microbiology

ORIGINAL RESEARCH
published: 14 December 2020
doi: 10.3389/fmicb.2020.596175

Metagenomic Next-Generation Sequencing of Cerebrospinal Fluid for the Diagnosis of External Ventricular and Lumbar Drainage-Associated Ventriculitis and Meningitis

Lingye Qian[1†], Yijun Shi[1†], Fangqiang Li[1†], Yufei Wang[1], Miao Ma[1], Yanfang Zhang[1], Yang W. Shao[2,3], Guanghui Zheng[1] and Guojun Zhang[1*]*

[1] Department of Clinical Diagnosis, Laboratory of Beijing Tiantan Hospital, Capital Medical University, Beijing, China, [2] Nanjing Geneseeq Technology Inc., Nanjing, China, [3] School of Public Health, Nanjing Medical University, Nanjing, China

Metagenomic next-generation sequencing (mNGS) has become a widely used technology that can accurately detect individual pathogens. This prospective study was performed between February 2019 and September 2019 in one of the largest clinical neurosurgery centers in China. The study aimed to evaluate the performance of mNGS on cerebrospinal fluid (CSF) from neurosurgical patients for the diagnosis of external ventricular and lumbar drainage (EVD/LD)-associated ventriculitis and meningitis (VM). We collected CSF specimens from neurosurgical patients with EVD/LD for more than 24 h to perform conventional microbiological studies and mNGS analyses in a pairwise manner. We also investigated the usefulness of mNGS of CSF for the diagnosis of EVD/LD-associated VM. In total, 102 patients were enrolled in this study and divided into three groups, including confirmed VM (cVM) (39), suspected VM (sVM) (49), and non-VM (nVM) (14) groups. Of all the patients, mNGS detected 21 Gram-positive bacteria, 20 Gram-negative bacteria, and five fungi. The three primary bacteria detected were *Staphylococcus epidermidis* (9), *Acinetobacter baumannii* (5), and *Staphylococcus aureus* (3). The mNGS-positive coincidence rate of confirmed EVD/LD-associated VM was 61.54% (24/39), and the negative coincidence rate of the nVM group was 100% (14/14). Of 15 VM pathogens not identified by mNGS in the cVM group, eight were negative with mNGS and seven were inconsistent with the conventional microbiological identification results. In addition, mNGS identified pathogens in 22 cases that were negative using conventional methods; of them, 10 patients received a favorable clinical treatment; thus, showing the benefit of mNGS-guided therapy.

Keywords: metagenomic next-generation sequencing (mNGS), diagnosis, EVD, LD, ventriculitis and meningitis

OPEN ACCESS

Edited by:
Kristin Hegstad, University Hospital of North Norway, Norway

Reviewed by:
Fabrizio Maggi, University of Pisa, Italy
Roland Lang, University Hospital Erlangen, Germany

***Correspondence:**
Guanghui Zheng
zgh999@yeah.net
Guojun Zhang
zgjlunwen@163.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Microbiology

Received: *18 August 2020*
Accepted: *13 November 2020*
Published: *14 December 2020*

Citation:
Qian L, Shi Y, Li F, Wang Y, Ma M, Zhang Y, Shao YW, Zheng G and Zhang G (2020) Metagenomic Next-Generation Sequencing of Cerebrospinal Fluid for the Diagnosis of External Ventricular and Lumbar Drainage-Associated Ventriculitis and Meningitis. Front. Microbiol. 11:596175. doi: 10.3389/fmicb.2020.596175

INTRODUCTION

External ventricular drains (EVDs) and, in some specific cases, lumbar drains (LDs) are essential to drain cerebrospinal fluid (CSF) and control intracranial pressure after neurosurgery (Brain Trauma Foundation et al., 2007; Carney et al., 2017). One of the risks surrounding the placement and maintenance of catheters is the infection of cerebrospinal fluid (CSF), with infection rates ranging from 2 to 45% (Lozier et al., 2008; Scheithauer et al., 2010; van de Beek et al., 2010; Dorresteijn et al., 2019). EVD/LD-associated ventriculitis and meningitis (VM) are correlated with high rates of morbidity and mortality, significantly prolonged lengths of hospitalization (LOS), elevated hospital costs, and negatively affected prognosis (Citerio et al., 2015; Hussein et al., 2019). However, the diagnosis and pathogen identification of VM samples can be difficult (Ross et al., 1988), causing delays in treatment, and the administration of inadequate/inappropriate antimicrobial therapies (Phan et al., 2016).

Metagenomic next-generation sequencing (mNGS) is a high-throughput sequencing technology that could overcome the limitations of conventional diagnostic tests and allow for hypothesis-free, culture-independent pathogen detection directly from biological specimens (Simner et al., 2018). mNGS provides a comprehensive method that accurately identifies all potential pathogens, including viruses, bacteria, fungi, and parasites in a single test (Simner et al., 2018). This approach is valuable in distinguishing pathogens that cause infections with non-specific and overlapping clinical manifestations (Miller et al., 2019). Recent advances in mNGS have facilitated the development of rapid and user-friendly data analysis tools, and the creation of accurate and comprehensive databases. Combined with its low cost, mNGS has become a first-line laboratory method in response to emerging infectious diseases and outbreaks. mNGS has also become increasingly popular in routine clinical practice for the rapid identification of pathogens in CNS infections from CSF (Guan et al., 2016; Fan et al., 2018; Wilson et al., 2019; Xing et al., 2019). However, most published studies use non-standardized methods and lack standards for verifications. In addition, inadequate CSF samples and the lack of relevant studies on EVD/LD patients also hinder the broad application of mNGS in the diagnosis of VM.

What remains to be explored is whether the diagnosis of EVD/LD-associated VM using mNGS is as effective and efficient as conventional microbiological methodologies for pathogen detection and identification. Here, we compare diagnostic methods in 102 neurosurgical patients with EVD/LD and assess the performance of mNGS in identifying CSF pathogens and predicting VM. We also evaluate whether mNGS can provide therapeutic guidance for EVD/LD-associated VM, using *Acinetobacter baumannii* and *Stenotrophomonas maltophilia*-induced VM as examples.

MATERIALS AND METHODS

This study was performed at the Beijing Tiantan Hospital and Capital Medical University between February 2019 and September 2019. Beijing Tiantan Hospital is a tertiary hospital with 1,850 beds and more than 15,000 annual surgeries. The protocol for this study was approved by the ethical committee of Beijing Tiantan Hospital and Capital Medical University (Approval Number: KY-2019-095-03).

Diagnostic Criteria and Data Abstraction

Neurosurgical patients 18 years of age and older with an EVD/LD for more than 24 h were eligible. Patients with EVDs or LDs inserted for cerebral or spinal infective processes were excluded. Patients who were discharged without a new surgery died within 7 days after neurosurgery, or those with an inadequate CSF volume (<0.5 ml) or incomplete medical records were also excluded from the study. All patients enrolled in the study were followed up with throughout the first 30 postoperative days.

Since there are no universally accepted definitions of device-related VM, for the purpose of this study, we defined the diagnosis criteria as follows (Leib et al., 1999; Horan et al., 2008; Citerio et al., 2015): (1) confirmed VM (cVM) was characterized by a positive CSF culture and a CSF leukocyte count >250 cells/μl; (2) suspected VM (sVM) was characterized by a CSF leukocyte count >1,000 cells/μl (neutrophil percentage >50%) or a CSF neutrophil count >250 cells/μl; and (3) non-VM (nVM) was defined by a negative CSF culture and CSF leukocyte count <250 cells/μl, with < 50% neutrophils.

All patients' daily progress notes were extracted from the standard database. Nineteen clinical features were included, such as patient characteristics (age, gender), hypertension, diabetes mellitus, fever (>38°C), LOS, time of infection clearance, Glasgow Coma Scale (GCS), CSF leakage, craniotomy, surgical wound classification, reoperation, assist mechanical ventilation, tumor, intensive care unit (ICU) admission, bacteremia, hospital-acquired pneumonia, mortality, and clinical treatment, including antibiotic prophylaxis, empirical use of antibiotics, and definitive therapy. In addition, we recorded 12 clinical laboratory indicators, including CSF cell count (C-Cell), CSF leukocyte count (C-Leu), CSF neutrophil percentage (C-Neu), CSF protein concentration (C-Pro), CSF chloride ion concentration (C-Cl), CSF glucose concentration (C-Glu), CSF lactate concentration (C-Lac), blood leukocyte count (B-Leu), blood glucose concentration (B-Glu), blood neutrophil percentage (B-Neu), the ratio of C-Glu to B-Glu (C/B-Glu), and procalcitonin level (PCT). All of the clinical laboratory parameters were obtained on the same day that EVD/LD-associated VM occurred.

mNGS of CSF

Sample Processing

CSF specimens were subjected to conventional microbiological studies (e.g., culture, serology) and pairwise mNGS. Specimens (0.5 to 1.2 ml) were stored at −80°C until use. Fresh CSF was extracted for routine clinical examination. Antibiotic resistance was determined from bacterial genome sequences using mNGS.

Nucleic Acid Extraction

CSF samples were collected from patients according to standard procedures. DNA was extracted using the TIANamp Magnetic DNA kit (Tiangen) according to the manufacturer's protocols.

The quantity and quality of DNA were assessed using Qubit (Thermo Fisher Scientific) and NanoDrop (Thermo Fisher Scientific) instruments, respectively.

Library Preparation and Sequencing

DNA libraries were prepared using the KAPA Hyper Prep kit (KAPA Biosystems) according to the manufacturer's protocols. An Agilent 2100 was used for quality control, and single-read sequencing of 75 bp DNA libraries was sequenced on an Illumina Next-Seq 550Dx (Illumina).

Bioinformatics Analysis

An in-house-developed bioinformatics pipeline was used for pathogen identification. Briefly, high-quality sequencing data were obtained following the removal of low-quality reads, adapter contamination, and duplicated and short (read length <36 bp) reads. Human host sequences were identified by mapping to the human reference genome (hs37d5) using bowtie2 software. Reads that could not be mapped to the human genome were retained and aligned with the NCBI microorganism genome database for pathogen identification. The microorganism genome database is composed of more than 20,000 NCBI reference genomes of bacteria, fungi, viruses, and parasites (download from ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/). The mapped reads were taxonomically classified using the Expectation Maximization algorithm.

Interpretation and Reporting

We used the following criteria for the positive results from mNGS: (1) For *Mycobacterium*, *Nocardia*, and *Legionella pneumophila*, the results were considered positive if more than one unique read mapped to the same species. (2) For bacteria (excluding *Mycobacterium*, *Nocardia*, and *Legionella pneumophila*), fungi, virus, and parasites, the result was considered positive for a specific species if at least three unique, non-overlapping reads mapped to that species. (3) Pathogens that were detected in the negative "no-template" control (NTC) were excluded if the number of detected reads was <10-fold of that in the NTC. The schematic of the mNGS workflow is shown in **Figure 1A**.

Analysis of Antibiotic Resistance Genes

Megahit v1.2.5 was used to assemble reads and remove the host sequences using default parameters. Identification of the resulting contigs was performed by matching them to CARD v3.0.5 using RGI v5.1.0. Potential resistance genes were filtered using the following parameters: identity of 84%, and query coverage of 90%.

Statistical Analysis

SAS 15.0 and EXCEL 2010 were used for statistical analyses. The distribution of continuous variables was described using the mean and standard deviation (SD), or median and interquartile range (IQR). Categorical variables were described as counts and percentages. For continuous quantitative variables, the Mann–Whitney *U*-test was performed to compare two groups and the Kruskal–Wallis H test was used to compare three or more groups. The Fisher's exact test was performed to compare categorical data. The original contributions presented in the study are publicly available. This data can be found at https://www.ncbi.nlm.nih.gov/sra/PRJNA660207.

RESULTS

Clinical and Laboratory Data Representation

A total of 120 patients who underwent neurosurgery and had an EVD/LD in place for more than 24 h were screened. Of those, one patient died within 7 days, nine patients had insufficient CSF volumes (<0.5 mL), six patients had incomplete medical records, and two patients were followed up for <30 days. A total number of 102 CSF specimens were obtained from the remaining 102 patients, including 39 cVM, 49 sVM, and 14 nVM patients (**Figure 1B**).

The median age of the 102 patients was 41 years, and 59 patients were male (57.84%). The median LOS was 20.5 days (range, 15–27), and the median time to infection clearance was 6 days (range, 3–10.5). Eighteen patients (17.65%) had hypertension, two (1.96%) had diabetes mellitus, 23 (22.55%) experienced fever, and 11 (10.78%) reported CSF leakage. The incidence of craniotomy surgery and reoperation was 68.32% (69/102) and 14.71% (15/102), respectively, and 62.75% of cases (66/102) had clean surgical wounds after the procedure, while the remainder had clean-contaminated surgical wounds. Of all patients, 80 (78.43%) had tumors and 29 (27.88%) were admitted to the ICU. Other common comorbidities included hospital-acquired pneumonia (8/102, 7.84%) and bacteremia (5/102, 4.90%). The majority of patients received antibiotics. During the 30-day follow-up period, two patients died of VM. The median C-cell count was 3,841.5 cells/μL (range, 1,432–8,897), median C-Leu count was 1,577.5 cells/μL (range, 289–3,651), and median C-Neu count was 83.45% (range, 74.5–89.3). All demographic and clinical characteristics are summarized in **Table 1**. Univariate analysis was conducted to evaluate the association of the above clinical features with the VM groups. Of those, 10 characteristics showed significant differences among the three groups, including age ($P = 0.03$), LOS ($P = 0.04$), craniotomy ($P < 0.01$), reoperation ($P = 0.03$), assisted mechanical ventilation ($P < 0.01$), bacteremia ($P = 0.02$), received definitive therapy ($P < 0.01$), C-Cell count ($P < 0.01$), C-Leu count ($P < 0.01$), and C-Neu count ($P = 0.01$).

Use of mNGS to Diagnose EVD/LD-Associated VM

To study the effectiveness and efficiency of mNGS for pathogen identification in VM patients, 102 CSF samples were tested by mNGS and conventional microbiological identification. Among the 88 patients diagnosed with cVM ($n = 39$) and sVM ($n = 49$), 24 (27.27%) were diagnosed by conventional identification and mNGS (**Figure 2A**), 22 (25.00%) were diagnosed by mNGS only (**Figure 2B**), 15 (17.05%) were diagnosed by conventional testing only (**Figure 2C**), and the remaining 27 (30.68%) were not diagnosed by either method. Fourteen nVM patients showed negativity using both methods. Notably, all mNGS results were

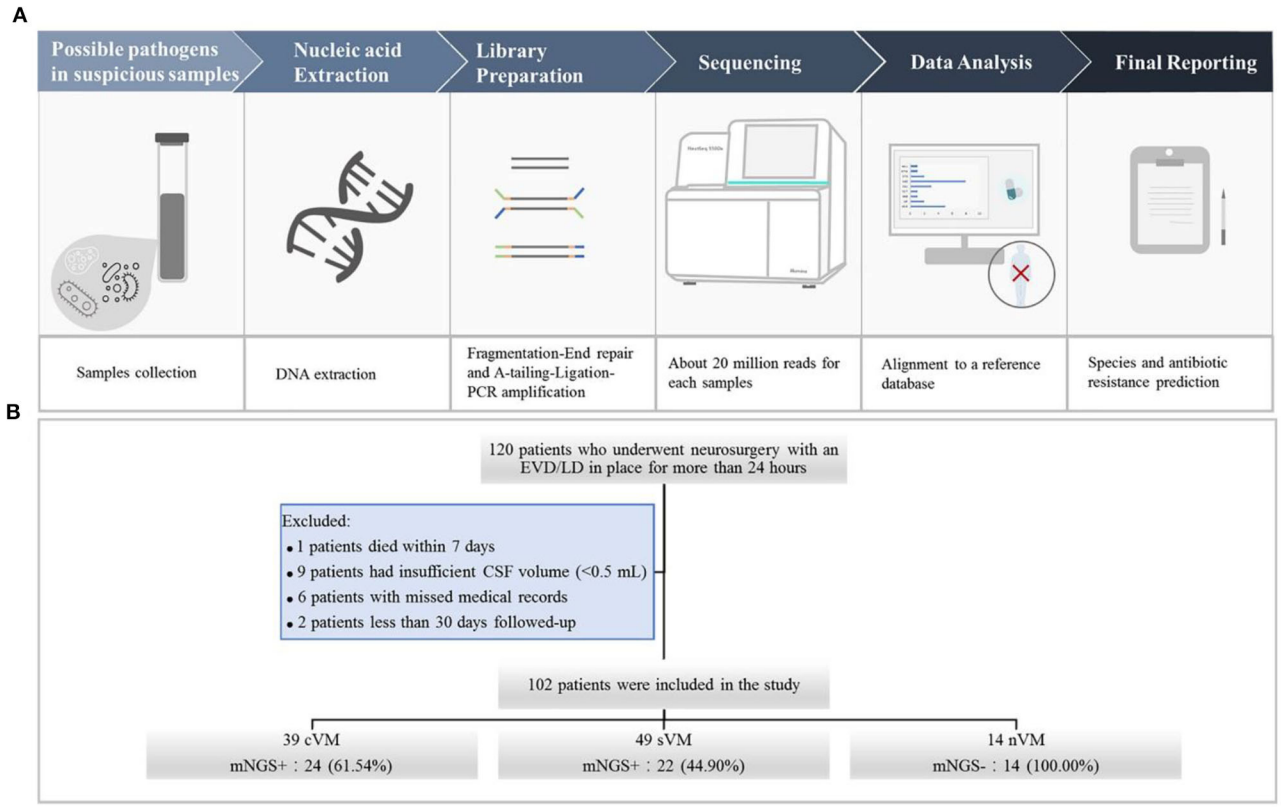


FIGURE 1 | (A) Schematic of the mNGS workflow. After collecting CSF samples, DNA was extracted using the TIANamp Magnetic DNA Kit (Tiangen). DNA libraries were prepared using the KAPA Hyper Prep kit (KAPA Biosystems) and fragmentation-end repair and a-tail-ligation PCR amplification. Libraries were quantified, pooled, and loaded onto the sequencer. Using an in-house-developed bioinformatics pipeline, high-quality sequencing data were obtained. Reads were aligned to the NCBI microorganism genome database for pathogen identification. Finally, we reported the results of mNGS according to the criteria. **(B)** Flowchart of study participants.

returned in <48 h, if not within 24 h. Such a turnaround was significantly shorter than that obtained using conventional microbiological tests, especially for positive samples.

Different types of infectious agents were identified by mNGS. Of the patients diagnosed by mNGS, 45.65% of VMs were caused by Gram-positive bacteria, 43.48% were caused by Gram-negative bacteria, and 10.87% were caused by fungi (**Figure 2D**). Furthermore, mNGS detected three fungal species, including *Candida albicans*, *Aspergillus sclerotioniger*, and *Malassezia restricta*. Among the 21 Gram-positive bacteria identified, *Staphylococcus epidermidis* (8, 38.10%) was the most common, followed by *Staphylococcus aureus* (3, 14.29%). Conversely, *A. baumannii* (5, 25.00%) was the most common Gram-negative bacterium identified. In addition, some pathogens were detected solely by mNGS, such as *Pseudomonas aeruginosa*, *Serratia marcescens*, *Acinetobacter johnsonii*, and *Burkholderia multivorans*, which highlighted the advantage of mNGS over conventional methods as such pathogens are not typically considered by clinicians during routine practice.

In the cVM group, the positive mNGS detection rate was 61.54% (24/39). Inconsistent positive agents revealed by conventional tests were seen in 7 (17.95%) patients and mainly included *Moraxella osloensis*, *Escherichia coli*, *Klebsiella Pneumoniae*, *Staphylococcus haemolyticus*, *Staphylococcus xylosus*, and *Staphylococcus hominis*. Moreover, mNGS effectively identified coinfections from multiple bacteria. Among 49 patients with sVM, mNGS identified bacterial pathogens in 22 patients (44.90%) that were negative based on the conventional method. Despite such discrepancy, 10 of those 22 patients showed probable clinical responses (**Figure 3**). All 27 CSF samples in the nVM group were concordant between the conventional method and mNGS (**Table 2**).

Unique Reads for Pathogens Detected by mNGS

In the cVM group, the median number of unique reads detected by mNGS was 248.5 (range, 39.5–8,043.5), while the median was 4 (range, 3–6) in patients with sVM, thus demonstrating a marked difference between these groups ($P < 0.01$; **Table 3**).

Antibiotic Resistance Gene Detection

Using mNGS, antibiotic resistance was identified in three CSF samples from cVM patients. Such resistance was imparted by *A. baumannii* and one instance of *S. maltophilia*. The results revealed the presence of several antibiotic resistance genes, including *adeIJK*, *adeFGH*, and *AbaQ*, which cause resistance

TABLE 1 | Demographic and clinical characteristics of the 102 patients.

Characteristics	All patients (102)	cVM (39)	sVM (49)	Non-VM (14)	*P*-value	P1	P2
Age (median, IQR, years)	41 (29, 55)	45 (39, 57)	41 (29, 55)	30 (15, 40)	0.03	<0.01	0.07
Gender (male, %)	59 (57.84%)	22 (56.41%)	26 (53.06%)	11 (78.57%)	0.25	0.20	0.13
Hypertension	18 (17.65%)	10 (25.64%)	7 (14.28%)	1 (7.14%)	0.26	0.25	0.67
Diabetes mellitus	2 (1.96%)	1 (2.46%)	1 (2.04%)	0 (0%)	1.00	1.00	1.00
Fever (b.t >38°C)	23 (22.55%)	13 (33.33%)	8 (16.32%)	2 (14.28%)	0.14	0.30	1.00
LOS (median, IQR)	20.5 (15, 27)	24(16, 31)	18 (15, 24)	22.5 (18, 27)	0.04	0.75	0.12
Time of cure of infection (median, IQR)	6 (3, 10.5)	6 (3, 11)	4 (1, 5)	6.5 (3.5, 12.0)	0.37	0.88	0.21
GCS (median, IQR)	5 (4, 9)	5 (4.0, 10.5)	–	7.1 (7.1, 7.1)	0.70	0.70	–
CSF leakage	11 (10.78%)	4 (10.25%)	7 (14.28%)	0 (0%)	0.45	0.56	0.33
Craniotomy	70 (68.63%)	18 (46.15%)	39 (79.59%)	13 (92.85%)	<0.01	<0.01	0.43
Surgical wound classification					0.78	1.00	1.00
Clean (I)	64 (62.75%)	26 (66.67%)	29 (59.18%)	9 (64.29%)			
Clean-contaminate (II)	38 (37.25%)	13 (33.33%)	20 (40.82%)	5 (35.71%)			
Reoperation	15 (14.71%)	10 (25.64%)	3 (6.12%)	2 (14.29%)	0.03	0.48	0.31
Assist mechanical ventilation	15 (14.71%)	13 (33.33%)	0 (0%)	2 (14.29%)	<0.01	0.30	0.05
Tumor	80 (78.43%)	31 (79.49%)	37 (75.51%)	12 (85.71%)	0.85	1.00	0.72
ICU admission	29 (27.88%)	14 (35.89%)	11 (22.45%)	4 (28.57%)	0.37	0.75	0.73
Bacteremia	5 (4.90%)	5 (12.82%)	0 (0%)	0 (0%)	0.02	0.31	\
Hospital-acquired pneumonia	8 (7.84%)	6 (15.38%)	1 (2.04%)	1 (7.14%)	0.05	0.66	0.40
Mortality	2 (1.96%)	2 (5.13%)	0 (0%)	0 (0%)	0.40	1.00	\
Antibiotics							
Antibiotic prophylaxis	84 (82.35%)	36 (92.31%)	36 (73.47%)	12 (85.71%)	1.00	0.60	0.49
Empirical therapy	91 (89.22%)	35 (89.73%)	43 (87.76%)	13 (92.86%)	1.00	1.00	1.00
Definitive therapy	82 (80.39%)	37 (94.87%)	34 (69.38%)	11 (78.57%)	<0.01	0.11	0.74
C-cell/μl (median, IQR)	3,841.5 (1,432, 8,897)	3,186 (671, 8,109)	4,671 (2,548, 14,442)	806.5 (125, 3,695)	<0.01	0.07	<0.01
C-leu/μl (median, IQR)	1,577.5 (289, 3,651)	586 (242, 2,110)	2,310 (1,632, 4,396)	171.5 (51,535)	<0.01	<0.01	<0.01
C-neu% (median, IQR)	83.45 (74.5, 89.3)	84 (66.9, 89.3)	85.7 (79.6, 89.9)	65.5 (27.7, 83.6)	0.01	0.07	<0.01
C-pro mg/dl (median, IQR)	121.305 (84.58, 212.03)	112.2 (84.6, 230.8)	139.6 (94.9, 194.9)	81.4 (65.8, 191.5)	0.12	0.17	0.04
C-Cl mmol/L (median, IQR)	119.5 (115, 123)	121.0 (115.0, 125.0)	118.7 (114.7, 120.6)	120.6 (117.3, 123.8)	0.16	0.81	0.20
C-GLU mmol/L (median, IQR)	2.4 (1.65, 3.15)	2.6 (1.9, 3.7)	2.1 (1.4, 3.1)	2.6 (1.2, 3.1)	0.14	0.44	0.83
C-lac mmol/L (median, IQR)	5.35 (3.8, 6.3)	5.4 (3.4, 7.0)	5.5 (4.8, 6.3)	4.6 (2.4, 5.3)	0.10	0.14	0.03
B-leu mg/L (median, IQR)	11.84 (9.02, 14.97)	11.7 (8.8, 15.3)	11.9 (9.2, 15.0)	12.1 (8.2, 14.1)	0.92	0.83	0.95
B-glu mmol/L (median, IQR)	6.13 (5.22, 7.38)	6.1 (5.0, 8.0)	6.3 (5.3, 7.4)	5.9 (5.3, 7.0)	0.79	0.98	0.62
B-neu% (median, IQR)	8.85 (7.15, 12.19)	8.9 (7.0, 12.2)	8.7 (7.5, 12.4)	9.4 (5.7, 11.8)	0.93	0.98	0.91
C/B-glu (median, IQR)	0.36 (0.2, 0.49)	0.4 (0.3, 0.6)	0.4 (0.2, 0.4)	0.4 (0.2, 0.5)	0.30	0.32	0.88
PCT (median, IQR)	0.11 (0.0, 0.59)	0.1 (0.0, 0.7)	0.1 (0.0, 0.1)	–	0.92	–	–

P1, Comparison between cVM and nVM; P2, Comparison between sVM and nVM.

to fluoroquinolones or tetracyclines. Additionally, *smeABC* and *smeDEF* genes that cause resistance to multiple antibiotics were also identified.

DISCUSSION

To our knowledge, the present study is the first to evaluate the use of mNGS for pathogen detection in a large prospective cohort of patients with EVD/LD-associated VM. Patients were divided into three subgroups based on routine clinical test results. The effectiveness and efficiency of mNGS were compared to a conventional microbiological identification method using CSF samples. In addition, we proposed new criteria for the determination of positive specimens according our detection methods. Our study showed that mNGS can provide much quicker and more accurate etiologic pathogen identification results than the conventional microbiological identification method.

VM is an acute complication associated with neurosurgery and may lead to permanent adverse outcomes if not managed properly (Beer et al., 2008). The incidence of VM rates was reported to range from 2 to 45% (Dorresteijn et al., 2019), and the predisposing factors for VM include non-adherence to rigid insertion and maintenance protocols, leakage of CSF, catheter irrigation, and the frequency of EVD or LD manipulation. Meanwhile, our study indicated the same results in cVM and

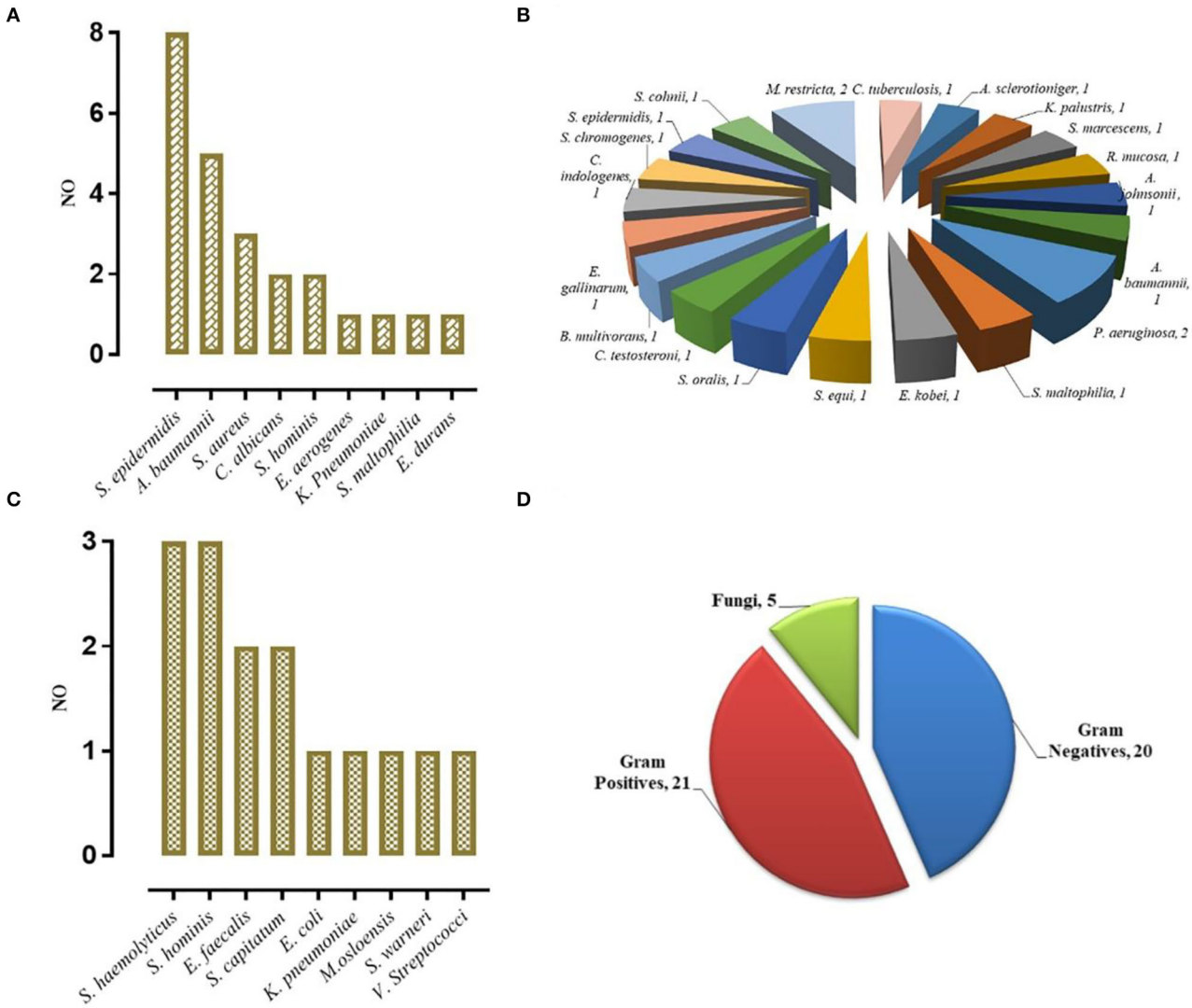
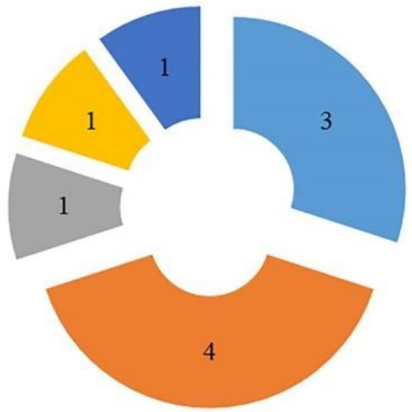


FIGURE 2 | VM Diagnosed by mNGS. **(A)** Concurrent diagnosis by mNGS and conventional microbiological testing (24). **(B)** Diagnosis by mNGS only (22). **(C)** Diagnosis by conventional testing only (15). **(D)** The category of pathogens diagnosed by mNGS.

sVM (39 and 49 of 102 patients, respectively). The prevention and management of EVD/LD-associated VM are great clinical challenges, particularly due to the difficulty in detecting pathogens in a timely manner and the re-emergence of infections caused by multidrug-resistant pathogens and emergent invasive neurosurgical procedures (Mayhall et al., 1984). Approximately 50% of meningoencephalitis cases remain undiagnosed, despite extensive clinical laboratory testing (Glaser et al., 2006). Thus, novel technologies, such as mNGS, are particularly important for the diagnosis and evaluation of VM.

In our study, untargeted shotgun mNGS analyses were performed to sequence the entirety of the DNA and/or RNA, rather than using specific primers or probes (Quince et al., 2017). This method was successfully performed on 102 samples. Although insufficient starting DNA concentrations in CSF samples might limit microbial detection due to insufficient sequencing data or sequencing failure, our optimized mNGS pipeline identified high numbers of pathogen reads from such samples. mNGS was also able to distinguish a broader spectrum of pathogens compared to the conventional culture method. As shown by several studies, mNGS has also successfully diagnosed rare (Wilson et al., 2014), novel (Hoffmann et al., 2015), and atypical infectious etiologies (Mongkolrattanothai et al., 2017; Wilson et al., 2018) of encephalitis. In this study, among the 22 samples of sVM diagnosed by mNGS, the causative pathogen was either not considered by clinicians or had been tested negative by conventional testing. Remarkably, 10 of those 22 patients showed mNGS-guided clinical responses. For example, a patient infected with *P. aeruginosa* (according to mNGS) experienced fever 10 days after neurosurgery. Empirical antimicrobial therapy



S. epidermidis, S. cohnii, S. chromogenes-removal of the EVD/LD in 3-10 days after neurosurgery; narrowing of antibiotic therapy to vancomycin; discontinuation of meropenem

A. baumannii, P. aeruginosa, A. johnsonii, E. kobei-removal of the EVD/LD in 3-8 days after neurosurgery; narrowing of antibiotic therapy to meropenem.

P. aeruginosa - narrowing of antibiotic therapy to sulbactam and cefoperazone; continuation of antibiotics for 4 weeks

S. marcescens-removal of the LD in 8 days after neurosurgery, changing of antibiotic therapy to sulfamethoxazole and trimethoprim;

E. gallinarum-changing of antibiotic therapy to ampicillin, and removal of the EVD in 8 days after neurosurgery.

FIGURE 3 | Clinical effect of the results of mNGS (10 of 22 cases diagnosed by mNGS only).

administered by clinicians did not relieve infection-related symptoms, such as headache and a low level of consciousness. However, the mNGS result guided the empirical antimicrobial therapy of this patient, indicating its clinical utility.

Moreover, there were 24 cases co-diagnosed by mNGS and the conventional method. The results of mNGS were valuable in reassuring the results from the conventional method, and in the detection or exclusion of coinfections, particularly in patients receiving high-grade prophylactic antibiotic administration. Those findings highlighted the outstanding advantage of mNGS, in that it does not rely on *a priori* selection of targeted pathogens, but rather, it detects several potential infectious agents in a single assay (Goldberg et al., 2015). Thus, mNGS was useful for diagnostic testing of CSF samples, regardless of sample volume and availability. Collectively, a combination of mNGS and conventional testing of CSF samples improved the detection of infections (Xing et al., 2020).

Among patients with cVM, mNGS reported negative results in eight cases. However, a negative mNGS sample did not necessarily mean the patient was negative for a clinical infection (Schlaberg et al., 2017), as the prior use of antibiotics could affect pathogen detection. In this study, the majority of patients with cVM received prophylactic antibiotics and/or empirical antibiotics before the CSF samples were obtained for mNGS testing. Additionally, for samples that had undergone multiple freeze–thaw cycles, the diagnostic yield could be decreased. Finally, we observed that low numbers of unique reads were usually observed in the false-negative results. In the cVM patient cohort, mNGS identified different pathogens in seven patients, potentially due to factors including the contamination of pathogen DNA across samples during mNGS library preparation, low-complexity sequences matching low-quality reads from the sample, misannotated species, or contaminants from database entries, sequencing adaptors, or vectors, colonization (Zhang et al., 2019). Elevated levels of human circulating free DNA (cfDNA) backgrounds in samples had only little effects on the sensitivity and precision, while the level of environmental contamination influenced sensitivity (Blauwkamp et al., 2019). It is worth noting that when using mNGS, repeated sequencing may be necessary when inconsistent results are observed.

Removing false-positive results was a primary challenge for mNGS analysis. False-positive results are likely to mislead therapeutic decision making, and clinicians are cautious when interpreting the sequences of bacteria that are commonly found in hospital environments. Thus, great effort was made to reduce the influence of false-positive results. The complexity of mNGS analysis required us to reduce the potential of false-positive results including in the sampling process, laboratory practices, reagents, environment, and skin or other commensal flora (Laurence et al., 2014; Lusk, 2014; Strong et al., 2014;

TABLE 2 | The results of mNGS and conventional microbiological testing.

cVM (39) /number	mNGS	Conventional microbiological testing	sVM (49) /number	mNGS
1	*A. baumannii*	*A. baumannii*	1	*A. sclerotioniger*
2	*S. epidermidis*	*S. epidermidis*	2	*K. palustris*
3	*S. epidermidis*	*S. epidermidis*	3	*S. marcescens*
4	*S. epidermidis*	*S. epidermidis*	4	*R. mucosa*
5	*S. epidermidis*	*S. epidermidis*	5	*A. johnsonii*
6	*S. epidermidis*	*S. epidermidis*	6	*A. baumannii*
7	*S. epidermidis*	*S. epidermidis*	7	*P. aeruginosa*
8	*S. epidermidis*	*S. epidermidis*	8	*P. aeruginosa*
9	*S. epidermidis*	*S. epidermidis*	9	*S. maltophilia*
10	*E. aerogenes*	*E. aerogenes*	10	*E. kobei*
11	*K. pneumoniae*	*K. pneumoniae*	11	*S. equi*
12	*S. aureus*	*S. aureus*	12	*S. oralis*
13	*S. aureus*	*S. aureus*	13	*C. testosteroni*
14	*S. aureus*	*S. aureus*	14	*B. multivorans*
15	*E. durans*	*E. durans*	15	*E. gallinarum*
16	*S. hominis*	*S. hominis*	16	*C. indologenes*
17	*C. albicans*	*C. albicans*	17	*S. chromogenes*
18	*C. albicans*	*C. albicans*	18	*S. epidermidis*
19	*S. maltophilia*	*S. maltophilia*	19	*S. cohnii*
20	*A. baumannii*	*A. baumannii*	20	*M. restricta*
21	*A. baumannii*	*A. baumannii*	21	*M. restricta*
22	*A. baumannii*	*A. baumannii*	22	*C. tuberculosis*
23	*S. hominis*	*S. hominis*		
24	*A. baumannii*	*A. baumannii*		
25	*P. bivia*	*M. osloensis*		
26	*P. buccalis*	*E. coli*		
27	*L. rhamnosus*	*K. pneumoniae*		
28	/	*E. faecalis*		
29	/	*E. faecalis*		
30	*C. testosteroni*	*S. haemolyticus*		
31	*C. amalonaticus*	*S. haemolyticus*		
32	/	*V. Streptococci*		
33	/	*S. shominis*		
34	/	*S. capitatum*		
35	/	*S. capitatum*		
36	/	*S. warneri*		
37	*H. parahaemolyticus*	*S. haemolyticus*		
38	*S. aureus*	*S. hominis*		
39	*P. bivia*	*S. hominis*		

TABLE 3 | Unique reads for pathogens detected by mNGS.

Microorganism	Unique reads	Microorganism	Unique reads	*P*-value
cVM-*A. baumannii*-1	250,078	sVM-*A. sclerotioniger*-1	10	$P < 0.01$
cVM-*S. epidermidis*-2	78	sVM-*K. palustris*-2	7	
cVM-*S. epidermidis*-3	149	sVM-*S. marcescens*-3	8	
cVM-*S. epidermidis*-4	13	sVM-*R. mucosa*-4	13	
cVM-*S. epidermidis*-5	3,496	sVM-*A. johnsonii*-5	3	
cVM-*S. epidermidis*-6	348	sVM-*A. baumannii*-6	3	
cVM-*S. epidermidis*-7	5,451	sVM-*P. aeruginosa*-7	4	
cVM-*S. epidermidis*-8	15,821	sVM-*P. aeruginosa*-8	3	
cVM-*S. epidermidis*-9	12	sVM-*S. maltophilia*-9	7	
cVM-*E. aerogenes*-10	955	sVM-*E. kobei*-10	4	
cVM-*K. Pneumoniae*-11	512,31	sVM-*S. equi*-11	3	
cVM-*S. aureus*-12	17	sVM-*S. oralis*-12	6	
cVM-*S. aureus*-13	434	sVM-*C. testosteroni*-13	3	
cVM-*S. aureus*-14	70	sVM-*B. multivorans*-14	5	
cVM-*E. durans*-15	35,244	sVM-*E. gallinarum*-15	3	
cVM-*S. hominis*-16	4	sVM-*C. indologenes*-16	3	
cVM-*C. albicans*-17	49	sVM-*S. chromogenes*-17	3	
cVM-*C. albicans*-18	41	sVM-*S. epidermidis*-18	5	
cVM-*S. maltophilia*-19	35,862	sVM-*S. cohnii*-19	6	
cVM-*A. baumannii*-20	120,122	sVM-M. restricta-20	3	
cVM-*A. baumannii*-21	3,785	sVM-*M. restricta*-21	4	
cVM-*A. baumannii*-22	43	sVM-*C. tuberculosis*-22	5	
cVM-*S. hominis*-16	35			
cVM-*A. baumannii*-24	24			
cVM-Median (IQR)	248.5 (39.5, 8,043.5)	sVM-Median (IQR)	4 (3, 6)	

Hocquet et al., 2016). For example, we took extreme care with sample handling to avoid cross-contamination (Chiu and Miller, 2019). We developed a strict laboratory workflow, whereby we required highly trained personnel to perform analyses and focused on minute amounts of exogenous nucleic acids introduced during specimen collection, aliquoting, nucleic acid extraction, and library preparation. We also ensured that laboratory surfaces, consumables, and reagents were DNA free. We also formulated strict bioinformatic criteria for calling positive results.

The number of unique mNGS reads in the sVM group was significantly lower than that of the cVM group, potentially due to the prior use of antibiotics or the development of diseases that affected pathogen concentration. As it is difficult to identify bacteria or fungi in extremely low concentrations by conventional microbiological measures, mNGS is useful due to its high sensitivity.

The most common causative bacterial pathogens of VM are skin commensals, such as coagulase-negative staphylococci, *S. aureus*, *E. coli*, *A. baumannii*, and *K. pneumoniae* (Dorresteijn et al., 2020). This was similar to the findings of our study, where *S. epidermidis*, *A. baumannii*, and *S. aureus* were the primary pathogens causing VM. The information provided by mNGS was leveraged to identify antibiotic resistance mechanisms from CSF samples to evaluate disease risk (Stefan et al., 2016). The genes, *adeIJK*, *adeFGH*, *AbaQ*, *adeABC*, and *adeL* were identified in three CSF samples where cVM was caused by *A. baumannii*. Those antibiotic resistance genes cause resistance to fluoroquinolones and tetracyclines (Damier-Piolle et al., 2008; Pérez-Varela et al., 2018). The *S. maltophilia* antibiotic resistance genes, *smeABC* and *smeDEF*, were also found in one CSF sample

(Li et al., 2002). This information facilitated the implementation of rational antibiotic administration guidelines. The detrimental effects of delays in the administration of antibiotics are widely known from the management of meningitis (van de Beek et al., 2012). Thus, if rapid bacterial identification is performed by mNGS, a timelier adjustment in antibiotics can be achieved and might lead to a better prognosis for VM patients.

Even though mNGS detected several pathogens that complemented conventional testing and played an important role in the guidance of clinical EVD/LD-associated VM treatment, there are some limitations to this technology. The financial cost of mNGS is higher than that of conventional testing. A number of factors also need to be ensured for accurate detection, including the quality of the available CSF samples, including the timing of their collection relative to symptom onset, whether the sample was adequately handled to ensure sterility, and whether the available DNA libraries offer wide coverage.

There are also several limitations to this research. First, this study was a single-center study with a single cohort of patients, thus resulting in a small sample size, especially after stratification of patients according to the VM criteria. Second, antibiotic resistance genes were only detected in four samples, which cannot indicate the ability of mNGS in resistance gene detection. Finally, we did not conduct real-time quantitative PCR to verify the mNGS results due to the insufficient volumes of CSF. Thus, clinical verification (clinical diagnosis, antibiotic treatment, etc.) was used as a substitute. With continuous development and optimization of mNGS technology, and stricter standards for sample collection and storage, we will perform a multicenter prospective study to better evaluate the use of mNGS in VM.

CONCLUSIONS

Our findings revealed that mNGS of CSF is a practical tool for the diagnostic evaluation of patients with VM. Compared with traditional diagnostic tools, mNGS is a rapid and accurate method for pathogen identification and disease diagnosis. This study highlights the feasibility of testing CSF samples with mNGS as a diagnostic tool for EVD/LD-associated VM to guide timely and targeted treatments.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: no restrictions. Requests to access these datasets should be directed to Guojun Zhang. zgjlunwen@163.com.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by this study was approved by the ethical committee of Beijing Tiantan Hospital and Capital Medical University (Approved Number: KY-2019-095-03). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

GZha and GZhe conceived and designed the study. LQ retrieved the literature and drafted the manuscript. YS and FL reviewed the literature and edited the manuscript. YW, MM, YZ, and YWS participated in the study design and editing of the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was supported by a research Grant No. QML20180502, Beijing Municipal Administration of Hospitals.

ACKNOWLEDGMENTS

On behalf of all authors, I want to thank the patients who participated in this study and Dinfectome Inc, Nanjing, who provided technical guidance and staff training. I am also grateful to those individuals who also provided valuable suggestions and corrections.

REFERENCES

Beer, R., Lackner, P., Pfausler, B., and Schmutzhard, E. (2008). Nosocomial ventriculitis and meningitis in neurocritical care patients. *J. Neurol.* 255, 1617–1624. doi: 10.1007/s00415-008-0059-8

Blauwkamp, T. A., Thair, S., Rosen, M. J., Blair, L., Lindner, M. S., Vilfan, I. D., et al. (2019). Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. *Nat. Microbiol.* 4, 663–674. doi: 10.1038/s41564-018-0349-6

Brain Trauma Foundation., American Association of Neurological Surgeons., Congress of Neurological Surgeons., Joint Section on Neurotrauma and Critical Care AANS/CN., Bratton, S. L., Chestnut, R. M., et al. (2007). Guidelines for the management of severe traumatic brain injury. *IV. infection prophylaxis. J. Neurotrauma* 24 (Suppl. 1), S26–31. doi: 10.1089/neu.2007.999

Carney, N., Totten, A. M., O'Reilly, C., Ullman, J. S., Hawryluk, G. W., Bell, M. J., et al. (2017). Guidelines for the management of severe traumatic brain injury, fourth edition. *Neurosurgery* 80, 6–15. doi: 10.1227/NEU.0000000000001432

Chiu, C. Y., and Miller, S. A. (2019). Clinical metagenomics. *Nat. Rev. Genet.* 20, 341–355. doi: 10.1038/s41576-019-0113-7

Citerio, G., Signorini, L., Bronco, A., Vargiolu, A., Rota, M., Latronico, N., et al. (2015). External ventricular and lumbar drain device infections in ICU patients: a prospective multicenter Italian study. *Crit. Care Med.* 43, 1630–1637. doi: 10.1097/CCM.0000000000001019

Damier-Piolle, L., Magnet, S., Brémont, S., Lambert, T., and Courvalin, P. (2008). AdeIJK, a resistance-nodulation-cell division pump effluxing multiple antibiotics in *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* 52, 557–562. doi: 10.1128/AAC.00732-07

Dorresteijn, K., Brouwer, M. C., Jellema, K., and van de Beek, D. (2020). Bacterial external ventricular catheter-associated infection. *Expert. Rev. Anti Infect. Ther.* 18, 219–229. doi: 10.1080/14787210.2020.1717949

Dorresteijn, K., Jellema, K., van de Beek, D., and Brouwer, M. C. (2019). Factors and measures predicting external CSF drain-associated ventriculitis: a review and meta-analysis. *Neurology* 93, 964–972. doi: 10.1212/WNL.0000000000008552

Fan, S., Qiao, X., Liu, L., Wu, H., Zhou, J., Sun, R., et al. (2018). Next-generation sequencing of cerebrospinal fluid for the diagnosis of neurocysticercosis. *Front. Neurol.* 9:471. doi: 10.3389/fneur.2018.00471

Glaser, C. A., Honarmand, S., Anderson, L. J., Schnurr, D. P., Forghani, B., Cossen, C. K., et al. (2006). Beyond viruses: clinical profiles and etiologies associated with encephalitis. *Clin. Infect. Dis.* 43, 1565–1577. doi: 10.1086/509330

Goldberg, B., Sichtig, H., Geyer, C., Ledeboer, N., and Weinstock, G. M. (2015). Making the leap from research laboratory to clinic: challenges and opportunities for next-generation sequencing in infectious disease diagnostics. *MBio* 6, e01888–e01815. doi: 10.1128/mBio.01888-15

Guan, H., Shen, A., Lv, X., Yang, X., Ren, H., Zhao, Y., et al. (2016). Detection of virus in CSF from the cases with meningoencephalitis by next-generation sequencing. *J. Neurovirol.* 22, 240–245. doi: 10.1007/s13365-015-0390-7

Hocquet, D., Muller, A., and Bertrand, X. (2016). What happens in hospitals does not stay in hospitals: antibiotic-resistant bacteria in hospital wastewater systems. *J. Hosp. Infect.* 93, 395–402. doi: 10.1016/j.jhin.2016.01.010

Hoffmann, B., Tappe, D., Höper, D., Herden, C., Boldt, A., Mawrin, C., et al. (2015). A variegated squirrel bornavirus associated with fatal human encephalitis. *N. Engl. J. Med.* 373, 154–162. doi: 10.1056/NEJMoa1415627

Horan, T. C., Andrus, M., and Dudeck, M. A. (2008). CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. *Am. J. Infect. Control.* 36, 309–332. doi: 10.1016/j.ajic.2008.03.002

Hussein, K., Rabino, G., Feder, O., Eghbaryeh, H., Zayyad, H., Sviri, G., et al. (2019). Risk factors for meningitis in neurosurgical patients with cerebrospinal fluid drains: prospective observational cohort study. *Acta Neurochir.* 161, 517–524. doi: 10.1007/s00701-019-03801-y

Laurence, M., Hatzis, C., and Brash, D. E. (2014). Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PLoS ONE* 9:e97876. doi: 10.1371/journal.pone.0097876

Leib, S. L., Boscacci, R., Gratzl, O., and Zimmerli, W. (1999). Predictive value of cerebrospinal fluid (CSF) lactate level versus CSF/blood glucose ratio for the diagnosis of bacterial meningitis following neurosurgery. *Clin. Infect. Dis.* 29, 69–74. doi: 10.1086/520184

Li, X. Z., Zhang, L., and Poole, K. (2002). SmeC, an outer membrane multidrug efflux protein of Stenotrophomonas maltophilia. *Antimicrob. Agents Chemother.* 46, 333–343. doi: 10.1128/AAC.46.2.333-343.2002

Lozier, A. P., Sciacca, R. R., Romagnoli, M. F., and Connolly, E. S. Jr. (2008). Ventriculostomy-related infections: a critical review of the literature. *Neurosurgery* 62 (Suppl. 2), 688–700. doi: 10.1227/01.neu.0000316273.35833.7c

Lusk, R. W. (2014). Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. *PLoS ONE* 9:e110808. doi: 10.1371/journal.pone.0110808

Mayhall, C. G., Archer, N. H., Lamb, V. A., Spadora, A. C., Baggett, J. W., Ward, J. D., et al. (1984). Ventriculostomy-related infections. *A prospective epidemiologic study. N. Engl. J. Med.* 310, 553–559. doi: 10.1056/NEJM198403013100903

Miller, S., Naccache, S. N., Samayoa, E., Messacar, K., Arevalo, S., Federman, S., et al. (2019). Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome Res.* 29, 831–842. doi: 10.1101/gr.238170.118

Mongkolrattanothai, K., Naccache, S. N., Bender, J. M., Samayoa, E., Pham, E., Yu, G., et al. (2017). Neurobrucellosis: unexpected answer from metagenomic next-generation sequencing. *J. Pediatr. Infect. Dis. Soc.* 6, 393–398. doi: 10.1093/jpids/piw066

Pérez-Varela, M., Corral, J., Aranda, J., and Barbé, J. (2018). Functional characterization of AbaQ, a novel efflux pump mediating quinolone resistance in acinetobacter baumannii. *Antimicrob. Agents Chemother.* 62, e00906–18. doi: 10.1128/AAC.00906-18

Phan, K., Schultz, K., Huang, C., Halcrow, S., Fuller, J., McDowell, D., et al. (2016). External ventricular drain infections at the canberra hospital: a retrospective study. *J. Clin. Neurosci.* 32, 95–98. doi: 10.1016/j.jocn.2016.03.019

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. (2017). Corrigendum: shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35:1211. doi: 10.1038/nbt1217-1211b

Ross, D., Rosegay, H., and Pons, V. (1988). Differentiation of aseptic and bacterial meningitis in postoperative neurosurgical patients. *J. Neurosurg.* 69, 669–674. doi: 10.3171/jns.1988.69.5.0669

Scheithauer, S., Bürgel, U., Bickenbach, J., Häfner, H., Haase, G., Waitschies, B., et al. (2010). External ventricular and lumbar drainage-associated meningoventriculitis: prospective analysis of time-dependent infection rates and risk factor analysis. *Infection* 38, 205–209. doi: 10.1007/s15010-010-0006-3

Schlaberg, R., Chiu, C. Y., Miller, S., Procop, G. W., Weinstock, G., Professional Practice Committee, et al. (2017). Validation of metagenomic next-generation sequencing tests for universal pathogen detection. *Arch. Pathol. Lab. Med.* 141, 776–786. doi: 10.5858/arpa.2016-0539-RA

Simner, P. J., Miller, S., and Carroll, K. C. (2018). Understanding the promises and hurdles of metagenomic next-generation sequencing as a diagnostic tool for infectious diseases. *Clin. Infect. Dis.* 66, 778–788. doi: 10.1093/cid/cix881

Stefan, C. P., Koehler, J. W., and Minogue, T. D. (2016). Targeted next-generation sequencing for the detection of ciprofloxacin resistance markers using molecular inversion probes. *Sci. Rep.* 6:25904. doi: 10.1038/srep25904

Strong, M. J., Xu, G., Morici, L., Splinter Bon-Durant, S., Baddoo, M., Lin, Z., et al. (2014). Microbial contamination in next generation sequencing: implications for sequence-based analysis of clinical samples. *PLoS Pathog.* 10:e1004437. doi: 10.1371/journal.ppat.1004437

van de Beek, D., Brouwer, M. C., Thwaites, G. E., and Tunkel, A. R. (2012). Advances in treatment of bacterial meningitis. *Lancet* 380, 1693–1702. doi: 10.1016/S0140-6736(12)61186-6

van de Beek, D., Drake, J. M., and Tunkel, A. R. (2010). Nosocomial bacterial meningitis. *N. Engl. J. Med.* 362, 146–154. doi: 10.1056/NEJMra0804573

Wilson, M. R., Naccache, S. N., Samayoa, E., Biagtan, M., Bashir, H., Yu, G., et al. (2014). Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N. Engl. J. Med.* 370, 2408–2417. doi: 10.1056/NEJMoa1401268

Wilson, M. R., O'Donovan, B. D., Gelfand, J. M., Sample, H. A., Chow, F. C., Betjemann, J. P., et al. (2018). Chronic meningitis investigated via metagenomic next-generation sequencing. *JAMA Neurol.* 75, 947–955. doi: 10.1001/jamaneurol.2018.0463

Wilson, M. R., Sample, H. A., Zorn, K. C., Arevalo, S., Yu, G., Neuhaus, J., et al. (2019). Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. *N. Engl. J. Med.* 380, 2327–2340. doi: 10.1056/NEJMoa1803396

Xing, X. W., Zhang, J. T., Ma, Y. B., He, M. W., Yao, G. E., Wang, W., et al. (2020). Metagenomic next-generation sequencing for diagnosis of infectious encephalitis and meningitis: a large, prospective case series of 213 patients. *Front. Cell Infect. Microbiol.* 10:88. doi: 10.3389/fcimb.2020.00088

Xing, X. W., Zhang, J. T., Ma, Y. B., Zheng, N., Yang, F., and Yu, S. Y. (2019). Apparent performance of metagenomic next-generation sequencing in the diagnosis of cryptococcal meningitis: a descriptive study. *J. Med. Microbiol.* 68, 1204–1210. doi: 10.1099/jmm.0.000994

Zhang, H. C., Ai, J. W., Cui, P., Zhu, Y. M., Hong-Long, W., Li, Y. J., et al. (2019). Incremental value of metagenomic next generation sequencing for the diagnosis of suspected focal infection in adults. *J. Infect.* 79, 419–425. doi: 10.1016/j.jinf.2019.08.012

Conflict of Interest: YWS was employed by Nanjing Geneseeq Technology Inc.

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

frontiers in Microbiology

ORIGINAL RESEARCH
published: 12 January 2021
doi: 10.3389/fmicb.2020.599524

Characterization of *Bacillus cereus* Group Isolates From Human Bacteremia by Whole-Genome Sequencing

*Angelica Bianco[1], Loredana Capozzi[1], Maria Rosa Monno[2], Laura Del Sambro[1], Viviana Manzulli[1], Graziano Pesole[3], Daniela Loconsole[4] and Antonio Parisi[1]**

[1] *Istituto Zooprofilattico Sperimentale della Puglia e della Basilicata, Foggia, Italy,* [2] *Department of Basic Medical Sciences, Neurosciences and Sense Organs, University of Bari "Aldo Moro", Bari, Italy,* [3] *Dipartimento di Bioscienze, Biotecnologie e Biofarmaceutica, University of Bari "A. Moro", Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies of the National Research Council and Consorzio Interuniversitario Biotecnologie, Bari, Italy,* [4] *Department of Biomedical Sciences and Human Oncology, Hygiene Unit, University of Bari "Aldo Moro", Bari, Italy*

Members of the *Bacillus cereus* group are spore-forming organisms commonly associated with food poisoning and intestinal infections. Moreover, some strains of the group (i.e., *B. cereus* sensu stricto and *Bacillus thuringiensis*) can cause bacteremia in humans, mainly in immunocompromised individuals. Here we performed the genetic characterization of 17 human clinical strains belonging to *B. cereus* group isolated from blood culture. The whole-genome sequencing (WGS) revealed that the isolates were closely related to *B. cereus* sensu stricto and *B. thuringiensis*–type strain. Multilocus sequence typing analysis performed on the draft genome revealed the genetic diversity of our isolates, which were assigned to different sequence types. Based on *panC* nucleotide sequence, the isolates were grouped in the phylogenetic groups III and IV. The *NHE*, *cer*, and *inhA* gene cluster, *entA*, *entFM*, *plcA*, and *plcB*, were the most commonly detected virulence genes. Although we did not assess the ability to generate biofilm by phenotypic tests, we verified the prevalence of biofilm associated genes using an *in silico* approach. A high prevalence of *pur* gene cluster, *xerC*, *clpY*, *codY*, *tasA*, *sipW*, *sinI*, and *sigB* genes, was found. Genes related to the resistance to penicillin, trimethoprim, and ceftriaxone were identified in most of the isolates. Intriguingly, the majority of these virulence and AMR genes appeared to be evenly distributed among *B. cereus* s.s. isolates, as well as closely related to *B. thuringiensis* isolates. We showed the WGS represents a good approach to rapidly characterize *B. cereus* group strains, being able to give useful information about genetic epidemiology, the presence of virulence and antimicrobial genes, and finally about the potential hazard related to this underestimated risk.

Keywords: *Bacillus cereus* group, whole-genome sequencing (WGS), ANIBlast, BTyper, virulence factors

OPEN ACCESS

Edited by:
Guido Werner,
Robert Koch Institute (RKI), Germany

Reviewed by:
Jasna Kovac,
Pennsylvania State University (PSU), United States
Christophe Nguyen-The,
Institut National de la Recherche Agronomique (INRA), France

****Correspondence:***
Antonio Parisi
antonio.parisi@izspb.it

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Microbiology

Received: *27 August 2020*
Accepted: *27 November 2020*
Published: *12 January 2021*

Citation:
Bianco A, Capozzi L, Monno MR, Del Sambro L, Manzulli V, Pesole G, Loconsole D and Parisi A (2021) Characterization of Bacillus cereus Group Isolates From Human Bacteremia by Whole-Genome Sequencing. Front. Microbiol. 11:599524. doi: 10.3389/fmicb.2020.599524

INTRODUCTION

Bacillus cereus sensu lato (*B. cereus* s.l.), known also as *B. cereus* group, consists of at least 12 spore-forming Gram-positive bacteria that are optionally motile and facultative anaerobic saprophyte (Liu et al., 2017). *B. cereus* group is widespread in nature as spores and vegetative cells. The spores are resistant to extreme environmental conditions (i.e., heat, freezing, drying, radiations)

and germinate when they come into contact with organic matter or within an animal host (Bottone, 2010). The group includes *B. cereus* sensu stricto (s.s.), which is responsible for both diarrheal and emetic human gastrointestinal syndromes and extraintestinal infections; *Bacillus thuringiensis*, an entomopathogen characterized by the production of crystal inclusions (containing insecticidal proteins); *Bacillus anthracis*, the agent of anthrax in humans and animals; *Bacillus mycoides* and *Bacillus pseudomycoides*, both of which are characterized by rhizoidal colonies on solid media and have not been described as food poisoning agents; *Bacillus weihenstephanensis*, a psychrotolerant bacterium; *Bacillus toyonensis*, which exhibits both probiotic and hemolytic properties; psychrotolerant and cytotoxic *Bacillus wiedmannii* (Miller et al., 2016); thermotolerant *Bacillus cytotoxicus*, which is responsible for occasional infections (Guinebretière et al., 2013); finally, the recently identified *Bacillus paranthracis*, *Bacillus pacificus*, *Bacillus tropicus*, *Bacillus albus*, *Bacillus mobilis*, *Bacillus Luti*, *Bacillus proteolyticus*, *Bacillus nitratireducens*, *Bacillus paramycoides*, *Bacillus gaemokensis*, *Bacillus manliponensis*, *Bacillus bingmayongensis*, and *Bacillus fungorum* (Liu et al., 2015, 2017, 2020). *B. cereus* s.l. is mainly responsible for two types of intoxication: the emetic gastrointestinal syndrome characterized by vomiting, strongly associated with rice and derived products (Johnson et al., 1983), and the diarrheal syndrome, characterized by aqueous diarrhea associated with abdominal pain. *B. cereus* s.l. is also involved in several non-gastrointestinal-tract clinical infections. The spectrum of syndromes includes fulminant septicemia, central nervous system involvement (meningitis and brain abscesses), gas gangrene–like infections (Bottone, 2010), progressive pneumonia (Hoffmaster et al., 2006), severe ocular infections such as endophthalmitis, and bacteremia in preterm neonates (Hilliard et al., 2003). Most commonly infected people are immunosuppressed patients (Goldstein and Abrutyn, 1985; Bryce et al., 1993), patients undergoing surgery, intravenous drug users, and patients with indwelling catheters (Hernaiz et al., 2003). The *B. cereus* catheter-related infections are generally caused by the formation of biofilm on biomedical devices (Ash et al., 1991; Kuroki et al., 2009; Liu et al., 2015). Hospital environment sources of *B. cereus* group include air filtration and ventilation equipment (Bryce et al., 1993), fiber-optic bronchoscopy equipment (Goldstein and Abrutyn, 1985), intravenous catheters (Hernaiz et al., 2003), and alcohol-based hand wash solutions (Hsueh et al., 1999). In recent years, it has been speculated that the gastrointestinal tract can act as a potential source of *B. cereus* strains acquired from an exogenous source (food, water, environment), which can invade the gastrointestinal tract, cause mucosal necrosis, and spread to other organs through the bloodstream (Bottone, 2010). Despite the low number of reports, also *B. thuringiensis* has been reported to be involved in gastrointestinal diseases (Jackson et al., 1995). Some *B. thuringiensis* strains are able to produce enterotoxins (Damgaard et al., 1997; Ghelardi et al., 2007) and possess genes known to be involved in the pathogenesis of *B. cereus* infections (Kreig and Lysenko, 1979; Hsieh et al., 1999).

The discrimination between pathogenic and non-pathogenic *B. cereus* group isolates has become a matter of public health. However, the close genetic relationship existing among the members of *B. cereus* group makes their identification to species level difficult, indicating that they have diverged from a common evolutionary lineage (Orrett, 2000; Liu et al., 2015).

Phenotypic and biochemical methods, as well as molecular methods, such as 16S rDNA or 23S rDNA sequencing, may not have sufficient discriminatory power to differentiate between members of the group (Kato et al., 2014; Yan et al., 2017). For these reasons, some other genetic loci have been selected as markers to differentiate between pathogenic and harmless *B cereus* group strains. Among these is the *rpo*B housekeeping gene (Caamaño-Antelo et al., 2015) or the pantoate-beta-alanine ligase gene (*pan*C) (Schmid et al., 2016; Warda et al., 2016), which classifies *B. cereus* isolates in seven phylogenetic groups (I to VII) (Guinebretière et al., 2008). Moreover, different schemes have been standardized for multilocus sequence typing (MLST), defined as TH (Tourasse et al., 2006), P (Priest et al., 2004), K, H (Helgason et al., 2004), and CS (Candelon et al., 2004; Sorokin et al., 2006)[1]. Recently, it has been shown that these methods are largely congruent in the *B. cereus* s.l. genomospecies attribution (Carroll et al., 2020). In order to evaluate the presence of the main virulence factors, generally polymerase chain reaction (PCR) amplifications are performed for the identification of seven enterotoxigenic among the *B. cereus* virulence genes: hemolysin BL (*hbl*A, *hbl*C, *hbl*D), enterotoxin non-hemolytic (*nhe*A, *nhe*B, *nhe*C); cytotoxin K (*cyt*K); enterotoxin FM (*ent*FM), enterotoxin S (*ent*S), and emetic toxin (*ces*) (Fricker et al., 2007; Owusu-Kwarteng et al., 2017).

The aim of this study was to characterize *B. cereus* s.l. isolated from 17 samples of blood cultures from hospitalized patients using different approaches. The presence of the genes associated with virulence and antimicrobial resistance (AMR) was checked by whole-genome sequencing (WGS) in all the isolates. Further, the *in vitro* sensitivity to antimicrobials of *B. cereus* isolates has also been evaluated.

MATERIALS AND METHODS

Seventeen Gram-positive *Bacillus* spp. isolated from blood culture collected from 17 epidemiologically non-related patients in the period 2004–2018 in a teaching hospital of Bari, Southern Italy, were studied. Blood culture analysis was performed by BacT/Alert system with FAN Plus Aerobic medium (bioMérieux, Marcy l'Etoile, France). When a positive bottle was flagged, a Gram stain of the broth was performed, and a portion of the fluid was subcultured on PolyViteX agar and on Columbia agar with 5% sheep blood (bioMérieux, Marcy l'Etoile, France). Identification of isolates was performed by VITEK 2 Automated system (bioMérieux, Marcy l'Etoile, France). For each patient, the strain belonging to *Bacillus* spp. was isolated from three separate blood cultures.

Matrix-Assisted Laser Desorption Ionization–Time of Flight Spectrometry

Prior to matrix-assisted laser desorption/ionization–time of flight (MALDI-TOF) analysis, the isolates were cultivated on Columbia

[1]http://mlstoslo.uio.no/

blood agar for 18–24 h at 37°C. After incubation, a sterile wooden tip was used to pick an isolated bacterial colony freshly grown and then smearing a thin film onto a 96-polished steel target plate (Bruker Daltonik GmbH, Germany) (direct transfer sample preparation procedure). Microbial films were overlaid with 1 μL α-cyano-4-hydroxycynnamic acid–matrix solution (Bruker Daltonik GmbH, Germany), prepared following the instruction for use and with final concentration of 10 mg/mL.

The sample-matrix mixture was dried at room temperature and subsequently inserted into the system for data acquisition. The mass spectra were generated using a MALDI-TOF system Microflex LT/SH^{TM} (Bruker Daltonik GmbH, Germany), which was operated in linear positive mode covering the molecular weight range of 2,000–20,000 Da. Each strain was applied to 10 spots, and each spot was hit 240 shots each in several points with a pulsed nitrogen laser beam operating at 337 nm, with a frequency equal to 60 Hz. Acceleration voltage was set at 20 kV, and the instrument was calibrated in the range of 3,637.8 and 16,953.3 Da using *Escherichia coli* DH5α (BTS, Bruker Daltonik GmbH, Germany). The data were processed automatically by the instrument software MBT Compass 4.1.70.1 database version 7.0.0.0 (Bruker Daltonik GmbH, Germany), and the spectra were compared with reference libraries for bacterial identification matching. The degree of correspondence between the test spectrum and the reference spectra in the database determines the attribution of the logarithmic score value (0–3.0). When a logarithmic score was < 1.7, the spectrum was reported as "not reliable identification," indicating that it could not identify the genus or species of the strain. A logarithmic score between 1.7 and 2.0 indicates that identification could be reliable only at the genus level, whereas a logarithmic score between 2.0 and 3.0 indicates that identification could be reliable at the species level of the organism.

Antibiotics Susceptibility Testing

The minimum inhibitory concentration (MIC) was used to determine antimicrobial susceptibility *in vitro* according to the Clinical and Laboratory Standards Institute (CLSI), as previously reported (Manzulli et al., 2019). The antibiotics tested were gentamicin, ceftriaxone, penicillin G, clindamycin, chloramphenicol, vancomycin, linezolid, cefotaxime, tetracycline, erythromycin, rifampin, amoxicillin, ciprofloxacin, doxycycline, and trimethoprim.

The CLSI breakpoints (μg/mL) for penicillin, ciprofloxacin, doxycicline, tetracycline, and cephems were those suggested for *B. cereus*, whereas for the other antimicrobials, interpretative criteria for *Staphylococcus* spp. were used according to CLSI guidelines M45-A2 (2011) and M100 (2017) (Sarker et al., 2007; Weinstein, 2018). *Staphylococcus aureus* ATCC 29213 and *E. coli* ATCC 25922 were used as control strains.

Whole-Genome Sequencing and Typing

Genomic DNA was extracted from the *B. cereus* s.l. isolates using DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany), according to the manufacturer's protocol. DNA quality and concentrations were estimated by Qubit Fluorometer using Qubit dsDNA HS Assay (Thermo Fisher Scientific). For each isolate, paired-end genomic libraries were prepared using Nextera DNA Flex Library preparation kit (Illumina, San Diego, CA, United States). Sequencing was performed using MiSeq Reagent Kit v2 (2×250 bp) on Illumina MiSeq platform (Illumina, San Diego, CA, United States). The paired-end raw reads were trimmed using Trimmomatic (Galaxy Version 0.36.6) (Bolger et al., 2014), and then the draft genomes were assembled by SPAdes 3.12.0 (Bankevich et al., 2012). Assembled genomes were submitted to BTyper tool (version 2.3.2) (Carroll et al., 2017), which performs *in silico* analysis to detect MLST profiles; *rpoB* allelic types (ATs), the belonging to the *panC* gene phylogenetic group; and identification of the closer strain, virulence factors, and AMR genes. Additionally, with the aim to identify the antibiotic resistance genes and plasmids, the draft genome of strains was analyzed using the software ABRicate (Galaxy Version 0.8), which includes different predownloaded databases [ARG-ANNOT (Gupta et al., 2014), NCBI AMRFinderPlus (Feldgarden et al., 2019), CARD (Jia et al., 2017), ResFinder (Zankari et al., 2012), and PlasmidFinder (Carattoli et al., 2014)]. In addition, the species identification was performed also by JSpecieWS online service (Richter et al., 2016)[2] using pairwise genome comparisons, which measures the average nucleotide identity (ANI) based on BLAST + (ANIb). We selected a total of 50 references genomes: 24 type strains and 26 genomes belonged to *B. cereus* group, of which 18 were defined as effective by Carroll et al. (2017). The draft genomes were also annotated using the software tool Prokka (version 1.13) (Seemann, 2014).

Nucleotide Sequence Accession Numbers

The draft genomes of *B. cereus* identified have been deposited in GenBank as BioProject PRJNA673333. The numbers of BioSample and accession ID are reported in **Table 1**.

RESULTS

Seventeen *B. cereus* s.l. isolated from human blood cultures of 17 patients were studied. MALDI-TOF mass spectrometry (MS) identified all strains as *B. cereus*. The draft genome sequence of the investigated *B. cereus* group isolates consisted of an average of 116 contigs comprising approximately 5,585,000 bp, with almost 5,737 predicted coding region sequences. The average coverage was estimated at ∼68.6*X*. The overall G + C content of 17 isolates was 35%. All data for each isolate were collected in **Supplementary Tables 1, 2**. The *rpoB* sequence revealed seven distinct ATs (**Table 1**): AT0380, AT0120, AT0092, AT0125, AT0481, AT0154, AT0125, and AT0463. The sequence of *panC* gene revealed that the isolates belonged to two phylogenetic group: clade III (IZSPB_BC106; IZSPB_BC107; IZSPB_BC109; IZSPB_BC110; IZSPB_BC114; IZSPB_BC115; IZSPB_BC210; IZSPB_BC211; IZSPB_BC213; IZSPB_BC214; IZSPB_BC217) and clade IV (IZSPB_BC108; IZSPB_BC111; IZSPB_BC112; IZSPB_BC212; IZSPB_BC215; IZSPB_BC216) (**Table 1**). Analysis of MLST genes from our isolate showed

[2]http://jspecies.ribohost.com/jspeciesws/

TABLE 1 | Characteristics of isolates *Bacillus cereus* s.l.

ID strain	BioSample ID	Accession ID	Closest strain	ST	CC	rpoB AT	MLST *Bacillus cereus*							panC phylogenetic groups
							glp	gmk	ilv	pta	pur	pyc	tpi	
IZSPB_BC106	SAMN16604380	JADNPV000000000	*B. thuringiensis*_YBT_020	2,163	–	AT0380	19	2	*368*	*333*	*181*	*3*	*275*	III
IZSPB_BC107	SAMN16604381	JADNPW000000000	*B. cereus*_03BB102	163	365	AT0120	44	1	32	16	18	33	24	III
IZSPB_BC108	SAMN16604382	JADNPX000000000	*B. cereus*_Rock1_15	73	142	AT0092	13	8	9	14	9	12	31	IV
IZSPB_BC109	SAMN16604383	JADNPY000000000	*B. cereus*_m1293	1066	205	AT0125	19	2	31	5	19	3	91	III
IZSPB_BC110	SAMN16604384	JADNPZ000000000	*B. cereus*_ATCC_10987	1,032	–	AT0481	117	4	123	118	43	6	3	III
IZSPB_BC111	SAMN16604385	JADNQA000000000	*B. thuringiensis*_T13001	24	–	AT0092	12	8	9	14	11	12	10	IV
IZSPB_BC112	SAMN16604386	JADNQB000000000	*B. thuringiensis*_T03a001	**2,096**	–	AT0154	91	8	14	11	2	36	7	IV
IZSPB_BC114	SAMN16604387	JADNQC000000000	*B. cereus*_m1293	462	–	AT0125	19	2	21	5	19	3	91	III
IZSPB_BC115	SAMN16604388	JADNQD000000000	*B_cereus*_Rock3_42	1,284	–	AT0463	247	1	83	1	239	37	43	III
IZSPB_BC210	SAMN16604389	JADNQE000000000	*B_cereus*_03BB102	365	–	AT0120	34	1	32	1	18	33	24	III
IZSPB_BC211	SAMN16604390	JADNQF000000000	*B_cereus*_ATCC_10987	1,032	–	AT0481	117	4	123	118	43	6	3	III
IZSPB_BC212	SAMN16604391	JADNQG000000000	*B_cereus*_Rock1_15	73	142	AT0092	13	8	9	14	9	12	31	IV
IZSPB_BC213	SAMN16604392	JADNQH000000000	*B_cereus*_03BB102	365	–	AT0120	34	1	32	1	18	33	24	III
IZSPB_BC214	SAMN16604393	JADNQI000000000	*B_cereus_0*3BB102	365	–	AT0120	34	1	32	1	18	33	24	III
IZSPB_BC215	SAMN16604394	JADNQJ000000000	*B_thuringiensis*_T13001	24	–	AT0092	12	8	9	14	11	12	10	IV
IZSPB_BC216	SAMN16604395	JADNQK000000000	*B_thuringiensis*_T13001	24	–	AT0092	12	8	9	14	11	12	10	IV
IZSPB_BC217	SAMN16604396	JADNQL000000000	*B_cereus*_03BB102	365	–	AT0120	34	1	32	1	18	33	24	III

Sequence type, clonal complex, rpoB allele type, MLST allele, and panC clade of B. cereus s.l. isolates. ST, sequence type; CC, clonal complex; AT, allele type. The sequence type in bold represent a new sequence type. The draft genomes were deposited as BioProject PRJNA673333.

different allelic combinations. In particular, 10 different STs were identified (**Table 1**), one of which (ST2096) resulted in a new ST and was submitted to the online MLST database[3]. Computational analysis performed by BTyper tool confirmed that our 17 isolates belonged to *B. cereus* group. In particular, among them, 12 (IZSPB_BC107, IZSPB_BC108, IZSPB_BC109, IZSPB_BC110, IZSPB_BC114, IZSPB_BC115, IZSPB_BC210, IZSPB_BC212, IZSPB_BC213, IZSPB_BC214, and IZSPB_BC217) were most closely to the type strain of *B. cereus* sensu stricto (*B. cereus* s.s.), and five (IZSPB_BC106, IZSPB_BC111, IZSPB_BC112, IZSPB_BC215, and IZSPB_BC216) were most closely to the type strain of *B. thuringiensis* (**Table 1**). Additionally, we performed species identification using 50 *B. cereus* species, including 26 type strains, using ANI by the online available service JSpacesWS. The results obtained showed a similar species attribution as provided by BTyper, with some exceptions: only two isolates (IZSPB_BC106 and IZSPB_BC112) were predicted as closer to *B. thuringiensis*, whereas the remaining isolates were all predicted as *B. cereus* s.s. (**Supplementary Table 3**). The presence of virulence factor genes was assessed by BTyper tool that identified a total of 28 genes (**Table 2**). Among them, 13 genes were identified in all of the isolates (100%; 17/17): two genes that codified for cereolysin proteins (*cerA* and *cerB*), two enterotoxin genes (*entA* and *entFM*), two immune inhibitor A precursor genes (*inhA1* and *inhA2*), the gene cluster of non-hemolytic enterotoxin (*nheA*, *nheB*, and *nheC*), two genes (*bpsE* and *bpsH*) of the gene cluster of exo-polysaccharide, the sphingomyelinase C gene (*sph*), and the phospholipase C (*plcB*). The pleiotropic regulator (PlcR) of extracellular virulence factor gene (*plcR*) was identified in all of our isolates, although the nucleotidic sequence of this gene matched with different *B. cereus* group species (**Table 2**). The genes *clo* and *plcA* were present in 94% (16/17) of isolates. The gene *bpsF* was present in 88% (15/17) of isolates; the gene *cytK2* was identified in 59% (10/17) of isolates; the gene *bpsD* was present in 41% (7/17) of isolates. The cluster genes of enterotoxins, hemolysin BL (*hblA*, *hblB*, *hblC*, and *hblD*) and the gene *hlyR*, were identified in 29% (5/17) of isolates (**Table 2**). Additionally, the annotation of the draft genome was performed for each isolate by the software Prokka, and a total of ∼5,630 genes were annotated (data not shown). Among them, 23 of 32 genes potentially involved in biofilm formation were identified in our isolates; in particular, among these genes, 17 were identified in 100% (17/17), 3 were identified in 97% (16/17), 1 was identified in 18% (3/17) of isolates, and 2 were identified in 12% (2/17) of isolates (**Table 3**). A total of 12 AMR genes were identified in the genome sequence of our isolates including (**Table 4**) the following: the vancomycin resistance genes: *Gly-vanR-M*, *Gly-vanZF-Pp*, and *vanR-B*, were identified in 100% (17/17), 88% (15/17), and 12% (2/17) of isolates, respectively; the beta-lactamase resistance genes: *BLA-1* and *BLA-2* and blaZ_12, were identified in 100% (17/17) and 6% (1/17) of isolates, respectively; the fosfomycin resistance gene: fosBx1, was identified in 100% (17/17) of isolates; the macrolide-lincosamide-streptogramin (*MLS-lsaB*) and the virginiamycin acetyltransferase (*vat-E*) were both identified in 41% (7/17) of isolates; the macrolide 2′-phosphotransferase II (*mph-B*) was identified in 35% (6/17) of isolates; the tetracycline resistance gene (*tetL*) and the resistance to macrolides, lincosamides, and streptogramin b (*erm-C*) were identified in 6% (1/17) of isolates, respectively. In addition to the *in silico* analysis, the antimicrobial susceptibility of *B. cereus s. l.* isolates to 15 antibacterial agents was determined, and the results are shown in **Table 5**. Among beta-lactam antibiotic class, only penicillin G resistance was confirmed in 100% (17/17) of isolates; eight isolates were resistant to ceftriaxone, whereas nine isolates showed intermediate resistance; six isolates were resistant, and 11 showed intermediate resistance to cefotaxime. Resistance to trimethoprim was observed in 100% of the isolates. The isolate IZSPB_BC210 showed resistance to clindamycin, whereas the remaining isolates showed intermediate resistance (24%; 4/17) or resulted susceptible (71%; 12/17). We found intermediate resistance to tetracycline and erythromycin in 24% (4/17) of isolate and to rifampicin in 12% (2/17) of isolates. All the isolates were susceptible to gentamicin, amoxicillin, chloramphenicol, vancomycin, linezolid, ciprofloxacin, and doxycycline.

Two plasmid replicons (*rep*) were detected: *rep12* belonging to a cryptic plasmid pBMB67 in IZSPB_BC107, IZSPB_BC108, and IZSPB_BC212, and *rep22* belonging to pUB110 in IZSPB_BC109 and IZSPB_BC210; additionally, in IZSPB_BC210, it was found repUS12 belonging to pUB110, also. Interestingly, the large plasmid that carries *cry* genes was not identified in none of the isolates that were predicted as closely to *B. thuringiensis*.

DISCUSSION

Bacillus species are widely distributed in nature and can colonize hospital environments; indeed, there is evidence that strains of *B. cereus* were found on the hands of nursing staff, in balloons used for manual ventilation and near ventilation system outlets (Kuroki et al., 2009). Recent reports suggest that *B. cereus* s.l. can cause nosocomial bacteremia via catheter-related infections caused by the formation of biofilm on biomedical devices (Jensen et al., 2003; Dohmae et al., 2008). Genetically, *B. cereus* s.s. is closely related to *B. thuringiensis* that is used extensively worldwide as pesticide in forestry and agriculture (Zhu et al., 2015). The aim of this study was to better understand the genetic characteristics of clinical *B. cereus* group isolates. In the past, several surveys using a variety of methods for detection of *B. cereus* group have been performed. However, some of them failed to discriminate between *B. cereus* group members at the species level (Seemann, 2014; Caamaño-Antelo et al., 2015; Zhu et al., 2015; Richter et al., 2016; Raymond and Federici, 2017). Generally, the identification and typing of *B. cereus* s.l. are based on MLST[4] or on the identification of virulence genes using PCR. Nevertheless, these methods are too expensive, time consuming, and labor intensive, and sometimes, because of the high genomic similarity within the group, they failed to identify or type correctly *B. cereus* spp. To date, different innovative techniques are available. Among

[3]https://pubmlst.org/bcereus/

[4]https://pubmlst.org/

TABLE 2 | Virulence factors of clinical isolates of *Bacillus cereus* s.l.

Virulence gene	IZSPB_BC106	IZSPB_BC107	IZSPB_BC108	IZSPB_BC109	IZSPB_BC110	IZSPB_BC111	IZSPB_BC112	IZSPB_BC114	IZSPB_BC115
Toxin									
cerA	95.05/100	95.76/100	100/100	95.05/100	95.05/100	100/100	100/100	95.05/100	95.05/100
cerB	87.45/93.39	92.02/71.47	95.18/93.39	87.78/93.39	87.46/93.39	94.86/93.39	94.86/93.39	87.78/93.39	91.41/87.39
clo	95.48/100	96.27/100	99.21/100	95.78/97.84	95.48/100	99.41/100	99.61/100	95.28/100	–/–
entA	94.93/100	94.26/100	99.32/100	94.93/100	94.93/100	99.32/100	97.97/100	94.93/100	93.58/100
entFM	95.58/100	93.26/100	97.44/100	94.42/100	95.58/100	97.44/100	91.4/100	94.88/100	92.56/100
nheA	97.40/100	96.89/100	99.74/100	97.41/100	97.41/100	99.74/100	98.96/100	97.41/100	96.89/100
nheB	99.25/100	99.50/100	100/100	99.5/100	99.25/100	100/100	99.5/100	99.5/100	99.25/100
nheC	97.497/100	94.43/100	99.72/100	95.54/100	97.49/100	100/100	99.16/100	95.54/100	94.43/100
cytK2	–/–	–/–	98.78/97.32	96.73/100	–/–	100/100	99.4/100	–/–	97.62/100
hblA	–/–	–/–	99.20/100	–/–	–/–	99.2/100	98.93/100	–/–	–/–
hblB	–/–	–/–	99.79/100	–/–	–/–	98.5/100	98.71/100	–/–	–/–
hblC	–/–	–/–	98.18/100	–/–	–/–	98.18/100	96.81/100	–/–	–/–
hblD	–/–	–/–	99.75/100	–/–	–/–	99.75/100	99.75/100	–/–	–/–
hlyR	–/–	–/–	–/–	–/–	–/–	73.13/100	–/–	–/–	99.5/100
Enzyme									
inhA1	94.03/100	96.32/100	99.87/100	96.07/100	94.04/100	99.87/100	96.07/100	96.07/100	96.32/100
inhA2	96.49/100	96.10/100	99.62/100	97/100	96.5/100	100/100	99.12/100	96.62/100	96.37/100
sph	97.92/100	99.22/100	92.01/100	99.11/100	97.93/100	92.31/100	91.42/100	99.11/100	99.7/100
bpsD	–/–	–/–	70.21/84.3	–/–	–/–	65.22/92.83	60.99/100	68.09/84.3	–/–
bpsE	89.16/93.89	89.89/93.90	90.29/94.24	79.79/95.59	89.17/93.9	88.85/94.24	88.85/94.24	89.57/94.24	87.36/93.90
bpsF	50/97.60		51.96/97.61	–/–	50/97.61	51.22/98.09	50.73/98.09	50/97.61	–/–
bpsH	67.33/98.68	66.78/100.00	78.95/100	67.32/100.66	67.33/98.68	79.28/100	79.61/100	79.28/100	67.65/100.66
Lipase									
plcA	93.31/100	94.83/100	100/100	94.53/100	93.31/100	100/100	97.26/100	94.53/100	94.53/100
plcB	95.05/100	95.76/100	100/100	95.05/100	95.05/100	100/100	100/100	95.05/100	95.05/100
Regulation									
plcR (*B. cereus* NC7401)	99.64/100	100/100	–/–	99.65/100	99.65/100	–/–	88.34/99.3	99.65/100	–/–
plcR (*B. thuringiensis*)	–/–	–/–	–/–	–/–	–/–	–/–	–/–	–/–	100/100
plcR (*B. cereus* ATCC 14579)	–/–	–/–	100/100	–/–	–/–	99.65/100	–/–	–/–	–/–
plcR (*B. anthracis*)	–/–	–/–	–/–	–/–	–/–	–/–	–/–	–/–	–/–

Virulence gene	IZSPB_BC210	IZSPB_BC211	IZSPB_BC212	IZSPB_BC213	IZSPB_BC214	IZSPB_BC215	IZSPB_BC216	IZSPB_BC217
Toxin								
cerA	95.05/100	95.05/100	100/100	95.05/100	95.05/100	100/100	100/100	95.05/100
cerB	91.41/87.39	88.42/93.39	95.18/93.39	91.41/87.39	91.41/87.39	94.86/93.39	94.86/93.39	91.41/87.39
clo	95.87/100	96.86/100	99.21/100	95.87/100	95.87/100	99.41/100	99.41/100	95.87/100
entA	93.58/100	95.27/100	99.32/100	93.58/100	93.58/100	99.32/100	99.32/100	93.58/100
entFM	92.56/100	95.35/100	97.44/100	92.56/100	92.56/100	97.44/100	97.44/100	92.56/100
nheA	96.11/100	97.15/100	99.74/100	96.11/100	96.11/100	99.74/100	99.74/100	96.11/100
nheB	99.5/100	99.5/100	100/100	99.5/100	99.5/100	100/100	100/100	99.5/100
nheC	94.71/100	94.71/100	99.72/100	94.71/100	94.71/100	100/100	100/100	94.71/100
cytK2	–/–	96.73/100	98.78/97.32	–/–	–/–	100/100	100/100	–/–
hblA	–/–	–/–	99.2/100	–/–	–/–	99.2/100	99.2/100	–/–
hblB	–/–	–/–	99.79/100	–/–	–/–	98.5/100	98.5/100	–/–
hblC	–/–	–/–	98.18/100	–/–	–/–	98.18/100	98.18/100	–/–
hblD	–/–	–/–	99.75/100	–/–	–/–	99.75/100	99.75/100	–/–
hlyR	–/–	–/–	73.13/100	–/–	–/–	73.13/100	73.13/100	–/–
Enzyme								
inhA1	96.19/100	95.56/100	99.87/100	96.19/100	96.19/100	99.87/100	99.87/100	96.19/100

(Continued)

TABLE 2 | Continued

Virulence gene	IZSPB_BC210	IZSPB_BC211	IZSPB_BC212	IZSPB_BC213	IZSPB_BC214	IZSPB_BC215	IZSPB_BC216	IZSPB_BC217
inhA2	96.25/100	96.37/100	99.62/100	96.25/100	96.25/100	100/100	100/100	96.25/100
sph	99.7/100	97.93/100	92.01/100	99.7/100	99.7/100	92.31/100	92.31/100	99.7/100
bpsD	–/–	–/–	70.21/84.3	–/–	–/–	65.22/92.83	65.22/92.83	–/–
bpsE	88.45/93.90	88.81/93.9	90.29/94.24	88.45/93.9	88.45/93.90	88.85/94.24	88.85/94.24	88.45/93.9
bpsF	50.49/97.61	51.22/98.09	51.96/97.61	50.49/97.61	50.49/97.61	51.22/98.09	51.22/98.09	50.49/97.61
bpsH	66.78/100	67.43/100	78.95/100	66.78/100	66.78/100	79.28/100	79.28/100	66.78/100
Lipase								
plcA	94.53/100	94.83/100	100/100	94.53/100	94.53/100	100/100	100/100	94.53/100
plcB	95.05/100	95.05/100	100/100	95.05/100	95.05/100	100/100	100/100	95.05/100
Regulation								
plcR (*B.cereus* NC7401)	–/–	–/–	–/–	–/–	–/–	–/–	–/–	–/–
plcR (*B. thuringiensis*)	100/100	–/–	–/–	100/100	100/100	–/–	–/–	100/100
plcR (*B. cereus* ATCC 14579)	–/–	–/–	100/100	–/–	–/–	99.65/100	99.65/100	–/–
plcR (*B. anthracis*)	–/–	97.89/100	–/–	–/–	–/–	–/–	–/–	–/–

The table reports the identity and coverage percentages for each virulence gene (Id%/Cov%).

these, MALDI-TOF MS is becoming an increasingly useful method for the rapid identification of bacteria and fungi. In fact, compared to conventional methods (phenotypic, genotypic, and immunological tests), this technology is fast and cheap and, for clinically significant bacteria, can provide accurate and reliable results from a single isolated colony within minutes. In our study, we compared the performance of two approaches for the capability to assign the isolates to the correct species: the MALDI-TOF MS and the ANIBlast method using the whole genome; for the last approach, we used two tools, BTyper and JSpeciesWS. Although both methods identified our isolates as *B. cereus* group members, the ANIBlast method was performing better, as it was able to identify two ANI groups, which included *B. cereus* s.s. and *B. thuringiensis* species, respectively. Even though the species predicted was the same using both ANIBlast tools, a different species attribution among our isolates was found: when we used Btyper, we found five *B. thuringiensis*, whereas when we used online available ANI Blast calculator, we found two B. *thuringiensis*; in both predictions, the remaining isolates were predicted as close to *B. cereus* s.s. However, *B. thuringiensis* virulence–associated genes and plasmids were not detected in any of our isolates. However, *B. thuringiensis* clones lacking Cry toxin have been described elsewhere (Zhu et al., 2015; Fayad et al., 2019) and defined as *B. thuringiensis-like* (Fayad et al., 2019). Thus, our results showed the limit in using the MALDI-TOF MS–based identification method, perhaps because some *B. cereus* group members, especially *B. cereus* s.s. and *B. thuringiensis*, do not have sufficient differences in their protein sequences, as they are genetically very similar (Guinebretière et al., 2008; Zheng et al., 2017).

With the aim of classifying potential pathogenic microorganisms quickly and effectively, WGS of 17 *B. cereus* group clinical isolates was performed. Based on the *rpoB* sequence, we identified seven different ATs, which showed a very high similarity (100%) to *B. cereus* s.s. Analysis sequence of *panC* gene revealed that the isolates belonged to phylogenetic groups III and IV. Interestingly, both groups include *B. cereus* s.s. isolates from hospitals and from patients (EFSA Panel on Biological Hazards, 2009), as well as more foodborne poisoning strains as reported in Guinebretière et al. (2008). The MLST analysis showed that the 17 isolates belonged to seven different STs, suggesting that the ability of *B. cereus* strains to cause human infection is not restricted to a specific ST or clonal group. About these, two sequence types, ST163 and ST73, were previously described as etiological cause of respiratory infection and pneumonia cases in Japan, respectively (Beecher and Wong, 2000). Because of the rarity of *B. cereus* infection and a paucity of genetic information, it is unclear whether particular genetic elements are associated with specific clinical manifestations. Generally, the pathogenicity of *B. cereus* has been associated with toxin production and putative virulence factors, such as enzymes and proteases, which are still poorly explored (EFSA Panel on Biological Hazards, 2016). Among the virulence factors, the toxins, such as hemolytic enterotoxin HBL, non-hemolytic enterotoxin NHE, cytotoxin K, and enterotoxin FM, have been associated with diarrheal diseases (Granum et al., 1999; Beecher and Wong, 2000; Lund et al., 2000; Hansen and Hendriksen, 2001; Fagerlund et al., 2004; Sergeev et al., 2006). The metalloprotease (*inhA*), the exo-polysaccharide (bpsX-H), and the phospholipases *sph* genes are important virulence factors as they make bacilli enable to escape innate and adaptive immune responses during infective phases (González-Zorn et al., 1999; Ramarao and Lereclus, 2005; Guillemet et al., 2010; Oh et al., 2011; Oda et al., 2012). Moreover, *sph* exhibits potent hemolytic

TABLE 3 | Genes that can play a role in biofilm formation detected in clinical isolates of *Bacillus cereus* s.l.

Gene	IZSPB_ BC106	IZSPB_ BC107	IZSPB_ BC108	IZSPB_ BC109	IZSPB_ BC110	IZSPB_ BC111	IZSPB_ BC112	IZSPB_ BC114	IZSPB_ BC115	IZSPB_ BC210	IZSPB_ BC211	IZSPB_ BC212	IZSPB_ BC213	IZSPB_ BC214	IZSPB_ BC215	IZSPB_ BC216	IZSPB_ BC217
xerC	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
clpY	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
clpQ	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
codY	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
purA	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
purB	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
purC	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+
purD	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
purE	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
purF	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
purH	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
purK	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+
purL	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
purM	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
purN	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
purQ	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
purS	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+
yezC	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
epsA	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
epsB	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
epsD	+	−	−	+	−	−	−	−	−	−	+	−	−	−	−	−	−
epsG	+	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−
epsK	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
epsM	+	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−
epsO	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
sipW	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
tasA	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
sinR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
sinI	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
tapA	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
aad	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
sigB	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

TABLE 4 | Presence of antimicrobial resistance genes.

AMR gene	IZSPB_ BC106	IZSPB_ BC107	IZSPB_ BC108	IZSPB_ BC109	IZSPB_ BC110	IZSPB_ BC111	IZSPB_ BC112	IZSPB_ BC114	IZSPB_ BC115	IZSPB_ BC210	IZSPB_ BC211	IZSPB_ BC212	IZSPB_ BC213	IZSPB_ BC214	IZSPB_ BC215	IZSPB_ BC216	IZSPB_ BC217
(Gly)vanZF-Pp| AF155139| 4339-4959| 621	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+
(Gly)vanR-M| FJ349556| 982-1680| 699	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
vanR-B	+	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−
(Bla)bla2| NG_047224| 101-874| 774	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
(Bla)BLA-1| AY453161| 501-1430| 930	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
blaZ_12	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−
(Fcyn)fosBx1| NG_050591| 101-517| 417	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Mph (B)	−	+	−	−	−	−	−	−	+	+	−	−	+	+	−	−	+
(MLS)lsa (B)| AJ579365| 4150-5628| 1479	−	−	+	−	−	+	+	−	+	−	−	+	−	−	+	+	−
(Tet)tetL| FN435329| 1-1377| 1377	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−
vat(E)	−	−	+	−	+	+	−	−	−	−	+	+	−	−	+	+	−
erm(C)	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−

TABLE 5 | Antimicrobial resistance pattern of clinical isolates of *Bacillus cereus* s.l.

		Interpretive Criteria (μg/mL)			IZSPB_ BC106	IZSPB_ BC107	IZSPB_ BC108	IZSPB_ BC109	IZSPB_ BC110	IZSPB_ BC111	IZSPB_ BC112	IZSPB_ BC114	IZSPB_ BC115	IZSPB_ BC210	IZSPB_ BC211	IZSPB_ BC212	IZSPB_ BC213	IZSPB_ BC214	IZSPB_ BC215	IZSPB_ BC216	IZSPB_ BC217
Antibioticrobial agent	**Range**	**S**	**I**	**R**																	
Gentamicin	0.008–16	≤ 4	8	≥ 16	S (1)	S (0.5)	S (1)	S (2)	S (1)	S (1)	S (1)	S (1)	S (0.5)	S (0.5)	S (2)	S (0.5)	S (1)	S (1)	S (1)	S (1)	S (1)
Penicillin G	0.06–128	≤ 0.12	–	≥ 0.25	R (> 128)	R (8)	R (> 128)	R (32)	R (64)	R (> 128)	R (> 128)	R (> 128)	R (16)	R (> 128)	R (64)	R (> 128)	R (> 128)	R (> 128)	R (128)	R (> 128)	R (> 128)
Amoxicillin	0.008–16	≤ 0.12	–	≥ 0.25	S (0.125)	S (0.125)	S (0.125)	S (0.125)	S (0.125)	S (0.125)	S (0.125)	S (0.06)	S (0.06)	S (0.06)	S (0.125)	S (0.06)	S (0.125)	S (0.06)	S (0.06)	S (0.125)	S (0.06)
Clindamycin	0.008–16	≤ 0.5	1– 2	≥ 4	S (0.125)	S (0.125)	S (0.25)	S (0.25)	S (0.25)	S (0.50)	S (0.25)	S (0.25)	I (1)	R (> 16)	S (0.50)	I (1)	S (0.50)	S (0.50)	I (1)	I (1)	S (0.50)
Chlo.ramphenicol	0.06–128	≤ 8	16	≥ 32	S (4)	S (4)	S (2)	S (4)	S (4)	S (4)	S (4)	S (4)	S (4)	S (4)	S (4)	S (2)	S (4)	S (2)	S (2)	S (1)	S (2)
Vancomycin	0.03– 64	≤ 4	–	–	S (2)	S (2)	S (2)	S (2)	S (2)	S (2)	S (2)	S (1)	S (2)	S (2)	S (1)	S (1)	S (2)	S (2)	S (1)	S (1)	S (2)
Linezolid	0.03– 64	≤ 4	–	≥ 8	S (2)	S (1)	S (1)	S (2)	S (2)	S (2)	S (1)	S (2)	S (2)	S (0.5)	S (1)	S (0.5)	S (1)	S (1)	S (1)	S (1)	S (1)
Erythromycin	0.008–16	≤ 0.5	1– 4	≥ 8	S (0.125)	I (4)	I (4)	S (0.06)	S (0.06)	S (0.06)	S (0.06)	S (0.125)	S (0.5)	I (4)	S (0.06)	I (4)	S (0.5)	S (0.5)	S (0.06)	S (0.25)	S (0.5)
Tetracycline	0.008–16	≤ 4	8	≥ 16	I (8)	S (4)	I (8)	I (8)	S (4)	S (4)	S (4)	S (1)	S (0.5)	S (0.5)	S (2)	I (8)	S (0.5)	S (0.5)	S (2)	S (2)	S (0.5)
Ciprofloxacin	0.004– 8	≤ 1	2	≥ 4	S (0.06)	S (0.06)	S (0.06)	S (0.06)	S (0.06)	S (0.06)	S (0.06)	S (0.03)	S (0.06)	S (0.06)	S (0.125)	S (0.25)	S (0.25)	S (0.06)	S (0.25)	S (0.125)	S (0.06)
Doxycycline	0.002– 4	≤ 4	8	≥ 16	S (0.06)	S (0.06)	S (0.25)	S (0.25)	S (0.125)	S (0.25)	S (0.25)	S (0.06)	S (0.03)	S (0.03)	S (0.25)	S (0.125)	S (0.06)	S (0.06)	S (0.125)	S (0.125)	S (0.03)
Rifampicin	0.004– 8	≤ 1	2	≥ 4	I (2)	S (0.5)	S (0.5)	I (2)	S (1)	S (1)	S (0.5)	S (1)	S (0.25)	S (0.06)	S (0.125)	S (0.125)	S (0.06)	S (0.125)	S (0.25)	S (0.25)	S (0.25)
Ceftriaxone	0.25–512	≤ 8	16– 32	≥ 64	I (32)	R (64)	I (32)	I (32)	R (64)	I (32)	R (512)	I (32)	I (32)	R (> 512)	I (32)	I (32)	R (> 512)	R (> 512)	I (32)	R (> 512)	R (> 512)
Cefotaxime	0.25–512	≤ 8	16– 32	≥ 64	I (32)	I (32)	I (16)	I (32)	I (32)	I (16)	R (128)	I (16)	I (16)	R (128)	I (32)	I (32)	R (128)	R (128)	I (32)	R (128)	R (128)
Trimethoprim	0.06–128	–	–	≥ 16	R (> 128)	R (> 128)	R (> 128)	R (> 128)	R (> 128)	R (> 128)	R (> 128)	R (> 128)	R (> 128)	R (> 128)	R (> 128)	R (> 128)	R (> 128)	R (> 128)	R (> 128)	R (> 128)	R (> 128)

All parameters were interpreted according to the NCCLS antimicrobial susceptibility standards for staphylococci. S, susceptible; I, intermediate; R, resistant. The breakpoints (μg/mL) for Staphylococcus spp. were used for linezolid, doxycicline and trimethoprim according to CLSI guidelines M100-S24 (2014), whereas for the other antimicrobials, the interpretative criteria for Bacillus spp. were used according to CLSI guidelines M45-A2 (2011). The MIC value (μg/mL) is shown in parentheses. S, susceptible; I, intermediate; R, resistant.

activity; thus, it has been associated as a virulence factor to septicemic infections (Oda et al., 2013). Also, the *CytK-2* protein is a hemolytic toxin, and it is able to form pores in planar lipid bilayers and to cause toxic effects on human intestinal cells (Fagerlund et al., 2004). Other enterotoxigenic factors are enterotoxin FM (*entFM*) and enterotoxin A (*entA*) genes. There is evidence that suggest *entFM* contributes to the severity of diarrheal illness (Castiaux et al., 2016). The phosphatidylcholine-preferring phospholipase and the sphingomyelinase constitute the hemolytic cereolysin AB complex; both play an important cytotoxic role in many infections, considering that they show the ability to hydrolyse membrane phospholipids (Titball, 1993). The physiological roles of the bacterial enzymes are not well understood, although it has been suggested that the phosphatidylinositol-specific phospholipase C (PI-PLC) are virulence factors in the human pathogens *Listeria monocytogenes* and *S. aureus* (Gässler et al., 1997). We also identified PlcR, a regulation protein, which is a well-known pleiotropic regulator of genes related to pathogenicity (Salamitou et al., 2000). Although the virulence factors associated with clinical non-gastrointestinal diseases are unclear, in our clinically isolates, we found the contemporary presence of genes encoding to *NHE*, *entA*, *entFM*, *sph*, *cerA*, *cerB*, *inhA*, *plcA*, and *plcB* virulence factors, in co-presence of the PlcR (Salamitou et al., 2000). Our findings show the high toxigenic potential of these bacteria. Interestingly, the isolates that in addition carry the genes that encoded enterotoxin hemolysin BL might be more virulent (Salamitou et al., 2000), whereas the absence of one of the virulence factors here described is not necessarily associated with a low pathogenicity power of *Bacillus* isolates. This assertion might be due to several reasons: in literature, a cytotoxic strain, *B. cereus* ATCC 10987, was reported that lacked the *HBL* operon, but produced a large amount of the *NHE* mRNA and exhibited a strong cytopathogenic activity in Vero cells (Lindbäck et al., 1999); the absence of one cytotoxic component may be compensated by the expression of other PlcR-regulated factors, which are still unknown. The gene *PlcR* plays an important role also in the biofilm formation (Ryu and Beuchat, 2005; Hsueh et al., 2006). The biofilm consists of a complex community and makes the *B. cereus* members group capable to colonize different environments (Majed et al., 2016). The isolates that possess *Plc*R may take advantage both in virulence genes regulation and in biofilm formation. Previous studies suggested that certain *B. cereus* strains were able to form different types of biofilms, either submerged, bottom-surface attached biofilms, floating pellicles, or pellicles attached to the side surfaces of the glass tubes. Different types of biofilms may require activities from different genetic determinants (Wijman et al., 2007; Caro-Astorga et al., 2015; Gao et al., 2015; Yan et al., 2017). In the closely related species *Bacillus subtilis*, an operon has been described, including three genes (*tasA*, *tapA*, *sipW*), which is required to form the biofilm (Candela et al., 2018). The transcription of these genes is promoted by *Sin*I and repressed by *SinR* (Kearns et al., 2005). Two orthologs of *tasA* have been described: the first is also named *tasA*, and it is found downstream of the signal peptidase gene *sipW*, in the *SinR*-regulated bicistronic operon *sipW-tasA* (Pflughoeft et al., 2011; Fagerlund et al., 2014; Caro-Astorga et al., 2015); the second is named *calY* and is located downstream from *sipW-tasA* (Caro-Astorga et al., 2015). In *B. cereus*, both *CalY* and *Tas* polymerize to form fibers in the matrix biofilm (Caro-Astorga et al., 2015). *TapA* contributes to the beginning and growth of the *tasA* fibers (Romero et al., 2014), but the strains that lack this gene contain all elements required for fiber assembly (Caro-Astorga et al., 2015). All of our isolates carried *calY*, *tasA*, and *sipW* genes. Although our isolate did not carry the gene *tapA*, which encodes for the accessory protein, the presence of the three genes mentioned above may be sufficient for the biofilm production. This statement is in accordance with a previous study where a similar condition was found (Caro-Astorga et al., 2015).

Another gene associated with biofilm formation in *B. cereus* is represented by global regulator *CodY* (Lindbäck et al., 2012). *CodY* gene resides in an operon with other genes, such as *xerC*, *clpY*, and *clpQ* (Slack et al., 1995). Altogether, these genes play a role in pellicle biofilm formation and swarming motility (Yan et al., 2017). Additionally, in *B. subtilis*, there are several *esp* genes (*epsA-O*) that are strongly expressed during biofilm formation (Vlamakis et al., 2013). We did not find *eps* gene cluster in our isolates, with few exceptions; a previous study noticed that in *B. cereus*, these genes did not appear to be important for pellicle formation, despite the important role described in equivalent product for *B. subtilis* (Gao et al., 2015). The *pur* gene cluster, including 11 genes, was required for purine biosynthesis (Vilain et al., 2009; Yan et al., 2017). Biofilm of several bacterial species, including *B. cereus*, has previously been shown to contain extracellular DNA as an integral component of extracellular polymeric substance (Vilain et al., 2009). In some bacteria, including *B. subtilis*, the alternative sigma factor σ^B, which is encoded by *sigB* gene, plays a role in stress conditions. *SigB* gives to the bacteria the ability to resist multiple stresses (van Schaik et al., 2004; Hecker et al., 2007). In the studied isolates, we investigated the presence of the determinants that may be related to biofilm formation, although we did not assess this ability *in vitro*. This aspect has a clinical importance, considering that the biofilm formed by pathogenic species is often associated with hospital-acquired infection (Lindbäck et al., 2012). Intriguingly, the majority of the virulence factors here identified appeared to be evenly distributed among *B. cereus* s.s. isolates, as well as strain close to type *B. thuringiensis* strains. This suggests that investigating the set of virulence factors regardless of right species identification could be more important to define the pathogenic power of strains belonging to *B. cereus* group.

Considering the emergence of antibiotic-resistant *B. cereus* strains, which may result in the failure of antibiotic treatment, it became highly relevant for public health to know the attitude of antibiotic resistance in *B. cereus*. In this study, we performed either *in silico* and *in vitro* analyses. Typically, *B. cereus* is resistant to penicillin G or other beta-lactam antibiotics (Citron and Appleman, 2006), and we found that all the isolates were resistant (penicillin G) and resistant or moderately resistant (i.e., ceftriaxone and cefotaxime) to beta-lactam antibiotics and carried genes related to this resistance. Similar to the results of other studies (Luna et al., 2007; Park et al., 2009; Raymond et al., 2010), our isolates showed susceptibility to ciprofloxacin, chloramphenicol, gentamicin, linezolid, and

doxycycline. In some isolates, we identified two genes (*ermC* and *mphB*) that confer resistance to macrolide drug family, but when we compared this result with phenotypic test, only for one isolate that the results agreed. All isolates carried genes associated with vancomycin resistance, but all isolates resulted susceptible to the phenotypic test. Additionally, all the isolates harbored the Fcyn-fosBx1 gene that probably confers resistance to fosfomycin, but we do not have phenotypic information, as there are no references on the interpretation. Similarly, we identified in two isolates the puB110 plasmid, which has been associated with kanamycin resistance, but no phenotypic information was available. On the contrary, the phenotypic test revealed that all isolates were resistant to trimethoprim, but in the genome of each isolate, we did not detect a specific genetic determinant that could clarify the resistance. This can be explained in many ways: there are several genetic mechanisms for resistance to a single antimicrobial agent; there are genetic mechanisms that can give rise to antibiotic resistance, i.e., mutations or acquisition of new genetic material (plasmids). There is insufficient knowledge about all genetic variations leading to reduced susceptibility for a given antimicrobial agent (Ellington et al., 2017). Thus, the AMR genomic data related are never static, as the genetic information is always moving. In conclusion, our AMR results underline the importance of combining the information detected by genome to those detected by phenotypic test and the necessity to update continually the databases used to this aim.

Taken together, this study suggests that the identification of *B. cereus* and of the several toxins produced by this bacterium should be considered essential to assess the risk for human health. However, the comprehensive risk characterization in clinical infections is difficult because of the underestimation of the risk, as the clinical laboratories do not necessarily complete species identification considering *B. cereus* as food-borne and/or environmental contaminants. Thus, we analyzed the molecular and genetic data in order to alert clinicians regarding the emerging threat that *B. cereus* can represent in clinical settings. Our results suggest that the analysis of WGS data, followed by appropriate data analysis strategies, could be a highly effective way to evaluate the pathogenic potential of *B. cereus*. The comprehensive molecular characterization of the isolates allowed identifying genetic diversity, as we identified different STs. Interestingly, we identified two members of *B. cereus* group, and the analyses reveal that strain close to *B. thuringiensis*, which lacks of cry plasmid, can carry similar virulence determinants as *B. cereus* s.s. Without specific information about the patients, for clinical treatment and for the human safety, it is essential to ensure an adequate antibiotic therapy. To this aim, our results showed that aminoglycosides, oxazolidinone, and lincosamide might be a good choice for treating *B. cereus* infections; on the opposite, penicillin and third-generation cephalosporins are not recommended. Certainly, more studies on clinical isolates are necessary to collect more information about pathogenicity of these strains with the aim to improve the genomic correlations that will help to identify the most pathogenic strains and take prompt action.

DATA AVAILABILITY STATEMENT

The datasets GENERATED for this study can be found in NCBI BioProject PRJNA673333.

ETHICS STATEMENT

Ethical approval was not provided for this study on human participants because the samples were collected during the last 3 years. After diagnostic routine, strains resulted from the biological material were stored to be processed for further analysis. No personal data or any other information than the type of material and the result of routine microbiology analysis were collected from each specimen, inhibiting any correlations of these fully anonymized samples to the respective patients. Thus, according to national regulations and the institutional rules for Good Scientific Practice, the requirement for submission to an ethical committee and for obtaining patients' informed consent was waived. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

AP and MM provided substantial contributions to the conception and design of the work and the acquisition, analysis, and interpretation of data and drafting the work. AB, LC, and LD performed the acquisition, analysis, and interpretation of data and drafting the work. VM, GP, and DL performed the acquisition, analysis, and interpretation of data and critical revising of the work for important intellectual content. All authors have approved the final version to be published. All authors are to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

This research was founded by Ministero della Salute – Ricerca corrente IZS PB 04/16 RC and by Consiglio Nazionale delle Ricerche – Flagship project Interomics.

ACKNOWLEDGMENTS

We thank Anna Giannico and Donato Ridolfi for technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.599524/full#supplementary-material

REFERENCES

Ash, C., Farrow, J. A., Dorsch, M., Stackebrandt, E., and Collins, M. D. (1991). Comparative analysis of *Bacillus anthracis*, *Bacillus cereus*, and related species on the basis of reverse transcriptase sequencing of 16S rRNA. *Int. J. Syst. Bacteriol.* 41, 343–346. doi: 10.1099/00207713-41-3-343

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

Beecher, D. J., and Wong, A. C. (2000). Tripartite haemolysin BL: isolation and characterization of two distinct homologous sets of components from a single *Bacillus cereus* isolate. *Microbiol. Read. Engl.* 146(Pt 6), 1371–1380. doi: 10.1099/00221287-146-6-1371

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

Bottone, E. J. (2010). *Bacillus cereus*, a volatile human pathogen. *Clin. Microbiol. Rev.* 23, 382–398. doi: 10.1128/CMR.00073-09

Bryce, E. A., Smith, J. A., Tweeddale, M., Andruschak, B. J., and Maxwell, M. R. (1993). Dissemination of *Bacillus cereus* in an intensive care unit. *Infect. Control Hosp. Epidemiol.* 14, 459–462.

Caamaño-Antelo, S., Fernández-No, I. C., Böhme, K., Ezzat-Alnakip, M., Quintela-Baluja, M., Barros-Velázquez, J., et al. (2015). Genetic discrimination of foodborne pathogenic and spoilage *Bacillus* spp. based on three housekeeping genes. *Food Microbiol.* 46, 288–298. doi: 10.1016/j.fm.2014.08.013

Candela, T., Fagerlund, A., Buisson, C., Gilois, N., Kolstø, A.-B., Økstad, O.-A., et al. (2018). CalY is a major virulence factor and a biofilm matrix protein. *Mol. Microbiol.* 111, 1416–1429. doi: 10.1111/mmi.14184

Candelon, B., Guilloux, K., Ehrlich, S. D., and Sorokin, A. (2004). Two distinct types of rRNA operons in the *Bacillus cereus* group. *Microbiol. Read. Engl.* 150, 601–611. doi: 10.1099/mic.0.26870-0

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

Caro-Astorga, J., Pérez-García, A., de Vicente, A., and Romero, D. (2015). A genomic region involved in the formation of adhesin fibers in *Bacillus cereus* biofilms. *Front. Microbiol.* 5:745. doi: 10.3389/fmicb.2014.00745

Carroll, L. M., Cheng, R. A., and Kovac, J. (2020). No assembly required: using BTyper3 to assess the congruency of a proposed taxonomic framework for the *Bacillus cereus* group with historical typing methods. *Front. Microbiol.* 11:580691. doi: 10.3389/fmicb.2020.580691

Carroll, L. M., Kovac, J., Miller, R. A., and Wiedmann, M. (2017). Rapid, High-throughput identification of anthrax-causing and emetic *Bacillus cereus* group genome assemblies via BTyper, a computational tool for virulence-based classification of *Bacillus cereus* group isolates by using nucleotide sequencing data. *Appl. Environ. Microbiol.* 83:e1096-17. doi: 10.1128/AEM.01096-17

Castiaux, V., Laloux, L., Schneider, Y.-J., and Mahillon, J. (2016). Screening of cytotoxic B. cereus on differentiated caco-2 cells and in co-culture with mucus-secreting (HT29-MTX) cells. *Toxins* 8:320. doi: 10.3390/toxins8110320

Citron, D. M., and Appleman, M. D. (2006). In vitro activities of Daptomycin, Ciprofloxacin, and other antimicrobial agents against the cells and spores of clinical isolates of *Bacillus* Species. *J. Clin. Microbiol.* 44, 3814–3818. doi: 10.1128/JCM.00881-06

Damgaard, P. H., Granum, P. E., Bresciani, J., Torregrossa, M. V., Eilenberg, J., and Valentino, L. (1997). Characterization of *Bacillus thuringiensis* isolated from infections in burn wounds. *FEMS Immunol. Med. Microbiol.* 18, 47–53. doi: 10.1111/j.1574-695X.1997.tb01026.x

Dohmae, S., Okubo, T., Higuchi, W., Takano, T., Isobe, H., Baranovich, T., et al. (2008). *Bacillus cereus* nosocomial infection from reused towels in Japan. *J. Hosp. Infect.* 69, 361–367. doi: 10.1016/j.jhin.2008.04.014

EFSA Panel on Biological Hazards, (2009). Safety and efficacy of the product Toyocerin® (*Bacillus cereus* var. toyoi) as feed additive for rabbit breeding does - scientific opinion of the panel on additives and products or substances used in animal feed. *EFSA J.* 7:913. doi: 10.2903/j.efsa.2009.913

EFSA Panel on Biological Hazards. (2016). Risks for public health related to the presence of *Bacillus cereus* and other *Bacillus* spp. including *Bacillus thuringiensis* in foodstuffs. *EFSA J.* 14:4524. doi: 10.2903/j.efsa.2016.4524

Ellington, M. J., Ekelund, O., Aarestrup, F. M., Canton, R., Doumith, M., Giske, C., et al. (2017). The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* 23, 2–22. doi: 10.1016/j.cmi.2016.11.012

Fagerlund, A., Dubois, T., Økstad, O.-A., Verplaetse, E., Gilois, N., Bennaceur, I., et al. (2014). SinR controls enterotoxin expression in *Bacillus thuringiensis* biofilms. *PLoS One* 9:e87532. doi: 10.1371/journal.pone.0087532

Fagerlund, A., Ween, O., Lund, T., Hardy, S. P., and Granum, P. E. (2004). Genetic and functional analysis of the cytK family of genes in *Bacillus cereus*. *Microbiol. Read. Engl.* 150, 2689–2697. doi: 10.1099/mic.0.26975-0

Fayad, N., Kallassy Awad, M., and Mahillon, J. (2019). Diversity of *Bacillus cereus* sensu lato mobilome. *BMC Genomics* 20:436. doi: 10.1186/s12864-019-5764-4

Feldgarden, M., Brover, V., Haft, D. H., Prasad, A. B., Slotta, D. J., Tolstoy, I., et al. (2019). Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob. Agents Chemother.* 63:e483-19. doi: 10.1128/AAC.00483-19

Fricker, M., Messelhäusser, U., Busch, U., Scherer, S., and Ehling-Schulz, M. (2007). Diagnostic real-time PCR assays for the detection of emetic *Bacillus cereus* strains in foods and recent food-borne outbreaks. *Appl. Environ. Microbiol.* 73, 1892–1898. doi: 10.1128/AEM.02219-2216

Gao, T., Foulston, L., Chai, Y., Wang, Q., and Losick, R. (2015). Alternative modes of biofilm formation by plant-associated *Bacillus cereus*. *MicrobiologyOpen* 4, 452–464. doi: 10.1002/mbo3.251

Gässler, C. S., Ryan, M., Liu, T., Griffith, O. H., and Heinz, D. W. (1997). Probing the roles of active site residues in phosphatidylinositol-specific phospholipase C from *Bacillus cereus* by site-directed mutagenesis. *Biochemistry* 36, 12802–12813. doi: 10.1021/bi971102d

Ghelardi, E., Celandroni, F., Salvetti, S., Fiscarelli, E., and Senesi, S. (2007). *Bacillus thuringiensis* pulmonary infection: critical role for bacterial membrane-damaging toxins and host neutrophils. *Microbes Infect.* 9, 591–598. doi: 10.1016/j.micinf.2007.02.001

Goldstein, B., and Abrutyn, E. (1985). Pseudo-outbreak of *Bacillus* species: related to fibreoptic bronchoscopy. *J. Hosp. Infect.* 6, 194–200.

González-Zorn, B., Domínguez-Bernal, G., Suárez, M., Ripio, M. T., Vega, Y., Novella, S., et al. (1999). The smcL gene of Listeria ivanovii encodes a sphingomyelinase C that mediates bacterial escape from the phagocytic vacuole. *Mol. Microbiol.* 33, 510–523.

Granum, P. E., O'Sullivan, K., and Lund, T. (1999). The sequence of the non-haemolytic enterotoxin operon from *Bacillus cereus*. *FEMS Microbiol. Lett.* 177, 225–229. doi: 10.1111/j.1574-6968.1999.tb13736.x

Guillemet, E., Cadot, C., Tran, S.-L., Guinebretière, M.-H., Lereclus, D., and Ramarao, N. (2010). The InhA metalloproteases of *Bacillus cereus* contribute concomitantly to virulence. *J. Bacteriol.* 192, 286–294. doi: 10.1128/JB.00264-09

Guinebretière, M.-H., Auger, S., Galleron, N., Contzen, M., De Sarrau, B., De Buyser, M.-L., et al. (2013). *Bacillus cytotoxicus* sp. nov. is a novel thermotolerant species of the *Bacillus cereus* Group occasionally associated with food poisoning. *Int. J. Syst. Evol. Microbiol.* 63, 31–40. doi: 10.1099/ijs.0.030627-0

Guinebretière, M.-H., Thompson, F. L., Sorokin, A., Normand, P., Dawyndt, P., Ehling-Schulz, M., et al. (2008). Ecological diversification in the *Bacillus cereus* Group. *Environ. Microbiol.* 10, 851–865. doi: 10.1111/j.1462-2920.2007.01495.x

Gupta, S. K., Padmanabhan, B. R., Diene, S. M., Lopez-Rojas, R., Kempf, M., Landraud, L., et al. (2014). ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob. Agents Chemother.* 58, 212–220. doi: 10.1128/AAC.01310-13

Hansen, B. M., and Hendriksen, N. B. (2001). Detection of enterotoxic *Bacillus cereus* and *Bacillus thuringiensis* strains by PCR analysis. *Appl. Environ. Microbiol.* 67, 185–189. doi: 10.1128/AEM.67.1.185-189.2001

Hecker, M., Pané-Farré, J., and Völker, U. (2007). SigB-dependent general stress response in Bacillus subtilis and related gram-positive bacteria. *Annu. Rev. Microbiol.* 61, 215–236. doi: 10.1146/annurev.micro.61.080706.093445

Helgason, E., Tourasse, N. J., Meisal, R., Caugant, D. A., and Kolstø, A.-B. (2004). Multilocus sequence typing scheme for bacteria of the *Bacillus cereus* group. *Appl. Environ. Microbiol.* 70, 191–201.

Hernaiz, C., Picardo, A., Alos, J. I., and Gomez-Garces, J. L. (2003). Nosocomial bacteremia and catheter infection by *Bacillus cereus* in an immunocompetent patient. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* 9, 973–975.

Hilliard, N. J., Schelonka, R. L., and Waites, K. B. (2003). *Bacillus cereus* bacteremia in a preterm neonate. *J. Clin. Microbiol.* 41, 3441–3444.

Hoffmaster, A. R., Hill, K. K., Gee, J. E., Marston, C. K., De, B. K., Popovic, T., et al. (2006). Characterization of *Bacillus cereus* isolates associated with fatal pneumonias: strains are closely related to *Bacillus anthracis* and harbor *B. anthracis* virulence genes. *J. Clin. Microbiol.* 44, 3352–3360. doi: 10.1128/JCM.00561-06

Hsieh, Y. M., Sheu, S. J., Chen, Y. L., and Tsen, H. Y. (1999). Enterotoxigenic profiles and polymerase chain reaction detection of *Bacillus cereus* group cells and *B. cereus* strains from foods and food-borne outbreaks. *J. Appl. Microbiol.* 87, 481–490.

Hsueh, P. R., Teng, L. J., Yang, P. C., Pan, H. L., Ho, S. W., and Luh, K. T. (1999). Nosocomial pseudoepidemic caused by *Bacillus cereus* traced to contaminated ethyl alcohol from a liquor factory. *J. Clin. Microbiol.* 37, 2280–2284.

Hsueh, Y.-H., Somers, E. B., Lereclus, D., and Wong, A. C. L. (2006). Biofilm formation by *Bacillus cereus* is influenced by PlcR, a pleiotropic regulator. *Appl. Environ. Microbiol.* 72, 5089–5092. doi: 10.1128/AEM.00573-06

Jackson, S. G., Goodbrand, R. B., Ahmed, R., and Kasatiya, S. (1995). *Bacillus cereus* and *Bacillus thuringiensis* isolated in a gastroenteritis outbreak investigation. *Lett. Appl. Microbiol.* 21, 103–105.

Jensen, G. B., Hansen, B. M., Eilenberg, J., and Mahillon, J. (2003). The hidden lifestyles of *Bacillus cereus* and relatives. *Environ. Microbiol.* 5, 631–640.

Jia, B., Raphenya, A. R., Alcock, B., Waglechner, N., Guo, P., Tsang, K. K., et al. (2017). CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 45, D566–D573. doi: 10.1093/nar/gkw1004

Johnson, K. M., Nelson, C. L., and Busta, F. F. (1983). Influence of temperature on germination and growth of spores of emetic and diarrheal strains of *Bacillus cereus* in a broth medium and in rice. *J. Food Sci.* 48, 286–287. doi: 10.1111/j.1365-2621.1983.tb14853.x

Kato, K., Matsumura, Y., Yamamoto, M., Nagao, M., Ito, Y., Takakura, S., et al. (2014). Seasonal trend and clinical presentation of *Bacillus cereus* bloodstream infection: association with summer and indwelling catheter. *Eur. J. Clin. Microbiol. Infect. Dis. Off. Publ. Eur. Soc. Clin. Microbiol.* 33, 1371–1379. doi: 10.1007/s10096-014-2083-1

Kearns, D. B., Chu, F., Branda, S. S., Kolter, R., and Losick, R. (2005). A master regulator for biofilm formation by Bacillus subtilis. *Mol. Microbiol.* 55, 739–749. doi: 10.1111/j.1365-2958.2004.04440.x

Kreig, A., and Lysenko, O. (1979). [Toxins and enzymes of several species of *Bacillus*, especially of the *B. cereus*-thuringiensis group (author's transl)]. *Zentralbl. Bakteriol. Naturwiss.* 134, 70–88.

Kuroki, R., Kawakami, K., Qin, L., Kaji, C., Watanabe, K., Kimura, Y., et al. (2009). Nosocomial bacteremia caused by biofilm-forming *Bacillus cereus* and *Bacillus thuringiensis*. *Intern. Med.* 48, 791–796.

Lindbäck, T., Mols, M., Basset, C., Granum, P. E., Kuipers, O. P., and Kovács, ÁT. (2012). CodY, a pleiotropic regulator, influences multicellular behaviour and efficient production of virulence factors in *Bacillus cereus*. *Environ. Microbiol.* 14, 2233–2246. doi: 10.1111/j.1462-2920.2012.02766.x

Lindbäck, T., Økstad, O. A., Rishovd, A.-L., and Kolstø, A.-B. (1999). Insertional inactivation of hblC encoding the L2 component of *Bacillus cereus* ATCC 14579 haemolysin BL strongly reduces enterotoxigenic activity, but not the haemolytic activity against human erythrocytes. *Microbiology* 145, 3139–3146. doi: 10.1099/00221287-145-11-3139

Liu, X., Wang, L., Han, M., Xue, Q., Zhang, G., Gao, J., et al. (2020). *Bacillus fungorum* sp. nov., a bacterium isolated from spent mushroom substrate. *Int. J. Syst. Evol. Microbiol.* 70, 1457–1462. doi: 10.1099/ijsem.0.003673

Liu, Y., Du, J., Lai, Q., Zeng, R., Ye, D., Xu, J., et al. (2017). Proposal of nine novel species of the *Bacillus cereus* group. *Int. J. Syst. Evol. Microbiol.* 67, 2499–2508. doi: 10.1099/ijsem.0.001821

Liu, Y., Lai, Q., Göker, M., Meier-Kolthoff, J. P., Wang, M., Sun, Y., et al. (2015). Genomic insights into the taxonomic status of the *Bacillus cereus* group. *Sci. Rep.* 5:14082. doi: 10.1038/srep14082

Luna, V. A., King, D. S., Gulledge, J., Cannons, A. C., Amuso, P. T., and Cattani, J. (2007). Susceptibility of *Bacillus anthracis*, *Bacillus cereus*, *Bacillus mycoides*, *Bacillus pseudomycoides* and *Bacillus thuringiensis* to 24 antimicrobials using Sensititre(R) automated microbroth dilution and Etest(R) agar gradient diffusion methods. *J. Antimicrob. Chemother.* 60, 555–567. doi: 10.1093/jac/dkm213

Lund, T., De Buyser, M. L., and Granum, P. E. (2000). A new cytotoxin from *Bacillus cereus* that may cause necrotic enteritis. *Mol. Microbiol.* 38, 254–261.

Majed, R., Faille, C., Kallassy, M., and Gohar, M. (2016). *Bacillus cereus* biofilms-same, only different. *Front. Microbiol.* 7:1054. doi: 10.3389/fmicb.2016.01054

Manzulli, V., Fasanella, A., Parisi, A., Serrecchia, L., Donatiello, A., Rondinone, V., et al. (2019). Evaluation of in vitro antimicrobial susceptibility of *Bacillus anthracis* strains isolated during anthrax outbreaks in Italy from 1984 to 2017. *J. Vet. Sci.* 20, 58–62. doi: 10.4142/jvs.2019.20.1.58

Miller, R. A., Beno, S. M., Kent, D. J., Carroll, L. M., Martin, N. H., Boor, K. J., et al. (2016). Bacillus wiedmannii sp. nov., a psychrotolerant and cytotoxic *Bacillus cereus* group species isolated from dairy foods and dairy environments. *Int. J. Syst. Evol. Microbiol.* 66, 4744–4753. doi: 10.1099/ijsem.0.001421

Oda, M., Fujita, A., Okui, K., Miyamoto, K., Shibutani, M., Takagishi, T., et al. (2013). *Bacillus cereus* sphingomyelinase recognizes ganglioside GM3. *Biochem. Biophys. Res. Commun.* 431, 164–168. doi: 10.1016/j.bbrc.2013.01.002

Oda, M., Hashimoto, M., Takahashi, M., Ohmae, Y., Seike, S., Kato, R., et al. (2012). Role of sphingomyelinase in infectious diseases caused by *Bacillus cereus*. *PLoS One* 7:e38054. doi: 10.1371/journal.pone.0038054

Oh, S.-Y., Budzik, J. M., Garufi, G., and Schneewind, O. (2011). Two capsular polysaccharides enable *Bacillus cereus* G9241 to cause anthrax-like disease. *Mol. Microbiol.* 80, 455–470. doi: 10.1111/j.1365-2958.2011.07582.x

Orrett, F. A. (2000). Fatal *Bacillus cereus* bacteremia in a patient with diabetes. *J. Natl. Med. Assoc.* 92, 206–208.

Owusu-Kwarteng, J., Wuni, A., Akabanda, F., Tano-Debrah, K., and Jespersen, L. (2017). Prevalence, virulence factor genes and antibiotic resistance of *Bacillus cereus* sensu lato isolated from dairy farms and traditional dairy products. *BMC Microbiol.* 17:65. doi: 10.1186/s12866-017-0975-9

Park, Y.-B., Kim, J.-B., Shin, S.-W., Kim, J.-C., Cho, S.-H., Lee, B.-K., et al. (2009). Prevalence, genetic diversity, and antibiotic susceptibility of *Bacillus cereus* strains isolated from rice and cereals collected in Korea. *J. Food Prot.* 72, 612–617. doi: 10.4315/0362-028X-72.3.612

Pflughoeft, K. J., Sumby, P., and Koehler, T. M. (2011). *Bacillus anthracis* sin locus and regulation of secreted proteases. *J. Bacteriol.* 193, 631–639. doi: 10.1128/JB.01083-10

Priest, F. G., Barker, M., Baillie, L. W. J., Holmes, E. C., and Maiden, M. C. J. (2004). Population structure and evolution of the *Bacillus cereus* group. *J. Bacteriol.* 186, 7959–7970. doi: 10.1128/JB.186.23.7959-7970.2004

Ramarao, N., and Lereclus, D. (2005). The InhA1 metalloprotease allows spores of the B. cereus group to escape macrophages. *Cell. Microbiol.* 7, 1357–1364. doi: 10.1111/j.1462-5822.2005.00562.x

Raymond, B., and Federici, B. A. (2017). In defence of *Bacillus thuringiensis*, the safest and most successful microbial insecticide available to humanity—a response to EFSA. *FEMS Microbiol. Ecol.* 93:fix084.

Raymond, B., Wyres, K. L., Sheppard, S. K., Ellis, R. J., and Bonsall, M. B. (2010). Environmental factors determining the epidemiology and population genetic structure of the *Bacillus cereus* group in the field. *PLoS Pathog.* 6:e1000905. doi: 10.1371/journal.ppat.1000905

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

Romero, D., Vlamakis, H., Losick, R., and Kolter, R. (2014). Functional analysis of the accessory protein TapA in *Bacillus subtilis* amyloid fiber assembly. *J. Bacteriol.* 196, 1505–1513. doi: 10.1128/JB.01363-13

Ryu, J.-H., and Beuchat, L. R. (2005). Biofilm formation and sporulation by *Bacillus cereus* on a stainless steel surface and subsequent resistance of vegetative cells and spores to chlorine, chlorine dioxide, and a peroxyacetic acid-based sanitizer. *J. Food Prot.* 68, 2614–2622.

Salamitou, S., Ramisse, F., Brehélin, M., Bourguet, D., Gilois, N., Gominet, M., et al. (2000). The plcR regulon is involved in the opportunistic properties of *Bacillus thuringiensis* and *Bacillus cereus* in mice and insects. *Microbiol. Read. Engl.* 146(Pt 11), 2825–2832. doi: 10.1099/00221287-146-11-2825

Sarker, S. D., Nahar, L., and Kumarasamy, Y. (2007). Microtitre plate-based antibacterial assay incorporating resazurin as an indicator of cell growth, and its application in the in vitro antibacterial screening of phytochemicals. *Methods* 42, 321–324. doi: 10.1016/j.ymeth.2007.01.006

Schmid, D., Rademacher, C., Kanitz, E. E., Frenzel, E., Simons, E., Allerberger, F., et al. (2016). Elucidation of enterotoxigenic *Bacillus cereus* outbreaks in Austria by complementary epidemiological and microbiological investigations, 2013. *Int. J. Food Microbiol.* 232, 80–86. doi: 10.1016/j.ijfoodmicro.2016.05.011

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

Sergeev, N., Distler, M., Vargas, M., Chizhikov, V., Herold, K. E., and Rasooly, A. (2006). Microarray analysis of *Bacillus cereus* group virulence factors. *J. Microbiol. Methods* 65, 488–502. doi: 10.1016/j.mimet.2005.09.013

Slack, F. J., Serror, P., Joyce, E., and Sonenshein, A. L. (1995). A gene required for nutritional repression of the Bacillus subtilis dipeptide permease operon. *Mol. Microbiol.* 15, 689–702.

Sorokin, A., Candelon, B., Guilloux, K., Galleron, N., Wackerow-Kouzova, N., Ehrlich, S. D., et al. (2006). Multiple-locus sequence typing analysis of *Bacillus cereus* and *Bacillus thuringiensis* reveals separate clustering and a distinct population structure of psychrotrophic strains. *Appl. Environ. Microbiol.* 72, 1569–1578. doi: 10.1128/AEM.72.2.1569-1578.2006

Titball, R. W. (1993). Bacterial phospholipases C. *Microbiol. Rev.* 57, 347–366.

Tourasse, N. J., Helgason, E., Økstad, O. A., Hegna, I. K., and Kolstø, A.-B. (2006). The *Bacillus cereus* group: novel aspects of population structure and genome dynamics. *J. Appl. Microbiol.* 101, 579–593. doi: 10.1111/j.1365-2672.2006.03087.x

van Schaik, W., Tempelaars, M. H., Wouters, J. A., de Vos, W. M., and Abee, T. (2004). The alternative sigma factor σB of *Bacillus cereus*: response to stress and role in heat adaptation. *J. Bacteriol.* 186, 316–325. doi: 10.1128/JB.186.2.316-325.2004

Vilain, S., Pretorius, J. M., Theron, J., and Brözel, V. S. (2009). DNA as an Adhesin: *Bacillus cereus* requires extracellular DNA to form biofilms. *Appl. Environ. Microbiol.* 75, 2861–2868. doi: 10.1128/AEM.01317-08

Vlamakis, H., Chai, Y., Beauregard, P., Losick, R., and Kolter, R. (2013). Sticking together: building a biofilm the Bacillus subtilis way. *Nat. Rev. Microbiol.* 11, 157–168. doi: 10.1038/nrmicro2960

Warda, A. K., Siezen, R. J., Boekhorst, J., Wells-Bennik, M. H. J., de Jong, A., Kuipers, O. P., et al. (2016). Linking *Bacillus cereus* genotypes and carbohydrate utilization capacity. *PLoS One* 11:e0156796. doi: 10.1371/journal.pone.0156796

Weinstein, M. P. (2018). *M100-Performance Standards for Antimicrobial Susceptibility Testing*, 28th Edn. Wayne, PA: Clinical and Laboratory Standards Institute.

Wijman, J. G. E., de Leeuw, P. P. L. A., Moezelaar, R., Zwietering, M. H., and Abee, T. (2007). Air-liquid interface biofilms of *Bacillus cereus*: formation, sporulation, and dispersion. *Appl. Env. Microbiol.* 73, 1481–1488. doi: 10.1128/AEM.01781-06

Yan, F., Yu, Y., Gozzi, K., Chen, Y., Guo, J., and Chai, Y. (2017). Genome-wide investigation of biofilm formation in *Bacillus cereus*. *Appl. Environ. Microbiol.* 83:e561-17. doi: 10.1128/AEM.00561-17

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

Zheng, J., Gao, Q., Liu, L., Liu, H., Wang, Y., Peng, D., et al. (2017). Comparative genomics of *Bacillus thuringiensis* reveals a path to specialized exploitation of multiple invertebrate hosts. *mBio* 8:e822-17. doi: 10.1128/mBio.00822-17

Zhu, L., Peng, D., Wang, Y., Ye, W., Zheng, J., Zhao, C., et al. (2015). Genomic and transcriptomic insights into the efficient entomopathogenicity of *Bacillus thuringiensis*. *Sci. Rep.* 5:14129. doi: 10.1038/srep14129

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

frontiers in Microbiology

ORIGINAL RESEARCH
published: 18 January 2021
doi: 10.3389/fmicb.2020.617651

Genome-Based Analysis of a Sequence Type 1049 Hypervirulent *Klebsiella pneumoniae* Causing Bacteremic Neck Abscess

Peng Lan[1,2,3†], Dongdong Zhao[1†], Jiong Gu[4], Qiucheng Shi[2‡], Rushuang Yan[2,3], Yan Jiang[2], Jiancang Zhou[3] and Yunsong Yu[1,2*]*

[1] Department of Infectious Diseases, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China, [2] Key Laboratory of Microbial Technology and Bioinformatics of Zhejiang Province, Hangzhou, China, [3] Department of Critical Care Medicine, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China, [4] Department of Infectious Diseases, The First Hospital of Jiaxing, Zhejiang, China

Hypervirulent *Klebsiella pneumoniae* (hvKP) has raised grave concerns in recent years and can cause severe infections with diverse anatomic locations including liver abscess, meningitis, and endophthalmitis. However, there is limited data about neck abscess caused by hvKP. A *K. pneumoniae* strain Kp_whw was isolated from neck abscess. We characterized the genetic background, virulence determinates of the strain by genomic analysis and dertermined the virulence level by serum resistance assay. Kp_whw belonged to sequence type (ST) 1049 K locus (KL) 5. Kp_whw showed hypermucoviscosity phenotype and was resistant to ampicillin but susceptible to the majority of the other antimicrobial agents. A pLVPK-like virulence plasmid and a chromosomal ICEKp5-like mobile genetic element were carried by Kp_whw, resulting in the risk of dissemination of hypervirulence. The strain exhibited relative higher level of core genome allelic diversity than accessory genome profile, in comparison to hvKP of K1/K2 serotype. Kp_whw was finally demonstrated as virulent as the ST23 K1 serotype hvKP strain NTUH-K2044 *in vitro*. In conclusion, this work elaborates the genetic background of a clinical hvKP strain with an uncommon ST, reinforcing our understanding of virulence mechanisms of hvKP.

Keywords: hypervirulent, *Klebsiella pneumoniae*, neck abscess, ST1049, comparative genomic analysis

OPEN ACCESS

Edited by:
Yang Wang,
China Agricultural University, China

Reviewed by:
Fupin Hu,
Fudan University, China
Haijian Zhou,
National Institute for Communicable Disease Control and Prevention (China CDC), China

****Correspondence:***
Yunsong Yu
yvys119@zju.edu.cn
Jiancang Zhou
jiancangzhou@zju.edu.cn

† These authors have contributed equally to this work

‡ Present address:
Qiucheng Shi,
Department of Clinical Laboratory, The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Microbiology

Received: *15 October 2020*
Accepted: *24 December 2020*
Published: *18 January 2021*

Citation:
Lan P, Zhao D, Gu J, Shi Q, Yan R, Jiang Y, Zhou J and Yu Y (2021) Genome-Based Analysis of a Sequence Type 1049 Hypervirulent Klebsiella pneumoniae Causing Bacteremic Neck Abscess. Front. Microbiol. 11:617651. doi: 10.3389/fmicb.2020.617651

INTRODUCTION

Hypervirulent *Klebsiella pneumoniae* (hvKP) has raised grave concerns in recent years due to its extraordinary invasiveness and hypervirulence, resulting in considerable risks for morbidity and mortality (Shon et al., 2013). Asia is the epidemic area for hvKP. K1/K2 accounted for 9.8% of all *K. pneumoniae* isolates in stools from healthy Chinese individuals in Asia countries (Lin et al., 2012). In recent years, hvKP has been increasingly reported in other continents. The estimated prevalence of hvKP in Canada and the United States was 8.2 and 6.3%, respectively (Peirano et al., 2013; Chou et al., 2016). According to different areas, the mortality of hvKP infections ranged from 29.2 to 55.1% (Lin Y.T. et al., 2010; Rafat et al., 2018; Namikawa et al., 2019). hvKp infection is often manifested as liver abscess (Liu et al., 1986; Siu et al., 2012; Joob and Wiwanitkit, 2017).

However, extrahepatic infections have been described for variable anatomic locations including endoophthalmitis (Liu et al., 1986), osteomyelitis (Prokesch et al., 2016) and meningitis (Melot et al., 2016). Originated in the Asian Pacific Rim in 1980s, hvKP has disseminated worldwide.

A number of virulence factors contribute to the pathogenicity of hvKP including lipopolysaccharide (LPS), capsule, fimbriae, siderophores, and virulence plasmids (Paczosa and Mecsas, 2016; Russo and Marr, 2019). Traditionally, hypermucoviscosity, resulted by over expression of capsular polysaccharides, was considered as a critical characteristic for hypervirulence (Catalan-Najera et al., 2017), as the mucoviscous shield significantly increased the resistance to immunological recognition and killing from the host (Lin et al., 2006; Lin J.C. et al., 2010). Recently, virulence plasmids were proved to have potential to transmit virulence genes among *K. pneumoniae* strains (Yang et al., 2019). Virulence plasmids encode two siderophores, aerobactin, and salmochelin, and RmpA (regulator of the mucoid phenotype). The virulence plasmid pK2044 harbored by K1 serotype strain NTUH-K2044 and pLVPK harbored by K2 serotype strain CG43 were well characterized and were reported associated with invasive syndrome. Besides, virulence genes located in chromosome can also be transmitted via integrative and conjugative element (ICE) (Lam et al., 2018). Virulence plasmids and ICEKps both harbor virulence loci encoding variable virulence factors and are involved in horizontal gene transfer (HGT), conferring the global dissemination of hypervirulence.

Clonal complex 23 (CC23) and some selected serotypes (e.g., K1 and K2) are commonly deemed as hypervirulent clones (Struve et al., 2015). Most of hvKP strains of K1 serotype belong to CC23, while K2 serotype strains are genetically more diverse and belong to multiple distinct multilocus sequencing types (MLSTs) (Struve et al., 2015). More than a half of severe *K. pneumoniae* infections such as bacteremia, liver abscess, and invasive extrahepatic infections are caused by K1/K2 serotype strains (Lin Y.T. et al., 2010). In addition, strains of serotype K5, K20, and K54 are also associated with hypervirulence phenotype (Wyres et al., 2020).

Clinically, hvKP usually causes liver abscess with extrahepatic complications including necrotizing fasciitis, bloodstream infection, meningitis, and endophthalmitis (Siu et al., 2012). Patients infected with hvKP are frequently relative healthy and immunosufficient. However, diabetes is a critical risk factor for hvKP infection, partly due to the suppression of the innate immune system (Alba-Loureiro et al., 2007; Hodgson et al., 2015).

Here, we described a potential hypervirulent *K. pneumoniae* strain belonging to ST1049, an uncommon ST, isolated from neck abscess. *K. pneumoniae* of ST1049 was previously reported to cause liver abscess and meningitis, and resulting in poor clinical outcomes (Ku et al., 2017; Zhang et al., 2019). However, no genetic evidence was showed that this infrequent ST was a hypervirulent clone. Hence, we will characterize the strain based on genetic background and virulence profile and thus to refresh our knowledge about hvKP pathogenicity.

MATERIALS AND METHODS

Isolations

The strain was isolated from a male patient diagnosed with neck abscess in 2018. It was identified to the species level via matrix-assisted laser desorption/ionization mass spectrometry and named as Kp_whw. The antimicrobial susceptibilities of the strain were determined by a VITEK-2 compact system and interpreted according to the M100-S26 guideline established by Clinical and Laboratory Standards Institute (CLSI). Since the clinical characteristics were extracted from the electronic record system and were de-identified, informed consent was waived. In addition, 92 hvKP strains of K1/K2 serotype from our previous study (Lan et al., 2020) plus NTUH-K2044 (Wu et al., 2009) and hvKP1 (Russo and Gill, 2013) were employed for comparative genomic analysis.

Whole Genome Sequencing

The strain was cultured to the mid-logarithmic phase in LB broth at 37°C. Genomic DNA was extracted using the QIAamp DNA Minikit (QIAGEN, Hilden, Germany) and was further purified using the PowerClean DNA cleanup kit (Mo Bio Laboratories, Carlsbad, United States), following the manufacturer's recommendations. The genome was sequenced on an Illumina HiSeq X Ten platform (Illumina, San Diego, United States) using a paired-end 2 × 150–base pair protocol by Tianke Company (Hangzhou, China). Derived short reads were *de novo* assembled using CLC Genomics Workbench 9.5.1 software. The sequences assemblies reported in this paper have been deposited in the European Nucleotide Archive database (accession nos. PRJEB38367 and PRJEB34922). The NCBI accession number for NTUH-K2044 and hvKP1 are AP006725 and AOIZ00000000.

Typing and Detection of Virulence Determinants

The virulence genes and *wzi* (a part of K-locus) alleles were identified using Institut Pasteur[1]. Multilocus sequence typing (MLST) analysis for sequence types of *K. pneumoniae* used Institut Pasteur MLST[2]. For MLST, 7 target genes (*gapA*, *infB*, *mdh*, *pgi*, *phoE*, *rpoB*, and *tonB*) were analyzed. To detect the virulence plasmid, the contigs of strain KP_whw were aligned with well-characterized virulence plasmid pLVPK (accession no. AY378100 in GenBank) using the BLAST Ring Image Generator (BRIG, version 0.95) (Alikhan et al., 2011). Putative ICE was identified by ICEberg[3] and then compared with previously reported ICEKps by Kleborate[4].

Phylogenetic Analysis

Ridom SeqSphere + software (version 5.0, Ridom GmbH, Germany) was used for the core genome multi-locus

[1]https://bigsdb.pasteur.fr/klebsiella/klebsiella.html

[2]https://bigsdb.pasteur.fr/cgi-bin/bigsdb/bigsdb.pl?db=pubmlst_klebsiella_seqdef

[3]http://db-mml.sjtu.edu.cn/ICEberg/

[4]https://github.com/katholt/Kleborate

sequence typing (cgMLST) analysis with the whole genome sequence. Compared with common typing method MLST of 7 housekeeping genes or pulse field gel electrophoresis (PFGE), cgMLST scheme shows higher discriminatory power (Pérez-Losada et al., 2018; Gona et al., 2020). cgMLST schemes consist of a fixed set of conserved genome-wide genes. Alleles are used instead of single nucleotide polymorphisms (SNP) or concatenated sequences to mitigate the effects of recombination and to enable for a global and public nomenclature. For *K. pneumoniae*, 2358 target genes named with allelic nomenclature are employed in a cgMLST scheme, with NTUH-K2044 as the reference genome[5]. Similarly, 2526 accessory target genes were also applied for accessory genome allelic analysis[6]. Neighbor-joining trees based on core- and accessory genome allelic profile were constructed and the corresponding pairwise distances were calculated.

Hypermucoviscosity Phenotype Identification

Hypermucoviscosity phenotype was determined by string test. As previously described, a string test result was determined to be positive when a viscous filament greater than 5 mm in length was generated by stretching a bacterial colony with a bacteriological inoculation loop on a blood agar plate (Hadano, 2013; Shon et al., 2013).

Serum Resistance Assay

Three independent cultures for each strain were grown overnight and diluted to 1:1000 in MH broth with or without 20% normal human serum. Three replicates of each culture were aliquoted into a flat-bottom 100-well plate. The plate was incubated at 37°C with agitation. The OD_{600} of each culture was recorded every 5 min for 10 h by using the Bioscreen C Automated Microbiology Growth Curve Analysis System (Oy Growth Curves Ab Ltd., Turku, Finland). NUTH-K2044 was employed as hvKP reference strain, while American Type Culture Collection (ATCC) 700603 as the non-hvKP reference.

Statistical Analysis

Comparison between continuous variables (allelic gene difference) was performed by the Mann-Whitney test, as they were not normally distributed. In serum resistance assay, records for OD_{600} at each time point were summarized as mean plus 95% confidence interval.

RESULTS

Characteristics of the Strain

A 74-year-old man, with predisposed type II diabetes mellitus and schistosomiasis cirrhosis, presented with fever and painful swelling of left neck for 1 week before admission. Computed tomographic scan revealed an abscess (66 mm × 42 mm) on the left side of neck (**Figure 1A**). During this hospitalization, both cultures from the neck abscess and blood yielded *K. pneumoniae*. The two strains were later determined as the same via whole genome sequencing. The patient was eventually discharged home in good condition.

The strain Kp_whw was resistant to ampicillin but susceptible to the majority of the other antimicrobial agents, including cephalosporins, quinolones and carbapenems (**Table 1**). KP_whw looked shiny and cream-colored on blood agar and showed hypermucoviscosity phenotype with positive string test (**Figure 1B**). The strain belonged to ST 1049 and KL 5. We compared the *wzi* locus of this strain (*wzi208*) with other known loci associated with hypervirulent serotype (*wzi1*, *wzi2*, and *wzi5*) (**Figure 2**). The length of *wzi208* sequence is 447 bp and there

[5]https://www.cgmlst.org/ncs/schema/schema/2187931/
[6]https://www.cgmlst.org/ncs/schema/2192830/

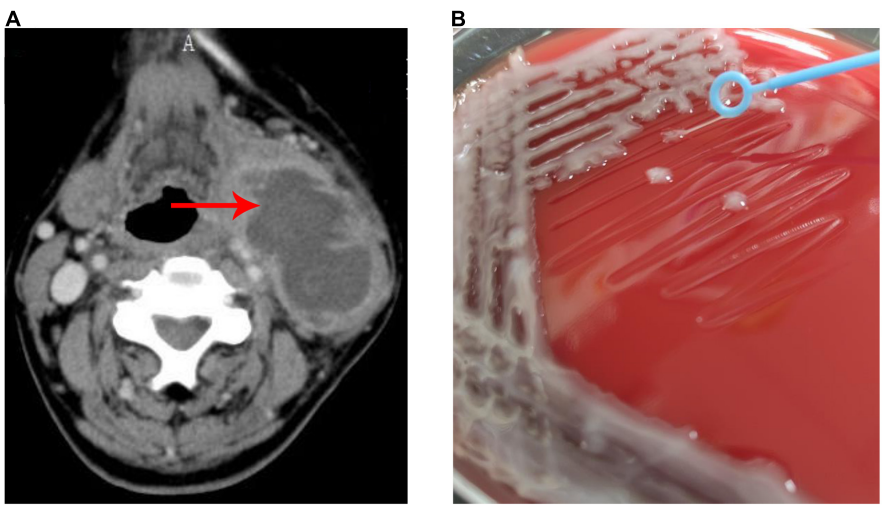


FIGURE 1 | Neck abscess caused by *K. pneumoniae* strain Kp_whw. **(A)** Computer tomography image showed an abscess (66 × 42 mm) on the left side of neck. **(B)** String test for the strain Kp_whw.

TABLE 1 | Susceptibility of Kp_whw to antimicrobial agents.

Antimicrobial agent	MIC (μg/ml)	Interpretation
Ampicillin	16	R
Piperacillin	≤4	S
Amoxicillin/Clavulanic acid	≤4/2	S
Ampicillin/Sulbactam	≤4/2	S
Piperacillin/Tazobactam	≤4/4	S
Ceftazidime	≤1	S
Cefotaxime	≤1	S
Cefepime	≤2	S
Aztreonam	≤2	S
Imipenem	≤1	S
Amikacin	≤8	S
Gentamycin	≤2	S
Ciprofloxacin	≤0.5	S
Levofloxacin	≤1	S
Sulfamethoxazole-trimethoprim	≤0.5/9.5	S
Chloramphenicol	≤4	S
Tetracycline	≤2	S
Cefperazone/Sulbactam	≤8	S
Cefuroxime	≤8	S
Meropenem	≤1	S

MIC, minimum inhibitory concentration; R, resistant; S, susceptible.

were 31 SNPs difference in comparison with *wzi1* and 24 SNPs with *wzi2*. However, *wzi208* is closest to *wzi5* (associated with K5 type), with only one SNP (C314A) (**Figure 2**).

hvKP of K1/K2 Serotype

The genetic similarity of strain Kp_whw with other K1/K2 serotype hvKP strains was analyzed. A total of 94 hvKP strains of K1/K2 serotype (K1, n = 64; K2, n = 30) were included. Of the 95 *K. pneumoniae* strains including Kp_whw, more than half of the strains were isolated from liver abscess (65/95, 67.4%) and 13 (13.7%) were from sputum (**Figure 3A**). Of note, the two isolated from bloodstream were strain NTUH-K2044 (ST23, K1 serotype, causing liver abscess and metastatic meningitis) (Wu et al., 2009) and strain hvKP1 (causing liver abscess with metastatic spread to the spleen, ST86, K2 serotype) (Russo and Gill, 2013). Similarly, Kp_whw caused invasive infections i.e., neck abscess and bloodstream infection. Minimum spanning tree (MST) based on MLST showed the relationship between strain Kp_whw and other K1/K2 serotype hvKP (**Figure 3B**). ST23, ST65, and ST86 were the top three clones of hvKP. Compared with ST374 (K2), strain Kp_whw had five variant loci.

Virulence Profile

Most K1 serotype strains were ST23 (56/64, 87.5%), while those K2 serotype showed more diversity, with 11 ST65 (36.7%), 10 ST86 (33.3%), 3 ST375 (10.0%), 3 ST380 (10.0%), 2 ST25 (6.7%), and 1 ST374 (3.3%) (**Figure 4**). Mucoviscosity-associated gene A (*magA*), which contributes largely to invasive infection but is specific to K1 serotype (Fang et al., 2004; Chuang et al., 2006), was not detected in Kp_whw. However, both regulator of mucoid phenotype A (*rmpA*) and *rmpA2* genes were present in Kp_whw. All the K2 type strains and Kp_whw showed positive string test, while less than a half of K1 type strain (46.9%, 30/64) exhibited hypermucoviscosity phenotype.

Mobile Genetic Elements Carrying Virulence Determinants

Virulence plasmid pLVPK was well characterized previously (Chen et al., 2004). Alignment of strain Kp_whw contigs to pLVPK showed that Kp_whw carried a plasmid that aligned well to most parts of the pLVPK plasmid, including the region in which *rmpA*, *rmpA2*, and *iroBCDN* (encoding salmochelin system) and *iucABCDiutA* (encoding aerobactin system) genes were located, as well as silver and tellurite resistance gene clusters (*silABCRS* and *terABCDEXYZW*) (**Figure 5A**). In addition, a chromosomal ICEKp5-like mobile genetic element was detected in the strain. A *ybt* locus encoding the biosynthesis of the siderophore yersiniabactin and its receptor was mobilized by this element (**Figure 5B**). These findings indicated that the strain Kp_whw harbored rich virulence factors and had the genetic potential to exacerbate the dissemination of hypervirulence.

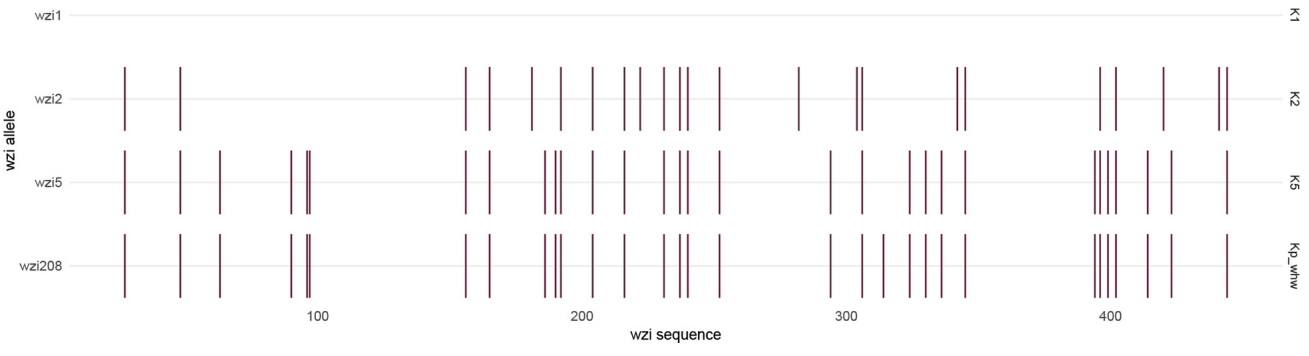


FIGURE 2 | Single nucleotide variant differences of *wzi* sequences, in relation to the reference sequence *wzi1*, which is associated with K1 serotype. Each variant is indicated by a small vertical line. The *wzi* sequences are 447 bp in length. The sequences with associated locus numbers were obtain from Institut Pasteur. Kp_whw is associated with *wzi208*.

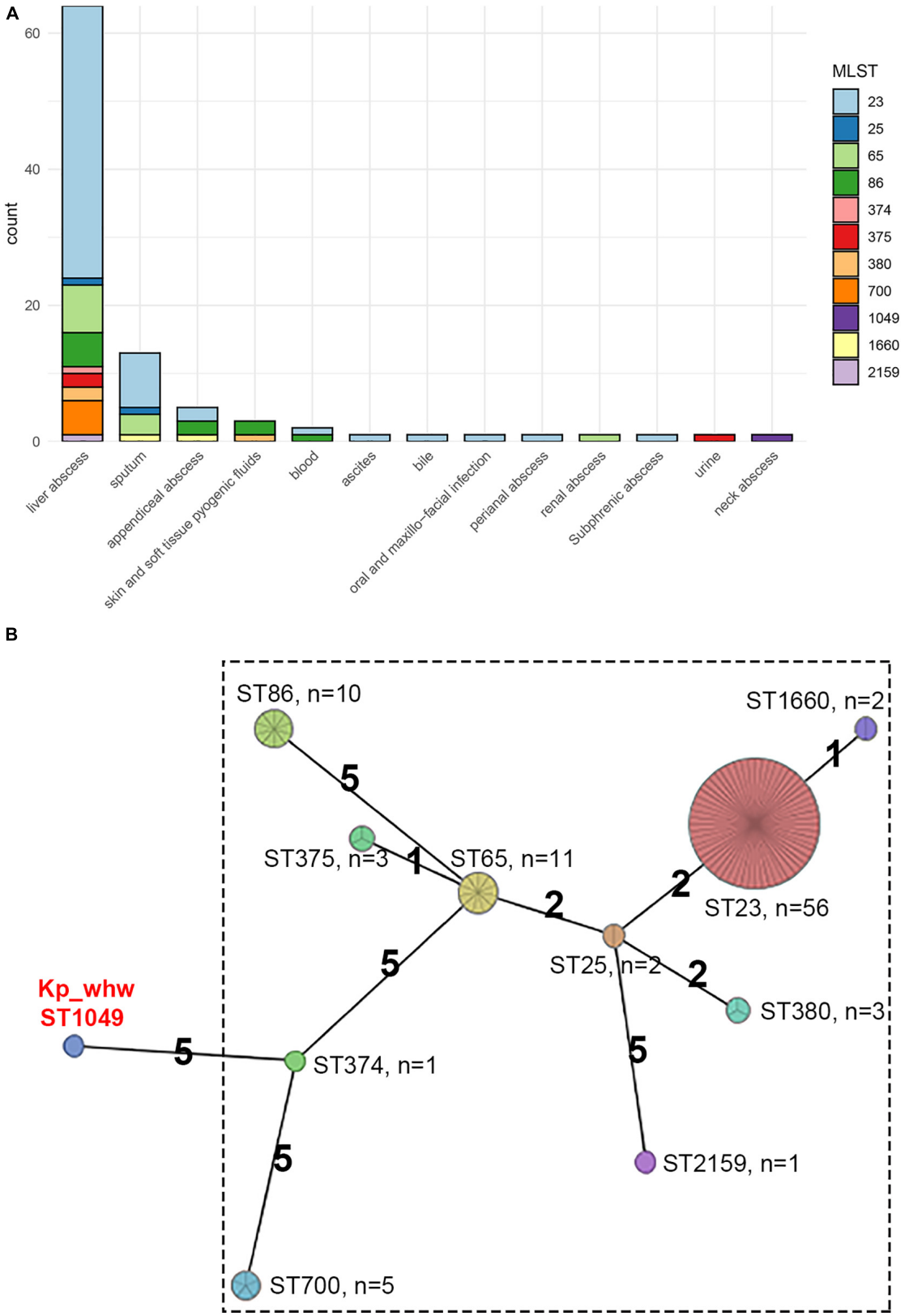


FIGURE 3 | Additional 94 K1/K2 serotype hvKP strains compared with Kp_whw. **(A)** Distribution of specimen sources of the 95 strains (including Kp_whw) according to sequence type. **(B)** Minimum spanning tree (MST) of the 95 strains based on multilocus sequence typing (MLST).

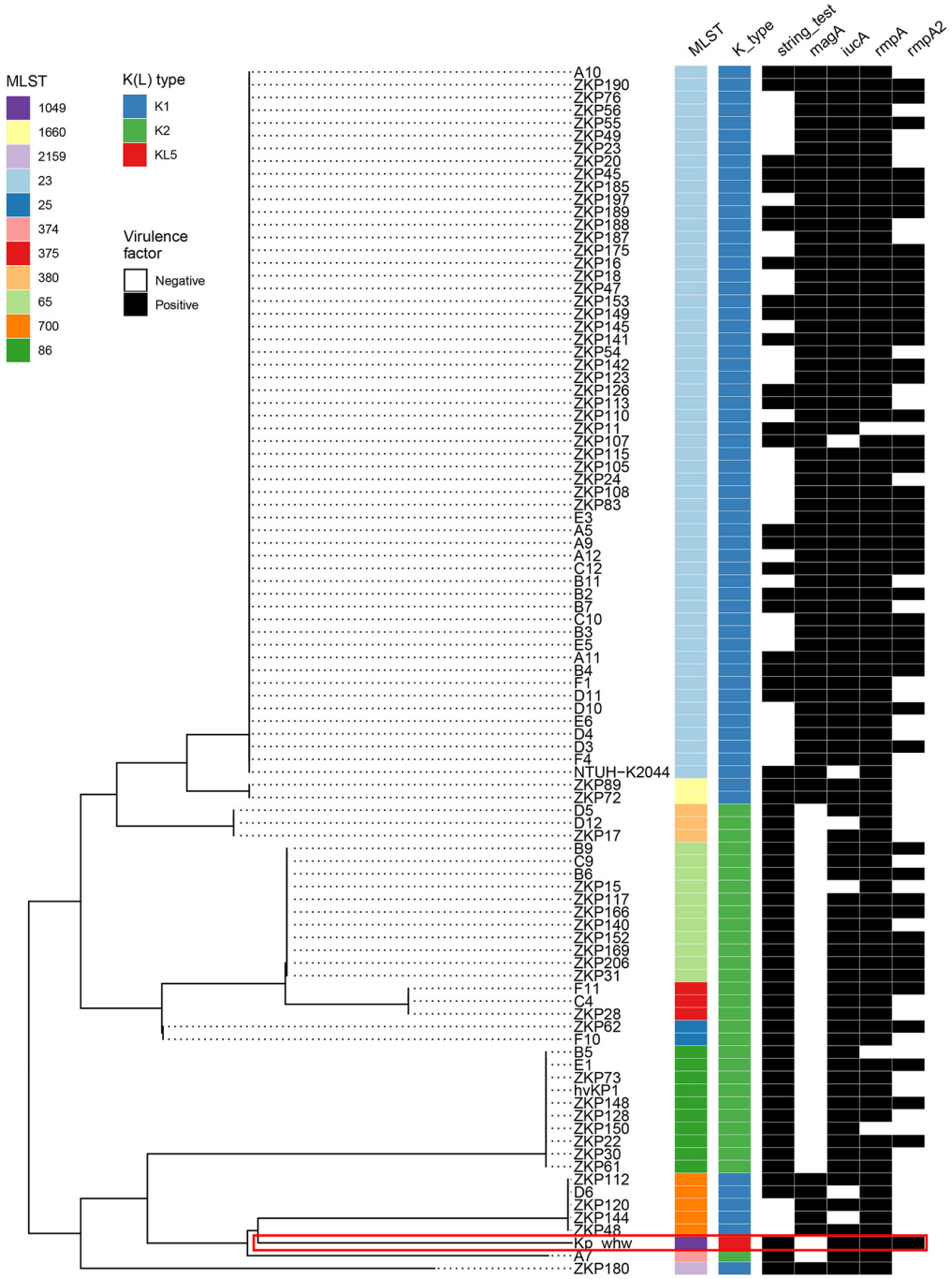


FIGURE 4 | Virulence profile of the 95 hvKP strains including Kp_whw. A MLST-based phylogenetic tree was constructed for all strains. ST and K(L) type were colored according to inset legend. The presence of the virulence factors including positive string test, *magA* gene, *iucA* gene, *rmpA*, and *rmpA2* genes were indicated as black rectangles. The position of Kp_whw was indicated with a red box.

Comparative Analysis and Phylogenetic Trees

The phylogenetic tree based on accessory gene allelic profile reflected almost completely the phylogeny of the core gene alleles (**Figure 6A**). One notable exception to this trend was identified in Kp_whw (ST1049), which was closed to strain ZKP180 (ST2159, K1 serotype) based on accessory gene profile but unexpectedly formed an individual branch according to core gene alleles (**Figure 6A**). This was in accordance with the finding that allelic differences between the strain Kp_whw and K1/K2 serotype hvKP strains based on core gene profile were significantly larger than those from accessory gene profile (**Figure 6B**).

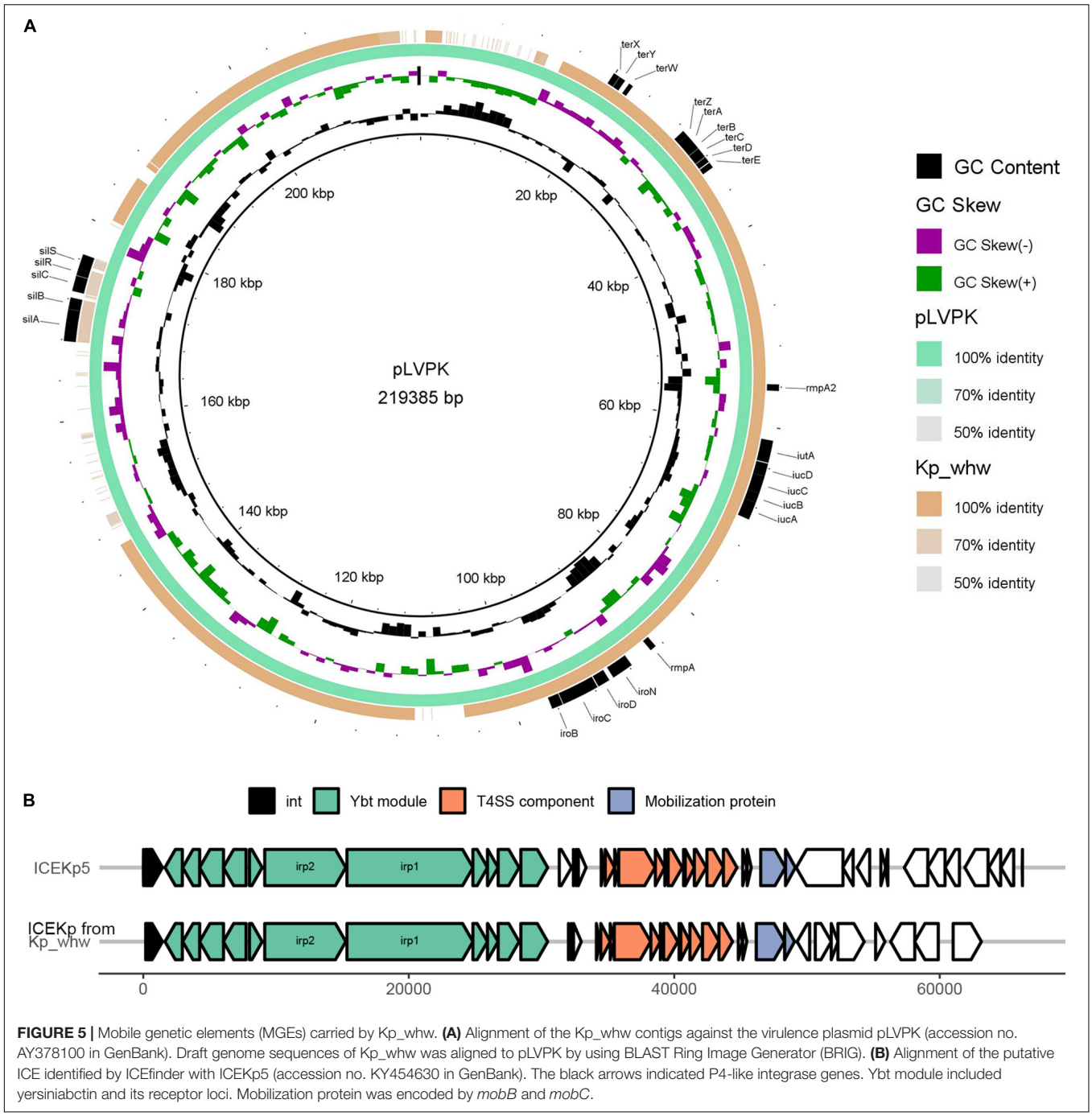


FIGURE 5 | Mobile genetic elements (MGEs) carried by Kp_whw. **(A)** Alignment of the Kp_whw contigs against the virulence plasmid pLVPK (accession no. AY378100 in GenBank). Draft genome sequences of Kp_whw was aligned to pLVPK by using BLAST Ring Image Generator (BRIG). **(B)** Alignment of the putative ICE identified by ICEfinder with ICEKp5 (accession no. KY454630 in GenBank). The black arrows indicated P4-like integrase genes. Ybt module included yersiniabctin and its receptor loci. Mobilization protein was encoded by *mobB* and *mobC*.

Virulence Assessment

To determine the virulence level of Kp_whw, serum resistant assay was performed. In MH broth without any supplement, ATCC 700603 outgrew NTUH-K2044 and Kp_whw (**Figure 7**). While in MH media with 20% normal human serum, however, Kp_whw and NTUH-K2044 were seldom influenced, yet ATCC700603 was dramatically inhibited to grow (**Figure 7**). Kp_whw and NTUH-K2044 had similar resistance to human serum, indicating that Kp_whw was as virulent as NTUH-K2044 *in vitro*.

DISCUSSION

Hypervirulent *K. pneumoniae* raised great concern recently. Here, we described a hvKP strain of ST1049, which was rarely reported, causing neck abscess, and invasive infection. hvKP is usually characterized as hypermucoviscosity, positive for *rmpA*, aerobactin, or virulence plasmid, and these features are frequently used to define hvKP. However, no single feature could accurately distinguish hvKP from non-hvKP (Lan et al., 2020). Clinical definition of hvKP based on the invasive liver abscess syndrome

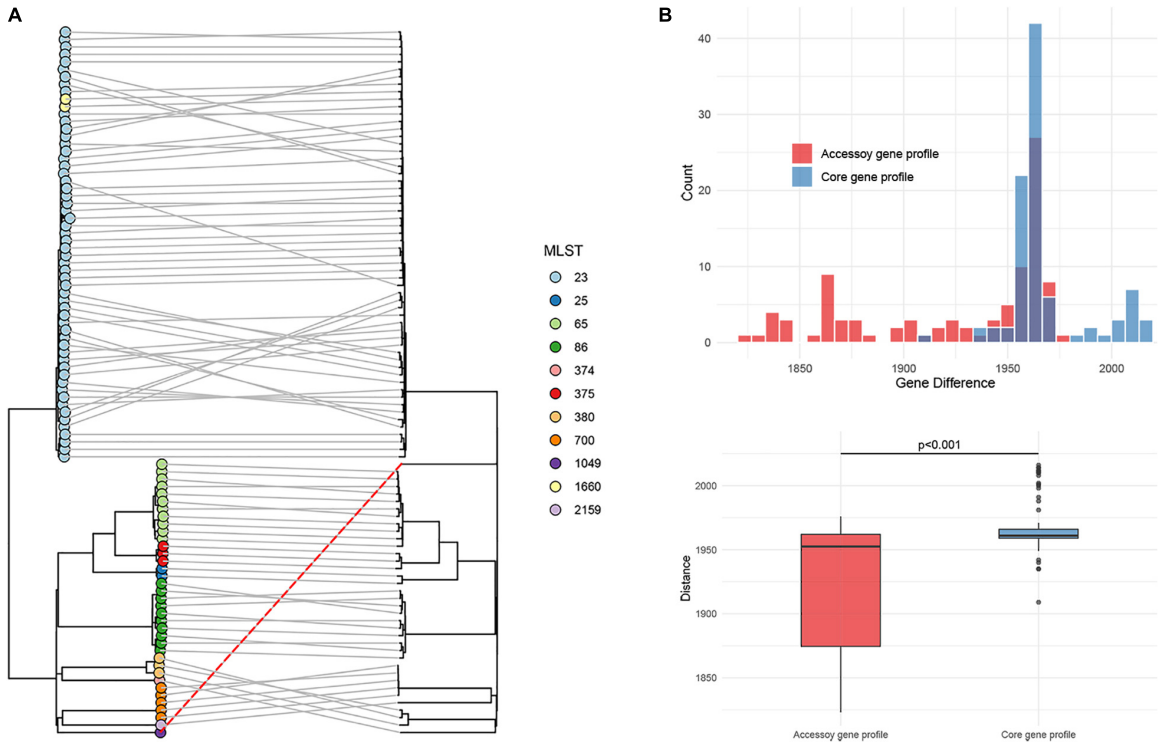


FIGURE 6 | Phylogenetic analysis of the 95 hvKP strains. **(A)** The phylogeny based on accessory gene allelic profile (left) was compared with the phylogeny based on the core genomic allelic profile. Kp_whw was indiacted with a dashed red line. Sequence types were colored according to inset legend. **(B)** Allelic gene differences on accessory genomic level between Kp_whw and K1/K2 strains were compared with those of core genomic level. Allelic gene differences were calculated by Ridom SeqSphere software.

and microbiological definition of K1/K2 serotype have been accepted in recent years (Siu et al., 2012).

Neck infection is usually associated with dental procedure or oral infection (Walia et al., 2014). Although neck abscess caused by *K. pneumoniae* have been occasionally reported (Nadasy et al., 2007; Frazee et al., 2009), the genetic background and virulence level for the causative strains were seldom determined. A large retrospective analysis about deep neck abscess in China showed that *K. pneumoniae* was the most predominant gram-negative pathogen (Yang et al., 2008). Though more than a half neck abscess tended to be polymicrobial (Yang et al., 2008), most invasive soft tissue infection caused by *K. pneumoniae* were found to be monomicrobial (Park et al., 2019; Rahim et al., 2019), partly due to the survival advantages of hvKP, such as hypermucoviscosity and high siderophore production (Rahim et al., 2019). Owing to the invasiveness, Kp_whw cuased neck abscess and bloodstream infection, which was also reported with classical hvKP NTUH-K2044 (Wu et al., 2009) and hvKP1 (Russo and Gill, 2013). hvKP as a monomicrobial pathogen in invasive soft tissue infection is recently increasing and causes high morbidity and mortality (Rahim et al., 2019). Diabetes mellitus was frequently associated with *K. pneumoniae* infection (Cheng et al., 2015) and was considered as a critical risk factor (Sharma et al., 2018). The other two cases of neck abscess caused by hvKP in the United States were also associated with diabetes (Nadasy et al., 2007; Frazee et al., 2009). This might be explained by the evidence that diabetic patients have defects on the neutrophil chemotactic and phagocytic activities (Alba-Loureiro et al., 2007). Of note, both the patients were Asia descent, and one had both neck abscess and blood positive culture for *K. pneumoniae* and the other developed metastatic complications.

Klebsiella pneumoniae of ST1049 was seldom reported globally. An epidemiological study on liver abscess caused by *K. pneumoniae* revealed that five cases were attributed to ST1049 *K. pneumoniae* (5/163, 3.1%) (Zhang et al., 2019). However, microbiological features, virulence level, and genomic information of these strains were not analyzed. In the present study, we analyzed the K-locus for Kp_whw and determined that it belonged to KL5. The *wzi208* (a region of K-locus) for Kp_whw is closed to *wzi5*, which is associated with K5 type. In Taiwan area of China, a ST1049 strain causing meningitis also belonged to K5 serotype (Ku et al., 2017). Actually, *K. pneumoniae* of K5 serotype is also a hypervirulent clone (Holt et al., 2015) and has frequently resulted in liver abscess (Lee et al., 2016), meningitis (Ku et al., 2017), and other invasive infections (Turton et al., 2010; Guo et al., 2017). From this perspective, Kp_whw possesses hypervirulent potential. On the other hand, Kp_whw was distant from K1/K2 serotype hvKP strains according to MLST and cgMLST. *K. pneumoniae* virulence factors are encoded by genes in both the core and accessory genomes (Martin and Bachman, 2018). LPS, siderophore enterobactin and

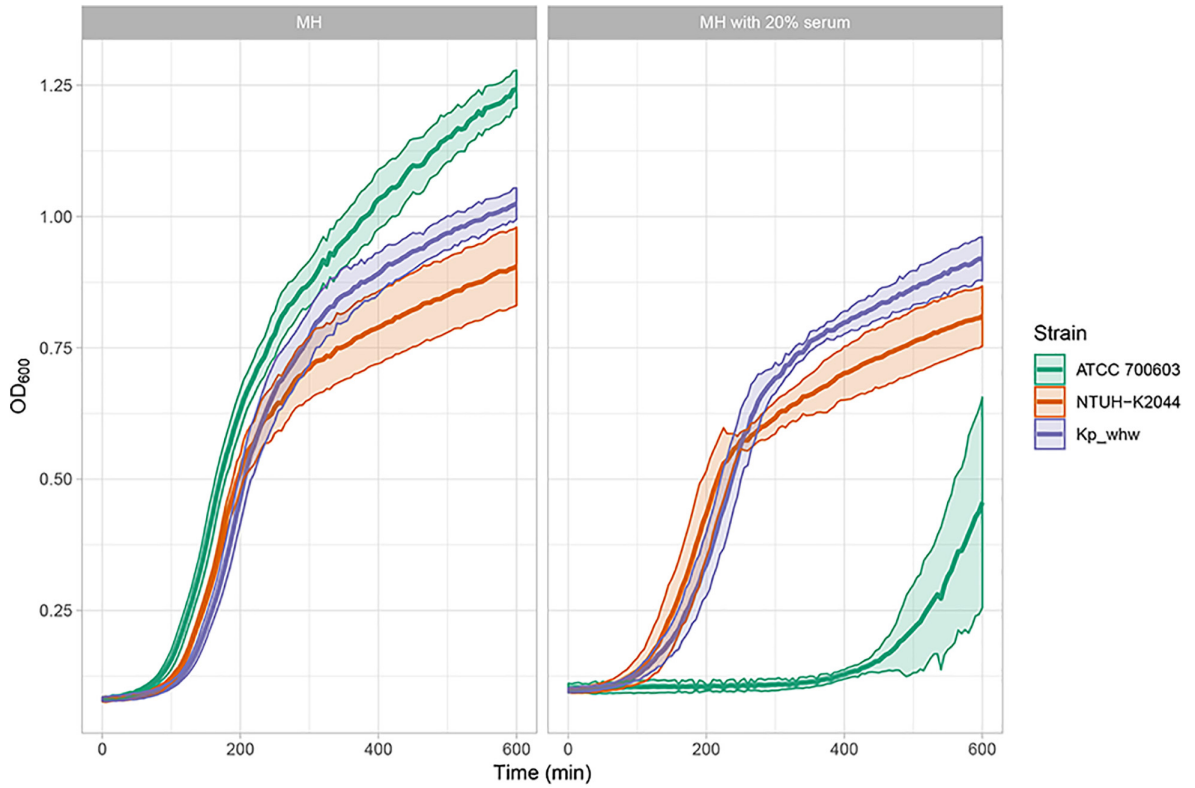


FIGURE 7 | Virulence level assessment *in vitro*. As microorganisms grow, they increase the turbidity of their growth medium. By measuring the turbidity of the medium over time, an OD_{600} curve can be generated. The curve reflects the growth (increased concentration) of the organism.

polysaccharide capsule synthesis are relative conserved and are primarily encoded by core genes while siderophore salmochelin, yersiniabactin and aerobactin are frequently ICE- or plasmid-encoded (Martin and Bachman, 2018). Compared with K1/K2 type hvKP, Kp_whw presented similar accessory genomic profile, with relative diverse allelic variation of conversed genes. These indicated that the relative higher level of core genome allelic diversity might determine the differences of genetic background between Kp_whw and hvKP strains of K1/K2 serotype.

Plasmids and ICEs are two key components contributing to HGT (Wozniak and Waldor, 2010). ICEKp is one of self-transmissible mobile genetic elements (MGEs) which encode the machinery for conjugation (type IV secretion system, T4SS) and intricate regulatory systems to control their excision from the chromosome in the host (Wozniak and Waldor, 2010). ICEKp mainly mobilizes *ybt* locus, which was first reported on the Yersinia high pathogenicity island (HPI), encoding yersiniabactin and its receptor. Yersiniabactin promoted the progression of bubonic and pneumonic plague, by scavenging iron directly from transferrin and lactoferrin (Fetherston et al., 2010). *In vivo*, yersiniabactin was demonstrated more important for *K. pneumoniae* growth than enterobactin under iron-limited conditions (Lawlor et al., 2007) and could facilitate microbes to colonize in the lungs and to cause disseminated infection (Lawlor et al., 2007; Bachman et al., 2011). Kp_whw carried an ICEKp5-like element, which containing a ybt14 locus, potentiating the transfer of yersiniabactin. Actually, additional siderophore like salmochelin (encoded by *iro* locus) can also be regulated by ICEKp (such as ICEKp1) (Lin et al., 2008). However, salmochelin, as well as aerobactin (encoded by *iuc* locus), are more frequently encoded by virulence plasmid, such as pLVPK (Chen et al., 2004). Though the affinity of aerobactin for iron is much lower than enterobactin (Brock et al., 1991), very low aerobactin concentrations are sufficient to stimulate bacterial growth (Garenaux et al., 2011), and it is essential for hvKP infection (Russo et al., 2015). For Kp_whw strain, *iuc* locus, *iro* locus and *rmpA*/*rmpA2* are simultaneously located in a single plasmid, which makes one concern about the acquisition of this type of plasmid by multi-drug resistant strains.

For a long time, virulence plasmid was deemed congenital. Carbapenem-resistant (CR) hvKP strains are increasingly emerging recently, with known mechanism of the acquisition of antimicrobial resistance genes by hvKP, but not the acquisition of virulence plasmid by CRKP, due to the non-conjugative nature of virulence plasmid (Yang et al., 2019). However, CRKP strains gaining a virulence plasmid have recently caused a lethal outbreak in China (Gu et al., 2018). It is particularly worrying that a newly reported virulence plasmid p15WZ-82_Vir could be conjugated to CRKP strains, leading to the expression of carbapenem resistance- and hypervirulence-associated phenotypes simultaneously

(Yang et al., 2019). Thus, it is now our urgency to pay attention to the convergence of hypervirulence and multi-drug resistance.

In conclusion, we characterized the genetic background and virulence level of a ST1049 *K. pneumoniae* strain Kp_whw isolated from neck abscess. Kp_whw is susceptible to commonly used antimicrobial agents except for being resistant to ampicillin. Kp_whw carried a pLVPK-like virulence plasmid and harbored a ICEKp5-like mobile genetic element. All results make clear that Kp_whw is a hvKP strain and has the potential to disseminate the phenotype, which poses a substantial public health threat. Although this is the first work characterizing the genetic virulence determinants of a ST1049 hvKP strain, only one strain was analyzed. In the future, isolates with diverse clinical characteristics should be collected for deep analysis.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ebi.ac.uk/ena, PRJEB38367 and https://www.ebi.ac.uk/ena, PRJEB34922.

AUTHOR CONTRIBUTIONS

PL and DZ conceived the idea and designed the experiments. JG isolated the strain. PL and RY performed the experiments. QS analyzed the data. YJ and JZ helped with materials and reagents. PL wrote the manuscript. YY reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China under Grant Nos. 81672067 and 81830069; and the Natural Science Foundation of Zhejiang Province, China under Grant No. LY17H190004.

REFERENCES

Alba-Loureiro, T. C., Munhoz, C. D., Martins, J. O., Cerchiaro, G. A., Scavone, C., Curi, R., et al. (2007). Neutrophil function and metabolism in individuals with diabetes mellitus. *Braz. J. Med. Biol. Res.* 40, 1037–1044. doi: 10.1590/s0100-879x2006005000143

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

Bachman, M. A., Oyler, J. E., Burns, S. H., Caza, M., Lepine, F., Dozois, C. M., et al. (2011). *Klebsiella pneumoniae* yersiniabactin promotes respiratory tract infection through evasion of lipocalin 2. *Infect. Immun.* 79, 3309–3316. doi: 10.1128/IAI.05114-11

Brock, J. H., Williams, P. H., Liceaga, J., and Wooldridge, K. G. (1991). Relative availability of transferrin-bound iron and cell-derived iron to aerobactin-producing and enterochelin-producing strains of *Escherichia coli* and to other microorganisms. *Infect. Immun.* 59, 3185–3190. doi: 10.1128/iai.59.9.3185-3190.1991

Catalan-Najera, J. C., Garza-Ramos, U., and Barrios-Camacho, H. (2017). Hypervirulence and hypermucoviscosity: two different but complementary *Klebsiella* spp. phenotypes? *Virulence* 8, 1111–1123. doi: 10.1080/21505594.2017.1317412

Chen, Y. T., Chang, H. Y., Lai, Y. C., Pan, C. C., Tsai, S. F., and Peng, H. L. (2004). Sequencing and analysis of the large virulence plasmid pLVPK of *Klebsiella pneumoniae* CG43. *Gene* 337, 189–198. doi: 10.1016/j.gene.2004.05.008

Cheng, N. C., Tai, H. C., Chang, S. C., Chang, C. H., and Lai, H. S. (2015). Necrotizing fasciitis in patients with diabetes mellitus: clinical characteristics and risk factors for mortality. *BMC Infect. Dis.* 15:417. doi: 10.1186/s12879-015-1144-0

Chou, A., Nuila, R. E., Franco, L. M., Stager, C. E., Atmar, R. L., and Zechiedrich, L. (2016). Prevalence of hypervirulent *Klebsiella pneumoniae*-associated genes rmpA and magA in two tertiary hospitals in Houston, TX, USA. *J. Med. Microbiol.* 65, 1047–1048. doi: 10.1099/jmm.0.000309

Chuang, Y. P., Fang, C. T., Lai, S. Y., Chang, S. C., and Wang, J. T. (2006). Genetic determinants of capsular serotype K1 of *Klebsiella pneumoniae* causing primary pyogenic liver abscess. *J. Infect. Dis.* 193, 645–654. doi: 10.1086/499968

Fang, C. T., Chuang, Y. P., Shun, C. T., Chang, S. C., and Wang, J. T. (2004). A novel virulence gene in *Klebsiella pneumoniae* strains causing primary liver abscess and septic metastatic complications. *J. Exp. Med.* 199, 697–705. doi: 10.1084/jem.20030857

Fetherston, J. D., Kirillina, O., Bobrov, A. G., Paulley, J. T., and Perry, R. D. (2010). The yersiniabactin transport system is critical for the pathogenesis of bubonic and pneumonic plague. *Infect. Immun.* 78, 2045–2052. doi: 10.1128/iai.01236-09

Frazee, B. W., Hansen, S., and Lambert, L. (2009). Invasive infection with hypermucoviscous *Klebsiella pneumoniae*: multiple cases presenting to a single emergency department in the United States. *Ann. Emerg. Med.* 53, 639–642. doi: 10.1016/j.annemergmed.2008.11.007

Garenaux, A., Caza, M., and Dozois, C. M. (2011). The Ins and outs of siderophore mediated iron uptake by extra-intestinal pathogenic *Escherichia coli*. *Vet. Microbiol.* 153, 89–98. doi: 10.1016/j.vetmic.2011.05.023

Gona, F., Comandatore, F., Battaglia, S., Piazza, A., Trovato, A., Lorenzin, G., et al. (2020). Comparison of core-genome MLST, coreSNP and PFGE methods for *Klebsiella pneumoniae* cluster analysis. *Microb. Genom.* 6:347. doi: 10.1099/mgen.0.000347

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

Guo, Y., Wang, S., Zhan, L., Jin, Y., Duan, J., Hao, Z., et al. (2017). Microbiological and clinical characteristics of hypermucoviscous *Klebsiella pneumoniae* isolates associated with invasive infections in China. *Front. Cell. Infect. Microbiol.* 7:24. doi: 10.3389/fcimb.2017.00024

Hadano, Y. (2013). String test. *BMJ Case Rep.* 2013:e008328. doi: 10.1136/bcr-2012-008328

Hodgson, K., Morris, J., Bridson, T., Govan, B., Rush, C., and Ketheesan, N. (2015). Immunological mechanisms contributing to the double burden of diabetes and intracellular bacterial infections. *Immunology* 144, 171–185. doi: 10.1111/imm.12394

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

Joob, B., and Wiwanitkit, V. (2017). *Klebsiella pneumoniae* invasive liver abscess syndrome and Endophthalmitis. *J. Emerg. Med.* 53:917. doi: 10.1016/j.jemermed.2017.03.050

Ku, Y. H., Chuang, Y. C., Chen, C. C., Lee, M. F., Yang, Y. C., Tang, H. J., et al. (2017). *Klebsiella pneumoniae* isolates from meningitis: epidemiology, virulence and antibiotic resistance. *Sci. Rep.* 7:6634. doi: 10.1038/s41598-017-06878-6

Lam, M. M. C., Wick, R. R., Wyres, K. L., Gorrie, C. L., Judd, L. M., Jenney, A. W. J., et al. (2018). Genetic diversity, mobilisation and spread of the yersiniabactin-encoding mobile element ICEKp in *Klebsiella pneumoniae* populations. *Microb. Genom.* 4:e000196. doi: 10.1099/mgen.0.000196

Lan, P., Shi, Q., Zhang, P., Chen, Y., Yan, R., Hua, X., et al. (2020). Core genome allelic profiles of clinical *Klebsiella pneumoniae* strains using a random forest algorithm based on multilocus sequence typing scheme for hypervirulence analysis. *J. Infect. Dis.* 221(Suppl._2), S263–S271. doi: 10.1093/infdis/jiz562

Lawlor, M. S., O'Connor, C., and Miller, V. L. (2007). Yersiniabactin is a virulence factor for *Klebsiella pneumoniae* during pulmonary infection. *Infect. Immun.* 75, 1463–1472. doi: 10.1128/IAI.00372-06

Lee, I. R., Molton, J. S., Wyres, K. L., Gorrie, C., Wong, J., Hoh, C. H., et al. (2016). Differential host susceptibility and bacterial virulence factors driving *Klebsiella* liver abscess in an ethnically diverse population. *Sci. Rep.* 6:29316. doi: 10.1038/srep29316

Lin, J. C., Chang, F. Y., Fung, C. P., Yeh, K. M., Chen, C. T., Tsai, Y. K., et al. (2010). Do neutrophils play a role in establishing liver abscesses and distant metastases caused by *Klebsiella pneumoniae*? *PLoS One* 5:e15005. doi: 10.1371/journal.pone.0015005

Lin, Y. T., Jeng, Y. Y., Chen, T. L., and Fung, C. P. (2010). Bacteremic community-acquired pneumonia due to *Klebsiella pneumoniae*: clinical and microbiological characteristics in Taiwan, 2001-2008. *BMC Infect. Dis.* 10:307. doi: 10.1186/1471-2334-10-307

Lin, J. C., Siu, L. K., Fung, C. P., Tsou, H. H., Wang, J. J., Chen, C. T., et al. (2006). Impaired phagocytosis of capsular serotypes K1 or K2 *Klebsiella pneumoniae* in type 2 diabetes mellitus patients with poor glycemic control. *J. Clin. Endocrinol. Metab.* 91, 3084–3087. doi: 10.1210/jc.2005-2749

Lin, T.-L., Lee, C.-Z., Hsieh, P.-F., Tsai, S.-F., and Wang, J.-T. (2008). Characterization of integrative and conjugative element ICEKp1-associated genomic heterogeneity in a *Klebsiella pneumoniae* strain isolated from a primary liver abscess. *J. Bacteriol.* 190, 515–526. doi: 10.1128/JB.01219-07

Lin, Y. T., Siu, L. K., Lin, J. C., Chen, T. L., Tseng, C. P., Yeh, K. M., et al. (2012). Seroepidemiology of *Klebsiella pneumoniae* colonizing the intestinal tract of healthy Chinese and overseas Chinese adults in Asian countries. *BMC Microbiol.* 12:13. doi: 10.1186/1471-2180-12-13

Liu, Y. C., Cheng, D. L., and Lin, C. L. (1986). *Klebsiella pneumoniae* liver abscess associated with septic endophthalmitis. *Arch. Intern. Med.* 146, 1913–1916. doi: 10.1001/archinte.146.10.1913

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

Melot, B., Brisse, S., Breurec, S., Passet, V., Malpote, E., Lamaury, I., et al. (2016). Community-acquired meningitis caused by a CG86 hypervirulent *Klebsiella pneumoniae* strain: first case report in the Caribbean. *BMC Infect. Dis.* 16:736. doi: 10.1186/s12879-016-2065-2

Nadasy, K. A., Domiati-Saad, R., and Tribble, M. A. (2007). Invasive *Klebsiella pneumoniae* syndrome in North America. *Clin. Infect. Dis.* 45, e25–e28. doi: 10.1086/519424

Namikawa, H., Yamada, K., Sakiyama, A., Imoto, W., Yamairi, K., Shibata, W., et al. (2019). Clinical characteristics of bacteremia caused by hypermucoviscous *Klebsiella pneumoniae* at a tertiary hospital. *Diagn. Microbiol. Infect. Dis.* 95, 84–88. doi: 10.1016/j.diagmicrobio.2019.04.008

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

Park, S. Y., Yu, S. N., Lee, E. J., Kim, T., Jeon, M. H., Choo, E. J., et al. (2019). Monomicrobial gram-negative necrotizing fasciitis: an uncommon but fatal syndrome. *Diagn. Microbiol. Infect. Dis.* 94, 183–187. doi: 10.1016/j.diagmicrobio.2018.12.013

Peirano, G., Pitout, J. D., Laupland, K. B., Meatherall, B., and Gregson, D. B. (2013). Population-based surveillance for hypermucoviscosity *Klebsiella pneumoniae* causing community-acquired bacteremia in Calgary, Alberta. *Can. J. Infect. Dis. Med. Microbiol.* 24, e61–e64. doi: 10.1155/2013/828741

Pérez-Losada, M., Arenas, M., and Castro-Nallar, E. (2018). Microbial sequence typing in the genomic era. *Infect. Genet. Evol.* 63, 346–359. doi: 10.1016/j.meegid.2017.09.022

Prokesch, B. C., TeKippe, M., Kim, J., Raj, P., TeKippe, E. M., and Greenberg, D. E. (2016). Primary osteomyelitis caused by hypervirulent *Klebsiella pneumoniae*. *Lancet Infect. Dis.* 16, e190–e195. doi: 10.1016/s1473-3099(16)30021-4

Rafat, C., Messika, J., Barnaud, G., Dufour, N., Magdoud, F., Billard-Pomarès, T., et al. (2018). Hypervirulent *Klebsiella pneumoniae*, a 5-year study in a French ICU. *J. Med. Microbiol.* 67, 1083–1089. doi: 10.1099/jmm.0.000788

Rahim, G. R., Gupta, N., Maheshwari, P., and Singh, M. P. (2019). Monomicrobial *Klebsiella pneumoniae* necrotizing fasciitis: an emerging life-threatening entity. *Clin. Microbiol. Infect.* 25, 316–323. doi: 10.1016/j.cmi.2018.05.008

Russo, T. A., and Gill, S. R. (2013). Draft genome sequence of the hypervirulent *Klebsiella pneumoniae* strain hvKP1, isolated in Buffalo, New York. *Genome Announc.* 1:e0006513. doi: 10.1128/genomeA.00065-13

Russo, T. A., and Marr, C. M. (2019). Hypervirulent *Klebsiella pneumoniae*. *Clin. Microbiol. Rev.* 32:e00001-19. doi: 10.1128/cmr.00001-19

Russo, T. A., Olson, R., MacDonald, U., Beanan, J., and Davidson, B. A. (2015). Aerobactin, but not yersiniabactin, salmochelin, or enterobactin, enables the growth/survival of hypervirulent (hypermucoviscous) *Klebsiella pneumoniae* ex vivo and in vivo. *Infect. Immun.* 83, 3325–3333. doi: 10.1128/iai.00430-15

Sharma, K., Das, D., Joshi, M., Barman, D., and Sarma, A. J. (2018). Deep neck space infections-A study in diabetic population in a tertiary care centre. *Indian J. Otolaryngol. Head. Neck Surg.* 70, 22–27. doi: 10.1007/s12070-017-1196-0

Shon, A. S., Bajwa, R. P., and Russo, T. A. (2013). Hypervirulent (hypermucoviscous) *Klebsiella pneumoniae*: a new and dangerous breed. *Virulence* 4, 107–118. doi: 10.4161/viru.22718

Siu, L. K., Yeh, K. M., Lin, J. C., Fung, C. P., and Chang, F. Y. (2012). *Klebsiella pneumoniae* liver abscess: a new invasive syndrome. *Lancet Infect. Dis.* 12, 881–887. doi: 10.1016/s1473-3099(12)70205-0

Struve, C., Roe, C. C., Stegger, M., Stahlhut, S. G., Hansen, D. S., Engelthaler, D. M., et al. (2015). Mapping the evolution of hypervirulent *Klebsiella pneumoniae*. *mBio* 6:e00630. doi: 10.1128/mBio.00630-15

Turton, J. F., Perry, C., Elgohari, S., and Hampton, C. V. (2010). PCR characterization and typing of *Klebsiella pneumoniae* using capsular type-specific, variable number tandem repeat and virulence gene targets. *J. Med. Microbiol.* 59(Pt 5), 541–547. doi: 10.1099/jmm.0.015198-0

Walia, I. S., Borle, R. M., Mehendiratta, D., and Yadav, A. O. (2014). Microbiology and antibiotic sensitivity of head and neck space infections of odontogenic origin. *J. Maxillofac. Oral. Surg.* 13, 16–21. doi: 10.1007/s12663-012-0455-6

Wozniak, R. A. F., and Waldor, M. K. (2010). Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat. Rev. Microbiol.* 8, 552–563. doi: 10.1038/nrmicro2382

Wu, K. M., Li, L. H., Yan, J. J., Tsao, N., Liao, T. L., Tsai, H. C., et al. (2009). Genome sequencing and comparative analysis of *Klebsiella pneumoniae* NTUH-K2044, a strain causing liver abscess and meningitis. *J. Bacteriol.* 191, 4492–4501. doi: 10.1128/jb.00315-09

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

Yang, S. W., Lee, M. H., See, L. C., Huang, S. H., Chen, T. M., and Chen, T. A. (2008). Deep neck abscess: an analysis of microbial etiology and the effectiveness of antibiotics. *Infect. Drug Resist.* 1, 1–8. doi: 10.2147/idr.s3554

Yang, X., Wai-Chi Chan, E., Zhang, R., and Chen, S. (2019). A conjugative plasmid that augments virulence in *Klebsiella pneumoniae*. *Nat. Microbiol.* 4, 2039–2043. doi: 10.1038/s41564-019-0566-7

Zhang, S., Zhang, X., Wu, Q., Zheng, X., Dong, G., Fang, R., et al. (2019). Clinical, microbiological, and molecular epidemiological characteristics of *Klebsiella pneumoniae*-induced pyogenic liver abscess in southeastern China. *Antimicrob. Resist. Infect. Control.* 8:166. doi: 10.1186/s13756-019-0615-2

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

frontiers in Microbiology

ORIGINAL RESEARCH
published: 15 February 2021
doi: 10.3389/fmicb.2021.639660

Genome-Wide Association Studies for the Detection of Genetic Variants Associated With Daptomycin and Ceftaroline Resistance in *Staphylococcus aureus*

*Robert E. Weber[1], Stephan Fuchs[2], Franziska Layer[1], Anna Sommer[1], Jennifer K. Bender[1], Andrea Thürmer[3], Guido Werner[1] and Birgit Strommenger[1]**

[1] Department of Infectious Diseases, Robert Koch-Institute, Wernigerode, Germany, [2] Methodology and Research Infrastructure, Bioinformatics, Robert Koch-Institute, Berlin, Germany, [3] Methodology and Research Infrastructure, Genome Sequencing, Robert Koch-Institute, Berlin, Germany

OPEN ACCESS

Edited by:
Karsten Becker, University Medicine Greifswald, Germany

Reviewed by:
Ben Pascoe, University of Bath, United Kingdom
Richard V. Goering, Creighton University, United States

***Correspondence:**
Birgit Strommenger
strommengerb@rki.de

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

Received: *11 December 2020*
Accepted: *22 January 2021*
Published: *15 February 2021*

Citation:
Weber RE, Fuchs S, Layer F, Sommer A, Bender JK, Thürmer A, Werner G and Strommenger B (2021) Genome-Wide Association Studies for the Detection of Genetic Variants Associated With Daptomycin and Ceftaroline Resistance in Staphylococcus aureus. Front. Microbiol. 12:639660. doi: 10.3389/fmicb.2021.639660

Background: As next generation sequencing (NGS) technologies have experienced a rapid development over the last decade, the investigation of the bacterial genetic architecture reveals a high potential to dissect causal loci of antibiotic resistance phenotypes. Although genome-wide association studies (GWAS) have been successfully applied for investigating the basis of resistance traits, complex resistance phenotypes have been omitted so far. For *S. aureus* this especially refers to antibiotics of last resort like daptomycin and ceftaroline. Therefore, we aimed to perform GWAS for the identification of genetic variants associated with DAP and CPT resistance in clinical *S. aureus* isolates.

Materials/methods: To conduct microbial GWAS, we selected cases and controls according to their clonal background, date of isolation, and geographical origin. Association testing was performed with PLINK and SEER analysis. By using *in silico* analysis, we also searched for rare genetic variants in candidate loci that have previously been described to be involved in the development of corresponding resistance phenotypes.

Results: GWAS revealed MprF P314L and L826F to be significantly associated with DAP resistance. These mutations were found to be homogenously distributed among clonal lineages suggesting convergent evolution. Additionally, rare and yet undescribed single nucleotide polymorphisms could be identified within *mprF* and putative candidate genes. Finally, we could show that each DAP resistant isolate exhibited at least one amino acid substitution within the open reading frame of *mprF*. Due to the presence of strong population stratification, no genetic variants could be associated with CPT resistance. However, the investigation of the staphylococcal cassette chromosome *mec* (SCC*mec*) revealed various *mecA* SNPs to be putatively linked with CPT resistance. Additionally, some CPT resistant isolates revealed no *mecA* mutations, supporting the hypothesis that further and still unknown resistance determinants are crucial for the development of CPT resistance in *S. aureus*.

Conclusion: We hereby confirmed the potential of GWAS to identify genetic variants that are associated with antibiotic resistance traits in *S. aureus*. However, precautions need to be taken to prevent the detection of spurious associations. In addition, the implementation of different approaches is still essential to detect multiple forms of variations and mutations that occur with a low frequency.

Keywords: GWAS, daptomycin, ceftaroline, *S. aureus*, antibiotic resistance, PLINK, SEER

INTRODUCTION

Staphylococcal aureus is a major human pathogen that is responsible for a large number of community- and hospital associated infections worldwide (Holden et al., 2013). It causes a variety of human maladies, from minor skin and soft tissue infections to systematic and life-threatening diseases such as endocarditis, pneumonia, and septicemia (Tong et al., 2015). The introduction of penicillinase-stable β-lactam antibiotics six decades ago gave rise to the selection and spread of methicillin-resistant *S. aureus* (MRSA) (Holden et al., 2013). This resistance phenotype is mediated by horizontal acquisition of the staphylococcal cassette chromosome *mec* (SCC*mec*) and, in particular *mecA* (or its homologues *mecB/mecC*), that encodes the alternative penicillin-binding protein 2A (PBP2a) which has significantly less affinity to methicillin (Robinson and Enright, 2003).

For many years, vancomycin (VAN) administration has been considered the reference standard for the treatment of invasive MRSA infections (Micek, 2007). However, owing to an overall increase in VAN minimal inhibitory concentrations (MIC), potential adverse consequences and concern of treatment failure, the role of VAN as first-line antibiotic has become controversial in modern therapeutics (Bruniera et al., 2015). Later on, new antibiotics such as daptomycin (DAP) and ceftaroline (CPT) have been developed, showing a promising anti-MRSA activity. The cyclic lipopeptide DAP was approved in Europe by the EMA for the treatment of skin and skin structure infections (SSSI), bacteremia and right sided endocarditis in 2006/2007 (Gonzalez-Ruiz et al., 2011). The proposed mechanism of action involves the calcium-dependent integration of DAP into the bacterial cell membrane that is triggered by the binding of DAP to the negatively charged phospholipid phosphatidylglycerol (PG) (Bayer et al., 2013). Although subsequent pore formation and ion leakage have frequently been described as the cause of cell death, recent studies suggest that the interaction between DAP and fluid membrane microdomains results in cell wall defects like membrane rigidification, depolarization and the delocalization of essential membrane proteins (Muller et al., 2016; Ernst and Peschel, 2019; Gray and Wenzel, 2020). DAP resistance of *S. aureus* has frequently been associated with single nucleotide polymorphisms (SNPs) in the multi-peptide resistance factor MprF, leading to an enhanced production and translocation of the positively charged lysyl-phosphotidylglycerol (L-PG) and thereby to a repulsion of the calcium-complexed DAP (Friedman et al., 2006; Mishra et al., 2011; Peleg et al., 2012; Bayer et al., 2013; Yang et al., 2013). However, results of recent studies suggest that these mutations might instead result in an extended substrate spectrum of MprF, thereby enabling the translocation of DAP itself or membrane proteins that are critical for the activity of DAP (Ernst and Peschel, 2019).

In 2012 CPT was licensed in Europe for treating serious skin and soft tissue infections and community acquired pneumonia (AstraZeneca, 2012). The fifth-generation cephalosporin binds with high affinity to the allosteric domain of PBP2a, thereby stimulating a multi-residue conformation change that predisposes the active site of PBP2a to be inactivated by a second CPT molecule (Kosowska-Shick et al., 2010; Otero et al., 2013). The irreversible acylation of the active serine causes the protein to lose its function and finally results in loss of membrane integrity and cell death (Otero et al., 2013; Peacock and Paterson, 2015). Treatment failure is expected to be caused by mutations in either the transpeptidase or allosteric domain of PBP2a that are accompanied with decreased binding capacity of CPT or impaired protein-protein interactions between PBP2a and PBP2 (Otero et al., 2013).

Although multiple genetic determinants of CPT and DAP resistance have been described, scientists still lack a deeper understanding of the involved resistance mechanisms, because a variety of factors and processes appear to affect the antimicrobial activity of these compounds (Şlusarczyk et al., 2018; Ernst and Peschel, 2019).

Since Next Generation Sequencing (NGS) technologies have experienced a rapid development in the past 10 years, the investigation of the bacterial genetic architecture reveals high potential do dissect causal loci of antimicrobial resistance traits. With the integration of whole genome sequencing in the clinical and public health setting, gene databases and tools have been developed to allow for the assessment of genes associated with antimicrobial resistance (Feldgarden et al., 2019; Wheeler et al., 2019; Alcock et al., 2020; Bortolaia et al., 2020). However, these tools quickly fail when facing multifactorial or yet unknown resistance mechanisms. In this case, genome-wide association studies (GWAS) have become a promising option for the prediction of resistance phenotypes based on genomic data (Farhat et al., 2013, 2019; Alam et al., 2014; Chen and Shapiro, 2015). For staphylococci, GWAS have also been successfully applied to identify genetic determinants significantly associated with virulence, pathogenicity and growth plasticity (Laabei et al., 2014; Meric et al., 2018; Rong et al., 2019; Young et al., 2019). As GWAS were originally developed for human studies, microbial GWAS had to be adopted to bacterial populations, addressing confounding factors like population stratification, linkage disequilibrium, and population structure that can lead to

the detection of spurious associations if not corrected properly (Falush and Bowden, 2006; Chen and Shapiro, 2015; Power et al., 2017; Farhat et al., 2019). Here we consider two strain collections of clinical *S. aureus* isolates to both identify genetic variants associated with DAP and CPT resistance and to highlight the chances and limitations of bacterial GWAS.

MATERIALS AND METHODS

Bacterial Isolates

All isolates originated from submissions to the German National Reference Centre (NRC) for Staphylococci and Enterococci. Species confirmation was conducted by colony morphology and positive plasma coagulase reaction. Isolates have been further characterized by *spa* typing as described previously (Strommenger et al., 2008). Since DAP and CPT resistance in *S. aureus* is still rare (Roch et al., 2017; Urban and Stone, 2019), the sample sizes remained comparatively small (Power et al., 2017). A total of 48 daptomycin resistant (DAP-R) and 47 daptomycin susceptible (DAP-S) *S. aureus* isolates were selected according to Robert and Weber (2017). The DAP breakpoint for *S. aureus* according to EUCAST version 7.1[1] was 1 mg/L (S $\leq$ 1 mg/L, R $>$ 1 mg/L). The final strain collection comprised clonal lineages CC5 [clonal complex; including sequence type (ST) ST5 (n = 3), ST225 (n = 21), ST149 (n = 2), ST4411 (n = 1), and ST228 (n = 1)], CC8 [including ST8 (n = 16), ST239 (n = 1), and ST241 (n = 1)], CC15 [including ST15 (n = 3) and ST4410 (n = 1)], CC45 [including ST45 (n = 5), ST4409 (n = 1), and ST4408 (n = 1)], ST7 (n = 4), ST22 (n = 30), ST30 (n = 2), and ST398 (n = 2) (**Supplementary Figure 1A**). Both, methicillin-susceptible *S. aureus* (MSSA, n = 32) and methicillin-resistant *S. aureus* (MRSA, n = 63) were considered. Isolates were defined as MRSA on the basis of oxacillin resistance (MIC_{OXA} $>$ 2 mg/L), cefoxitin resistance (MIC_{CIX} $>$ 4 mg/L; CIX MICs not available for isolates collected prior to 2014) and *mecA* carriage. For CPT analysis, we selected a total of 44 CPT resistant (CPT-R) and 43 CPT susceptible (CPT-S) *S. aureus* isolates. The CPT breakpoint for *S. aureus* according to EUCAST version 7.1 was 1 mg/L (S $\leq$ 1 mg/L, R $>$ 1 mg/L). The final strain collection included clonal lineages CC5 [including ST5 (n = 12), ST225 (n = 13), ST146 (n = 1), ST228 (n = 44), ST111 (n = 7), and ST4511 (n = 1)], CC8 [including ST8 (n = 3), ST113 (n = 1), ST239 (n = 9), ST14407 (n = 1), and ST1465 (n = 1)], ST398 (n = 1), ST30 (n = 6), and CC22 [including ST22 (n = 19) and ST4406 (n = 1)] (**Supplementary Figure 1B**). The **Supplementary Tables 1, 2** contain detailed information for both strain collections.

Susceptibility Testing

DAP, CPT, and OXA MICs were determined by the use of broth microdilution (BMD) according to EUCAST guidelines version 7.1 as described before (Strommenger et al., 2015; Robert and Weber, 2017). ATCC 29213 was used as quality control strain. In case of ambiguous CPT MIC results, measurements were repeated up to three times and a median was calculated. Additionally, the following clinically or epidemiologically relevant antibiotics were routinely tested: benzylpenicillin (BEN), Cefoxitin (CXI, since 2014) gentamicin (GEN), erythromycin (ERY), clindamycin (CLI), tetracycline (TET), vancomycin (VAN), teicoplanin (TEI), ciprofloxacin (CIP), trimethoprim-sulfamethoxazole (TRS), fusidic acid (FUS), rifampicin (RIF), mupirocin (MUP), phosphomycin (FOS), linezolid (LIN), moxifloxacin (MOX), and tigecycline (TIG).

[1]https://eucast.org/clinical_breakpoints/

Whole Genome Sequencing and Quality Control

Mueller Hinton Bouillon (MHB) was used as growth medium for bacterial cultivation. Isolation of genomic DNA from an overnight culture of *S. aureus* was performed using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) with a lysostaphin pre-treatment in 20 mM Tris/HCl, pH 8.0, 2 mM sodium EDTA, 1.2% Triton X-100 (100 μg/ml final concentration of lysostaphin). For determining purity and quantity of nucleic acids, the Eppendorf BioPhotometer (Eppendorf AG, Hamburg, Germany) and the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA, United States) were used in line with the manufacturer's instructions. Sequencing libraries were generated with the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, United States) and whole genome sequencing was carried out in paired-end on a MiSeq instrument using the 2 × 300-cycle version 3 kit as recommended by the manufacturer (Illumina, San Diego, CA, United States). Subsequently, quality of raw sequence data was checked using FastQC version 0.11.5 (Andrews et al., 2012). To clean up raw reads we excluded poor-quality and undersized reads by applying Trimmomatic (Bolger et al., 2014) with LEADING/TAILING 3 and MINLEN 36 as default parameters. Additionally, in case of read mapping and *de novo* assembly, we used SLIDINGWINDOW 4:15 and MAXINFO 15:0.8, respectively. The quality of trimmed reads was rechecked using FastQC.

Phylogenetic Analysis

Reference sequences were selected by the use of *refRank* version 3.0.0 (Becker et al., 2018) with a set of 142 complete *S. aureus* genome sequences available at NCBI (03/2017). *S. aureus* COL (MRSA, ST250, acc. no. NC_002951) and *S. aureus* ECT-R2 (MSSA, ST5, acc. no. NC_017343) were used as a reference strain for phylogenetic reconstruction of the DAP and CPT strain collections, respectively. As recombination events are well-known confounders of tree topology, we screened annotations of the reference genomes *S. aureus* COL and *S. aureus* ECT-R2 for mobile genetic elements and genes associated with drug resistance which were then removed (cut out) from the reference genomes (**Supplementary Table 3**). In the following, we will refer to these sequences as modified reference sequences. Subsequently, trimmed paired-end reads were read-aligned to the corresponding modified reference genomes using our in-house pipeline *batchMap* as described previously (Halbedel et al., 2018). Based on the generated multiple consensus alignment SNPs were filtered using *SNPFilter* version 3.0.0 (Becker et al., 2018) with an exclusion distance of 150 bp. By setting an exclusion

distance SNPs within spatial proximity (150 bp) to each other were excluded to additionally prevent recombination events to alter the phylogeny. Resulting SNP-alignments were used with Geneious version 8.1.7 (Biomatters Ltd., Auckland, New Zealand) to calculate neighbor joining consensus trees with bootstrap support (100 replicates). For illustration purpose iTOL v3.2.4 was used (Letunic and Bork, 2016).

Genome-Wide Association Studies (GWAS)

Genome Reconstruction and SNP Calling

For the identification of SNPs, trimmed Illumina reads of isolates belonging to the DAP strain collection were aligned to *S. aureus* COL by the use of *batchMap*. Since *S. aureus* ECT-R2 lacks SCC*mec*, *S. aureus* 04-02981 [acc. no. NC_017340, MRSA, ST225, SCC*mec* type II (2A)] was used as a reference sequence for CPT PLINK analysis. Alignments of consensus sequences were reduced to variant positions using *SNPFilter* without applying an exclusion distance. As SEER is an alignment-free method, high paired-end reads were *de novo* assembled using A5-miseq 20150522 with default parameters (Coil et al., 2015).

PLINK Analysis

Multi-sequence alignments were used to create input files for PLINK v1.90b3.31[2] (Purcell et al., 2007). Since PLINK is unable to utilize triallelic SNPs, we replaced the minor with the major variant. Prior to PLINK analysis, we screened for isolates with more than 10% missing genotypes and excluded SNPs on the basis of minor allele frequency (MAF $< 5\%$) or missing genotype data (call rate $< 90\%$). Furthermore, a Z score was calculated to detect population outliers, which were subsequently excluded from further analysis (standard deviation > 3; **Supplementary Table 4**). To account for population stratification, subpopulations within the overall population were identified with hierBAPS v6.0 (Cheng et al., 2013) and covariates were used with the Cochran-Mantel-Haenszel test (CMH) for $2 \times 2 \times K$ stratified tables. In order to search for the evidence of systematic bias a quantile-quantile (Q-Q) plot was constructed that compares the observed and expected *p*-value ($-\log_{10}$ transformed) under the null hypothesis of no true association. In case of substantial deviation of the observed *p*-values from the null hypothesis, we used genomic control (GC) to control for the confounding effects of population stratification. Finally, we created a Manhattan Plot illustrating the genome wide significance levels of corresponding SNPs in relation to their genome position. For illustration purposes we excluded SNPs with a $-\log_{10}$ *p*-value of 0.

Sequence Element Enrichment Analysis (SEER)

SEER identifies k-mers that are significantly enriched in a phenotype of interest and includes an alignment-free correction to account for population structure (Lees et al., 2016). For association analysis, we used SEER v.1.4.1. K-mers were counted from *de novo* assembled contigs by the use of the single-core implementation of frequency-based substring mining (fsm-lite; options -s 5 and -S 95)[3]. To create a matrix representing population structure, kmds (control for population structure) was performed with the *no filtering* option and a total of 200.000 k-mers. The matrix was then used to run SEER association statistics. Subsequently, we excluded k-mers with a MAF $< 5\%$, a negative beta ($ß_1 \leq 0$) and a length < 20 bp. Post-association filtered k-mers were used for the construction of Q-Q plots. In general, we used the likelihood ratio test (lrt) *p*-value for further downstream analysis. In case of strong population stratification, we preferred the Wald-test over the LRT. Those reaching the Bonferroni corrected significance thresholds were finally mapped to annotated reference genomes (see above).

Searching for Low Frequency and Rare SNPs

To detect rare mutations (MAF $< 5\%$) in candidate genes, we used ancestral state reconstruction in Mesquite v2.75[4] (Maddison and Maddison, 2017).

SCC*mec* Analysis

Elements of the staphylococcal cassette chromosome (SCC*mec*) were identified using SCC*mec*Finder v1.0[5] (Kaya et al., 2018). The web-based tool uses *de novo* assembled reads (contigs) and defines the distinct position of each SCC*mec* associated gene. Therefore, we were able to extract corresponding regions for each isolate which we subsequently aligned with progressive Mauve algorithm (Darling et al., 2004).

RESULTS AND DISCUSSION

Identification of DAP-R Associated Mutations in *S. aureus*

Phylogenetic Reconstruction and Prevention of Population Stratification

For the identification of DAP-R associated mutations, we established a strain collection containing 95 *S. aureus* isolates, of which 48 were resistant to DAP (**Supplementary Table 1**). The strain collection comprised 4 CCs and 17 STs that predominantly mirrored the distribution of CCs and STs in Europe and, in particular, in Germany (Grundmann et al., 2010; Asadollahi et al., 2018; **Supplementary Figure 1A**). Within most CCs and STs we observed a well-balanced distribution of resistance phenotypes, that could be achieved by selecting and matching isolates, whenever possible, according to their clonal background (ST/CC), isolation date and geographical origin (**Supplementary Figures 1A, 2**). In microbial GWAS, a homogeneous distribution of cases and controls reduces the risk of detecting spurious associations (associations that are due to relatedness rather than a true association with the phenotype of interest). These types of associations usually occur as a consequence of population stratification (PS) which refers to a situation in which members of a group of interest are on average more closely related

[2]https://www.cog-genomics.org/plink2

[3]https://github.com/nvalimak/fsm-lite

[4]http://mesquiteproject.org/

[5]https://cge.cbs.dtu.dk/services/SCCmecFinder-1.0/

to each other, than to the rest of the wider population (Chen and Shapiro, 2015).

To address systematic bias caused by population stratification in PLINK analysis, subpopulations within the overall population were inferred with hierBAPS (Cheng et al., 2013). Altogether, five hierBAPS sub-clusters could be identified that mirrored the phylogenetic relatedness of 95 *S. aureus* isolates (**Supplementary Figure 2**). These sub-clusters were then used as covariates with the CMH and a total of 66.667 SNPs (relative to the reference genome *S. aureus* COL), allowing to test for associations conditional on the population structure (Chen and Shapiro, 2015). This approach has already shown to reliably reflect the phylogenetic relatedness of bacterial populations and to efficiently control for population stratification in microbial GWAS (Chewapreecha et al., 2014; Howell et al., 2014; Chen and Shapiro, 2015; Weinert et al., 2015; Mobegi et al., 2017). To test whether systematic inflation occurred due to population stratification, we constructed a Q-Q plot and calculated the genomic inflation factor lambda (λ_{GC}). In a Q-Q plot, most *p*-values should follow a uniform distribution, with few SNPs producing significant *p*-values at the end of the line (Power et al., 2017). Furthermore, the λ_{GC} should be around 1 as values higher than 1.05 are seen as genome-wide inflation (Price et al., 2010). The constructed Q-Q plot of 14.525 adjusted PLINK variants revealed that most of the *p*-values fit the expected distribution ($\lambda_{GC} = 1$), with few SNPs deviating at the end of the tail (**Supplementary Figure 4A**).

In contrast to PLINK, SEER controls for clonal population structure by distance matrix construction and subsequent multidimensional scaling (Lees et al., 2016). This method is analogs to the principal component analysis (PCA) used in human association studies, but with the advantage of being directly applicable to k-mer counting instead of relying on core gene alignment or SNP calling (Price et al., 2010; Lees et al., 2016). With the ability to capture various types of genetic variations, SEER has successfully been used to identify genetic markers in *Burkholderia pseudomallei* (Chewapreecha et al., 2017) and to investigate host adaption of *Streptococcus suis* (Weinert et al., 2019). By applying fsm-lite we counted 10 M k-mers that were then tested for association. As recommended by Lees[6], we used the likelihood ratio test (lrt) *p*-value for further downstream analysis. The Q-Q plot of the resultant *p*-value distribution of 3 M quality-filtered k-mers demonstrated no severe deviation of the observed from the expected *p*-values (**Supplementary Figure 4C**), indicating that correction for clonal population structure was appropriate.

PLINK GWAS for the Identification of SNPs Associated With DAP Resistance

Out of 14.525 SNPs, PLINK identified MprF P314L (*mprF* locus tag: SACOL_RS07105, $p = 2.39 \times 10^{-8}$) and L826F ($p = 1.25 \times 10^{-6}$) to be associated with DAP resistance (**Figure 1**). Although not reaching the suggestive association threshold, we additionally detected MprF S337L ($p = 4.61 \times 10^{-5}$) (**Figure 1**). Both P314L and S337L are located in the central bifunctional domain of MprF that is involved in L-PG synthesis and flipping, while MprF L826F is found in the C-terminal catalytic synthase domain of the protein. In previous studies, these amino acid substitutions (AAS) were suspected to be linked with a gain-of-function phenotype, either in terms of increased synthesis or enhanced translocation of L-PG, resulting in the electrostatic repulsion of calcium-complexed DAP (Friedman et al., 2006; Bayer et al., 2013, 2015; Steed et al., 2013; Heidary et al., 2017). These findings were, however, not consistent and other seemingly MprF-independent mechanisms, such as the increase in D-alanylation of teichoic acid, have been described in clinical DAP-R isolates (Bertsche et al., 2011; Mishra et al., 2014). By engineering *mprF* substitution mutants in *S. aureus,* Ernst et al. (2018) showed that the *mprF* signature mutations S295L, P314L, S337L, I420N, and L826F had no impact on cell surface charge and were not sufficient to confer DAP resistance on their own. Therefore one may suggest that additive genetic variations contribute to DAP resistance that might have evolved prior to or acquired during antibiotic treatment as could be shown in early passaging experiments by Friedman et al. (2006). Interestingly, a recent study suggests that the *mprF* signature mutations might lead to an extended substrate spectrum of MprF by weakening the intramolecular interaction between the flippase and the synthase domain of the protein (Ernst and Peschel, 2019). These structural changes may enable the translocation of DAP or critical membrane-embedded molecules from fluid microdomains, thereby conferring DAP resistance. In a clinical context, although rarely reported, few case reports describe the *in vivo* acquisition of daptomycin resistance in *S. aureus* during daptomycin therapy with subsequent clinical failure for bacteremia/endocarditis that could be associated with mutations in *mprF* (including T345A and L826F) (Murthy et al., 2008; Sotillo et al., 2016). In addition, a study conducted by Richards et al. (2015) suggests *mprF* mutations to be involved in *S. aureus* persistence during complex bacteremia by enhancing bacterial fitness and immune evasion.

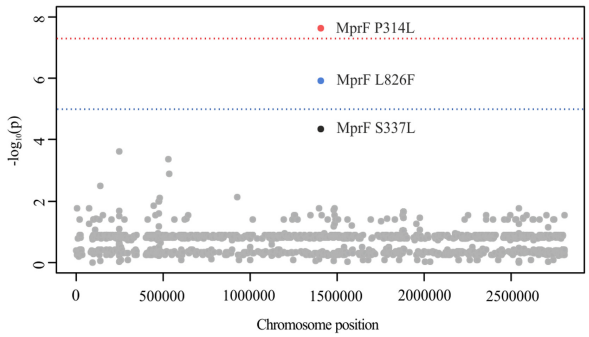


FIGURE 1 | Manhattan plot illustrating DAP PLINK results. The $-\log_{10}$ unadjusted *p*-values of identified SNPs are plotted against the entire length of the chromosome of *S. aureus* COL. The blue line indicates suggestive association ($p = 1 \times 10^{-5}$), while the red line indicates the genome-wide significance threshold ($p = 5 \times 10^{-8}$). For illustration purposes, SNPs with a $-\log_{10}$ *p*-value of 0 have been excluded ($n = 14.527$).

[6]GitHub. SEER. https://github.com/johnlees/seer/wiki/Usage [Accessed November 30, 2020].

SEER GWAS for the Identification of k-mers Associated With DAP Resistance

Due to the high degree of genome plasticity in bacterial populations as well as the need for reference sequences, PLINK is limited in the detection of causal genetic variations that are associated with differences in gene content, recombination events or variable promotor architectures (Lees et al., 2016). To overcome typical shortcomings of SNP based association tools, we additionally used SEER as an alignment free tool. This procedure is also recommended by San et al. (2019) who suggest testing for multiple forms of variations, especially when the type of variation responsible for the phenotype of interest is not known *a priori*.

Using SEER, we identified 198 significant k-mers that were mapped to the reference sequence of *S. aureus* COL. All k-mers were found to be located within *mprF* exhibiting either AAS P314L ($p = 1.2 \times 10^{-6}$) or L826F ($p = 2.4 \times 10^{-8}$) (**Figure 2**; Ng and Henikoff, 2003). The AAS P314L did not reach genome wide significance (**Figure 2**). This is most likely due to lineage specific mutations within *mprF* which prevent "identical" k-mers from accumulating at specific regions. Isolates from CC45 (carrying MprF P314L) have a significantly different *mprF* nucleotide sequence when compared to isolates of the wider population (**Supplementary Figure 7**). Therefore, k-mers harboring P314L were split into separate clusters, thus leading to a shift in significance levels. Additionally, we examined SEER results for additional MprF AAS and found S337L with k-mer *p*-values close to the suggestive association threshold ($p = 2.4 \times 10^{-5}$).

Distribution of MprF AAS

As can be seen from **Figure 3**, MprF P314L, S337L, and L826F were predominantly restricted to DAP-R isolates and showed a fairly homogeneous distribution among clonal clusters (**Figure 3**). The independent emergence of identical mutations on multiple ancestral branches points to a convergent evolution, which is a strong indicator for positive selection, e.g., by antibiotic pressure (Shapiro et al., 2009). Despite the even distribution, we observed a slight tendency toward an association of MprF AAS with certain clonal lineages (e.g., P314L with isolates of CC22 or L826F with isolates of CC5) (**Figure 3**). This phenomenon has already been described in a previous study in which the predominance of genetic modifications associated with the vancomycin-intermediate *S. aureus* (VISA) phenotype could be linked with specific genetic backgrounds (McGuinness et al., 2017). Thus, although associated mutations may explain the phenotype of interest in the studied population, the results cannot necessarily be applied to the wider population.

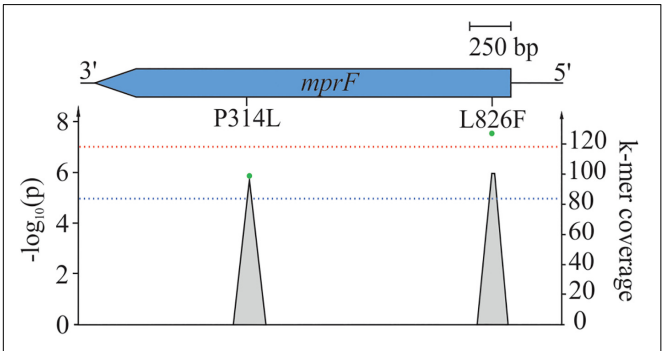


FIGURE 2 | Mapping of significant k-mers against *S. aureus* COL. One **hundred ninety-eight** significant k-mers were mapped against the reference *S. aureus* COL. Gray-shaded elements illustrate k-mer coverage at corresponding loci while green dots represent adjusted (lrt) *p*-values of mapped k-mers. The blue line indicates suggestive association ($p = 1 \times 10^{-5}$), while the red line indicates the genome-wide significance threshold ($p = 5 \times 10^{-8}$).

Searching for Rare Variants Associated With DAP Resistance

For some isolates no *mprF* mutations could be identified by GWAS ($n = 13$; **Figure 3**). Therefore, we used the software package Mesquite (Maddison and Maddison, 2017) to manually search for rare genetic variants in *mprF*, *cls2* (SACOL_RS10875), *rpoB* (SACOL_RS03050), *rpoC* (SACOL_RS03055), *yycF* (SACOL_RS00120), *yycG* (SACOL_RS00125), *pgsA* (SACOL_RS06645), and *ddl* (SACOL_RS10850), as these loci have previously been described to be associated with DAP resistance (Friedman et al., 2006; Peleg et al., 2012; Bayer et al., 2015; Berti et al., 2015).

At least one MprF related AAS could be identified in each DAP-R isolate (**Figure 3**), three of which, A315T, D317N, and L432F, have not been described until now. This observation emphasizes the importance of MprF in the development of DAP resistance in *S. aureus*. Additionally, our results suggest that *mprF* mutations do not accumulate (**Figure 3**) but rather occur as individual hot-spot mutations. These findings are supported by Yang et al. (2018) who showed that an accumulation of point mutations paradoxically caused a reduction in DAP MICs, positive surface charge and L-PG synthesis.

For Cls2 we detected the AAS I298T (15-00711) and S466P (15-01065) that are located in close proximity to the putative cadiolipin synthase domains at residues 229–256 and 407–434 (**Figure 3**; Peleg et al., 2012). AAS in Cls2 and varying levels of *cls2* transcription have already been suggested to be involved in the development of DAP resistance in *S. aureus*, either alone or in combination with *mprF* mutations (Camargo et al., 2008; Peleg et al., 2012; Jiang et al., 2019; Lasek-Nesselquist et al., 2019). Interestingly, within our strain collection we found a DAP-R isolate (MIC = 8 mg/L; median of all DAP-R isolates = 4 mg/L) harboring MprF S337L, Cls2 S466P, and RpoB Q468K (15-01065; **Figure 3**), indicating that a combination of mutations in different loci might lead to synergistic effects. The role of *cls2* in establishing DAP resistance was also supported by studies of clinical *Enterococcus* isolates in which *cls2* mutations could be associated with elevated DAP MICs (Arias et al., 2011; Palmer et al., 2011). These changes most likely result in an altered protein function and shifted PG:cardiolipin ratios in the bacterial cell membrane, thereby leading to impaired DAP penetration and membrane disruption (Miller et al., 2016).

As could be seen for the ST22 isolate 15-01065, *rpoB* mutations might be involved in the development of DAP resistance (**Figure 3**). In a study conducted by Friedman et al. (2006) multiple genetic changes in laboratory-derived *S. aureus* could

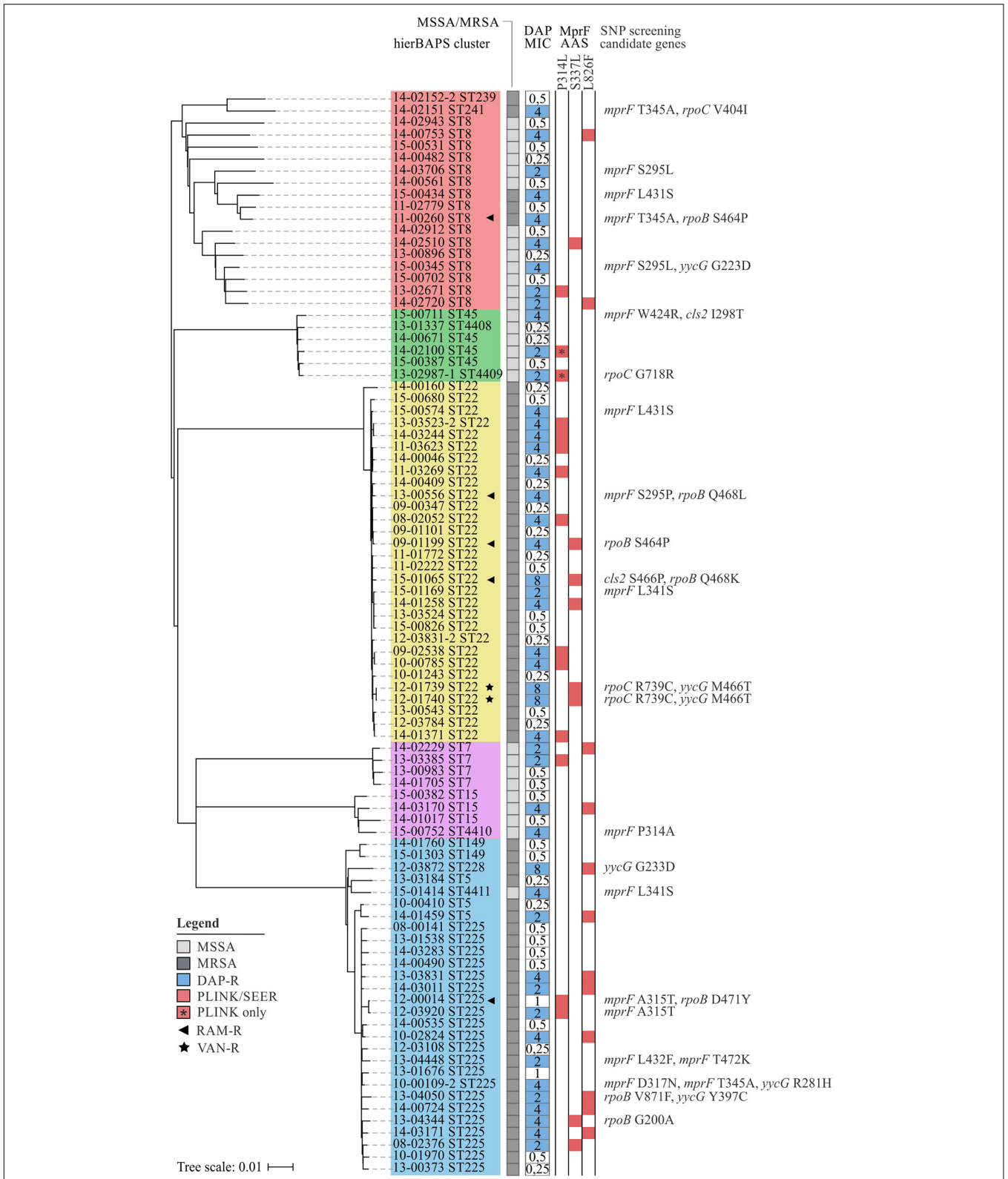


FIGURE 3 | Phylogenetic relation and genotypes of DAP-R *S. aureus* included in this study. On the left, a neighbor-joining tree based on 7.342 SNPs represents the phylogenetic relatedness of 90 *S. aureus* isolates. Identified hierBAPS subclusters are color-coded. Dark and light shaded grayish boxes represent MRSA and MSSA isolates, respectively. Blue boxes point to DAP resistance. Corresponding numbers refer to DAP MICs as determined by BMD. Red-shaded squares indicate the presence of AAS that could be identified by SEER and/or PLINK analysis. On the right, AAS in candidate genes are shown as identified by the use Mesquite analysis.

be identified that correlated with increased DAP MICs. After serial passaging, the authors observed not only mutations in *mprF, yycG* and *rpoC* but also *rpoB*. Furthermore, Cui et al. (2006) described a *rpoB*-mediated resistance conversion that was accompanied by an increase in DAP and vancomycin (VAN) MICs. In our study, we detected the RpoB AAS G200A (13-04344), S464P (11-00260, 09-01199), Q468K (15-01065), Q468L (13-00556), D471Y (12-00014), and V871F (13-04050) (**Figure 3**). While G200A and V871F have not been described until now S464P, Q468K/L, and D471Y are known to be located within the cluster 1 of the rifampicin resistance mutation site and could thus be associated with rifampicin resistance (**Figure 3** and **Supplementary Table 1**). A correlation of the latter mutations with DAP resistance is therefore unlikely (Wichelhaus et al., 1999).

Within our strain collection, we observed two DAP-R isolates that additionally were VAN resistant (VAN MIC = 4 mg/L) (**Figure 3**). Increased DAP MICs are often seen in VISA isolates that are commonly characterized by thickened cell walls (Cui et al., 2006; Julian et al., 2007). Because DAP is large in molecular size, thickened peptidoglycan layers may hinder the lipopeptide from reaching its antimicrobial target and might therefore facilitate the development of a dual resistance to VAN and DAP (Cui et al., 2006). In addition to mutations in *mprF*, both isolates showed genetic variations within *rpoC* (R739C) and *yycG* (M466T) (**Figure 3**). The two-component regulator YycFG (also known as WalKR) is known to be involved in the positive regulation of genes associated with cell wall metabolism. Several studies have shown that mutations within *yycFG* contribute to thickening of cell walls, which are thought to result from reduced expression of the major autolysins AtlA and LytM (Friedman et al., 2006; Howden et al., 2011; Patel et al., 2011; Shoji et al., 2011; Hafer et al., 2012). As could be seen for *rpoB* mutations, genetic variations in *rpoC* may indirectly influence the expression of genes involved in cell wall biosynthesis, thereby leading to altered cell surface charges and cell wall thickening (Cui et al., 2006). A recently published study described increased cell wall thickness for a clinical *S. aureus* isolate with cross-reduced susceptibility to DAP and VAN that was likely to be associated with *mprF* W424R (Thitiananpakorn et al., 2020). Although we detected this AAS in one DAP-R isolate (**Figure 3**), no increased VAN MICs could be observed (VAN MIC $\leq$ 1 mg/L). Since cell wall thickness has not been described as a common feature of clinical DAP-R isolates, Thitiananpakorn et al. (2020) further postulated that *mprF* mediated alterations in surface charges directly affect susceptibility to both DAP and VAN. These results are somewhat contradictory to our observations, as elevated VAN MICs were observed in only two DAP-R isolates (**Figure 3**). Consequently, it remains to be investigated to what extent *mprF* mutations—alone or in combination with *yycFG* mutations—are effectively related to this cross-resistance phenotype. Of note, as no data on therapy was available, we are unable to judge whether these mutations were due to previous VAN or DAP therapy in the affected patients.

Genes *ddl*, *pgsA*, and *yycF* exhibited no mutations previously associated with DAP resistance.

Interestingly, MSSA and MRSA isolates showed identical MprF AAS with no characteristic patterns detectable (**Figure 3**). In contrast to MRSA-related infections, physicians have numerous treatment options available for MSSA-related infections (David and Daum, 2017). Thus, MSSA isolates were unlikely to be exposed to DAP therapeutically. Consequently, MSSA might have undergone selection pressures that trigger the same mechanisms of resistance as DAP. Previous studies suggest, that both cationic host defense peptides (CHDPs) and DAP trigger the same mechanisms of resistance and that the exposure of *S. aureus* to CHDPs is likely to facilitate the development of DAP resistance (Mishra et al., 2011, 2012). In addition, Renzoni et al. (2017) showed that exposure of *S. aureus* to the antiseptic polyhexanide resulted in the selection of mutants possessing *mprF* mutations and thus in cross-resistance between the antiseptic agent and clinically used antibiotics.

Identification of CPT-R Associated Mutations in *S. aureus*

Phylogenetic Reconstruction and Detection of Population Stratification

The established strain collection comprised a total of 44 CPT-R and 43 CPT-S isolates (**Supplementary Table 2**) that belonged to 3 CCs and 15 STs (**Supplementary Figure 1B**). Similar to the DAP strain collection, the involved CCs and STs predominantly mirrored the distribution of clonal lineages in Europe and, more precisely, in Germany. As indicated by phylogenetic analysis, we observed the formation of distinct subclusters that lacked sufficient amounts of susceptible counterparts (**Supplementary Figure 3**). This is due to the fact, that isolates frequently had to be matched on the basis on CCs when no STs could be derived from *spa*-typing. Thus, although we observed a homogenous distribution of CPT resistance phenotypes within the overall CC5 clade (**Supplementary Figure 1**), the clustering of CPT-R isolates belonging to ST228 (CC5) and ST111 (CC5) was obscured by our sampling strategy (**Supplementary Figure 1**). Therefore, future studies should focus on adopting sampling strategies as described by Farhat et al., in order to both minimizing the impact of population structure and increasing the power of association studies (Farhat et al., 2014).

In order to identify SNPs associated with CPT resistance, a total of 45.989 SNPs (relative to *S. aureus* 04-02981) were analyzed with the CMH and five hierBAPS sub-clusters (**Supplementary Figure 3**). Although few studies described the CMH to occasionally overcorrect for population structure, we observed genome-wide inflation in the Q-Q plot of 6.543 adjusted PLINK variants (λ_{GC} = 2.11) (**Supplementary Figure 5A**) (Chen and Shapiro, 2015; Lees et al., 2016). Therefore, we used GC adjusted *p*-values for further downstream analysis (**Supplementary Figure 4B**). This method normalizes all *p*-values by the single inflation factor λ, which is the observed median chi-square divided by the expected median chi-square with 1 degree of freedom (Chen and Shapiro, 2015). The high inflation is most likely due to the close relationship of CPT-R isolates within the ST228 and ST111 clusters (**Supplementary Figure 3**). Previous studies have already shown that these lineages

are usually associated with elevated CPT MICs and that CPT resistance in ST228 isolates had already been observed prior to the introduction of CPT in the clinical setting (Kelley et al., 2015; Strommenger et al., 2015). Therefore, the implementation of successful clonal lineages poses a unique challenge for microbial GWAS, as these phenotypic lineage-level differences need to be accounted for Lees (2017) and Power et al. (2017). This is even more relevant when investigating highly clonal pathogens such as *Mycobacterium tuberculosis*, where the entire genome is in strong linkage, further preventing the fine-mapping of causal loci (Chen and Shapiro, 2015).

Using SEER, we counted 10 M k-mers of which 4.2 M remained after association testing and post-association filtering. As we detected an early separation of the observed from the expected lrt *p*-values in the constructed Q-Q plot (**Supplementary Figure 5B**), we favored the Wald- of lrt-correction. Although most of the Wald-corrected *p*-values follow a uniform distribution, GWAS power was reduced significantly (**Supplementary Figure 4D**).

Population Stratification Prevents GWAS From Identifying Causal Variants

Out of 6.543 SNPs, PLINK analysis identified GrlA S80F (GC adjusted $p = 9.05 \times 10^{-8}$, locus tag: SA2981_RS06930) to be suggestively associated with CPT resistance (**Figure 4**). Close to the genome-wide significance threshold, we additionally observed GyrA S84L (GC adjusted $p = 3.71 \times 10^{-7}$, locus tag: SA2981_RS00030) (**Figure 4**). As these mutations are known to mediate ciprofloxacin (CIP) resistance in *S. aureus* (Jones et al., 2000; Kwak et al., 2013), we screened all isolates for the corresponding resistance phenotype. Indeed, we identified 95% ($n = 42$) of CPT-R and 40% ($n = 16$) of CPT-S isolates to be CIP-R, carrying either S80F or S84L (**Supplementary Table 5**). These observations are in line with results of previous studies, showing that dominant hospital-associated MRSA lineages are almost universally resistant to CIP and that this resistance phenotype contributes to the selection and survival of *S. aureus* (Knight et al., 2012; Layer et al., 2019). These results stress the importance of carefully verifying putative variants in order to avoid the detection of false positive correlations caused by genetically linked features as can be frequently seen in hospital-associated pathogens like *S. aureus*, *E. coli*, *E. faecium*, *M. tuberculosis*, and *A. baumanii* (Struelens, 1998; Cornejo-Juarez et al., 2015; Tacconelli et al., 2018). Using SEER analysis, 10 M k-mers were tested for CPT-R association. However, due to the stringent correction for population stratification, no significant k-mers remained ($p < 1 \times 10^{-5}$).

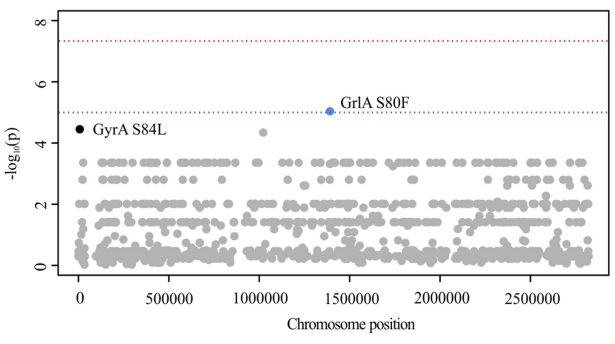


FIGURE 4 | Manhattan plot illustrating CPT PLINK results. The $-\log_{10}$ GC-corrected *p*-values of identified SNPs are plotted against the chromosome position of *S. aureus* 04-02981. The blue line indicates suggestive association ($p = 1 \times 10^{-5}$), while the red line indicates genome-wide significance threshold ($p = 5 \times 10^{-8}$).

Searching for Rare Variants in Essential SCC*mec* Elements

To further search for genetic variants that are putatively associated with elevated CPT MICs, the nucleotide sequences of essential SCC*mec* elements (*mecA, mecR1,* Δ*mecRI*, and *mecI*) were extracted from *de novo* assembled contigs. Detailed sequence analysis revealed multiple SNPs within *mecA* (D139N, N146K, E150K, N204K, T235I, E239K/R, G246E, and K281R; **Figure 5**), but only few were found within the regulatory genes *mecR1*, Δ*mecRI*, and *mecI* (**Supplementary Table 6**). The majority of AAS observed in *mecA* were restricted to the non-penicillin-binding (nPBD) domain of PBP2a (AAS 27–326) (Lim and Strynadka, 2002). These observations are in line with investigations of Lahiri et al. (2015) who found CPT-R associated substitutions to be predominantly located in or close to a structural groove of the nPBD (including D139N, N146K, E150K, N204K, T235I, and E239K). Already back in 2013, Otero et al. (2013) have shown that CPT can bind to this allosteric domain, resulting in a multiresidue conformational change of PBP2a, the opening of the active site and the irreversible acylation of the active serine by a second CPT molecule. Consequently, it has been presumed that mutations within the allosteric domain of PBP2a can lead to a disruption of salt bridge interactions that are crucial for the allosteric response, leaving the active site occluded (Otero et al., 2013). Furthermore, Alm et al. (2014) postulated that mutations in the nPBD of PBP2a might lead to a destabilization of protein-protein interactions between PBP2a and PBP2, presumably promoting the interaction with alternative PBPs, like PBP4, which has been shown to have a low affinity for CPT (Moisan et al., 2010). The *mecA* AAS G246E could not only be detected within CPT-R ($n = 13$) but also CPT-S ($n = 7$) isolates (**Figure 5**). Consequently, we hypothesize that this AAS on its own is unlikely to be involved in the development of CPT resistance in *S. aureus.* This assumption is supported by previous studies, showing that *S. aureus* with *mecA* G246E did not exhibit increased CPT MICs (Alm et al., 2014; Lahiri et al., 2015; Schaumburg et al., 2016). The G246E AAS in particular demonstrates that identical *mecA* mutations are often associated not only with one but several SCC*mec* types (**Figure 5**), suggesting that these mutations emerged after the acquisition of SCC*mec* rather than being linked with the dissemination of specific SCC*mec* types. The only change that is located near the cephalosporin-binding pocket of the penicillin-binding domain of PBP2a is the AAS E447K (**Figure 5**). This substitution is presumed to directly influence the binding of CPT by forming a new salt bridge with the neighboring $Gluc_{460}$ residue (Alm et al., 2014). The diversity of *mecA* mutations clearly illustrates

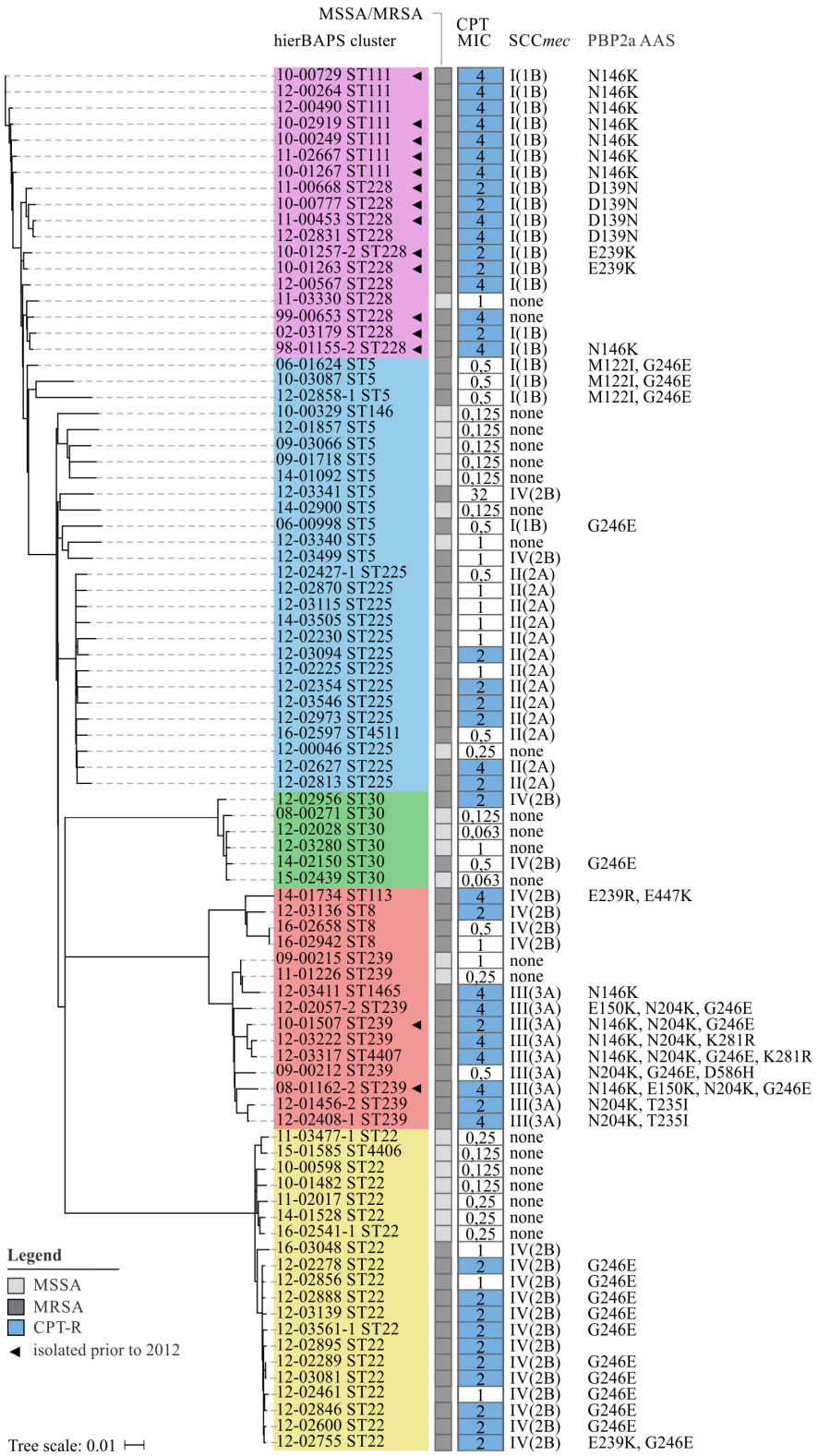


FIGURE 5 | Phylogenetic relation and genotypes of CPT-R *S. aureus* included in this study. On the left, a neighbor-joining tree based on 4.344 SNPs represents the phylogenetic relatedness of 86 *S. aureus* isolates. Identified hierBAPS subclusters are color-coded. Dark and light shaded grayish boxes represent MRSA and MSSA isolates, respectively. Blue boxes point to CPT resistance. Corresponding numbers refer to CPT MICs as determined by BMD. On the right, AAS in PBP2a are shown as identified by the use of manual *in silico* analysis.

why SEER was limited in the detection of significant associations. In this case, k-mers were split into clusters of lower frequency, which drastically reduced significance levels and prevented the detection of corresponding mutations.

For some isolates we neither detected mutations within *mecA* nor within the regulatory elements of SCC*mec* (**Figure 5**). Therefore, we performed *in silico* analysis to search for mutations in *pbp1* (SACOL_RS06115), *pbp2* (SACOL_RS07590), *pbp3* (SACOL_RS08205), *pbp4* (SACOL_RS03595), *gdpP* (SACOL_RS00090), *arcB* (SACOL_RS13910), *pp2C* (SACOL_RS10765), *clpX* (SACOL_RS08790), and *rho* (SACOL_RS11055), as these loci have either been described to bind ceftaroline or to play a functional role in the development of β-lactam resistance in *S. aureus* (Banerjee et al., 2010; Moisan et al., 2010; Chan et al., 2015; Greninger et al., 2016). However, with the exception of PBP1 P360S and PBP2 A382T (12-02627; ST225), no putative resistance-associated AAS could be detected. Thus, still unknown genetic variants seem to influence the development of CPT resistance in *S. aureus* which remain to be identified.

We detected 15 CPT-R *S. aureus* isolates that were collected prior to the licensing of CPT in Europe which mainly belonged to clonal lineages ST111 and ST228 (**Figure 5**). Of these, eight isolates have already been described in a study by Strommenger et al. (2015). Also Kelley et al. (2015) reporteded CPT resistance in *S. aureus* at least as far back as 1998 (predominantly in ST228). Therefore it is likely that these lineages express CPT resistance as a result of natural variation (and thus by chance) and/or were selected by environmental factors other than CPT. Since CPT itself has a low potential to select for resistance, Kelley et al. (2015) also speculated that other agents like β-lactam antibiotics may have contributed to the selection of this resistance phenotype (Knight et al., 2012; Mobegi et al., 2017). Supporting this hypothesis, we observed a positive correlation between the levels of CPT and OXA MICs ($p < 0.001$; $R2 = 0.065$; **Supplementary Figure 6**).

CONCLUSION

With this study, we confirmed the potential of microbial GWAS to identify genetic variants that are significantly associated with antibiotic resistance in *S. aureus*. However, due to the clonal population structure of bacterial populations, it remains challenging to control for the detection of spurious associations. Therefore, profound sampling strategies are necessary to investigate the genetic architecture of bacterial phenotypes. The ongoing development and refinement of new and existing tools will enable researchers to account for stratification more precisely and to use multiple variants in association testing. A key challenge in microbial GWAS has been and still is the requirement for larger and more complex sample collections but is often restricted by the number of well-characterized clinical isolates available. A large collection not only increases the statistical power of association studies but also enables the investigation of traits that are linked to lineage-specific features and/or variants that occur with a low frequency.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ebi.ac.uk/ena, PRJEB41643.

AUTHOR CONTRIBUTIONS

BS, GW, and SF: conception and design of the study. RW: strain characterization and manuscript writing. RW, JB, and AT: WGS sequencing. RW and AS: data curation and analyses. All authors: manuscript editing and reviewing.

FUNDING

This study was partially supported with a Ph.D. grant of the Robert Koch Institute. The German Reference Centre for Staphylococci and Enterococci was funded by the German Federal Ministry of Health.

ACKNOWLEDGMENTS

We thank Mike Henkel, Franziska Erdmann, Birgit Pasemann, Petra Vilbrandt, and the staff at the RKI central sequencing lab for excellent technical assistance. We are grateful to a network of collaborating laboratories for providing strains continuously.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.639660/full#supplementary-material

Supplementary Figure 1 | Distribution of resistance phenotypes for the DAP **(A)** and CPT **(B)** strain collection. Clonal lineages with corresponding numbers of included isolates are shown below. Resistant strains are highlighted as dark shaded columns, while susceptible isolates are shown as light shaded columns. The distribution of phenotypes is given in percent. DAP-R, daptomycin-resistant; DAP-S, daptomycin-susceptible; CPT-R, ceftaroline-resistant; CPT-S, ceftaroline-susceptible; ST, sequence type; CC, clonal cluster.

Supplementary Figure 2 | Phylogenetic analysis of 95 *S. aureus* strains. Neighbor-Joining tree based on 7.342 SNPs describing the linear evolution among 95 *S. aureus* isolates. SNPs were called with an exclusion distance of 150 bp. Modified NC-002951 was used as a reference genome. hierBAPS clusters are represented by color-shaded boxes. Bold isolate identifiers indicate resistance towards DAP. Bootstraps are color-coded (minimum = red, maximum = green).

Supplementary Figure 3 | Phylogenetic analysis of 87 *S. aureus* isolates. Neighbor-Joining tree based on 4.344 SNPs describing the linear evolution among 87 *S. aureus* isolates. SNPs were called with an exclusion distance of 150 bp. Modified NC-017343 was used as a reference genome. hierBAPS clusters are represented by color-shaded boxes. Bold isolate identifiers indicate resistance against CPT. Bootstraps are color-coded (minimum = red, maximum = green).

Supplementary Figure 4 | Q-Q plot illustrating population stratification in GWAS. The plot compares the observed and expected *p*-value under the null hypothesis of no true association. Deviation from the X = Y reference line indicates the presence of a systematic bias. **(A)** DAP Q-Q plot with unadjusted PLINK *p*-values.

(B) CPT Q-Q plot with GC-adjusted PLINK *p*-values. **(C)** DAP Q-Q plot with lrt corrected SEER *p*-values. **(D)** CPT Q-Q plot with Wald-corrected SEER *p*-values.

Supplementary Figure 5 | Q-Q plot illustrating population stratification CPT GWAS. The plot compares the observed and expected *p*-value under the null hypothesis of no true association. Deviation from the X = Y reference line indicates the presence of a systematic bias. CPT Q-Q plot with unadjusted PLINK *p*-values **(A)** and lrt corrected SEER *p*-values **(B)**.

Supplementary Figure 6 | Linear regression model fitted on the $\log_2$ values of CPT and OXA MICs of 98 *S. aureus* isolates (p-value < 0.001; $R^2 = 0.065$). The horizontal and vertical dotted lines represent the OXA and CPT breakpoint for *S. aureus* respectively (according to EUCAST v. 7.1; OXA: S $\leq$ 2 mg/L, R $>$ 2 mg/L; CPT: S $\leq$ 1 mg/L, R $>$ 1 mg/L).

Supplementary Figure 7 | Mauve genome alignment for the illustration of *mprF* sequence similarity. To illustrate the *mprF* sequence similarity between different clonal lineages, one representative DAP-R isolate was chosen from each *hier*BAPS subcluster. The *mprF* sequence was identified and extracted from *de novo* assembled contigs. To compute a genome alignment, we used Mauve and Geneious Prime v. 11.0.4. The following isolates were used: 14-02100 (ST45), 10-00785 (ST22), 13-04344 (ST225), 13-02671 (ST8), and 13-03385 (ST7). Nucleotide differences are highlighted as dark shaded boxes.

Supplementary Table 1 | Details on the daptomycin strain collection.

Supplementary Table 2 | Details on the ceftaroline strain collection.

Supplementary Table 3 | Genetic elements excised from the reference sequences of *S. aureus* COL and *S. aureus* ECT-R2.

Supplementary Table 4 | PLINK defined population outliers.

Supplementary Table 5 | Detection of GrlA S80F and GyrA S84L.

Supplementary Table 6 | Detection of *mecI*, Δ*mecRI*, and *mecRI* mutations.

REFERENCES

Alam, M. T., Petit, R. A. III, Crispell, E. K., Thornton, T. A., Conneely, K. N., Jiang, Y., et al. (2014). Dissecting vancomycin-intermediate resistance in *staphylococcus aureus* using genome-wide association. *Genome Biol. Evol.* 6, 1174–1185. doi: 10.1093/gbe/evu092

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

Alm, R. A., McLaughlin, R. E., Kos, V. N., Sader, H. S., Iaconis, J. P., and Lahiri, S. D. (2014). Analysis of *Staphylococcus aureus* clinical isolates with reduced susceptibility to ceftaroline: an epidemiological and structural perspective. *J. Antimicrob. Chemother.* 69, 2065–2075. doi: 10.1093/jac/dku114

Andrews, S., Krueger, F., Segonds-Pichon, F., Biggins, L., Krueger, C., and Wingett, S. (2012). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Cambridge: Babraham Institude.

Arias, C. A., Panesso, D., McGrath, D. M., Qin, X., Mojica, M. F., Miller, C., et al. (2011). Genetic basis for in vivo daptomycin resistance in enterococci. *N. Engl. J. Med.* 365, 892–900. doi: 10.1056/nejmoa1011138

Asadollahi, P., Farahani, N. N., Mirzaii, M., Khoramrooz, S. S., van Belkum, A., Asadollahi, K., et al. (2018). Distribution of the most prevalent spa types among clinical isolates of methicillin-resistant and -susceptible *Staphylococcus aureus* around the world: a review. *Front. Microbiol.* 9:163. doi: 10.3389/fmicb.2018.00163

AstraZeneca, (2012). *AstraZeneca Plc (AZN) Granted Marketing Authorisation in Europe for Zinforo. Clinical Space News*. Cambridge: AstraZeneca.

Banerjee, R., Gretes, M., Harlem, C., Basuino, L., and Chambers, H. F. (2010). A mecA-negative strain of methicillin-resistant *Staphylococcus aureus* with high-level beta-lactam resistance contains mutations in three genes. *Antimicrob. Agents Chemother.* 54, 4900–4902. doi: 10.1128/aac.00594-10

Bayer, A. S., Mishra, N. N., Chen, L., Kreiswirth, B. N., Rubio, A., and Yang, S. J. (2015). Frequency and distribution of single-nucleotide polymorphisms within mprf in methicillin-resistant *Staphylococcus aureus* clinical isolates and their role in cross-resistance to daptomycin and host defense antimicrobial peptides. *Antimicrob. Agents Chemother.* 59, 4930–4937. doi: 10.1128/aac.00970-15

Bayer, A. S., Schneider, T., and Sahl, H. G. (2013). Mechanisms of daptomycin resistance in *Staphylococcus aureus*: role of the cell membrane and cell wall. *Ann. N. Y. Acad. Sci.* 1277, 139–158. doi: 10.1111/j.1749-6632.2012.06819.x

Becker, L., Fuchs, S., Pfeifer, Y., Semmler, T., Eckmanns, T., Korr, G., et al. (2018). Whole genome sequence analysis of CTX-M-15 Producing Klebsiella Isolates allowed dissecting a polyclonal outbreak scenario. *Front. Microbiol.* 9:322. doi: 10.3389/fmicb.2018.00322

Berti, A. D., Baines, S. L., Howden, B. P., Sakoulas, G., Nizet, V., Proctor, R. A., et al. (2015). Heterogeneity of genetic pathways toward daptomycin nonsusceptibility in *Staphylococcus aureus* determined by adjunctive antibiotics. *Antimicrob. Agents Chemother.* 59, 2799–2806. doi: 10.1128/aac.04990-14

Bertsche, U., Weidenmaier, C., Kuehner, D., Yang, S. J., Baur, S., Wanner, S., et al. (2011). Correlation of daptomycin resistance in a clinical *Staphylococcus aureus* strain with increased cell wall teichoic acid production and D-alanylation. *Antimicrob. Agents Chemother.* 55, 3922–3928. doi: 10.1128/aac.01226-10

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

Bortolaia, V., Kaas, R. S., Ruppe, E., Roberts, M. C., Schwarz, S., Cattoir, V., et al. (2020). ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.* 75, 3491–3500. doi: 10.1093/jac/dkaa345

Bruniera, F. R., Ferreira, F. M., Saviolli, L. R., Bacci, M. R., Feder, D., da Luz Goncalves Pedreira, M., et al. (2015). The use of vancomycin with its therapeutic and adverse effects: a review. *Eur. Rev. Med. Pharmacol. Sci.* 19, 694–700.

Camargo, I. L., Neoh, H. M., Cui, L., and Hiramatsu, K. (2008). Serial daptomycin selection generates daptomycin-nonsusceptible *Staphylococcus aureus* strains with a heterogeneous vancomycin-intermediate phenotype. *Antimicrob. Agents Chemother.* 52, 4289–4299. doi: 10.1128/aac.00417-08

Chan, L. C., Basuino, L., Diep, B., Hamilton, S., Chatterjee, S. S., and Chambers, H. F. (2015). Ceftobiprole- and ceftaroline-resistant methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 59, 2960–2963. doi: 10.1128/aac.05004-14

Chen, P. E., and Shapiro, B. J. (2015). The advent of genome-wide association studies for bacteria. *Curr. Opin. Microbiol.* 25, 17–24. doi: 10.1016/j.mib.2015.03.002

Cheng, L., Connor, T. R., Siren, J., Aanensen, D. M., and Corander, J. (2013). Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol. Biol. Evol.* 30, 1224–1228. doi: 10.1093/molbev/mst028

Chewapreecha, C., Holden, M. T., Vehkala, M., Valimaki, N., Yang, Z., Harris, S. R., et al. (2017). Global and regional dissemination and evolution of Burkholderia pseudomallei. *Nat. Microbiol.* 2:16263.

Chewapreecha, C., Marttinen, P., Croucher, N. J., Salter, S. J., Harris, S. R., Mather, A. E., et al. (2014). Comprehensive identification of single nucleotide polymorphisms associated with beta-lactam resistance within pneumococcal mosaic genes. *PLoS Genet.* 10:e1004547. doi: 10.1371/journal.pgen.1004547

Coil, D., Jospin, G., and Darling, A. E. (2015). A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics.* 31, 587–589. doi: 10.1093/bioinformatics/btu661

Cornejo-Juarez, P., Vilar-Compte, D., Perez-Jimenez, C., Namendys-Silva, S. A., Sandoval-Hernandez, S., and Volkow-Fernandez, P. (2015). The impact of hospital-acquired infections with multidrug-resistant bacteria in an oncology intensive care unit. *Int. J. Infect. Dis.* 31, 31–34. doi: 10.1016/j.ijid.2014.12.022

Cui, L., Tominaga, E., Neoh, H. M., and Hiramatsu, K. (2006). Correlation between reduced daptomycin susceptibility and vancomycin resistance in vancomycin-intermediate *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 50, 1079–1082. doi: 10.1128/aac.50.3.1079-1082.2006

Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704

David, M. Z., and Daum, R. S. (2017). Treatment of *Staphylococcus aureus* Infections. *Curr. Top Microbiol. Immunol.* 409, 325–383. doi: 10.1007/82_2017_42

Ernst, C. M., and Peschel, A. (2019). MprF-mediated daptomycin resistance. *Int. J. Med. Microbiol.* 309, 359–363. doi: 10.1016/j.ijmm.2019.05.010

Ernst, C. M., Slavetinsky, C. J., Kuhn, S., Hauser, J. N., Nega, M., Mishra, N. N., et al. (2018). Gain-of-function mutations in the phospholipid flippase mprf confer specific daptomycin resistance. *mBio.* 9:e01659-18. doi: 10.1128/mBio.01659-18

Falush, D., and Bowden, R. (2006). Genome-wide association mapping in bacteria? *Trends Microbiol.* 14, 353–355. doi: 10.1016/j.tim.2006.06.003

Farhat, M. R., Freschi, L., Calderon, R., Ioerger, T., Snyder, M., Meehan, C. J., et al. (2019). GWAS for quantitative resistance phenotypes in Mycobacterium tuberculosis reveals resistance genes and regulatory regions. *Nat. Commun.* 10:2128.

Farhat, M. R., Shapiro, B. J., Kieser, K. J., Sultana, R., Jacobson, K. R., Victor, T. C., et al. (2013). Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. *Nat. Genet.* 45, 1183–1189. doi: 10.1038/ng.2747

Farhat, M. R., Shapiro, B. J., Sheppard, S. K., Colijn, C., and Murray, M. (2014). A phylogeny-based sampling strategy and power calculator informs genome-wide associations study design for microbial pathogens. *Genome Med.* 6:101.

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

Friedman, L., Alder, J. D., and Silverman, J. A. (2006). Genetic changes that correlate with reduced susceptibility to daptomycin in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 50, 2137–2145. doi: 10.1128/aac.00039-06

Gonzalez-Ruiz, A., Beiras-Fernandez, A., Lehmkuhl, H., Seaton, R. A., Loeffler, J., and Chaves, R. L. (2011). Clinical experience with daptomycin in Europe: the first 2.5 years. *J. Antimicrob. Chemother.* 66, 912–919. doi: 10.1093/jac/dkq528

Gray, D. A., and Wenzel, M. (2020). More than a pore: a current perspective on the In Vivo mode of action of the lipopeptide antibiotic daptomycin. *Antibiotics (Basel)* 9:17. doi: 10.3390/antibiotics9010017

Greninger, A. L., Chatterjee, S. S., Chan, L. C., Hamilton, S. M., Chambers, H. F., and Chiu, C. Y. (2016). Whole-genome sequencing of methicillin-resistant *Staphylococcus aureus* resistant to fifth-generation cephalosporins reveals potential Non-mecA mechanisms of resistance. *PLoS One* 11:e0149541. doi: 10.1371/journal.pone.0149541

Grundmann, H., Aanensen, D. M., van den Wijngaard, C. C., Spratt, B. G., Harmsen, D., Friedrich, A. W., et al. (2010). Geographic distribution of *Staphylococcus aureus* causing invasive infections in Europe: a molecular-epidemiological analysis. *PLoS Med.* 7:e1000215. doi: 10.1371/journal.pmed.1000215

Hafer, C., Lin, Y., Kornblum, J., Lowy, F. D., and Uhlemann, A. C. (2012). Contribution of selected gene mutations to resistance in clinical isolates of vancomycin-intermediate *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 56, 5845–5851. doi: 10.1128/aac.01139-12

Halbedel, S., Prager, R., Fuchs, S., Trost, E., Werner, G., and Flieger, A. (2018). Whole-genome sequencing of recent listeria monocytogenes isolates from Germany reveals population structure and disease clusters. *J. Clin. Microbiol.* 56, e00119–18.

Heidary, M., Khosravi, A. D., Khoshnood, S., Nasiri, M. J., Soleimani, S., and Goudarzi, M. (2017). Daptomycin. *J. Antimicrob. Chemother.* 73, 1–11.

Holden, M. T., Hsu, L. Y., Kurt, K., Weinert, L. A., Mather, A. E., Harris, S. R., et al. (2013). A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res.* 23, 653–664.

Howden, B. P., McEvoy, C. R., Allen, D. L., Chua, K., Gao, W., Harrison, P. F., et al. (2011). Evolution of multidrug resistance during *Staphylococcus aureus* infection involves mutation of the essential two component regulator WalKR. *PLoS Pathog.* 7:e1002359. doi: 10.1371/journal.ppat.1002359

Howell, K. J., Weinert, L. A., Chaudhuri, R. R., Luan, S. L., Peters, S. E., Corander, J., et al. (2014). The use of genome wide association methods to investigate pathogenicity, population structure and serovar in Haemophilus parasuis. *BMC Genomics* 15:1179. doi: 10.1186/1471-2164-15-1179

Jiang, J. H., Bhuiyan, M. S., Shen, H. H., Cameron, D. R., Rupasinghe, T. W. T., Wu, C. M., et al. (2019). Antibiotic resistance and host immune evasion in *Staphylococcus aureus* mediated by a metabolic adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 116, 3722–3727.

Jones, M. E., Boenink, N. M., Verhoef, J., Kohrer, K., and Schmitz, F. J. (2000). Multiple mutations conferring ciprofloxacin resistance in *Staphylococcus aureus* demonstrate long-term stability in an antibiotic-free environment. *J. Antimicrob. Chemother.* 45, 353–356. doi: 10.1093/jac/45.3.353

Julian, K., Kosowska-Shick, K., Whitener, C., Roos, M., Labischinski, H., Rubio, A., et al. (2007). Characterization of a daptomycin-nonsusceptible vancomycin-intermediate *Staphylococcus aureus* strain in a patient with endocarditis. *Antimicrob. Agents Chemother.* 51, 3445–3448. doi: 10.1128/aac.00559-07

Kaya, H., Hasman, H., Larsen, J., Stegger, M., Johannesen, T. B., Allesoe, R. L., et al. (2018). SCCmecFinder, a web-based tool for typing of staphylococcal cassette chromosome mec in *Staphylococcus aureus* using whole-genome sequence data. *mSphere* 3, e00612–17.

Kelley, W. L., Jousselin, A., Barras, C., Lelong, E., and Renzoni, A. (2015). Missense mutations in PBP2A Affecting ceftaroline susceptibility detected in epidemic hospital-acquired methicillin-resistant *Staphylococcus aureus* clonotypes ST228 and ST247 in Western Switzerland archived since 1998. *Antimicrob. Agents Chemother.* 59, 1922–1930. doi: 10.1128/aac.04068-14

Knight, G. M., Budd, E. L., Whitney, L., Thornley, A., Al-Ghusein, H., Planche, T., et al. (2012). Shift in dominant hospital-associated methicillin-resistant *Staphylococcus aureus* (HA-MRSA) clones over time. *J. Antimicrob. Chemother.* 67, 2514–2522. doi: 10.1093/jac/dks245

Kosowska-Shick, K., McGhee, P. L., and Appelbaum, P. C. (2010). Affinity of ceftaroline and other beta-lactams for penicillin-binding proteins from *Staphylococcus aureus* and Streptococcus pneumoniae. *Antimicrob. Agents Chemother.* 54, 1670–1677. doi: 10.1128/aac.00019-10

Kwak, Y. G., Truong-Bolduc, Q. C., Bin Kim, H., Song, K. H., Kim, E. S., and Hooper, D. C. (2013). Association of norB overexpression and fluoroquinolone resistance in clinical isolates of *Staphylococcus aureus* from Korea. *J. Antimicrob. Chemother.* 68, 2766–2772. doi: 10.1093/jac/dkt286

Laabei, M., Recker, M., Rudkin, J. K., Aldeljawi, M., Gulay, Z., Sloan, T. J., et al. (2014). Predicting the virulence of MRSA from its genome sequence. *Genome Res.* 24, 839–849. doi: 10.1101/gr.165415.113

Lahiri, S. D., McLaughlin, R. E., Whiteaker, J. D., Ambler, J. E., and Alm, R. A. (2015). Molecular characterization of MRSA isolates bracketing the current EUCAST ceftaroline-susceptible breakpoint for *Staphylococcus aureus*: the role of PBP2a in the activity of ceftaroline. *J. Antimicrob. Chemother.* 70, 2488–2498. doi: 10.1093/jac/dkv131

Lasek-Nesselquist, E., Lu, J., Schneider, R., Ma, Z., Russo, V., Mishra, S., et al. (2019). Insights into the evolution of *Staphylococcus aureus* daptomycin resistance from an in vitro bioreactor model. *Front. Microbiol.* 10:345. doi: 10.3389/fmicb.2019.00345

Layer, F., Strommenger, B., Cuny, C., Noll, S., Klingenberg, A., and Werner, G. H. (2019). äufigkeit und verbreitung von MRSA in Deutschland – update 2017/2018. *Epid. Bull.* 42, 437–444.

Lees, J. (2017). The background of bacterial GWAS. figshare. Thesis. doi: 10.6084/m9.figshare.5550037.v1

Lees, J. A., Vehkala, M., Valimaki, N., Harris, S. R., Chewapreecha, C., Croucher, N. J., et al. (2016). Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. *Nat. Commun.* 7:12797.

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245.

Lim, D., and Strynadka, N. C. (2002). Structural basis for the beta lactam resistance of PBP2a from methicillin-resistant *Staphylococcus aureus*. *Nat. Struct. Biol.* 9, 870–876.

Maddison, W. P., and Maddison, D. R. (2017). *Mesquite: A Modular System for Evolutionary Analysis.* Version 2.75. Available online at: http://mesquiteproject.org (accessed May 29, 2017).

McGuinness, W. A., Malachowa, N., and DeLeo, F. R. (2017). Vancomycin Resistance in *Staphylococcus aureus*. *Yale J. Biol. Med.* 90, 269–281.

Meric, G., Mageiros, L., Pensar, J., Laabei, M., Yahara, K., Pascoe, B., et al. (2018). Disease-associated genotypes of the commensal skin bacterium Staphylococcus epidermidis. *Nat. Commun.* 9:5034.

Micek, S. T. (2007). Alternatives to vancomycin for the treatment of methicillin-resistant *Staphylococcus aureus* infections. *Clin. Infect. Dis.* 45(Suppl. 3), S184–S190.

Miller, W. R., Bayer, A. S., and Arias, C. A. (2016). Mechanism of action and resistance to daptomycin in *Staphylococcus aureus* and enterococci. *Cold Spring Harb. Perspect. Med.* 6:a026997. doi: 10.1101/cshperspect.a026997

Mishra, N. N., Bayer, A. S., Moise, P. A., Yeaman, M. R., and Sakoulas, G. (2012). Reduced susceptibility to host-defense cationic peptides and daptomycin coemerge in methicillin-resistant *Staphylococcus aureus* from daptomycin-naive bacteremic patients. *J. Infect. Dis.* 206, 1160–1167. doi: 10.1093/infdis/jis482

Mishra, N. N., Bayer, A. S., Weidenmaier, C., Grau, T., Wanner, S., Stefani, S., et al. (2014). Phenotypic and genotypic characterization of daptomycin-resistant methicillin-resistant *Staphylococcus aureus* strains: relative roles of mprF and dlt operons. *PLoS One* 9:e107426. doi: 10.1371/journal.pone.0107426

Mishra, N. N., McKinnell, J., Yeaman, M. R., Rubio, A., Nast, C. C., Chen, L., et al. (2011). *In vitro* cross-resistance to daptomycin and host defense cationic antimicrobial peptides in clinical methicillin-resistant *Staphylococcus aureus* isolates. *Antimicrob. Agents Chemother.* 55, 4012–4018. doi: 10.1128/aac.00223-11

Mobegi, F. M., Cremers, A. J., de Jonge, M. I., Bentley, S. D., van Hijum, S. A., and Zomer, A. (2017). Deciphering the distance to antibiotic resistance for the pneumococcus using genome sequencing data. *Sci. Rep.* 7: 42808.

Moisan, H., Pruneau, M., and Malouin, F. (2010). Binding of ceftaroline to penicillin-binding proteins of *Staphylococcus aureus* and Streptococcus pneumoniae. *J. Antimicrob. Chemother.* 65, 713–716. doi: 10.1093/jac/dkp503

Muller, A., Wenzel, M., Strahl, H., Grein, F., Saaki, T. N. V., Kohl, B., et al. (2016). Daptomycin inhibits cell envelope synthesis by interfering with fluid membrane microdomains. *Proc. Natl. Acad. Sci. U.S.A.* 113, E7077–E7086.

Murthy, M. H., Olson, M. E., Wickert, R. W., Fey, P. D., and Jalali, Z. (2008). Daptomycin non-susceptible meticillin-resistant *Staphylococcus aureus* USA 300 isolate. *J. Med. Microbiol.* 57, 1036–1038. doi: 10.1099/jmm.0.2008/000588-0

Ng, P. C., and Henikoff, S. (2003). Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31, 3812–3814. doi: 10.1093/nar/gkg509

Otero, L. H., Rojas-Altuve, A., Llarrull, L. I., Carrasco-Lopez, C., Kumarasiri, M., Lastochkin, E., et al. (2013). How allosteric control of *Staphylococcus aureus* penicillin binding protein 2a enables methicillin resistance and physiological function. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16808–16813. doi: 10.1073/pnas.1300118110

Palmer, K. L., Daniel, A., Hardy, C., Silverman, J., and Gilmore, M. S. (2011). Genetic basis for daptomycin resistance in enterococci. *Antimicrob. Agents Chemother.* 55, 3345–3356. doi: 10.1128/aac.00207-11

Patel, D., Husain, M., Vidaillac, C., Steed, M. E., Rybak, M. J., Seo, S. M., et al. (2011). Mechanisms of in-vitro-selected daptomycin-non-susceptibility in *Staphylococcus aureus*. *Int. J. Antimicrob. Agents* 38, 442–446. doi: 10.1016/j.ijantimicag.2011.06.010

Peacock, S. J., and Paterson, G. K. (2015). Mechanisms of methicillin resistance in *Staphylococcus aureus*. *Annu. Rev. Biochem.* 84, 577–601.

Peleg, A. Y., Miyakis, S., Ward, D. V., Earl, A. M., Rubio, A., Cameron, D. R., et al. (2012). Whole genome characterization of the mechanisms of daptomycin resistance in clinical and laboratory derived isolates of *Staphylococcus aureus*. *PLoS One* 7:e28316. doi: 10.1371/journal.pone.0028316

Power, R. A., Parkhill, J., and de Oliveira, T. (2017). Microbial genome-wide association studies: lessons from human GWAS. *Nat. Rev. Genet.* 18, 41–50. doi: 10.1038/nrg.2016.132

Price, A. L., Zaitlen, N. A., Reich, D., and Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* 11, 459–463.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Renzoni, A., Von Dach, E., Landelle, C., Diene, S. M., Manzano, C., Gonzales, R., et al. (2017). Impact of exposure of methicillin-resistant *Staphylococcus aureus* to Polyhexanide *In Vitro* and *In Vivo*. *Antimicrob. Agents Chemother.* 61, e00272–17.

Richards, R. L., Haigh, R. D., Pascoe, B., Sheppard, S. K., Price, F., Jenkins, D., et al. (2015). Persistent *Staphylococcus aureus* isolates from two independent cases of bacteremia display increased bacterial fitness and novel immune evasion phenotypes. *Infect. Immun.* 83, 3311–3324. doi: 10.1128/iai.00255-15

Robert, E., and Weber, F. L. (2017). Ingo Klare, Guido Werner and Birgit Strommenger. Comparative evaluation of VITEK 2® and three commercial gradient strip assays for daptomycin susceptibility testing of *Staphylococcus aureus*. *J. Antimicrob. Chemother.* 72, 3059–3062. doi: 10.1093/jac/dkx255

Robinson, D. A., and Enright, M. C. (2003). Evolutionary models of the emergence of methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 47, 3926–3934. doi: 10.1128/aac.47.12.3926-3934.2003

Roch, M., Gagetti, P., Davis, J., Ceriana, P., Errecalde, L., Corso, A., et al. (2017). Daptomycin resistance in clinical MRSA strains is associated with a high biological fitness cost. *Front. Microbiol.* 8:2303. doi: 10.3389/fmicb.2017.02303

Rong, M., Zheng, X., Ye, M., Bai, J., Xie, X., Jin, Y., et al. (2019). Phenotypic Plasticity of *Staphylococcus aureus* in liquid medium containing vancomycin. *Front. Microbiol.* 10:809. doi: 10.3389/fmicb.2019.00809

San, J. E., Baichoo, S., Kanzi, A., Moosa, Y., Lessells, R., Fonseca, V., et al. (2019). Current affairs of microbial genome-wide association studies: approaches, bottlenecks and analytical pitfalls. *Front. Microbiol.* 10:3119. doi: 10.3389/fmicb.2019.03119

Schaumburg, F., Peters, G., Alabi, A., Becker, K., and Idelevich, E. A. (2016). Missense mutations of PBP2a are associated with reduced susceptibility to ceftaroline and ceftobiprole in African MRSA. *J. Antimicrob. Chemother.* 71, 41–44. doi: 10.1093/jac/dkv325

Shapiro, B. J., David, L. A., Friedman, J., and Alm, E. J. (2009). Looking for Darwin's footprints in the microbial world. *Trends Microbiol.* 17, 196–204. doi: 10.1016/j.tim.2009.02.002

Shoji, M., Cui, L., Iizuka, R., Komoto, A., Neoh, H. M., Watanabe, Y., et al. (2011). walK and clpP mutations confer reduced vancomycin susceptibility in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 55, 3870–3881.

Ślusarczyk, R., Bielejewska, A., Bociek, A., and Bociek, M. (2018). Resistance to ceftaroline-2018 review. *Eur. J. Biol. Res.* 8, 112–120.

Sotillo, A., Pano-Pardo, J. R., Lopez-Quintana, B., and Gomez-Gil, R. (2016). Development of daptomycin resistance during therapy in a patient with methicillin-resistant *Staphylococcus aureus* endocarditis: a case report. *Enferm. Infecc. Microbiol. Clin.* 34, 534–535. doi: 10.1016/j.eimc.2015.11.006

Steed, M. E., Hall, A. D., Salimnia, H., Kaatz, G. W., Kaye, K. S., and Rybak, M. J. (2013). Evaluation of daptomycin non-susceptible *Staphylococcus aureus* for stability, population profiles, mprf mutations, and daptomycin activity. *Infect. Dis. Ther.* 2, 187–200. doi: 10.1007/s40121-013-0021-7

Strommenger, B., Braulke, C., Heuck, D., Schmidt, C., Pasemann, B., Nubel, U., et al. (2008). spa Typing of *Staphylococcus aureus* as a frontline tool in epidemiological typing. *J. Clin. Microbiol.* 46, 574–581. doi: 10.1128/jcm.01599-07

Strommenger, B., Layer, F., Klare, I., and Werner, G. (2015). Pre-use susceptibility to ceftaroline in clinical *Staphylococcus aureus* Isolates from Germany: Is there a non-susceptible pool to be selected? *PLoS One* 10:e0125864. doi: 10.1371/journal.pone.0125864

Struelens, M. J. (1998). The epidemiology of antimicrobial resistance in hospital acquired infections: problems and possible solutions. *BMJ* 317, 652–654. doi: 10.1136/bmj.317.7159.652

Tacconelli, E., Carrara, E., Savoldi, A., Harbarth, S., Mendelson, M., Monnet, D. L., et al. (2018). Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect. Dis.* 18, 318–327.

Thitiananpakorn, K., Aiba, Y., Tan, X. E., Watanabe, S., Kiga, K., Sato'o, Y., et al. (2020). Association of mprF mutations with cross-resistance to daptomycin and vancomycin in methicillin-resistant *Staphylococcus aureus* (MRSA). *Sci. Rep.* 10:16107.

Tong, S. Y., Davis, J. S., Eichenberger, E., Holland, T. L., and Fowler, V. G. Jr. (2015). *Staphylococcus aureus* infections: epidemiology, pathophysiology, clinical manifestations, and management. *Clin. Microbiol. Rev.* 28, 603–661. doi: 10.1128/cmr.00134-14

Urban, E., and Stone, G. G. (2019). Impact of EUCAST ceftaroline breakpoint change on the susceptibility of methicillin-resistant *Staphylococcus aureus* isolates collected from patients with complicated skin and soft-tissue infections. *Clin. Microbiol. Infect.* 25, e1–e4.

Weinert, L. A., Chaudhuri, R. R., Wang, J., Peters, S. E., Corander, J., Jombart, T., et al. (2015). Genomic signatures of human and animal disease in the zoonotic pathogen Streptococcus suis. *Nat. Commun.* 6:6740.

Weinert, L. A., Chaudhuri, R. R., Wang, J., Peters, S. E., Corander, J., Jombart, T., et al. (2019). Publisher Correction: genomic signatures of human and animal disease in the zoonotic pathogen Streptococcus suis. *Nat. Commun.* 10: 5326.

Wheeler, N. E., Reuter, S., Chewapreecha, C., Lees, J. A., Blane, B., Horner, C., et al. (2019). Contrasting approaches to genome-wide association studies impact the detection of resistance mechanisms in *Staphylococcus aureus*. *bioRxiv [Preprint]*

Wichelhaus, T. A., Schafer, V., Brade, V., and Boddinghaus, B. (1999). Molecular characterization of rpoB mutations conferring cross-resistance to rifamycins on methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 43, 2813–2816. doi: 10.1128/aac.43.11.2813

Yang, S. J., Mishra, N. N., Kang, K. M., Lee, G. Y., Park, J. H., and Bayer, A. S. (2018). Impact of multiple single-nucleotide polymorphisms within mprf on daptomycin resistance in *Staphylococcus aureus*. *Microb. Drug Resist.* 24, 1075–1081. doi: 10.1089/mdr.2017.0156

Yang, S. J., Mishra, N. N., Rubio, A., and Bayer, A. S. (2013). Causal role of single nucleotide polymorphisms within the mprF gene of *Staphylococcus aureus* in daptomycin resistance. *Antimicrob. Agents Chemother.* 57, 5658–5664. doi: 10.1128/aac.01184-13

Young, B. C., Earle, S. G., Soeng, S., Sar, P., Kumar, V., Hor, S., et al. (2019). Panton-Valentine leucocidin is the key determinant of *Staphylococcus aureus* pyomyositis in a bacterial GWAS. *Elife* 8:e42486.

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

frontiers in Microbiology

ORIGINAL RESEARCH
published: 04 March 2021
doi: 10.3389/fmicb.2021.636788

The Surge of Hypervirulent ST398 MRSA Lineage With Higher Biofilm-Forming Ability Is a Critical Threat to Clinics

Huiying Lu[1†], Lin Zhao[1†], Yuanguo Si[2], Ying Jian[1], Yanan Wang[1], Tianming Li[1], Yingxin Dai[1], Qian Huang[1], Xiaowei Ma[1], Lei He[1] and Min Li[1*]*

[1] Department of Laboratory Medicine, School of Medicine, Renji Hospital, Shanghai Jiao Tong University, Shanghai, China,
[2] Department of Laboratory Medicine, Qingdao Hiser Medical Center, Qingdao, China

The global increase of community-associated (CA) infections with methicillin-resistant *Staphylococcus aureus* (MRSA) is a major healthcare problem. Although sequence type (ST) 398 MRSA was first described as a livestock-associated (LA) lineage, human-adapted MRSA (HO-MRSA) ST398 without livestock contact has subsequently been reported from China in our previous study and other later research. The proportion of ST398 HO-MRSA has also remarkably increased in recent years in China. Based on 3878 *S. aureus* isolates that were collected in a general hospital between 2008 and 2018, we identified 56 ST398 HO-MRSA isolates. The four early appearing isolates of them have been sequenced by whole-genome sequencing (WGS) in our previous study. Here, by usage of WGS on the later-appearing 52 isolates and analyzing the phylogenetic dynamics of the linage, we found that 50 isolates clustered together with the former 4 isolates, making it a main clade out of MSSA clones and other MRSA clones, although ST398 HO-MRSA evolved with multiple origins. Drug resistance and virulence gene analysis based on the WGS data demonstrated that ST398 HO-MRSA main clade exhibited a similar pattern in both parts. Furthermore, they all carried a conserved variant of prophage 3 to guarantee virulence and a short SCC*mec* type V element of class D to maintain considerable lower methicillin resistance. Further phenotypical research verified that the epidemic HO-MRSA ST398 displayed enhanced biofilm formation ability when keeping high virulence. The dual advantages of virulence and biofilm formation in the HO-MRSA ST398 subtype promote their fitness in the community and even in the healthcare environment, which poses a serious threat in clinical *S. aureus* infections. Therefore, further surveillance is required to prevent and control the problematic public health impact of HO-MRSA ST398 in the future.

Keywords: methicillin-resistant *staphylococcus aureus*, sequence type 398, phylogenetic analysis, whole-genome sequencing, virulence

OPEN ACCESS

Edited by:
Yang Wang,
China Agricultural University, China

Reviewed by:
Jiachang Cai,
Zhejiang University, China
Yi-Wei Tang,
Cepheid, United States

***Correspondence:**
Lei He
buningweishi_1985@126.com
Min Li
ruth_limin@126.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Microbiology

Received: *02 December 2020*
Accepted: *27 January 2021*
Published: *04 March 2021*

Citation:
Lu H, Zhao L, Si Y, Jian Y, Wang Y, Li T, Dai Y, Huang Q, Ma X, He L and Li M (2021) The Surge of Hypervirulent ST398 MRSA Lineage With Higher Biofilm-Forming Ability Is a Critical Threat to Clinics. Front. Microbiol. 12:636788. doi: 10.3389/fmicb.2021.636788

INTRODUCTION

Methicillin-resistant *Staphylococcus aureus* (MRSA) is not only a human pathogen causing a variety of infections, such as skin and soft tissue infection (SSTI), pneumonia, and sepsis, but it also can colonize and cause diseases in multifarious animals, known as livestock-associated MRSA (LA-MRSA) (Chuang and Huang, 2015; Chen and Huang, 2018). MRSA sequence type (ST) 398 was

first identified as a LA-MRSA and also the most dominant clone of LA-MRSA globally, which was first found as prevalent in Europe and then identified outside Europe, including the North America and some Asian countries (Lewis et al., 2008; Larsen et al., 2017).

LA-MRSA has been found more and more in humans and is associated with serious diseases and even death. In the past decade, several infections caused by ST398 MRSA have been reported in the community, ranging from mild skin infections to serious invasive infections and even death, both with and without livestock contact. Human colonization with LA-MRSA ST398, which is genetically identical with LA-MRSA, was first recognized among swine farmers in France and The Netherlands in the early 2000s (Armand-Lefevre et al., 2005; Voss et al., 2005) and showed rapid emergence in Europe (Khanna et al., 2008; Smith et al., 2009, 2013), yet human cases of LA-MRSA ST398 were rarely reported from Asian countries. In contrast, human cases of MRSA ST398 without livestock contact, named host-adapted MRSA ST398 (HO-MRSA ST398), have been gradually reported as emerging in China. Recently, we reported the emergence in Shanghai, China, of several high-virulence HO-MRSA ST398 isolates, which evolved from methicillin-sensitive *Staphylococcus aureus* (MSSA) and are genetically different from LA-MRSA with the characteristics of the lack of *tetM* resistance determinants and the presence of a variant of prophage 3 (He et al., 2018). Soon after, two cases of surgical site infections caused by highly virulent HO-MRSA ST398 were reported in Zhejiang, China (Sun et al., 2019). Subsequently, HO-MRSA ST398 strains has been reported to spread locally in sanatoriums in Taiwan District, probably after entrancing from mainland China (Huang and Chen, 2020).

We continuously tracked the epidemiology of all *S. aureus* isolates in the general teaching hospital and found a rapid increase of ST398 in China while the previously dominant HA-MRSA ST239 clones showed a significant decrease, leading to a marked decrease in the prevalence of MRSA over the past decade (Dai et al., 2019). This prompted us to study those strains to better understand the cause of and the prospect for the rise of HO-MRSA ST398 in China. Here, we report the detection, phylogenetic analysis, epidemiological information, and the phenotypic characteristics of the epidemic HO-MRSA ST398 isolates identified by our clinical laboratory, providing evidence to inform and benefit the clinical prevention and control of *S. aureus* infections.

RESULTS

Surveillance of MRSA Isolates at a General Teaching Hospital in 2008–2018

A total of 2,588 MRSA isolates were identified among 3878 *S. aureus* infections at an affiliated tertiary hospital from 2008 to 2018 in Shanghai, China. The proportion of MRSA infections decreased from 83.5% (643 of 770 total *S. aureus* infections) in 2008 to 49.1% (195 of 397) in 2018. The prevalence of one predominant healthcare-associated MRSA (HA-MRSA) lineage, ST239, significantly decreased (from 48.5 to 4.1%, 2008–2018), which has been proved to be responsible for the marked decrease in the prevalence of MRSA over the past decade. The majority of the 195 MRSA isolates belonged to ST5 (n = 99), followed by the community-acquired MRSA (CA-MRSA) clones ST59 (n = 26), ST1 (n = 21), and ST398 (n = 14) in 2018. The proportion of another predominant HA-MRSA lineage, ST5, maintained stability between 2008 (50.1%) and 2018 (50.8%). Contrastingly, the CA-MRSA clones ST59, ST1, and ST398 increased significantly from 0.6% (4 of 643) to 13.3% (26 of 195), 0.5% (3 of 643) to 10.8% (21 of 195), and 0.0% (0 of 643) to 7.18% (14 of 195), respectively, from 2008 to 2018. ST398 MRSA ranked as the third most prevalent CA-MRSA lineage in all SA isolates in 2018. The majority of the total SA isolates belonged to ST5 (n = 114) and ST398 (n = 34), followed by ST59 (n = 32) in 2018. The proportion of ST5 isolates in all SA isolates maintained relative stability between 2008 (41.9%) and 2018 (28.7%), whereas the proportion of ST398 isolates increased obviously from 1.8% (14 of 770) in 2008 to 8.6% (34 of 397) in 2018 ($P < 0.001$), in which it ranked first in prevalence among CA-SA lineage in all SA isolates in 2018 (**Figure 1** and **Supplementary Table S1**). The above data indicate that the increasing emerging dominant MRSA ST398 could be a factor responsible for the remaining stubborn MRSA in recent years.

Phylogenetic Tree of ST398 MRSA Isolates

A total of 56 MRSA ST398 isolates were collected in this study, while 4 of them that were collected during 2012–2014 had already been analyzed in our previous study (He et al., 2018). Therefore, 52 ST398 MRSA isolates collected during 2015–2018 were whole genome sequenced in this study. The Illumina sequences generated in this study are deposited and available in the Sequence Read Archive (SRA)[1] under the study accession number PRJNA624723 with the SRA accession numbers SRR11526821 to SRR11526872. We also included our previous genomic data of 61 ST398 isolates (sample accession numbers: SRR5054902 to SRR5054977 and SRR5062006 in the SRA of NCBI). All the characteristic data of 113 ST398 *S. aureus* isolates used for phylogenetic analysis are listed in **Supplementary Table S2**. The core-genome SNPs were applied for phylogenetic tree reconstruction based on the data of 52 current MRSA ST398 genome sequences and our previously published data of 61 ST398 genome sequences (7 HO-MRSA isolates and 54 HO-MSSA isolates during 2010–2015) using maximum likelihood estimation (**Figure 2**). All ST398 isolates were divided into two major phylogenetic clades, with each supported by 100% bootstrapping. Most of the HO-MSSA ST398 strains were in Clade 1. However, there were also five human-adapted MRSA ST398 strains in Clade 1, including 17-398-18, 16-398-12, HO-MRSA-5, HO-MRSA-6, and HO-MRSA-7. Human-adapted MRSA ST398 were mainly in Clade 2, including 4 HO-MRSA ST398 isolates from our previous study, clustered with the four isolates of HO-MSSA ST398 isolates. Clade 1 was a mixture of samples collected from both humans and livestock, while Clade 2 was a pure clade of sample from humans, which accounted for 91.7% (55/60) of HO-MRSA ST398 isolates sampled, suggesting

[1] http://www.ncbi.nlm.nih.gov/sra

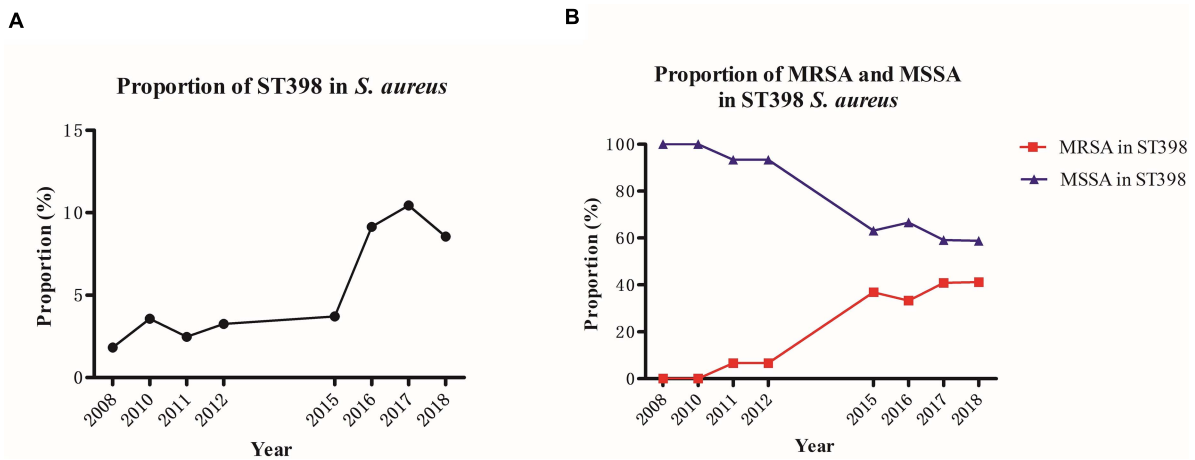


FIGURE 1 | Prevalence of the total ST398 **(A)** and the increasingly emerging dominant MRSA ST398 clone **(B)** from different clinical specimens of patients, 2008–2018.

that the epidemic MRSA ST398 mainly originated from Clade 2. The above data showed that the increasing emerging dominant MRSA ST398 in Shanghai, China, mainly evolved to a subclone.

Clinical Characteristics of ST398 HO-MRSA

Among the 56 MRSA ST398 isolates collected in this study, 96.42% (54 of 56) were identified as HO-MRSA according to the phylogenetic tree. Only two isolates (17-398-18 and 16-398-12) were finally judged as LA-MRSA originated, which clustered with S0385, and clinical data also showed that these two patients ever had livestock contact. Interestingly, 3 of 56 infection cases (18-398-11, 18-398-21, and 18-398-32) by ST398 MRSA that clustered in Clade 2 were found to occur > 48 h after hospitalization in 2018, which suggested the onset of high-virulence MRSA ST398 in the healthcare environment.

Antibiotic Susceptibility

We profiled the 113 ST398 isolates with *in vitro* susceptibility tests to 13 common antibiotics. Almost all the ST398 *S. aureus* isolates were susceptible to RD, TEC, LZD, and VA and resistant to P (antibiotics abbreviations provided in the "Materials and Methods" section). Compared to the 54 ST398 MSSA, the 56 ST398 MRSA collected in this study together with 3 ST398 MRSA from other hospitals displayed total resistance to FOX and high resistance rates for CZ (41 of 59; 69.49%) (**Supplementary Table S2**), whereas ST398 MSSA exhibited higher resistance rates for CN (14 of 54, 25.93%) and FOS (6 of 54, 11.11%), compared with ST398 MRSA (1 of 59, 1.69% and 0 of 59, 0.00%).

Virulence Factors and Resistance Gene Analysis

We further mapped virulence and antibiotic susceptibility information for each isolate to the phylogeny (**Figure 3**). In all ST398 strains, 20 virulence genes were found, including *hlgA*, *hlgB*, *hlgC*, *scn*, *sak*, *hla*, *aur*, *icaA*, *icaB*, *icaC*, *icaD*, *clfA*, etc. Among the 20 virulence genes, *hlgA*, *hlgB*, and *hlgC* encode γ-hemolysin; staphylococcal complement inhibitor (*scn*) is associated with immune evasion; *sak* encodes staphylokinase, contributing to spread of the bacteria; and *aur* encodes aureolysin, which belongs to staphylococcal exoenzyme. All the HO-MRSA ST398 strains in Clade 2 harbored the *scn*, staphylokinase (*sak*), and chemotaxis inhibitory protein (*chp*) genes, the three of which are in common with the conserved variant of prophage 3. Almost none of the above strains carried the *pvl* gene or enterotoxin genes or *cna* gene. Nevertheless, other HO-MRSA ST398 strains scattered in Clade 1 carried only *scn* and *sak* genes, and the *chp* gene was always absent in this clade.

Antibiotic resistance genes were rarely found in all ST398 strains. All the HO-MRSA ST398 isolates in Clade 2 carried *mecA* (methicillin resistance gene). Almost all of them harbored *blaZ* (penicillin resistance gene) and *erm*(C) (macrolide resistance gene). Besides the above genes, the HO-MRSA ST398 isolates in Clade 2 carried very few other antibiotic resistance genes. The strains in Clade 1 also harbored very few antibiotic resistance genes besides *blaZ*.

The above data suggested that all the HO-MRSA ST398 strains in Clade 2 had similar antibiotic resistance and virulence pattern background with the conserved variant of prophage 3, containing the immune evasion complex (IEC) genes encoding the chemotaxis inhibitory protein (CHIP), staphylococcal complement inhibitor (SCIN), and staphylokinase (SAK) together, which was usually different from that of HO-MRSA ST398 in Clade 1, thus probably facilitating the clone transmission of this specific ST398 MRSA.

SCC*mec* Analysis

The most interesting feature of the genome of all HO-MRSA ST398 isolates in Clade 2 and one HO-MRSA ST398 isolate in Clade 1 (HO-MRSA-6) is the SCC*mec* type V element of class D in *S. aureus* only described in our previous study

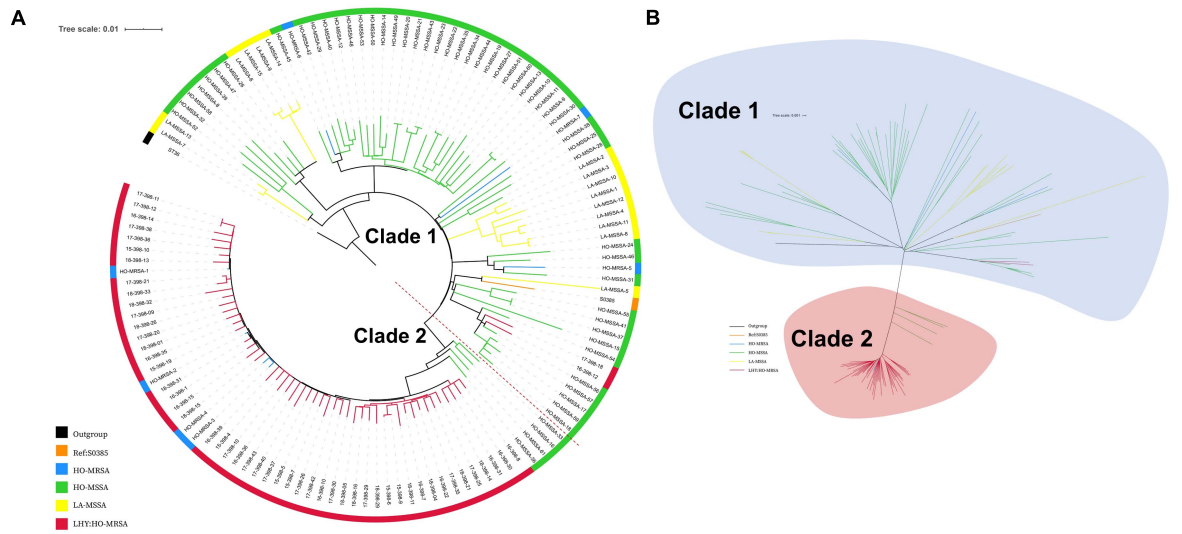


FIGURE 2 | Phylogenetic structure of ST398 *S. aureus* isolates. **(A)** Maximum-likelihood phylogenetic tree of ST398. A total of 129 *S. aureus* ST398 isolates (114 from adult patients and 15 from livestock) were used for phylogenetic reconstruction. Relationships are shown with respect to the genome of the standard LA-MRSA strain (S0385). Branches are colored by geographic origin of isolates. The tree is rooted with ST36 as an outgroup. **(B)** Unrooted tree of ST398 *S. aureus* isolates. Branches in this tree could be classified into two main clades, including Clade I and Clade II. Branches are colored by geographic origin of isolates. Black branch: outgroup. Orange branch: reference isolate. Blue branch: human-adapted (HO-MRSA) isolates sequenced in our previous study. Green branch: HO-MSSA isolates in our previous study. Yellow branch: livestock-adapted (LA-MSSA) isolates in our previous study. Red branch: HO-MRSA isolates sequenced in this study.

(He et al., 2018), characterized by an IS431-*mecA*-*mecR1*' composition and one copy of the *ccrC* recombinase gene, which has previously only been found in CoNS like *S. caprae* (Katayama et al., 2001). However, other HO-MRSA ST398 isolates in Clade 1 have different SCC*mec* elements, respectively. One HO-MRSA ST398 strain (17-398-18) harbors SCC*mec* type V (5C2&5) (composition IS431-*mecA*-*mecR1*'-IS431 and two *ccrC* copies), which belong to most LA-MRSA ST398 strains, including the reference strain S0385. Another HO-MRSA ST398 strain (HO-MRSA-5) contains SCC*mec* type IVa (2B) with the following gene complexes: *ccrA2*-*ccrB2* and *mecA*-*mecR1*. Furthermore, the HO-MRSA ST398 strain (16-398-12) harbors SCC*mec* type V (5C2) with the following gene complexes: *ccrC* and *mecA* (**Figure 4A**). Obviously, the SCC*mec* type V element of class D is shorter than other kinds of SCC*mec* elements in Clade 1.

Considering previous findings (Pozzi et al., 2012; Rudkin et al., 2012) that have shown repression of virulence by the methicillin resistance-encoding *mecA* gene and our previous results (He et al., 2018) showing that striking and highly conserved differences of MICs to β-lactams and *mecA* expression exist between HA-MRSA clones and CA-MRSA clones, here we further sought for the possible influence of this kind of SCC*mec* to MIC of oxacillin and *mecA* expression. Six HO-MRSA ST398 isolates with short SCC*mec* V were randomly selected to compare the MIC of oxacillin with the isolates with other SCC*mec*. Isolates with the short SCC*mec* V showed lower oxacillin MIC levels compared to other CA-MRSA-characteristic SCC*mec* types (**Figure 4B**) in accord with our previous finding (He et al., 2018), showing that this SCC*mec* element in epidemic HO-MRSA ST398 confers only very low-level methicillin resistance to fit both in the community and healthcare environment where the former epidemic HA-MRSA ST239 with high methicillin resistance is significantly decreasing now.

The above data indicated that the short SCC*mec* V element of class D could be a factor accounting for the fitness of HO-MRSA ST398 strains in Clade 2 both in the community and healthcare setting.

HO-MRSA ST398 Epidemic Isolates Exhibit High Virulence

We randomly selected four representative HO-MSSA ST398 isolates from Clade 1 and four representative HO-MRSA ST398 isolates from Clade 2, together with four HO-MRSA ST398 isolates in Clade 1. We next measured cytolytic potential by lysing analysis of human erythrocytes and determined expression of RNAIII and α-toxin as important core genome-encoded virulence determinants (**Figures 5A–C**). HO-MRSA ST398 in Clade 2 displayed high erythrocyte lysis capacity and also exhibited elevated RNAIII and α-toxin expression at the transcriptional level in comparison with the other two groups in Clade 1. All these analyses showed that the HO-MRSA ST398 isolates in Clade 2 have higher virulence compared to the HO-MRSA ST398 isolates and the closely related MSSA isolates in Clade 1.

HO-MRSA ST398 Epidemic Isolates Exhibit Higher Biofilm Formation Ability

In addition to the observed high virulence, semi-quantitative biofilm tests of the above 12 representative ST398 isolates revealed that compared with the 4 HO-MRSA and 4 HO-MSSA

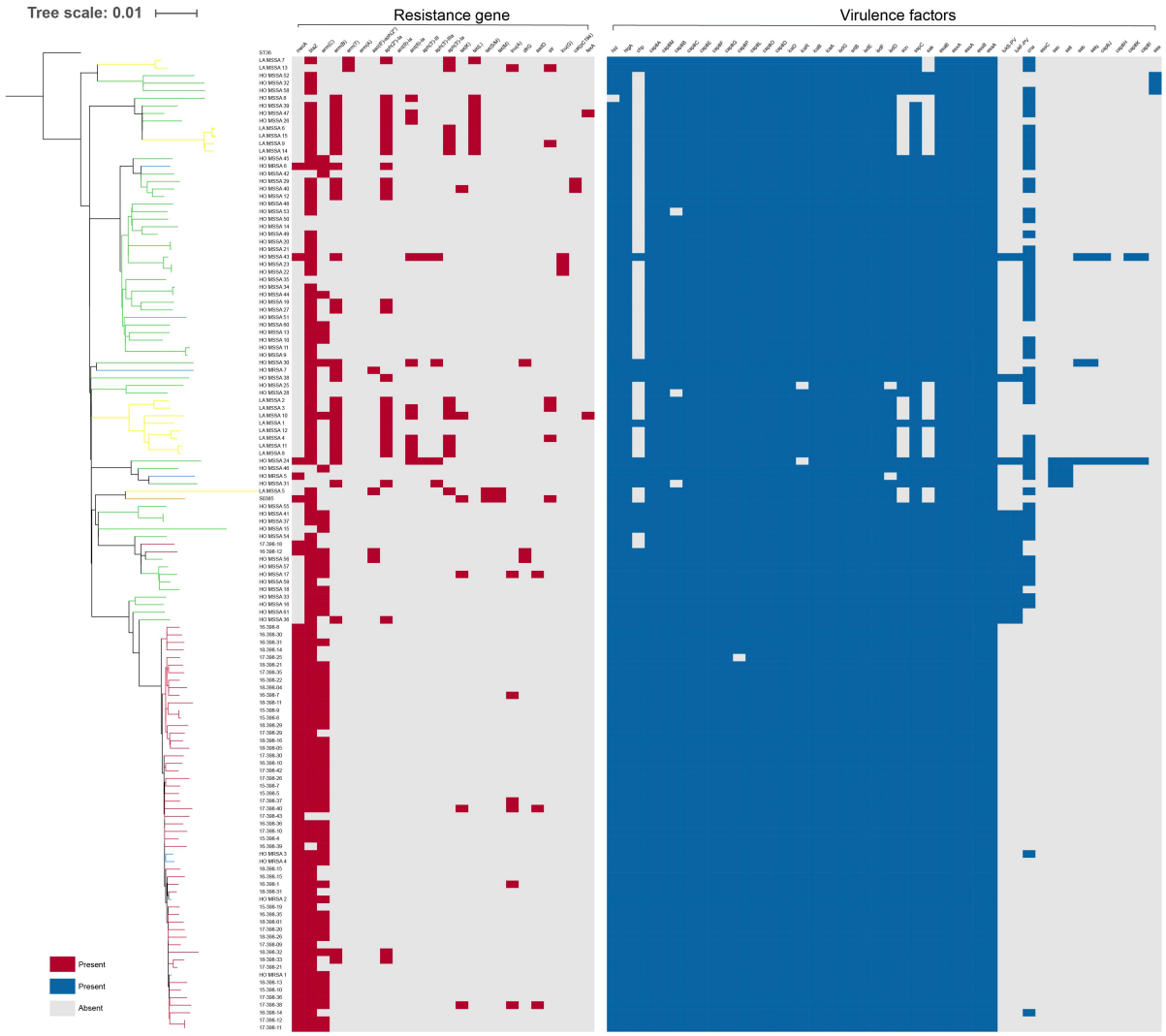


FIGURE 3 | Presence or absence of antibiotic resistance genes and virulence genes among ST398 isolates. The distribution of antibiotic resistance genes and virulence genes is plotted against the core genome phylogeny; the presence of an antibiotic resistance or virulence gene in an isolate is shown in red and blue, respectively, and the absence is shown in gray.

ST398 isolates in Clade 1, the 4 HO-MRSA ST398 isolates in Clade 2 demonstrated significantly higher biofilm formation, the key process in the occurrence and development of persistent infections (**Figure 5D**).

DISCUSSION

To our knowledge, this is the first report focusing on the epidemiologically important emerging host-adapted MRSA ST398 in China. Based on our previous study and this study, the first ST398 isolate was identified in 2011 from a 65 year-old male patient whose sputum and blood cultures grew MRSA ST398, which suggests that ST398 existed in Shanghai since 2011 or earlier. Also, the population of both ST398 and MRSA ST398 gradually increased annually in Shanghai, China.

MRSA ST398, first identified as a LA-MRSA, was prevalent in Europe and the North America. However, it was rarely reported in Asian countries (Lewis et al., 2008; Chuang and Huang, 2015; Larsen et al., 2017; Chen and Huang, 2018). Human-adapted ST398 isolates, which are different from LA clones in genetics, often cause life-threatening infections and have some common genetic characteristics in adapting to human as host (Price et al., 2012; Larsen et al., 2016; He et al., 2018; Sun et al., 2019). Most previous reports found that human-adapted ST398 isolates are MSSA (Zhao et al., 2012). In our recent report (He et al., 2018), eight cases of ST398 CA-MRSA infection were reported in China, of which six cases were severe and two cases died. Genome analysis showed that these highly infectious strains of ST398 CA-MRSA evolved from human-adapted methicillin-sensitive clones, which were then called ST398 HO-MRSA. Similar severe and even fatal cases have also been reported

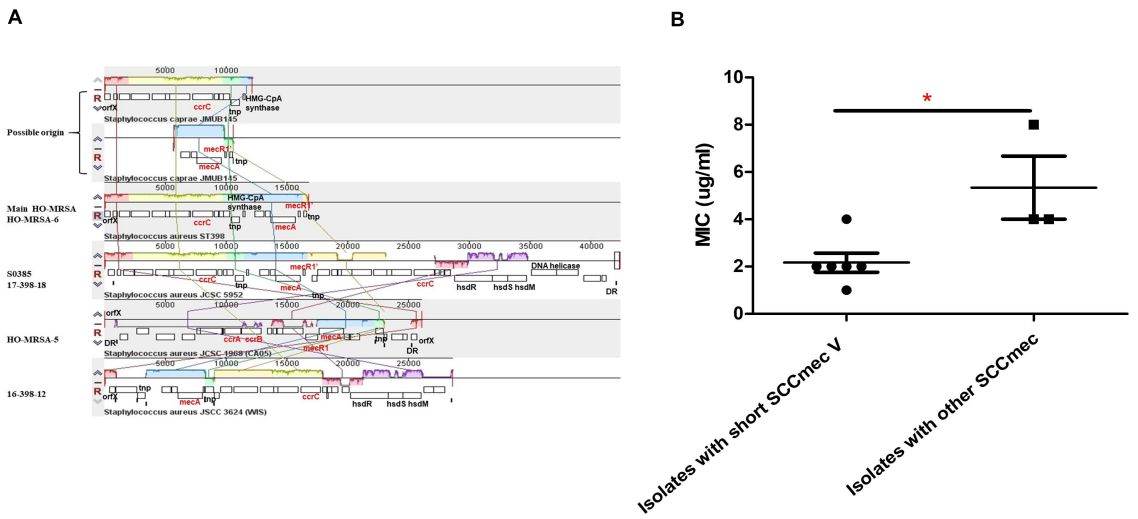


FIGURE 4 | Analysis of HO-MRSA ST398 SCC*mec*. **(A)** SCC*mec* comparison. **(B)** Comparison of oxacillin MIC. $p \leq 0.05$ (*).

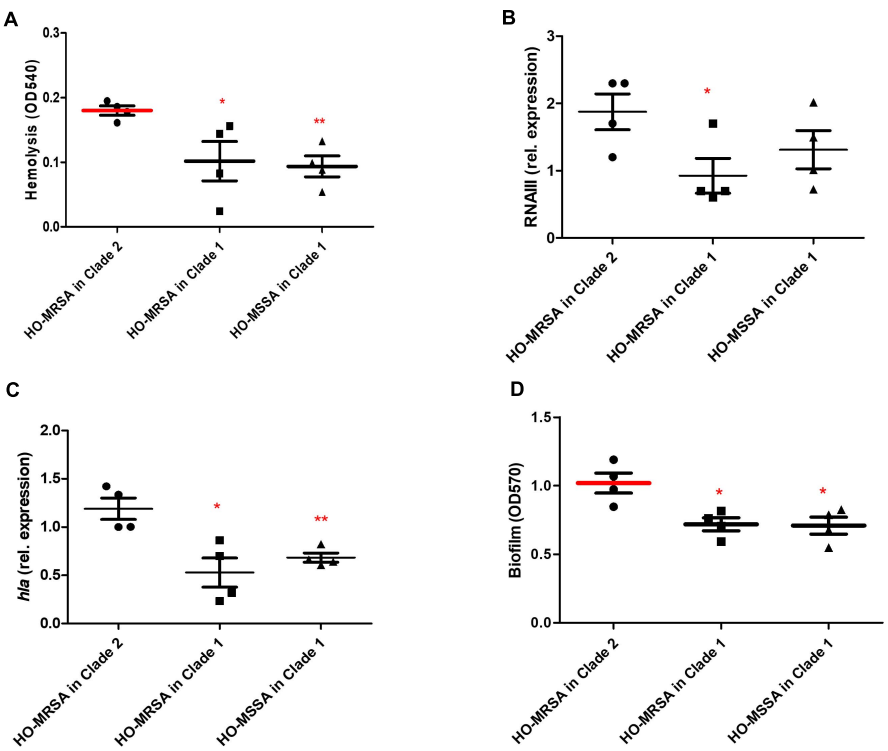


FIGURE 5 | Phenotypic differences between HO-MRSA ST398 in Clade 1 and Clade 2. **(A)** Hemolysis (erythrocyte lysis) of four representative HO-MRSA ST398 isolates, respectively, in Clade 1 and Clade 2. Hemolytic activities were determined by incubating culture filtrates with human red blood cells. **(B,C)** Expression of *RNAIII*/*hla* was determined by qRT-PCR. **(D)** Biofilm elaboration measured using polystyrene microtiter plate assay. Statistical analysis was performed using Student's *t*-test. The data shown are mean ± SD of three independent experiments. Statistical significances: $p > 0.05$ (ns), $p \leq 0.05$ (*), $p < 0.001$ (**).

in Japan and Australia (Koyama et al., 2015; Coombs et al., 2019). Subsequently, 12 strains of MRSA ST398 were identified as colonization strains in Taiwan (Huang and Chen, 2020). The strain appears to be spreading internationally and needs further monitoring.

Phylogenetic reconstructions of the dominating pandemics of healthcare-associated ST239 MRSA and LA ST398 MRSA have been well described (Harris et al., 2010; Price et al., 2012; Hsu et al., 2015), while the evolutionary dynamics of the increasing ST398 HO-MRSA has rarely been studied. We identified 52

ST398 MRSA during 2015 and 2018 in the general hospital in Shanghai, China, together with six ST398 MRSA during 2011–2014 in our previous study. Next, the phylogeny of ST398 MRSA linages using the above 58 individual ST398 MRSA isolates was reconstructed, proving that ST398 MRSA isolates evolved with multiple origins and 54 of them evolved together as an emerging main clade that differed from those MSSA isolates in clinical *S. aureus* infections. ST398 MRSA isolates in the main Clade (Clade 2) definitely exhibited differences both genetically and phenotypically in comparison to ST398 MSSA and other ST398 MRSA isolates clustered in Clade 1.

Herein, analysis of the virulence genes and the resistance genes revealed that ST398 MRSA isolates in Clade 2 exhibited uniform virulence and antibiotic resistance pattern. Discrepancy in the virulence genes between ST398 MRSA in Clade 2 and ST398 MRSA and ST398 MSSA in Clade 1 mainly focused on the *chp*, which is conserved in Clade 2, while all of them harbor the *scn* and *sak genes*. The CHIP encoded by *chp* is a 14 kDa protein that blocks neutrophil chemotaxis via binding the formylated peptide receptor and the C5a receptor on neutrophils (Postma et al., 2004, 2005). Higher carriage rate of *chp* in these isolates might indicate a higher ability of immune escape of ST398 HO-MRSA main clade during host defense in *S. aureus* infections. In addition to the high presence of virulence genes, ST398 HO-MRSA isolates displayed higher hemolysis capacity, making the ST398 HO-MRSA in Clade 2 a hypervirulent linage in clinical *S. aureus* infections. On the other hand, ST398 HO-MRSA isolates exhibited high sensitivity to several common antibiotics as well as ST398 HO-MSSA isolates. Additionally, the type of all the ST398 HO-MRSA in Clade 2 is the short SCC*mec* type V element of class D, which maintained the much lower methicillin resistance.

Generally speaking, hypervirulent isolates exhibited higher susceptibility to the most common antibiotics, while multi-drug resistant (MDR) isolates would be less virulent, which is named "fitness cost" in related research (Foucault et al., 2009; Maher et al., 2012; Nielsen et al., 2012). However, strains with high virulence have appeared under scrutiny recently in the healthcare environment (Rasko and Sperandio, 2010). Herein, this study shows that ST398 HO-MRSA in Clade 2 displayed lower resistance to methicillin and several common antibiotics while exhibiting higher virulence in comparison to ST398 HO-MSSA and ST398 HO-MRSA in Clade 1. Moreover, ST398 HO-MRSA isolates in Clade 2 formed thicker biofilm compared to the isolates in Clade 1, which was more common in healthcare-associated *S. aureus* infections, promoting the fitness of the ST398 HO-MRSA in the community and even hospital. Thus, more attention should be paid to the ST398 HO-MRSA isolates in the prevention and control of infections.

As for the limitations of this study, since all the ST398 *S. aureus* isolates identified and characterized in this study were only from an affiliated tertiary hospital in Shanghai, further studies are needed to determine the whole perspective of HO-MRSA ST398 *S. aureus* in China.

In conclusion, our data provide important insight into the current epidemic status, pathogenicity, transmission, and phylogenetic relationship of the human-adapted MRSA ST398 in Shanghai, China. Genomic analyses presented here, in conjunction with the epidemiological data, suggest that the epidemic transmission of HO-MRSA ST398 is strongly related to the short SCC*mec* type V element maintaining the much lower methicillin resistance and the conserved variant of prophage 3, as well as containing the complete IEC genes, guaranteeing virulence. Further phenotypical research verified that the epidemic HO-MRSA ST398 showed higher biofilm formation ability when maintaining high virulence, promoting their fitness in the community and even in the healthcare environment. Additional research and surveillance are required to predict the public health impact of HO-MRSA ST398 in the future.

MATERIALS AND METHODS

Bacterial Strains, Growth Conditions, and Clinical Definitions

S. aureus strains were grown in tryptic soy broth (TSB) (Oxoid) with 0.25% glucose or on tryptic soy agar plates at 37°C. We collected and analyzed a total of 3,878 clinical isolates from adult patients at a comprehensive teaching hospital in Shanghai, China, between 2008 and 2018, from which the epidemiological data of 3695 clinical isolates between 2008 and 2017 have been published in our previous study (Dai et al., 2019).

ST398 MRSA and MSSA isolates were further investigated in the present study after initial characterization. Healthcare-associated SA (HA-SA) was defined as SA infection that occurred > 48 h after hospitalization. Community-associated SA (CA-SA) was defined as an isolate that was obtained either from an outpatient or from an inpatient (including those from general and urgent care and emergency rooms) $\leq$ 48 h after hospital admission. Human-adapted SA (HO-SA) was defined as human-originated and adapted SA without livestock contact, which is genetically different from LA-SA (He et al., 2018).

An infection was considered invasive when isolates were isolated from otherwise sterile body sites. Clinical data were obtained from patient electronic medical records. The information collected included the location of the patient at the time of sample collection, date of sample collection, date of isolation, and body site of the sample. The clinical syndrome of MRSA infection was classified into syndrome categories (SSTI, respiratory infection, bacteremia, CSF, or other sterile body fluid).

Antimicrobial Resistance Profiles

Antibiograms were determined by standard disc diffusion on Mueller–Hinton agar in accordance with the Clinical and Laboratory Standards Institute (CLSI) guidelines. Fourteen antimicrobial agents tested included gentamycin (CN), penicillin (P), cefazolin (CZ), erythromycin (E), clindamycin (DA), sulfamethoxazole-trimethoprim (SXT), fosfomycin (FOS), rifampicin (RD), levofloxacin (LEV), cefuroxime (CXM), teicoplanin (TEC), linezolid (LZD), vancomycin (VA), and cefoxitin (FOX). *S. aureus* ATCC29213 was used as a quality control.

Molecular Typing

Molecular typing was performed using multilocus sequence typing (MLST) as previously described (Maiden et al., 1998). The sequences of the polymerase chain reaction (PCR) products were compared with the existing sequences available at the MLST website[2].

Whole-Genome Sequencing of ST398 Isolates and Genome Comparison

Chromosomal DNA of 52 *S. aureus* ST398 isolates was extracted by a standard phenol–chloroform extraction procedure. *S. aureus* whole-genome sequencing was performed on an Illumina HiSeq 2500 sequencer (Illumina, San Diego, CA, United States) with 150 bp paired-end reads. The data generated from the Illumina platform were analyzed after quality control was performed.

Original sequencing reads were exported to Fastq files, and then snippy (BWA, SAMtools, SnpEff, and Freebayes) was used to align reads to the S0385 ST398 chromosome as a reference [GenBank:AM990992] to generate the SNPs of the core genome. ST36 [GenBank:BX571856], determined by Price et al. as the most closely related non-CC398 ST (Price et al., 2012), was used as an outgroup. The repeat regions were removed by TRF and BLASTN. Gubbins was used to remove the recombinant region in the genome, and the total number of SNPs was 28,644, and then SNPs only caused by outgroup were eliminated.

Fastq files of 76 ST398 samples including 7 HO-MRSA, 54 HO-MSSA, and 15 LA-MSSA ST398 isolates and in our previous study (He et al., 2018) were also downloaded from GenBank, and the variants were also called using the preceding strategy. Therefore, the total number of SNPs was 7205.

Phylogenetic Analysis

The maximum likelihood tree was constructed based on the 7205 SNPs in the core genome after duplicate and recombination reads were removed, using the GTR + G model in the RAxML software by 100 bootstraps. The phylogenetic results were displayed by ITOL[3] (Letunic and Bork, 2019).

Genome Assembly and Detecting the Presence of Virulence-Associated Genes and Antibiotic Resistant Genes

Pre-processed reads were *de novo* assembled using CLC Genomics Workbench 12.0 (Qiagen) using the default options. Then, the generated *de novo* assembled contigs were analyzed separately via the pipelines of BLASTing the drug resistant gene database (Resfinde database, 2020-05-28) (Zankari et al., 2012; Liu et al., 2019) and at the Virulence FactorsDatabase (Larsen et al., 2012; Liu et al., 2019) in the CLC software.

SCC*mec* Identification

The SCC*mec* element is the defining feature of MRSA isolates and encodes the single determinant for methicillin resistance, the *mecA* gene. In order to assign SCC*mec* type, we use SCC*mec*Finder, which identifies SCC*mec* elements in sequenced *S. aureus* isolates[4].

Lysis of Erythrocytes by Culture Filtrates

Supernatants were collected from bacterial cultures grown for 15 h. Hemolytic activities were determined by incubating samples with human red blood cells (2% v/v in Dulbecco's phosphate-buffered saline, DPBS) for 1 h at 37°C. Hemolysis was determined by measuring the optical density at 540 nm using an enzyme-linked immunosorbent assay (ELISA) reader. The assay was performed in triplicate.

Quantitative Reverse-Transcription (qRT)-PCR

Overnight cultures were diluted 1:100 into 50 ml of TSB and incubated at 37°C with shaking at 200 rpm until grown to mid-exponential phase (4 h). Complementary DNA was synthesized from total RNA using the QuantiTect Reverse Transcription Kit (Qiagen) according to the manufacturer's instructions. Oligonucleotide primers were designed using Primer Express. The primers used are listed in **Supplementary Table S1**.

The resulting complementary DNA and negative control samples were amplified using the QuantiTect SYBR Green PCR Kit (Qiagen). Reactions were performed in a MicroAmp Optical 96-well reaction plate using a 7500 Sequence Detector (Applied Biosystems). Relative messenger RNA (mRNA) levels were calculated using *gyrB* as a control. All qRT-PCR experiments were performed in duplicate.

Semi-Quantification of Biofilms

Crystal violet staining was applied to semi-quantify biofilm formation of *S. aureus* strains. Briefly, overnight bacterial cultures were diluted into TSBg to a final optical density of 0.05. The diluted cultures were aliquoted to 96-well flat-bottom tissue culture plates (200 μl/well) and incubated at 37°C for 24 h. Wells were washed with PBS after gentile removal of culture supernatants. Bouin fixative was added onto the bottom of the wells to treat biofilm for 1 h. The fixative was gently aspirated out and wells were washed three times with PBS and then stained with 0.4% (wt/vol) crystal violet. Biofilm formation was measured by a MicroELISA autoreader (BioTeK, United States) at 570 nm.

Growth Curve

Growth curves were performed as previously described. Overnight cultures were diluted by 100-fold into fresh TSB media and incubated at 37°C under shaking conditions for 24 h. OD_{600} was measured every 2 h.

Statistics

Statistical analysis was performed using GraphPad Prism v6.0. For the comparison, unpaired, two-tailed *t*-tests were used. All error bars depict the standard deviation. Lines depict the mean.

[2]https://pubmlst.org/multilocus-sequence-typing

[3]https://itol.embl.de/

[4]https://cge.cbs.dtu.dk/services/SCCmecFinder/

DATA AVAILABILITY STATEMENT

The Illumina sequences generated in this study are deposited and available in the Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra) under the study accession number PRJNA624723 with the SRA accession numbers SRR11526821 to SRR11526872.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

ML and LH contributed to the conception and design of the study, and wrote the first draft of the manuscript. HL, LH, YS, and XM performed all the experiments. HL and LZ analyzed the statistics and plotted the figures and tables in this work. YJ, YW, TL, QH, and YD wrote sections of the manuscript. All authors contributed to manuscript revision and read and approved the submitted version.

FUNDING

This work was supported by the Innovative Research Team of High-level Local Universities in Shanghai (grant no. 81861138043), the National Natural Science Foundation of China (grant nos. 81873957 and 81974311), the Shanghai Committee of Science and Technology (grant no. 19JC1413005), and Shanghai Pujiang Program (grant no. 2019PJD026).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.636788/full#supplementary-material

REFERENCES

Armand-Lefevre, L., Ruimy, R., and Andremont, A. (2005). Clonal comparison of *Staphylococcus aureus* isolates from healthy pig farmers, human controls, and pigs. *Emerg. Infect. Dis.* 11, 711–714. doi: 10.3201/eid1105.040866

Chen, C. J., and Huang, Y. C. (2018). Emergence of livestock-associated methicillin-resistant *Staphylococcus aureus*: should it be a concern? *J. Formos. Med. Assoc.* 117, 658–661. doi: 10.1016/j.jfma.2018.04.004

Chuang, Y. Y., and Huang, Y. C. (2015). Livestock-associated meticillin-resistant *Staphylococcus aureus* in Asia: an emerging issue? *Int. J. Antimicrob. Agents* 45, 334–340. doi: 10.1016/j.ijantimicag.2014.12.007

Coombs, G. W., Pang, S., Daley, D. A., Lee, Y. T., Abraham, S., and Leroi, M. (2019). Severe disease caused by community-associated MRSA ST398 type V, Australia, 2017. *Emerg. Infect. Dis.* 25, 190–192. doi: 10.3201/eid2501.181136

Dai, Y., Liu, J., Guo, W., Meng, H., Huang, Q., He, L., et al. (2019). Decreasing methicillin-resistant *Staphylococcus aureus* (MRSA) infections is attributable to the disappearance of predominant MRSA ST239 clones, Shanghai, 2008-2017. *Emerg. Microbes. Infect.* 8, 471–478. doi: 10.1080/22221751.2019.1595161

Foucault, M. L., Courvalin, P., and Grillot-Courvalin, C. (2009). Fitness cost of VanA-type vancomycin resistance in methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 53, 2354–2359. doi: 10.1128/aac.01702-08

Harris, S. R., Feil, E. J., Holden, M. T., Quail, M. A., Nickerson, E. K., Chantratita, N., et al. (2010). Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 327, 469–474. doi: 10.1126/science.1182395

He, L., Zheng, H. X., Wang, Y., Le, K. Y., Liu, Q., Shang, J., et al. (2018). Detection and analysis of methicillin-resistant human-adapted sequence type 398 allows insight into community-associated methicillin-resistant *Staphylococcus aureus* evolution. *Genome Med.* 10:5. doi: 10.1186/s13073-018-0514-9

Hsu, L. Y., Harris, S. R., Chlebowicz, M. A., Lindsay, J. A., Koh, T. H., Krishnan, P., et al. (2015). Evolutionary dynamics of methicillin-resistant *Staphylococcus aureus* within a healthcare system. *Genome Biol.* 16:81. doi: 10.1186/s13059-015-0643-z

Huang, Y. C., and Chen, C. J. (2020). Detection and phylogeny of *Staphylococcus aureus* sequence type 398 in Taiwan. *J. Biomed. Sci.* 27:15. doi: 10.1186/s12929-019-0608-8

Katayama, Y., Ito, T., and Hiramatsu, K. (2001). Genetic organization of the chromosome region surrounding mecA in clinical staphylococcal strains: role of IS431-mediated mecI deletion in expression of resistance in mecA-carrying, low-level methicillin-resistant Staphylococcus haemolyticus. *Antimicrob. Agents Chemother.* 45, 1955–1963. doi: 10.1128/AAC.45.7.1955-1963.2001

Khanna, T., Friendship, R., Dewey, C., and Weese, J. S. (2008). Methicillin resistant *Staphylococcus aureus* colonization in pigs and pig farmers. *Vet. Microbiol.* 128, 298–303. doi: 10.1016/j.vetmic.2007.10.006

Koyama, H., Sanui, M., Saga, T., Harada, S., Ishii, Y., Tateda, K., et al. (2015). A fatal infection caused by sequence type 398 methicillin-resistant *Staphylococcus aureus* carrying the Panton-Valentine leukocidin gene: a case report in Japan. *J. Infect. Chemother.* 21, 541–543. doi: 10.1016/j.jiac.2015.03.013

Larsen, J., Petersen, A., Larsen, A. R., Sieber, R. N., Stegger, M., Koch, A., et al. (2017). Emergence of livestock-associated methicillin-resistant *Staphylococcus aureus* bloodstream infections in Denmark. *Clin. Infect. Dis.* 65, 1072–1076. doi: 10.1093/cid/cix504

Larsen, J., Stegger, M., Andersen, P. S., Petersen, A., Larsen, A. R., Westh, H., et al. (2016). Evidence for human adaptation and foodborne transmission of livestock-associated methicillin-resistant *Staphylococcus aureus*. *Clin. Infect. Dis.* 63, 1349–1352. doi: 10.1093/cid/ciw532

Larsen, M. V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R. L., et al. (2012). Multilocus sequence typing of total-genome-sequenced bacteria. *J. Clin. Microbiol.* 50, 1355–1361. doi: 10.1128/jcm.06094-11

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

Lewis, H. C., Mølbak, K., Reese, C., Aarestrup, F. M., Selchau, M., Sørum, M., et al. (2008). Pigs as source of methicillin-resistant *Staphylococcus aureus* CC398 infections in humans. Denmark. *Emerg. Infect. Dis.* 14, 1383–1389. doi: 10.3201/eid1409.071576

Liu, B., Zheng, D., Jin, Q., Chen, L., and Yang, J. (2019). VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res.* 47, D687–D692. doi: 10.1093/nar/gky1080

Maher, M. C., Alemayehu, W., Lakew, T., Gaynor, B. D., Haug, S., Cevallos, V., et al. (2012). The fitness cost of antibiotic resistance in *Streptococcus pneumoniae*: insight from the field. *PLoS One* 7:e29407. doi: 10.1371/journal.pone.0029407

Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., et al. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* 95, 3140–3145. doi: 10.1073/pnas.95.6.3140

Nielsen, K. L., Pedersen, T. M., Udekwu, K. I., Petersen, A., Skov, R. L., Hansen, L. H., et al. (2012). Fitness cost: a bacteriological explanation for the demise of the first international methicillin-resistant *Staphylococcus aureus* epidemic. *J. Antimicrob. Chemother.* 67, 1325–1332. doi: 10.1093/jac/dks051

Postma, B., Kleibeuker, W., Poppelier, M. J., Boonstra, M., Van Kessel, K. P., Van Strijp, J. A., et al. (2005). Residues 10-18 within the C5a receptor N terminus compose a binding domain for chemotaxis inhibitory protein of *Staphylococcus aureus*. *J. Biol. Chem.* 280, 2020–2027. doi: 10.1074/jbc.M412230200

Postma, B., Poppelier, M. J., van Galen, J. C., Prossnitz, E. R., van Strijp, J. A., de Haas, C. J., et al. (2004). Chemotaxis inhibitory protein of *Staphylococcus aureus* binds specifically to the C5a and formylated peptide receptor. *J. Immunol.* 172, 6994–7001. doi: 10.4049/jimmunol.172.11.6994

Pozzi, C., Waters, E. M., Rudkin, J. K., Schaeffer, C. R., Lohan, A. J., Tong, P., et al. (2012). Methicillin resistance alters the biofilm phenotype and attenuates virulence in *Staphylococcus aureus* device-associated infections. *PLoS Pathog.* 8:e1002626. doi: 10.1371/journal.ppat.1002626

Price, L. B., Stegger, M., Hasman, H., Aziz, M., Larsen, J., Andersen, P. S., et al. (2012). *Staphylococcus aureus* CC398: host adaptation and emergence of methicillin resistance in livestock. *mBio* 3:e00305-11. doi: 10.1128/mBio.00305-11

Rasko, D. A., and Sperandio, V. (2010). Anti-virulence strategies to combat bacteria-mediated disease. *Nat. Rev. Drug Discov.* 9, 117–128. doi: 10.1038/nrd3013

Rudkin, J. K., Edwards, A. M., Bowden, M. G., Brown, E. L., Pozzi, C., Waters, E. M., et al. (2012). Methicillin resistance reduces the virulence of healthcare-associated methicillin-resistant *Staphylococcus aureus* by interfering with the agr quorum sensing system. *J. Infect. Dis.* 205, 798–806. doi: 10.1093/infdis/jir845

Smith, T. C., Gebreyes, W. A., Abley, M. J., Harper, A. L., Forshey, B. M., Male, M. J., et al. (2013). Methicillin-resistant *Staphylococcus aureus* in pigs and farm workers on conventional and antibiotic-free swine farms in the USA. *PLoS One* 8:e63704. doi: 10.1371/journal.pone.0063704

Smith, T. C., Male, M. J., Harper, A. L., Kroeger, J. S., Tinkler, G. P., Moritz, E. D., et al. (2009). Methicillin-resistant *Staphylococcus aureus* (MRSA) strain ST398 is present in midwestern U.S. swine and swine workers. *PLoS One* 4:e4258. doi: 10.1371/journal.pone.0004258

Sun, L., Chen, Y., Wang, D., Wang, H., Wu, D., Shi, K., et al. (2019). Surgical site infections caused by highly virulent methicillin-resistant *Staphylococcus aureus* sequence type 398. China. *Emerg. Infect. Dis.* 25, 157–160. doi: 10.3201/eid2501.171862

Voss, A., Loeffen, F., Bakker, J., Klaassen, C., and Wulf, M. (2005). Methicillin-resistant *Staphylococcus aureus* in pig farming. *Emerg. Infect. Dis.* 11, 1965–1966. doi: 10.3201/eid1112.050428

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

Zhao, C., Liu, Y., Zhao, M., Liu, Y., Yu, Y., Chen, H., et al. (2012). Characterization of community acquired *Staphylococcus aureus* associated with skin and soft tissue infection in Beijing: high prevalence of PVL+ ST398. *PLoS One* 7:e38577. doi: 10.1371/journal.pone.0038577

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

frontiers in Microbiology

MINI REVIEW
published: 16 March 2021
doi: 10.3389/fmicb.2021.637656

Genomic Insights Into Last-Line Antimicrobial Resistance in Multidrug-Resistant *Staphylococcus* and Vancomycin-Resistant *Enterococcus*

*Adrianna M. Turner[1], Jean Y. H. Lee[1,2], Claire L. Gorrie[1,3], Benjamin P. Howden[1,3,4] and Glen P. Carter[1,3]**

[1] Department of Microbiology and Immunology, Doherty Institute, The University of Melbourne, Melbourne, VIC, Australia, [2] Department of Infectious Diseases, Monash Health, Melbourne, VIC, Australia, [3] Antimicrobial Reference and Research Unit, Microbiological Diagnostic Unit Public Health Laboratory, Department of Microbiology and Immunology, Doherty Institute, The University of Melbourne, Melbourne, VIC, Australia, [4] Department of Infectious Diseases, Austin Health, Melbourne, VIC, Australia

OPEN ACCESS

Edited by:
Guido Werner,
Robert Koch Institute (RKI), Germany

Reviewed by:
Ana P. Tedim,
Institute of Health Sciences Studies of Castilla y León (IECSCYL), Spain
Ana R. Freitas,
University of Porto, Portugal

***Correspondence:**
Glen P. Carter
glen.carter@unimelb.edu.au

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

Received: 11 December 2020
Accepted: 25 February 2021
Published: 16 March 2021

Citation:
Turner AM, Lee JYH, Gorrie CL, Howden BP and Carter GP (2021) Genomic Insights Into Last-Line Antimicrobial Resistance in Multidrug-Resistant Staphylococcus and Vancomycin-Resistant Enterococcus. Front. Microbiol. 12:637656. doi: 10.3389/fmicb.2021.637656

Multidrug-resistant *Staphylococcus* and vancomycin-resistant *Enterococcus* (VRE) are important human pathogens that are resistant to most clinical antibiotics. Treatment options are limited and often require the use of 'last-line' antimicrobials such as linezolid, daptomycin, and in the case of *Staphylococcus*, also vancomycin. The emergence of resistance to these last-line antimicrobial agents is therefore of considerable clinical concern. This mini-review provides an overview of resistance to last-line antimicrobial agents in *Staphylococcus* and VRE, with a particular focus on how genomics has provided critical insights into the emergence of resistant clones, the molecular mechanisms of resistance, and the importance of mobile genetic elements in the global spread of resistance to linezolid.

Keywords: linezolid, daptomycin, vancomycin, genomics, *Enterococcus*, *Staphylococcus*

INTRODUCTION

Staphylococcus and *Enterococcus* are Gram-positive cocci that are recognized as globally important opportunistic pathogens that can cause serious infections in humans, especially in hospitalized patients (Otto, 2009; Gilmore et al., 2013; Lee A.S. et al., 2018). Over recent decades, there has been a significant increase in the rates of acquired antimicrobial resistance (AMR) in these species, either through the acquisition of resistance determinants by horizontal gene transfer of mobile genetic elements (MGEs), or through mutations that alter gene expression or binding sites in native genes. This has resulted in the emergence of polyclonal lineages that are resistant to front-line therapeutic agents. In staphylococci, this includes the development and global spread of methicillin-resistant *Staphylococcus aureus* (MRSA) (Enright et al., 2002) and the recent emergence of multidrug-resistant *Staphylococcus epidermidis* (MDRSE) (Lee J.Y.H. et al., 2018). In enterococci, the emergence and dissemination of vancomycin-resistant enterococci (VRE) (Arias and Murray, 2012), particularly in two healthcare-associated species, vancomycin-resistant *Enterococcus faecalis* (VREfs) and vancomycin-resistant *Enterococcus faecium* (VREfm), is of particular clinical concern.

As resistance to different antimicrobials increases in MRSA, MDRSE, and VRE, effective and appropriate treatment becomes increasingly difficult. Limited "last-line" therapeutic options, such as linezolid, daptomycin, and, in the case of staphylococci, vancomycin, remain to treat these infections. However, the high prevalence of these species in healthcare settings (Weiner et al., 2016; Weiner-Lastinger et al., 2020), the increasing clinical use of these last-resort antimicrobials, and the ability of these species to readily develop resistance, provides strong selective pressure for the emergence of clones that are resistant to these agents. Accordingly, resistance to these last-line agents with resultant treatment failure has been increasingly reported. Individual and multiple resistance to linezolid, daptomycin, and vancomycin have been reported in clinical MRSA and MDRSE. While combined resistance to linezolid and daptomycin has recently been reported in VREfm (Wardenburg et al., 2019).

Genomic data can be used to understand the evolution of resistance and investigate the underlying resistance mechanisms. Phylogenetic analyses based on whole-genome sequencing (WGS) can also identify the presence of clones associated with specific resistance determinants, and track their emergence and global spread. The utilization of WGS in hospital infection control to identify outbreaks and transmission events associated with resistant strains is increasingly common. In this mini-review we will provide an overview of the application of genomic analyses to provide critical insights into last-line AMR in MRSA, MDRSE, and VRE, with a particular focus on the emergence of resistant clones, the molecular mechanisms of resistance, and the importance of MGEs in the global spread of linezolid resistance.

LINEZOLID RESISTANCE

Linezolid was the first oxazolidinone approved for clinical use due to its bacteriostatic activity against Gram-positive species. Its mechanism of action inhibits protein synthesis by preventing the formation of the ternary complex between $tRNA^{fMet}$, mRNA, and the ribosome (Hashemian et al., 2018). Although preliminary research indicated that resistance to linezolid should be rare, resistance in clinical enterococci and staphylococci have been increasingly reported since 2001 (Zurenko et al., 1996; Xiong et al., 2000; Gu et al., 2013; Bender et al., 2018a; Kosecka-Strojek et al., 2020). Early studies using PCR amplicon sequencing in clinical VRE and *S. aureus* isolates identified mutations in the 23S rRNA genes, that form the binding pocket of the ribosomal peptidyl transferase centre (PTC) to which linezolid binds (Howe et al., 2003; Meka et al., 2004; Livermore et al., 2007). A number of resistance-associated mutations, some specific to staphylococci, others conserved across staphylococci and enterococci, have been identified within and outside the PTC; the G2576T (*E. coli* nucleotide numbering) mutation is most common (**Figure 1**). Conferring high-level resistance in MRSA, MDRSE, and VRE, the G2576T 23S rRNA mutation is associated with linezolid treatment failure. Additional single nucleotide polymorphisms (SNPs) in the ribosomal proteins L3 and L4 have been identified through WGS of linezolid-resistant clinical staphylococci (Locke et al., 2009a,b, 2010; Endimiani et al., 2011). Mutations in the L3 and L4 proteins appear to be less common in enterococci, but several reports have identified SNPs in L3 (S133L) and L4 (T35A, N79D, I98V, and N130K) (Mendes et al., 2014; Hua et al., 2019). Further, mutations (A138G, C141T, and G166A) in ribosomal protein L22 have also been identified in linezolid-resistant staphylococci (Bender et al., 2015), as well as a S77T mutation identified in one linezolid-intermediate *E. faecium* (Lee et al., 2017). The L3, L4, and L22 proteins are in close proximity to the linezolid binding site in the ribosomal PTC, with identified mutations typically adjacent to the PTC, which may affect linezolid binding (Long and Vester, 2012). Using WGS, several studies have documented the local emergence of linezolid resistance in response to linezolid treatment during individual infections (Seedat et al., 2006; Wong et al., 2010; Chen et al., 2018). These studies identified common mutations in 23S rRNA and/or L3, L4, or L22 proteins that independently and repeatedly arise in different species of staphylococci and enterococci from different patients, demonstrating the conserved nature of mutational linezolid resistance and suggesting that convergent evolution is occurring.

Identification of Transferable Linezolid Resistance

Transferable or acquired resistance to linezolid has been identified in both staphylococcal and enterococcal clinical isolates. The first transferable resistance gene identified was the multi-resistance gene *cfr*, that encodes a rRNA methyltransferase (Arias et al., 2008). Cfr catalyzes the post-transcriptional methylation of nucleotide A2503 in the 23S rRNA, which confers combined resistance to five different classes of antimicrobials: phenicols, lincosamides, oxazolidinones, pleuromutilins, and Streptogramin A; that bind at overlapping non-identical sites at the PTC, known as the $PhLOPS_A$ phenotype (Long et al., 2006). The first linezolid-resistant clinical isolate bearing *cfr* was reported in 2005, in which *cfr* had chromosomally integrated into a MRSA strain, within an IS21-558 MGE together with *ermB*, resulting in resistance to all clinically relevant antibiotics that inhibit protein synthesis (Toh et al., 2007). The presence of a *cfr* gene has subsequently been identified in staphylococci, including MRSA (Locke et al., 2010; Morales et al., 2010; Antonelli et al., 2016) and MDRSE (Bonilla et al., 2010; Lazaris et al., 2017), and in *Enterococcus* (Liu et al., 2014; Bender et al., 2016; Fioriti et al., 2020; Ruiz-Ripa et al., 2020) isolates worldwide, although the overall prevalence and distribution in different sequence types (STs) is poorly characterized. In MDRSE, phylogenetic analysis suggests that *cfr* carriage is most commonly associated with healthcare-associated, clonal complex 2 strains (ST22, ST2, ST5, and ST168), where carriage is associated with a multidrug-resistant phenotype (Mendes et al., 2012).

Several recent studies have identified novel variants of the *cfr* gene [*cfr*(B) (Bender et al., 2016), *cfr*(C) (Candela et al., 2017), and *cfr*(D) (Guerin et al., 2020)], which are present within a variety of MGEs including multiple plasmids, insertion sequences, and transposons (summarized in Sadowy, 2018 for enterococci). Furthermore, a diverse range of resistance determinants have been co-located with *cfr*, suggesting that

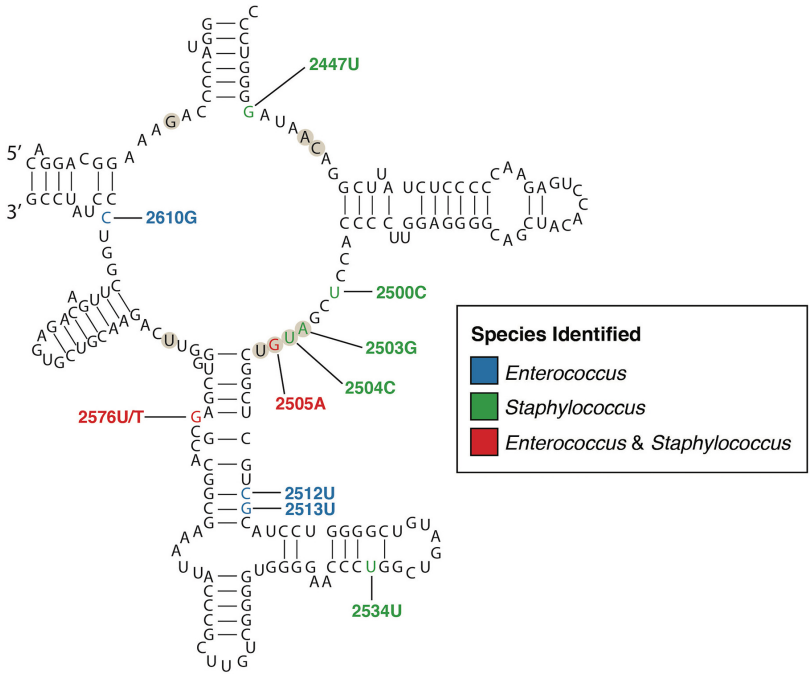


FIGURE 1 | Secondary structure of the peptidyl transferase loop of domain V of 23S rRNA (*E. coli* numbering). The nucleotides that form the linezolid binding pocket are marked with circles. Nucleotide positions where mutations confer linezolid resistance are colored according to the species identified, blue for *Enterococcus*, green for *Staphylococcus*, and red for both. Only mutations with a published relationship in clinical isolates have been included (Wong et al., 2010; Long and Vester, 2012; Mendes et al., 2012; Chen et al., 2018; Wardenburg et al., 2019).

broader antimicrobial use may co-select for the spread of these MGEs. Variation in the genes surrounding *cfr* in clinical strains suggests that *cfr*-mediated linezolid resistance has been independently acquired on multiple occasions in clinical staphylococci and enterococci isolates worldwide. Although the reservoir for *cfr* remains unknown, one study using WGS data from *cfr*-positive *E. faecium* suggests that *E. faecium* may be an important reservoir of *cfr*(D), which has clear implications for infection control (Guerin et al., 2020). Alternatively, *cfr* variants may originate from *Bacillales*, since three *cfr*-like genes have been identified (Hansen et al., 2012).

Resistance to linezolid also arises through the acquisition of the transferable ribosomal protection genes, *optrA,* and *poxtA*. OptrA and PoxtA are part of the F-lineage in the ATP-binding cassette (ABC) superfamily of proteins that are associated with AMR (Jones et al., 2009). These ABC-F proteins mediate resistance to several different classes of anti-ribosomal antimicrobials. For instance, *optrA* confers resistance to oxazolidinones and phenicols. While, *poxtA* confers resistance to phenicols, oxazolidinones, and tetracyclines. The family of ABC-F proteins have been shown to mediate resistance by displacing the antibiotic from the ribosome to provide ribosomal protection within the *Staphylococcus* genus (Sharkey et al., 2016). OptrA was first detected in 2009, on pE349 conjugative plasmids found within three linezolid-resistant *E. faecalis* clinical isolates (Wang et al., 2015). A phylogenetic analysis of 43 *optrA*-carrying enterococci revealed nine different variants of the *optrA* gene (*optrA*$_{WT}$, *optrA*1-5, *optrA*6, *optrA*7, and *optrA*8). These variants were distributed throughout diverse genetic backgrounds of *E. faecium* and *E. faecalis*, indicating that multiple independent acquisitions of *optrA* have occurred in these species (Bender et al., 2018b). The presence of *optrA* has been subsequently observed in both vancomycin-resistant and vancomycin-susceptible *E. faecium* and *E. faecalis* from clinical and food-producing animal isolates globally; but only reported in food-producing animal isolates of *S. aureus* (Zhu et al., 2020). This suggests that food-producing animals may serve as a reservoir for *optrA* transfer. Phylogenetic analysis of *optrA*-carrying enterococci, including some strains from a hospital outbreak, in Europe and South America led to the frequent identification of ST480 *E. faecalis*, suggesting a potential association between ST480 *E. faecalis* strains and *optrA* carriage (Sassi et al., 2019; Egan et al., 2020b; Freitas et al., 2020a; Moure et al., 2020). OptrA has also been reported in several different clonal backgrounds of *E. faecalis* (Bender et al., 2019) and has been identified in pets carrying *E. faecalis* (Wu et al., 2019). In *E. faecium*, *optrA* has been identified in several different lineages, including ST17, ST80, ST117, ST412, and ST650, indicating that *optrA* carriage is present in phylogenetically diverse strains in *E. faecium* (Morroni et al., 2018; Sassi et al., 2019).

The *poxtA* linezolid resistance gene was first detected in a MRSA clinical strain in 2016 (Antonelli et al., 2018). The *poxtA* gene exhibits genetic similarity (32% identity and 95% coverage) to *optrA* and possesses conserved features with other members of the ABC-F family (Antonelli et al., 2018). Carriage of the *poxtA* gene appears to occur sporadically in MRSA and VREfm

clinical isolates (Papagiannitsis et al., 2019), with the gene most commonly identified in food-producing animals (Huang et al., 2017; Elghaieb et al., 2019; Hao et al., 2019; Lei et al., 2019; Fioriti et al., 2020; Freitas et al., 2020b; Na et al., 2020).

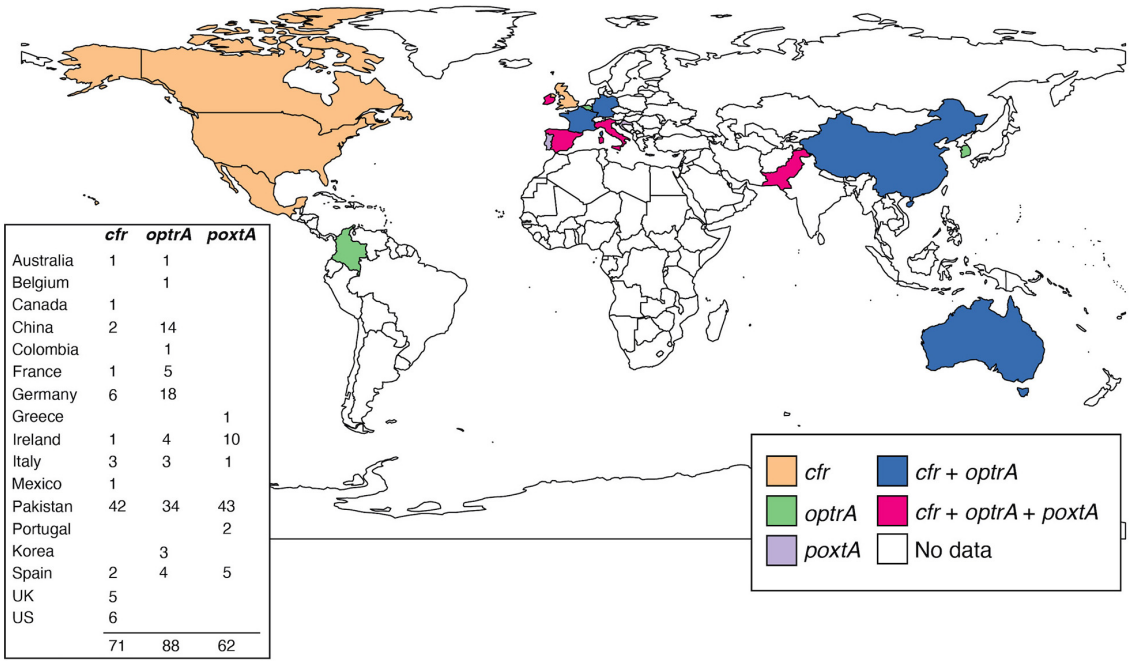
Phylogenetics of Transferable Linezolid Resistance

Since their initial discovery, the prevalence of *cfr*, *optrA*, and *poxtA* genes in clinical multidrug-resistant enterococcal and staphylococcal isolates has increased (**Figure 2**; Bi et al., 2018). To date, the highest prevalence of *optrA* and/or *poxtA* (35/154; 23%) were reported in *E. faecium* and *E. faecalis* in a retrospective Irish study from 2016 to 2019, suggesting significant selection pressure for gene maintenance (Egan et al., 2020b). Genomics has provided critical insights into the genetic relatedness of resistant strains, the structure of MGEs and their role in the mobilization of linezolid resistance determinants. Examples include: an analysis of *optrA*-carrying plasmids from *E. faecalis*, which found *optrA* flanked by two active copies of IS*1216* that can facilitate its spread (Shang et al., 2019); the widespread identification of *optrA* located in a Tn*6674*-like element in *E. faecalis* isolated from different countries and sources (Freitas et al., 2020b); IS*1216E* elements that have been shown to mediate the insertion of *poxtA* into a pT-E1077-31 plasmid in *E. faecium* (Shan et al., 2020); while in MRSA, a Tn*6349* composite transposon carrying both *poxtA* and *cfr* was found to possess a modular structure consisting of fragments from other previously described MGEs (D'Andrea et al., 2019). The presence of these MGEs in *cfr*, *poxtA,* and *optrA* carrying multi-resistance plasmids likely aids in the persistence and dissemination of linezolid resistance genes among these Gram-positive pathogens.

Recent studies have also used genomics to identify linezolid-resistant strains for surveillance and infection control purposes. These studies utilized different bioinformatic tools, including LRE-finder (Hasman et al., 2019) and the Resistance Gene Identifier using the Comprehensive Antibiotic Resistance Database (Alcock et al., 2020), to detect *cfr*, *optrA*, and *poxtA* resistance genes in WGS data, enabling rapid genotypic surveillance (Wardenburg et al., 2019; Eisenberger et al., 2020; Moure et al., 2020) and phylogenetic inference (Kerschner et al., 2019). For instance, WGS of *E. faecium* during an Irish hospital outbreak identified 19 nearly identical ST80 clinical isolates (allelic difference range < 10 SNPs) all carrying *optrA*-encoded within conjugative plasmid pEfmO_03. Subsequent implementation of enhanced infection control interventions successfully eliminated the outbreak (Egan et al., 2020a).

GENOMIC INSIGHTS INTO DAPTOMYCIN RESISTANCE

Shortly after FDA approval of daptomycin for use in MRSA and in the context of off-label use in VRE, resistance, and



	cfr	*optrA*	*poxtA*
Australia	1	1	
Belgium		1	
Canada	1		
China	2	14	
Colombia		1	
France	1	5	
Germany	6	18	
Greece			1
Ireland	1	4	10
Italy	3	3	1
Mexico	1		
Pakistan	42	34	43
Portugal			2
Korea		3	
Spain	2	4	5
UK	5		
US	6		
	71	88	62

FIGURE 2 | Snapshot of reported *cfr*, *optrA*, and *poxtA* genes in human, clinical, linezolid-resistant *E. faecium*. Each country is colored based on which gene(s) have been identified. The number of reported strains for each country is shown. It should be noted that the identification of each gene within regions may not accurately represent the genomic epidemiology in *E. faecium*, due to both collection and availability biases. Only published reports have been included (Bender et al., 2016, 2018b; Lazaris et al., 2017; Wardenburg et al., 2019; Egan et al., 2020a,b; Moure et al., 2020). Map was generated in R (v. 4.0.3) using the ggplot2 package (v. 1.3.0).

corresponding clinical treatment failures were reported in both species (Skiest, 2006; Arias et al., 2007; Boyle-Vavra et al., 2011). The development of daptomycin resistance commonly coincides with prior patient exposure, indicating that therapeutic use of daptomycin is likely the main driver of resistance. In *S. aureus*, the emergence of daptomycin resistance has also been associated with the use of un-related antimicrobials, including vancomycin (Tran et al., 2015) and rifampicin (Guérillot et al., 2018b). Importantly, horizontally acquired and transferable resistance to daptomycin has not been reported to date. The worldwide prevalence of daptomycin resistance is poorly defined since few large-scale surveillance studies exist. A recent meta-analysis in *S. aureus*, suggested that only 0.1% of *S. aureus* and 0.1% of MRSA were daptomycin-resistant (Shariati et al., 2020). In enterococci, the prevalence of daptomycin resistance was estimated to be 2.6% in VREfm and 0.1% in VREfs, based on a systemic review of previous studies (Kelesidis et al., 2011). Little is known about which lineages of staphylococci and enterococci are associated with the development daptomycin resistance. In *S. aureus*, genomic analyses indicate that daptomycin resistance has arisen in ST5, ST398, and ST22 isolates, based on data from the United States and South America (Dabul and Camargo, 2014; Damasco et al., 2019). In *S. epidermidis*, no comprehensive phylogenetic analyses of daptomycin-resistant strains have been completed, but resistance has been reported in ST2, internationally (Lee J.Y.H. et al., 2018). Similarly, no phylogenetic studies have been completed for *E. faecalis*, but a novel ST736 clone associated with daptomycin resistance has been identified for *E. faecium* (Wang et al., 2014) in some United States hospitals (Wang et al., 2018), as has daptomycin-resistant ST80 *E. faecium* (Udaondo et al., 2020).

Daptomycin's mode of action involves binding to the anionic phosphoglycerol and various undecaprenyl-coupled cell envelope precursors within the bacterial cell membrane in the presence of Ca^{2+}, blocking cell wall synthesis and triggering delocalization of biosynthetic proteins from the membrane, resulting in bacterial cell death (Grein et al., 2020). Although the specific mechanisms leading to resistance are not fully understood, comparative genomic studies have been pivotal in identifying genetic determinants of daptomycin resistance. Genetically, a critical step toward developing a resistant phenotype appears to involve mutations in two groups of genes: (1) bacterial regulatory systems that function in cell wall homeostasis and (2) enzymes involved in membrane phospholipid metabolism.

In *S. aureus* and *S. epidermidis*, mutations in *mprF* (S295L, S337L, T345L, I420N, and L826F) and *walKR* (*yycFG*) ($WalK^{V500F}$ and $WalR^{R216S}$) are associated with the development of daptomycin resistance in clinical isolates and laboratory-derived strains (Howden et al., 2011; Yang et al., 2009, 2013; Jiang et al., 2019). The MprF enzyme produces cationic phospholipid lysylphophatidyl-glyercol (LysPG) and flips it from the inner to outer leaflet of the cytoplasmic membrane, which decreases the negative cell surface charge (Oku et al., 2004). Although not characterized in *S. epidermidis*, the mutations in *S. aureus* likely increase the enzymatic activity of MprF since daptomycin-resistant isolates display increased levels of LysPG in their membrane, which variably leads to a more cationic cell surface that electrostatically repels the daptomycin-Ca^{2+} complex (Mishra et al., 2011). For mutations in *walKR*, the mechanism leading to resistance is distinct in *S. aureus* and *S. epidermidis*, since *S. aureus* strains carrying mutations in *walKR* display increased surface charge while *S. epidermidis* strains display decreased biofilm formation, compared to the wild-type strains (Jiang et al., 2019). Comparative genomic studies in *S. aureus* have also identified mutations in the cardiolipin synthase gene *cls*2 and RNA polymerase β and β' subunits (*rpoB* and *rpoC*) that are associated with resistance to daptomycin, although these are less frequently observed than mutations in *mprF* and *walKR* (Miller et al., 2016). Transposon-sequencing in *S. aureus* has also identified multi-component sensing genes, specifically *graRS* and *vraFG* (which regulate cell envelope processes), as being associated with daptomycin resistance, with these genes being important for survival following daptomycin stress (Rajagopal et al., 2016; Coe et al., 2019). However, *graRS* and *vraFG* mutations have so far not been described in clinical daptomycin-resistant *S. aureus* isolates.

In enterococci, comparative genomic studies have shown that mutations in the three-component regulatory system *liaFSR* and cardiolipin synthase *cls* are associated with the development of clinical daptomycin resistance. Mutations within the LiaFSR three-component system, which regulates the cell envelope response to cationic host defense peptides and antimicrobial stress, have been frequently identified in daptomycin-resistant clinical isolates, with the $LiaS^{T120A}$ and $LiaR^{W73C}$ mutations being of particular importance (Arias et al., 2011). A genomic analysis of *E. faecium* clinical isolates found that isolates with higher daptomycin MICs ($\geq$ 4 mg/L) usually contained mutations in the *liaFSR* system (Diaz et al., 2014). Mutations in Cls (H215R and R218Q), encoding a transmembrane protein that catalyzes the synthesis of cardiolipin from phosphoglycerol, are also frequently associated with daptomycin resistance (Arias et al., 2011; Palmer et al., 2011). However, molecular studies have thus far not demonstrated a causal relationship between *cls* and daptomycin resistance in *E. faecium* (Tran et al., 2013). Since *cls* mutations are often found in conjunction with substitutions in the *liaFSR* regulon, they likely contribute to the progression of an isolate developing a daptomycin-resistant phenotype rather than independently causing resistance (Davlieva et al., 2013). Recent genomic studies in enterococci have also identified daptomycin-resistant isolates that contain wild-type *liaFSR* and *cls* alleles, indicating that other unknown genetic determinants of daptomycin resistance exist that require further investigation (Wang et al., 2018; Prater et al., 2019). Indeed, putative associations with daptomycin resistance and mutations within *mprF, cfa, pgsA*, and *dlt* have been suggested, but have not yet been verified.

GENOMIC INSIGHTS INTO VANCOMYCIN RESISTANCE IN *STAPHYLOCOCCUS*

Vancomycin is a critical last-line antibiotic for treating infections caused by multidrug-resistant *Staphylococcus*. Thirty-nine years

after FDA approval of vancomycin, the first instances of *S. aureus* with intermediate vancomycin resistance (MICs 4 to 8 mg/L) were reported (Sieradzki et al., 1999; Smith et al., 1999). Soon after, the first clinical case of a vancomycin-resistant *S. aureus* isolate (MIC 1024 mg/L), carrying the *vanA* vancomycin resistance operon, was reported (Chang et al., 2003). Subsequent molecular studies demonstrated that the *vanA* operon originated from a co-infecting VREfs strain, raising significant concerns about the potential dissemination of vancomycin resistance in *S. aureus* from VRE reservoirs (de Niederhäusern et al., 2011). Fortunately, this has not eventuated. Detailed genomic studies have demonstrated that the transfer of the *van* operon to *S. aureus* rarely occurs, with very few reports worldwide (McGuinness et al., 2017). The rare occurrence of VRSA may be due to the fitness cost associated with carriage of the *van* genes, as a longer lag phase was observed during *in vitro* growth with vancomycin, and the plasmid carrying the *van* genes was found to be genetically unstable. Further, a MRSA PBP2a strain was unable to utilize lipid II with the D-Alanyl-D-lactate alteration conferred by the *van* operon (Foster, 2017). A single study identified *S. aureus* clonal complex 5 (CC5) as the genetic background in which the *van* operon is most commonly acquired. This study found that CC5 vancomycin-resistant *S. aureus* (VRSA) strains were genetically distinct, indicating independent acquisition of the *vanA* operon had occurred (Kos et al., 2012). This suggests that the spread of *van* operons within *S. aureus* may occur should appropriate selective conditions develop. Continued vigilance is therefore paramount.

Vancomycin-intermediate *S. aureus* (VISA) resulting from the progressive accumulation of mutations in key essential and regulatory genes are clinically more common than true vancomycin-resistant *S. aureus*. VISA isolates are commonly characterized by a thickened cell wall, slower growth, and increased autolysis (reviewed in Howden et al., 2010). Genetically, VISA strains typically contain mutations in two-component regulators *vraRS* (Cui et al., 2009), *graRS* (Cui et al., 2009), and *walKR* (Howden et al., 2011), which function in cell-wall synthesis, but also may contain mutations in the rifampicin resistance determining region of *rpoB* (Cui et al., 2010). A convergent evolution analysis of 7,099 *S. aureus* genomes suggested that clinical rifampicin use may promote the emergence of multidrug-resistant lineages of *S. aureus* that contain rifampicin resistance mutations in *rpoB* that also confer intermediate resistance to vancomycin (Guérillot et al., 2018a). Although VISA can be found in any genetic background, ST5 is most commonly associated with VISA (Howden et al., 2014), although ST239 isolates tend to display the highest vancomycin MICs (Holmes et al., 2014). Genome-wide association studies (GWAS) performed in ST239 *S. aureus* have identified mutations in *rpoB* (H481Y/L/N) (Alam et al., 2014) and the *walKR* genes (Baines et al., 2015) as being most strongly associated with the development of the VISA phenotype.

In *S. epidermidis*, the global dissemination of three nearly pan-drug-resistant lineages (two ST2 and one ST23) with heteroresistance to vancomycin was recently demonstrated (Lee J.Y.H. et al., 2018). Analyses of WGS data from 419 clinical isolates from 96 institutions in 24 countries, sampling at least 77 ST types, identified dual D471E and I527M RpoB mutations to be the most common cause of rifampicin resistance in *S. epidermidis*, accounting for 86.6% of mutations. Of note, the D471E and I527M combination occurred almost exclusively in isolates from the ST2 and ST23 lineages. Through site-specific mutagenesis, the dual RpoB mutation was shown to confer high-level rifampicin resistance as well as reduced susceptibility to vancomycin and teicoplanin. The presence of these same mutations in multiple genetically diverse backgrounds was consistent with their independent emergence with subsequent fixation, presumably due to their advantageous antimicrobial resistant phenotype. Furthermore, acquisition of additional resistance to linezolid (mediated by *cfr* carrying plasmids and/or mutations in 23S rRNA, L3 and L4) and daptomycin (substitutions in *mprF*) rendered some European *S. epidermidis* essentially untreatable. Further examples of convergent evolution include the presence of H481Y/L/N substitutions in RpoB that confer vancomycin resistance in both *S. aureus* (Guérillot et al., 2018a) and *S. epidermidis* (Gao et al., 2013; Lee J.Y.H. et al., 2018; Dao et al., 2020).

CONCLUSION

Collectively, MRSA, MDRSE, and VRE pose a significant clinical and economic healthcare burden. This has been compounded by increasing resistance to the key last-line agents daptomycin, linezolid and in the case of *Staphylococcus*, vancomycin. Genomic analyses have been pivotal in providing evolutionary insights into the underlying genetic mechanisms of resistance that have arisen in all three species. In particular, the identification of transferable and acquired linezolid resistance determinants through genomic studies, has been critical for understanding the global dissemination and genetic mobility of these genes. Genomic analyses using WGS data is therefore a powerful method for understanding last-line AMR in these and other clinically relevant multidrug-resistant species.

AUTHOR CONTRIBUTIONS

GC and AT conceptualized the mini-review topic. All authors contributed to writing the mini-review and editing the final version.

FUNDING

This work was supported by the Australian Government Research Training Program (AMT). GC was supported by a National Health and Medical Research Council (NHMRC) of Australia Ideas grant (GNT1185213). BH was supported by a NHMRC Practitioner Fellowship (GNT1105905).

REFERENCES

Alam, M. T., Petit, R. A., Crispell, E. K., Thornton, T. A., Conneely, K. N., Jiang, Y., et al. (2014). Dissecting vancomycin-intermediate resistance in *Staphylococcus aureus* using genome-wide association. *Genome Biol. Evol.* 6, 1174–1185. doi: 10.1093/gbe/evu092

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

Antonelli, A., D'Andrea, M. M., Brenciani, A., Galeotti, C. L., Morroni, G., Pollini, S., et al. (2018). Characterization of poxtA, a novel phenicol–oxazolidinone–tetracycline resistance gene from an MRSA of clinical origin. *J. Antimicrob. Chemother.* 73, 1763–1769. doi: 10.1093/jac/dky088

Antonelli, A., D'Andrea, M. M., Galano, A., Borchi, B., Brenciani, A., Vaggelli, G., et al. (2016). Linezolid-resistant cfr-positive MRSA, Italy. *J. Antimicrob. Chemother.* 71, 2349–2351. doi: 10.1093/jac/dkw108

Arias, C. A., and Murray, B. E. (2012). The rise of the *Enterococcus*: beyond vancomycin resistance. *Nat. Rev. Microbiol.* 10, 266–278. doi: 10.1038/nrmicro2761

Arias, C. A., Panesso, D., McGrath, D. M., Qin, X., Mojica, M. F., Miller, C., et al. (2011). Genetic basis for in vivo daptomycin resistance in *Enterococci*. *N. Engl. J. Med.* 365, 892–900. doi: 10.1056/NEJMoa1011138

Arias, C. A., Torres, H. A., Singh, K. V., Panesso, D., Moore, J., Wanger, A., et al. (2007). Failure of daptomycin monotherapy for endocarditis caused by an *Enterococcus faecium* strain with vancomycin-resistant and vancomycin-susceptible subpopulations and evidence of in vivo loss of the vanA gene cluster. *Clin. Infect. Dis.* 45, 1343–1346. doi: 10.1086/522656

Arias, C. A., Vallejo, M., Reyes, J., Panesso, D., Moreno, J., Castañeda, E., et al. (2008). Clinical and microbiological aspects of linezolid resistance mediated by the cfr gene encoding a 23S rRNA methyltransferase. *J. Clin. Microbiol.* 46, 892–896. doi: 10.1128/JCM.01886-07

Baines, S. L., Holt, K. E., Schultz, M. B., Seemann, T., Howden, B. O., Jensen, S. O., et al. (2015). Convergent adaptation in the dominant global hospital clone ST239 of methicillin-resistant *Staphylococcus aureus*. *mBio* 6:e00080-15. doi: 10.1128/mBio.00080-15

Bender, J. K., Cattoir, V., Hegstad, K., Sadowy, E., Coque, T. M., Westh, H., et al. (2018a). Update on prevalence and mechanisms of resistance to linezolid, tigecycline and daptomycin in enterococci in Europe: towards a common nomenclature. *Drug Resist. Updat.* 40, 25–39. doi: 10.1016/j.drup.2018.10.002

Bender, J. K., Fleige, C., Klare, I., Fiedler, S., Mischnik, A., Mutters, N. T., et al. (2016). Detection of a cfr(B) variant in German *Enterococcus faecium* clinical isolates and the impact on linezolid resistance in *Enterococcus* spp. *PLoS One* 11:e0167042. doi: 10.1371/journal.pone.0167042

Bender, J. K., Fleige, C., Klare, I., and Werner, G. (2019). Development of a multiplex-PCR to simultaneously detect acquired linezolid resistance genes cfr, optrA and poxtA in enterococci of clinical origin. *J. Microbiol. Methods* 160, 101–103. doi: 10.1016/j.mimet.2019.03.025

Bender, J. K., Fleige, C., Lange, D., Klare, I., and Werner, G. (2018b). Rapid emergence of highly variable and transferable oxazolidinone and phenicol resistance gene optrA in German *Enterococcus* spp. clinical isolates. *Int. J. Antimicrob. Agents* 52, 819–827. doi: 10.1016/j.ijantimicag.2018.09.009

Bender, J. K., Strommenger, B., Steglich, M., Zimmermann, O., Fenner, I., Lensing, C., et al. (2015). Linezolid resistance in clinical isolates of *Staphylococcus epidermidis* from German hospitals and characterization of two cfr-carrying plasmids. *J. Antimicrob. Chemother.* 70, 1630–1638. doi: 10.1093/jac/dkv025

Bi, R., Qin, T., Fan, W., Ma, P., and Gu, B. (2018). The emerging problem of linezolid-resistant enterococci. *J. Global Antimicrob. Resist.* 13, 11–19. doi: 10.1016/j.jgar.2017.10.018

Bonilla, H., Huband, M. D., Seidel, J., Schmidt, H., Lescoe, M., McCurdy, S. P., et al. (2010). Multicity outbreak of linezolid-resistant *Staphylococcus epidermidis* associated with clonal spread of a cfr-containing strain. *Clin. Infect. Dis.* 51, 796–800. doi: 10.1086/656281

Boyle-Vavra, S., Jones, M., Gourley, B. L., Holmes, M., Ruf, R., Balsam, A. R., et al. (2011). Comparative genome sequencing of an isogenic pair of USA800 clinical methicillin-resistant *Staphylococcus aureus* isolates obtained before and after daptomycin treatment failure. *Antimicrob. Agents Chemother.* 55, 2018–2025. doi: 10.1128/AAC.01593-10

Candela, T., Marvaud, J.-C., Nguyen, T. K., and Lambert, T. (2017). A cfr-like gene cfr(C) conferring linezolid resistance is common in *Clostridium difficile*. *Int. J. Antimicrob. Agents* 50, 496–500. doi: 10.1016/j.ijantimicag.2017.03.013

Chang, S., Sievert, D. M., Hageman, J. C., Boulton, M. L., Tenover, F. C., Downes, F. P., et al. (2003). Infection with vancomycin-resistant *Staphylococcus aureus* containing the vanA resistance gene. *N. Engl. J. Med.* 348, 1342–1347. doi: 10.1056/NEJMoa025025

Chen, M., Pan, H., Lou, Y., Wu, Z., Zhang, J., Huang, Y., et al. (2018). Epidemiological characteristics and genetic structure of linezolid-resistant *Enterococcus faecalis*. *Infect. Drug Resist.* 11, 2397–2409. doi: 10.2147/IDR.S181339

Coe, K. A., Lee, W., Stone, M. C., Komazin-Meredith, G., Meredith, T. C., Grad, Y. H., et al. (2019). Multi-strain Tn-Seq reveals common daptomycin resistance determinants in *Staphylococcus aureus*. *PLoS Pathog.* 15:e1007862. doi: 10.1371/journal.ppat.1007862

Cui, L., Isii, T., Fukuda, M., Ochiai, T., Neoh, H., Camargo, I. L. B., et al. (2010). An RpoB mutation confers dual heteroresistance to daptomycin and vancomycin in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 54, 5222–5233. doi: 10.1128/AAC.00437-10

Cui, L., Neoh, H., Shoji, M., and Hiramatsu, K. (2009). Contribution of vraSR and graSR point mutations to vancomycin resistance in vancomycin-intermediate *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 53, 1231–1234. doi: 10.1128/AAC.01173-08

Dabul, A. N. G., and Camargo, I. L. B. C. (2014). Molecular characterization of methicillin-resistant *Staphylococcus aureus* resistant to tigecycline and daptomycin isolated in a hospital in Brazil. *Epidemiol. Infect.* 142, 479–483. doi: 10.1017/S0950268813001325

Damasco, A. P., Costa, T. M., da Morgado, P. G. M., Guimarães, L. C., Cavalcante, F. S., Nouér, S. A., et al. (2019). Daptomycin and vancomycin non-susceptible methicillin-resistant *Staphylococcus aureus* clonal lineages from bloodstream infection in a Brazilian teaching hospital. *Braz. J. Infect. Dis.* 23, 139–142. doi: 10.1016/j.bjid.2019.03.003

D'Andrea, M. M., Antonelli, A., Brenciani, A., Di Pilato, V., Morroni, G., Pollini, S., et al. (2019). Characterization of Tn6349, a novel mosaic transposon carrying poxtA, cfr and other resistance determinants, inserted in the chromosome of an ST5-MRSA-II strain of clinical origin. *J. Antimicrob. Chemother.* 74, 2870–2875. doi: 10.1093/jac/dkz278

Dao, T. H., Alsallaq, R., Parsons, J. B., Ferrolino, J., Hayden, R. T., Rubnitz, J. E., et al. (2020). Vancomycin heteroresistance and clinical outcomes in bloodstream infections caused by coagulase-negative staphylococci. *Antimicrob. Agents Chemother.* 64:e00944-20. doi: 10.1128/AAC.00944-20

Davlieva, M., Zhang, W., Arias, C. A., and Shamoo, Y. (2013). Biochemical characterization of cardiolipin synthase mutations associated with daptomycin resistance in enterococci. *Antimicrob. Agents Chemother.* 57, 289–296. doi: 10.1128/AAC.01743-12

de Niederhäusern, S., Bondi, M., Messi, P., Iseppi, R., Sabia, C., Manicardi, G., et al. (2011). Vancomycin-resistance transferability from VanA enterococci to *Staphylococcus aureus*. *Curr. Microbiol.* 62, 1363–1367. doi: 10.1007/s00284-011-9868-6

Diaz, L., Tran, T. T., Munita, J. M., Miller, W. R., Rincon, S., Carvajal, L. P., et al. (2014). Whole-genome analyses of *Enterococcus faecium* isolates with diverse daptomycin MICs. *Antimicrob. Agents Chemother.* 58, 4527–4534. doi: 10.1128/AAC.02686-14

Egan, S. A., Corcoran, S., McDermott, H., Fitzpatrick, M., Hoyne, A., McCormack, O., et al. (2020a). A hospital outbreak of linezolid-resistant and vancomycin-resistant ST80 *Enterococcus faecium* harbouring an optrA-encoding conjugative plasmid investigated by whole-genome sequencing. *J. Hosp. Infect.* 105, 726–735. doi: 10.1016/j.jhin.2020.05.013

Egan, S. A., Shore, A. C., O'Connell, B., Brennan, G. I., and Coleman, D. C. (2020b). Linezolid resistance in Enterococcus faecium and Enterococcus faecalis from hospitalized patients in Ireland: high prevalence of the MDR genes optrA and poxtA in isolates with diverse genetic backgrounds. *J. Antimicrob. Chemother.* 75, 1704–1711. doi: 10.1093/jac/dkaa075

Eisenberger, D., Tuschak, C., Werner, M., Bogdan, C., Bollinger, T., Hossain, H., et al. (2020). Whole-genome analysis of vancomycin-resistant *Enterococcus faecium* causing nosocomial outbreaks suggests the occurrence of few endemic clonal lineages in Bavaria, Germany. *J. Antimicrob. Chemother.* 75, 1398–1404. doi: 10.1093/jac/dkaa041

Elghaieb, H., Freitas, A. R., Abbassi, M. S., Novais, C., Zouari, M., Hassen, A., et al. (2019). Dispersal of linezolid-resistant enterococci carrying poxtA or optrA in retail meat and food-producing animals from Tunisia. *J. Antimicrob. Chemother.* 74, 2865–2869. doi: 10.1093/jac/dkz263

Endimiani, A., Blackford, M., Dasenbrook, E. C., Reed, M. D., Bajaksouszian, S., Hujer, A. M., et al. (2011). Emergence of linezolid-resistant *Staphylococcus aureus* after prolonged treatment of cystic fibrosis patients in Cleveland, Ohio. *Antimicrob. Agents Chemother.* 55, 1684–1692. doi: 10.1128/AAC.01308-10

Enright, M. C., Robinson, D. A., Randle, G., Feil, E. J., Grundmann, H., and Spratt, B. G. (2002). The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). *Proc. Natl. Acad. Sci. U.S.A.* 99, 7687–7692. doi: 10.1073/pnas.122108599

Fioriti, S., Morroni, G., Coccitto, S. N., Brenciani, A., Antonelli, A., Di Pilato, V., et al. (2020). Detection of oxazolidinone resistance genes and characterization of genetic environments in Enterococci of swine origin, Italy. *Microorganisms* 8:2021. doi: 10.3390/microorganisms8122021

Foster, T. J. (2017). Antibiotic resistance in *Staphylococcus aureus*. current status and future prospects. *FEMS Microbiol. Rev.* 41, 430–499. doi: 10.1093/femsre/fux007

Freitas, A. R., Tedim, A. P., Novais, C., Lanza, V. F., and Peixe, L. (2020a). Comparative genomics of global optrA-carrying *Enterococcus faecalis* uncovers a common chromosomal hotspot for optrA acquisition within a diversity of core and accessory genomes. *Microb. Genom.* 6:e000350. doi: 10.1099/mgen.0.000350

Freitas, A. R., Tedium, A. P., Duarte, B., Elghaieb, H., Abbassi, M. S., Hassen, A., et al. (2020b). Linezolid-resistant (Tn6346::fexB-poxtA) *Enterococcus faecium* strains colonizing humans and bovines on different continents: similarity without epidemiological link. *J. Antimicrob. Chemother.* 75, 2416–2423. doi: 10.1093/jac/dkaa227

Gao, W., Cameron, D. R., Davies, J. K., Kostoulias, X., Stepnell, J., Tuck, K. L., et al. (2013). The RpoB H481Y rifampicin resistance mutation and an active stringent response reduce virulence and increase resistance to innate immune responses in *Staphylococcus aureus*. *J. Infect. Dis.* 207, 929–939. doi: 10.1093/infdis/jis772

Gilmore, M. S., Lebreton, F., and van Schaik, W. (2013). Genomic transition of Enterococci from gut commensals to leading causes of multidrug-resistant hospital infection in the antibiotic era. *Curr. Opin. Microbiol.* 16, 10–16. doi: 10.1016/j.mib.2013.01.006

Grein, F., Müller, A., Scherer, K. M., Liu, X., Ludwig, K. C., Klöckner, A., et al. (2020). Ca 2+-daptomycin targets cell wall biosynthesis by forming a tripartite complex with undecaprenyl-coupled intermediates and membrane lipids. *Nat. Commun.* 11, 1455. doi: 10.1038/s41467-020-15257-1

Gu, B., Kelesidis, T., Tsiodras, S., Hindler, J., and Humphries, R. M. (2013). The emerging problem of linezolid-resistant *Staphylococcus*. *J. Antimicrob. Chemother.* 68, 4–11. doi: 10.1093/jac/dks354

Guérillot, R., Gonçalves da Silva, A., Monk, I., Giulieri, S., Tomita, T., Alison, E., et al. (2018a). Convergent evolution driven by rifampin exacerbates the global burden of drug-resistant *Staphylococcus aureus*. *mSphere* 3, e00550-17. doi: 10.1128/mSphere.00550-17

Guérillot, R., Li, L., Baines, S., Howden, B., Schultz, M., Seemann, T., et al. (2018b). Comprehensive antibiotic-linked mutation assessment by resistance mutation sequencing (RM-seq). *Genome Med.* 10:63. doi: 10.1186/s13073-018-0572-z

Guerin, F., Sassi, M., Dejoies, L., Zouari, A., Schutz, S., Potrel, S., et al. (2020). Molecular and functional analysis of the novel cfr(D) linezolid resistance gene identified in *Enterococcus faecium*. *J. Antimicrob. Chemother.* 75, 1699–1703. doi: 10.1093/jac/dkaa125

Hansen, L. H., Planellas, M. H., Long, K. S., and Vester, B. (2012). The order *Bacillales* hosts functional homologs of the worrisome cfr antibiotic resistance gene. *Antimicrob. Agents Chemother.* 56, 3563–3567. doi: 10.1128/AAC.00673-12

Hao, W., Shan, X., Li, D., Schwarz, S., Zhang, S.-M., Li, X.-S., et al. (2019). Analysis of a *poxtA*- and *optrA*-co-carrying conjugative multiresistance plasmid from *Enterococcus faecalis*. *J. Antimicrob. Chemother.* 74, 1771–1775. doi: 10.1093/jac/dkz109

Hashemian, S. M. R., Farhadi, T., and Ganjparvar, M. (2018). Linezolid: a review of its properties, function, and use in critical care. *Drug Des. Dev. Ther.* 12, 1759–1767. doi: 10.2147/DDDT.S164515

Hasman, H., Clausen, P. T. L. C., Kaya, H., Hansen, F., Knudsen, J. D., Wang, M., et al. (2019). LRE-Finder, a web tool for detection of the 23S rRNA mutations and the optrA, cfr, cfr(B) and poxtA genes encoding linezolid resistance in Enterococci from whole-genome sequences. *J. Antimicrob. Chemother.* 74, 1473–1476. doi: 10.1093/jac/dkz092

Holmes, N. E., Turnidge, J. D., Munckhof, W. J., Robinson, J. O., Korman, T. M., O'Sullivan, M. V. N., et al. (2014). Genetic and molecular predictors of high vancomycin MIC in *Staphylococcus aureus* bacteremia isolates. *J. Clin. Microbiol.* 52, 3384–3393. doi: 10.1128/JCM.01320-14

Howden, B. P., Davies, J. K., Johnson, P. D. R., Stinear, T. P., and Grayson, M. L. (2010). Reduced vancomycin susceptibility in *Staphylococcus aureus*, including vancomycin-intermediate and heterogeneous vancomycin-intermediate strains: resistance mechanisms, laboratory detection, and clinical implications. *Clin. Microbiol. Rev.* 23, 99–139. doi: 10.1128/CMR.00042-09

Howden, B. P., McEvoy, C. R. E., Allen, D. L., Chua, K., Gao, W., Harrison, P. F., et al. (2011). Evolution of multidrug resistance during *Staphylococcus aureus* infection involves mutation of the essential two component regulator WalKR. *PLoS Path.* 7:e1002359. doi: 10.1371/journal.ppat.1002359

Howden, B. P., Peleg, A. Y., and Stinear, T. P. (2014). The evolution of vancomycin intermediate *Staphylococcus aureus* (VISA) and heterogenous-VISA. *Infect. Genet. Evol.* 21, 575–582. doi: 10.1016/j.meegid.2013.03.047

Howe, R. A., Wootton, M., Noel, A. R., Bowker, K. E., Walsh, T. R., and MacGowan, A. P. (2003). Activity of AZD2563, a novel oxazolidinone, against *Staphylococcus aureus* strains with reduced susceptibility to vancomycin or linezolid. *Antimicrob. Agents. Chemother.* 47, 3651–3652. doi: 10.1128/AAC.47.11.3651-3652.2003

Hua, R., Xia, Y., Wu, W., Yang, M., and Yan, J. (2019). Molecular epidemiology and mechanisms of 43 low-level linezolid-resistant *Enterococcus faecalis* strains in Chongqing. *China. Ann. Lab. Med.* 39, 36–42. doi: 10.3343/alm.2019.39.1.36

Huang, J., Chen, L., Wu, Z., and Wang, L. (2017). Retrospective analysis of genome sequences revealed the wide dissemination of optrA in Gram-positive bacteria. *J. Antimicrob. Chemother.* 72, 614–616. doi: 10.1093/jac/dkw488

Jiang, J.-H., Dexter, C., Cameron, D. R., Monk, I. R., Baines, S. L., Abbott, I. J., et al. (2019). Evolution of daptomycin resistance in coagulase-negative staphylococci involves mutations of the essential two-component regulator WalKR. *Antimicrob. Agents Chemother.* 63:e01926-18. doi: 10.1128/AAC.01926-18

Jones, P. M., O'Mara, M. L., and George, A. M. (2009). ABC transporters: a riddle wrapped in a mystery inside an enigma. *Trends Biochem. Sci.* 34, 520–531. doi: 10.1016/j.tibs.2009.06.004

Kelesidis, T., Humphries, R., Uslan, D. Z., and Pegues, D. A. (2011). Daptomycin nonsusceptible Enterococci: an emerging challenge for clinicians. *Clin. Infect. Dis.* 52, 228–234. doi: 10.1093/cid/ciq113

Kerschner, H., Cabal, A., Hartl, R., Machherndl-Spandl, S., Allerberger, F., Ruppitsch, W., et al. (2019). Hospital outbreak caused by linezolid resistant *Enterococcus faecium* in Upper Austria. *Antimicrob. Resist. Infect. Control* 8:150. doi: 10.1186/s13756-019-0598-z

Kos, V. N., Desjardins, C. A., Griggs, A., Cerqueira, G., Tonder, A. V., Holden, M. T. G., et al. (2012). Comparative genomics of vancomycin-resistant *Staphylococcus aureus* strains and their positions within the clade most commonly associated with methicillin-resistant *S. aureus* hospital-acquired infection in the United States. *mBio* 3:e00112-12. doi: 10.1128/mBio.00112-12

Kosecka-Strojek, M., Sadowy, E., Gawryszewska, I., Klepacka, J., Tomasik, T., Michalik, M., et al. (2020). Emergence of linezolid-resistant *Staphylococcus epidermidis* in the tertiary children's hospital in Cracow, Poland. *Eur. J. Clin. Microbiol. Infect. Dis.* 39, 1717–1725. doi: 10.1007/s10096-020-03893-w

Lazaris, A., Coleman, D. C., Kearns, A. M., Pichon, B., Kinnevey, P. M., Earls, M. R., et al. (2017). Novel multiresistance cfr plasmids in linezolid-resistant methicillin-resistant *Staphylococcus epidermidis* and vancomycin-resistant *Enterococcus faecium* (VRE) from a hospital outbreak: co-location of cfr and optrA in VRE. *J. Antimicrob. Chemother.* 72, 3252–3257. doi: 10.1093/jac/dkx292

Lee, A. S., de Lencastre, H., Garau, J., Kluytmans, J., Malhotra-Kumar, S., Peschel, A., et al. (2018). Methicillin-resistant *Staphylococcus aureus*. *Nat. Rev. Dis.* 4, 1–23. doi: 10.1038/nrdp.2018.33

Lee, J. Y. H., Monk, I. R., da Silva, A. G., Seemann, T., Chua, K. Y. L., Kearns, A., et al. (2018). Global spread of three multidrug-resistant lineages of *Staphylococcus epidermidis*. *Nat. Microbiol.* 3, 1175–1185. doi: 10.1038/s41564-018-0230-7

Lee, S. M., Huh, J. H., Song, D. J., Shim, H. J., Park, K. S., Kang, C. I., et al. (2017). Resistance mechanisms of linezolid-nonsusceptible Enterococci in Korea: low rate of 23S rRNA mutation in *Enterococcus faecium*. *J. Med. Microbiol.* 66, 1730–1735. doi: 10.1099/jmm.0.00063

Lei, C.-W., Kang, Z.-Z., Wu, S.-K., Chen, Y.-P., Kong, L.-H., and Wang, H.-N. (2019). Detection of the phenicol–oxazolidinone–tetracycline resistance gene poxtA in *Enterococcus faecium* and *Enterococcus faecalis* of food-producing animal origin in China. *J. Antimicrob. Chemother.* 74, 2459–2461. doi: 10.1093/jac/dkz198

Liu, Y., Wang, Y., Schwarz, S., Wang, S., Chen, L., Wu, C., et al. (2014). Investigation of a multiresistance gene cfr that fails to mediate resistance to phenicols and oxazolidinones in *Enterococcus faecalis*. *J. Antimicrob. Chemother.* 69, 892–898. doi: 10.1093/jac/dkt459

Livermore, D. M., Warner, M., Mushtaq, S., North, S., and Woodford, N. (2007). In vitro activity of the oxazolidinone RWJ-416457 against linezolid-resistant and -susceptible Staphylococci and Enterococci. *Antimicrob. Agents Chemother.* 51, 1112–1114. doi: 10.1128/AAC.01347-06

Locke, J. B., Hilgers, M., and Shaw, K. J. (2009a). Mutations in ribosomal protein L3 are associated with oxazolidinone resistance in staphylococci of clinical origin. *Antimicrob. Agents Chemother.* 53, 5275–5278. doi: 10.1128/AAC.01032-09

Locke, J. B., Hilgers, M., and Shaw, K. J. (2009b). Novel ribosomal mutations in *Staphylococcus aureus* strains identified through selection with the oxazolidinones linezolid and torezolid (TR-700). *Antimicrob. Agents Chemother.* 53, 5265–5274. doi: 10.1128/AAC.00871-09

Locke, J. B., Morales, G., Hilgers, M., C, K. G., Rahawi, S., Picazo, J. J., et al. (2010). Elevated linezolid resistance in clinical cfr-positive *Staphylococcus aureus* isolates is associated with co-occurring mutations in ribosomal protein L3. *Antimicrob. Agents Chemother.* 54, 5352–5355. doi: 10.1128/AAC.00714-10

Long, K. S., Poehlsgaard, J., Kehrenberg, C., Schwarz, S., and Vester, B. (2006). The Cfr rRNA methyltransferase confers resistance to phenicols, lincosamides, oxazolidinones, pleuromutilins, and streptogramin A antibiotics. *Antimicrob. Agents Chemother.* 50, 2500–2505. doi: 10.1128/AAC.00131-06

Long, K. S., and Vester, B. (2012). Resistance to linezolid caused by modifications at its binding site on the ribosome. *Antimicrob. Agents Chemother.* 56, 603–612. doi: 10.1128/AAC.05702-11

McGuinness, W. A., Malachowa, N., and DeLeo, F. R. (2017). Vancomycin resistance in *Staphylococcus aureus*. *Yale J. Biol. Med.* 90, 269–281.

Meka, V. G., Pillai, S. K., Sakoulas, G., Wennersten, C., Venkataraman, L., DeGirolami, P. C., et al. (2004). Linezolid resistance in sequential *Staphylococcus aureus* isolates associated with a T2500A mutation in the 23S rRNA gene and loss of a single copy of rRNA. *J. Infect. Dis.* 190, 311–317. doi: 10.1086/421471

Mendes, R. E., Deshpande, L. M., Costello, A. J., and Farrell, D. J. (2012). Molecular epidemiology of *Staphylococcus epidermidis* clinical isolates from U.S. Hospitals. *Antimicrob. Agents Chemother.* 56, 4656–4661. doi: 10.1128/AAC.00279-12

Mendes, R. E., Deshpande, L. M., and Jones, R. N. (2014). Linezolid update: stable in vitro activity following more than a decade of clinical use and summary of associated resistance mechanisms. *Drug Resist. Updat.* 17, 1–12. doi: 10.1016/j.drup.2014.04.002

Miller, W. R., Bayer, A. S., and Arias, C. A. (2016). Mechanism of action and resistance to daptomycin in *Staphylococcus aureus* and Enterococci. *Cold Spring Harb. Perspect. Med.* 6:a026997. doi: 10.1101/cshperspect.a026997

Mishra, N. N., McKinnell, J., Yeaman, M. R., Rubio, A., Nast, C. C., Chen, L., et al. (2011). In vitro cross-resistance to daptomycin and host defense cationic antimicrobial peptides in clinical methicillin-resistant *Staphylococcus aureus* isolates. *Antimicrob. Agents Chemother.* 55, 4012–4018. doi: 10.1128/AAC.00223-11

Morales, G., Picazo, J. J., Baos, E., Candel, F. J., Arribi, A., Peláez, B., et al. (2010). Resistance to linezolid is mediated by the cfr gene in the first report of an outbreak of linezolid-resistant *Staphylococcus aureus*. *Clin. Infect. Dis.* 50, 821–825. doi: 10.1086/650574

Morroni, G., Brenciani, A., Antonelli, A., Maria, D., Andrea, M., Di Pilato, V., et al. (2018). Characterization of a multiresistance plasmid carrying the optrA and cfr resistance genes from an *Enterococcus faecium* clinical isolate. *Front. Microbiol.* 9:2189. doi: 10.3389/fmicb.2018.02189

Moure, Z., Lara, N., Marín, M., Sola-Campoy, P. J., Bautista, V., Gómez-Bertomeu, F., et al. (2020). Interregional spread in Spain of linezolid-resistant *Enterococcus* spp. isolates carrying the optrA and poxtA genes. *Int. J. Antimicrob. Agents* 55, 105977. doi: 10.1016/j.ijantimicag.2020.105977

Na, S. H., Moon, D. C., Kim, M. H., Kang, H. Y., Kim, S. J., Choi, J. H., et al. (2020). Detection of the phenicol-oxazolidinone resistance gene poxtA in *Enterococcus faecium* and *Enterococcus faecalis* from food-producing animals during 2008-2018 in Korea. *Microorganisms* 8:1839. doi: 10.3390/microorganisms8111839

Oku, Y., Kurokawa, K., Ichihashi, N., and Sekimizu, K. (2004). Characterization of the *Staphylococcus aureus* mprF gene, involved in lysinylation of phosphatidylglycerol. *Microbiology* 150, 45–51. doi: 10.1099/mic.0.26706-0

Otto, M. (2009). *Staphylococcus epidermidis* — the "accidental" pathogen. *Nat. Rev. Microbiol.* 7, 555–567. doi: 10.1038/nrmicro2182

Palmer, K. L., Daniel, A., Hardy, C., Silverman, J., and Gilmore, M. S. (2011). Genetic basis for daptomycin resistance in Enterococci. *Antimicrob. Agents Chemother.* 55, 3345–3356. doi: 10.1128/AAC.00207-11

Papagiannitsis, C. C., Tsilipounidaki, K., Malli, E., and Petinaki, E. (2019). Detection in Greece of a clinical *Enterococcus faecium* isolate carrying the novel oxazolidinone resistance gene poxtA. *J. Antimicrob. Chemother.* 74, 2461–2462. doi: 10.1093/jac/dkz155

Prater, A. G., Mehta, H. H., Kosgei, A. J., Miller, W. R., Tran, T. T., Arias, C. A., et al. (2019). Environment shapes the accessible daptomycin resistance mechanisms in *Enterococcus faecium*. *Antimicrob. Agents Chemother.* 63:e00790-19. doi: 10.1128/AAC.00790-19

Rajagopal, M., Martin, M. J., Santiago, M., Lee, W., Kos, V. N., Meredith, T., et al. (2016). Multidrug intrinsic resistance factors in *Staphylococcus aureus* identified by profiling fitness within high-diversity transposon libraries. *mBio* 7:e00950-16. doi: 10.1128/mBio.00950-16

Ruiz-Ripa, L., Feßler, A. T., Hanke, D., Eichhorn, I., Azcona-Gutiérrez, J. M., Pérez-Moreno, M. O., et al. (2020). Mechanisms of linezolid resistance among Enterococci of clinical origin in Spain – detection of optrA- and cfr(D)-carrying *E. faecalis*. *Microorganisms* 8:1155. doi: 10.3390/microorganisms8081155

Sadowy, E. (2018). Linezolid resistance genes and genetic elements enhancing their dissemination in Enterococci and streptococci. *Plasmid* 99, 89–98. doi: 10.1016/j.plasmid.2018.09.011

Sassi, M., Guérin, F., Zouari, A., Beyrouthy, R., Auzou, M., Fines-Guyon, M., et al. (2019). Emergence of optrA-mediated linezolid resistance in Enterococci from France, 2006–16. *J. Antimicrob. Chemother.* 74, 1469–1472. doi: 10.1093/jac/dkz097

Seedat, J., Zick, G., Klare, I., Konstabel, C., Weiler, N., and Sahly, H. (2006). rapid emergence of resistance to linezolid during linezolid therapy of an *Enterococcus faecium* infection. *Antimicrob. Agents Chemother.* 50, 4217–4219. doi: 10.1128/AAC.00518-06

Shan, X., Li, X.-S., Wang, N., Schwarz, S., Zhang, S.-M., Li, D., et al. (2020). Studies on the role of IS1216E in the formation and dissemination of poxtA-carrying plasmids in an *Enterococcus faecium* clade A1 isolate. *J. Antimicrob. Chemother.* 75, 3126–3130. doi: 10.1093/jac/dkaa325

Shang, Y., Li, D., Shan, X., Schwarz, S., Zhang, S.-M., Chen, Y.-X., et al. (2019). Analysis of two pheromone-responsive conjugative multiresistance plasmids carrying the novel mobile optrA locus from *Enterococcus faecalis*. *Infect. Drug Resist.* 12, 2355–2362. doi: 10.2147/IDR.S206295

Shariati, A., Dadashi, M., Chegini, Z., van Belkum, A., Mirzaii, M., Khoramrooz, S. S., et al. (2020). The global prevalence of daptomycin, tigecycline, quinupristin/dalfopristin, and linezolid-resistant *Staphylococcus aureus* and coagulase–negative staphylococci strains: a systematic review and meta-analysis. *Antimicrob. Resist. Infect. Control.* 9:56. doi: 10.1186/s13756-020-00714-9

Sharkey, L. K. R., Edwards, T. A., and O'Neill, A. J. (2016). ABC-F proteins mediate antibiotic resistance through ribosomal protection. *mBio* 7:e01975. doi: 10.1128/mBio.01975-15

Sieradzki, K., Roberts, R. B., Haber, S. W., and Tomasz, A. (1999). The development of vancomycin resistance in a patient with methicillin-resistant *Staphylococcus aureus* infection. *N. Engl. J. Med.* 340, 517–523. doi: 10.1056/NEJM199902183400704

Skiest, D. J. (2006). treatment failure resulting from resistance of *Staphylococcus aureus* to daptomycin. *J. Clin. Microbiol.* 44, 655–656. doi: 10.1128/JCM.44.2.655-656.2006

Smith, T. L., Pearson, M. L., Wilcox, K. R., Cruz, C., Lancaster, M. V., Robinson-Dunn, B., et al. (1999). Emergence of vancomycin resistance in *Staphylococcus aureus*. *N. Engl. J. Med.* 340, 493–501. doi: 10.1056/NEJM199902183400701

Toh, S.-M., Xiong, L., Arias, C. A., Villegas, M. V., Lolans, K., Quinn, J., et al. (2007). Acquisition of a natural resistance gene renders a clinical strain of methicillin-resistant *Staphylococcus aureus* resistant to the synthetic antibiotic linezolid. *Mol. Microbiol.* 64, 1506–1514. doi: 10.1111/j.1365-2958.2007.05744.x

Tran, T. T., Munita, J. M., and Arias, C. A. (2015). Mechanisms of drug resistance: daptomycin resistance. *Ann. N. Y. Acad. Sci.* 1354, 32–53. doi: 10.1111/nyas.12948

Tran, T. T., Panesso, D., Gao, H., Roh, J. H., Munita, J. M., Reyes, J., et al. (2013). Whole-genome analysis of a daptomycin-susceptible *Enterococcus faecium* strain and its daptomycin-resistant variant arising during therapy. *Antimicrob. Agents Chemother.* 57, 261–268. doi: 10.1128/AAC.01454-12

Udaondo, Z., Jenjaroenpun, P., Wongsurawat, T., Meyers, E., Anderson, C., Lopez, J., et al. (2020). Two cases of vancomycin-resistant *Enterococcus faecium* bacteremia with development of daptomycin-resistant phenotype and its detection using oxford nanopore sequencing. *Open Forum Infect. Dis.* 7:ofaa180. doi: 10.1093/ofid/ofaa180

Wang, G., Kamalakaran, S., Dhand, A., Huang, W., Ojaimi, C., Zhuge, J., et al. (2014). Identification of a novel clone, ST736, among *Enterococcus faecium* clinical isolates and its association with daptomycin nonsusceptibility. *Antimicrob. Agents Chemother.* 58, 4848–4854. doi: 10.1128/AAC.02683-14

Wang, G., Yu, F., Lin, H., Murugesan, K., Huang, W., Hoss, A. G., et al. (2018). Evolution and mutations predisposing to daptomycin resistance in vancomycin-resistant *Enterococcus faecium* ST736 strains. *PLoS One* 13:e0209785. doi: 10.1371/journal.pone.0209785

Wang, Y., Lv, Y., Cai, J., Schwarz, S., Cui, L., Hu, Z., et al. (2015). A novel gene, optrA, that confers transferable resistance to oxazolidinones and phenicols and its presence in *Enterococcus faecalis* and *Enterococcus faecium* of human and animal origin. *J. Antimicrob. Chemother.* 70, 2182–2190. doi: 10.1093/jac/dkv116

Wardenburg, K. E., Potter, R. F., D'Souza, A. W., Hussain, T., Wallace, M. A., Andleeb, S., et al. (2019). Phenotypic and genotypic characterization of linezolid-resistant *Enterococcus faecium* from the USA and Pakistan. *J. Antimicrob. Chemother.* 74, 3445–3452. doi: 10.1093/jac/dkz367

Weiner, L. M., Webb, A. K., Limbago, B., Dudeck, M. A., Patel, J., Kallen, A. J., et al. (2016). Antimicrobial-resistant pathogens associated with healthcare-associated infections: summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2011–2014. *Infect. Control Hosp. Epidemiol.* 37, 1288–1301. doi: 10.1017/ice.2016.174

Weiner-Lastinger, L. M., Abner, S., Edwards, J. R., Kallen, A. J., Karlsson, M., Magill, S. S., et al. (2020). Antimicrobial-resistant pathogens associated with adult healthcare-associated infections: summary of data reported to the National Healthcare Safety Network, 2015–2017. *Infect. Control Hosp. Epidemiol.* 41, 1–18. doi: 10.1017/ice.2019.296

Wong, A., Reddy, S. P., Smyth, D. S., Aguero-Rosenfeld, M. E., Sakoulas, G., and Robinson, D. A. (2010). Polyphyletic emergence of linezolid-resistant staphylococci in the United States. *Antimicrob. Agents. Chemother.* 54, 742–748. doi: 10.1128/AAC.00621-09

Wu, Y., Fan, R., Wang, Y., Lei, L., Feßler, A. T., Wang, Z., et al. (2019). Analysis of combined resistance of oxazolidinones and phenicols among bacteria from dogs fed with raw meat/vegetables and the respective food items. *Sci. Rep.* 9:15500. doi: 10.1038/s41598-019-51918-y

Xiong, L., Kloss, P., Douthwaite, S., Andersen, N. M., Swaney, S., Shinabarger, D. L., et al. (2000). Oxazolidinone resistance mutations in 23S rRNA of *Escherichia coli* reveal the central region of domain V as the primary site of drug action. *J. Bacteriol.* 182, 5325–5331. doi: 10.1128/JB.182.19.5325-5331.2000

Yang, S.-J., Mishra, N. N., Rubio, A., and Bayer, A. S. (2013). Causal role of single nucleotide polymorphisms within the mprF gene of *Staphylococcus aureus* in daptomycin resistance. *Antimicrob. Agents Chemother.* 57, 5658–5664. doi: 10.1128/AAC.01184-13

Yang, S.-J., Xiong, Y. Q., Dunman, P. M., Schrenzel, J., François, P., Peschel, A., et al. (2009). Regulation of mprF in Daptomycin-nonsusceptible *Staphylococcus aureus* strains. *Antimicrob. Agents Chemother.* 53, 2636–2637. doi: 10.1128/AAC.01415-08

Zhu, Y., Zhang, W., Wang, C., Liu, W., Yang, Q., Luan, T., et al. (2020). Identification of a novel optrA-harbouring transposon, Tn6823, in *Staphylococcus aureus*. *J. Antimicrob. Chemother.* 75, 3395–3397. doi: 10.1093/jac/dkaa323

Zurenko, G. E., Yagi, B. H., Schaadt, R. D., Allison, J. W., Kilburn, J. O., Glickman, S. E., et al. (1996). In vitro activities of U-100592 and U-100766, novel oxazolidinone antibacterial agents. *Antimicrob. Agents Chemother.* 40, 839–845. doi: 10.1128/AAC.40.4.839

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

frontiers in Microbiology

ORIGINAL RESEARCH
published: 26 March 2021
doi: 10.3389/fmicb.2021.628740

Horizontal Transfer of Different *erm*(B)-Carrying Mobile Elements Among *Streptococcus suis* Strains With Different Serotypes

*Li Chen[1], Jinhu Huang[1], Xinxin Huang[2], Yuping He[2], Junjie Sun[1], Xingyang Dai[1], Xiaoming Wang[1], Muhammad Shafiq[1] and Liping Wang[1]**

[1] *Ministry of Education (MOE) Joint International Research Laboratory of Animal Health and Food Safety, College of Veterinary Medicine, Nanjing Agricultural University, Nanjing, China,* [2] *Technical Center for Animal, Plant and Food Inspection and Quarantine of Shanghai Customs, Shanghai, China*

Macrolide-resistant *Streptococcus suis* is highly prevalent worldwide. The acquisition of the *erm*(B) gene mediated by mobile genetic elements (MGEs) in particular integrative and conjugative elements (ICEs) is recognized as the main reason for the rapid spread of macrolide-resistant streptococcal strains. However, knowledge about different *erm*(B)-carrying elements responsible for the widespread of macrolide resistance and their transferability in *S. suis* remains poorly understood. In the present study, two *erm*(B)- and *tet*(O)-harboring putative ICEs, designated as ICE*Ssu*YSB17_*rplL* and ICE*Ssu*YSJ15_*rplL*, and a novel *erm*(B)- and *aadE-spw*-like-carrying genomic island (GI), named GI*Ssu*JHJ17_*rpsI*, were identified to be excised from the chromosome and transferred among *S. suis* strains with different serotypes. ICE*Ssu*YSB17_*rplL* and ICE*Ssu*YSJ15_*rplL* were integrated downstream the *rplL* gene, a conserve locus of the ICE*Sa*2603 family. GI*Ssu*JHJ17_*rpsI*, with no genes belonging to the conjugation module, was integrated into the site of *rpsI*. All transconjugants did not exhibit obvious fitness cost by growth curve and competition assays when compared with the recipient. The results demonstrate that different *erm*(B)-carrying elements were presented and highlight the role of these elements in the dissemination of macrolide resistance in *S. suis*.

Keywords: *erm*(B), ICEs, GIs, horizontal transfer, *S. suis*

OPEN ACCESS

Edited by:
Yang Wang,
China Agricultural University, China

Reviewed by:
Sophie Payot,
INRA Centre Nancy-Lorraine, France
Haruyoshi Tomita,
Gunma University, Japan

***Correspondence:**
Liping Wang
wlp71@163.com

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

Received: *12 November 2020*
Accepted: *03 February 2021*
Published: *26 March 2021*

Citation:
Chen L, Huang J, Huang X, He Y, Sun J, Dai X, Wang X, Shafiq M and Wang L (2021) Horizontal Transfer of Different erm(B)-Carrying Mobile Elements Among Streptococcus suis Strains With Different Serotypes. Front. Microbiol. 12:628740. doi: 10.3389/fmicb.2021.628740

INTRODUCTION

The rapid increase of macrolide resistance in *Streptococcus* has been reported worldwide from both pig and human isolates during the past two decades (Princivalli et al., 2009; Palmieri et al., 2011; Vela et al., 2017). Although numerous resistance genes have been reported since the early 1980's[1] (Roberts, 2008), macrolide resistance in streptococci is primarily due to the ribosomal alteration of the 23S rRNA target site by methylases encoded by the *erm* genes, predominantly *erm*(B), which mediate resistance to macrolides, lincosamides, and streptogramin B (MLS_B) antimicrobials,

[1] faculty.washington.edu/marilynr/

and active efflux by the *mef* and *msr* genes (Wilson, 2014; Fyfe et al., 2016). These resistance genes are frequently carried by mobile genetic elements (MGEs), such as plasmids, transposons, prophages, and more recently, integrative and conjugative elements (ICEs) (Horaud et al., 1985; Woodbury et al., 2008; Varaldo et al., 2009; Huang et al., 2016b,c; Feßler et al., 2018; Libante et al., 2019). ICEs primarily reside in the bacterial chromosome and can excise from the donor chromosome to form a circular molecule that can be horizontally self-transferred to a recipient cell by conjugation (Bellanger et al., 2014). Other chromosomal elements,including integrative and mobilizable elements (IMEs), which encode a recombinase and only some conjugation proteins, and some genomic islands (GIs), which encode a recombinase but do not encode any conjugation proteins, were recently found to be mobilized in *trans* by ICEs (Daccord et al., 2010) and might have played crucial roles in bacterial evolution.

The *erm*(B) gene was originally identified on a 5,266 bp transposon Tn*917* from *Enterococcus faecalis* (Tomich et al., 1979). In human streptococci strains, the *erm*(B)-containing Tn*917* was usually integrated into Tn*916* (designated as Tn*3872*), which also carries the tetracycline resistance gene *tet*(M) (Brenciani et al., 2007; Varaldo et al., 2009). Further, two other *erm*(B)-containing elements, *erm*(B) element and macrolide–aminoglycoside–streptothricin element, were frequently inserted into *tet*(M)-carrying Tn*916*-like structure (e.g., Tn*6002*/Tn*6003*, Tn*1545*, Tn*2009*/Tn*2010*) (Varaldo et al., 2009; Marosevic et al., 2017). This genetic linkage between *erm*(B) and *tet*(M) on different MGEs was considered to be the primary mechanism for the spread of streptococcal bacteria that are resistant to both macrolide and tetracycline antimicrobials (Brenciani et al., 2007; Cochetti et al., 2008; Xu et al., 2010). However, in the zoonotic pathogen *Streptococcus suis*, the linkage between *erm*(B) and *tet*(O) was more frequently detected in different countries (Martel et al., 2005; Gurung et al., 2015; Huang et al., 2015; Bojarska et al., 2016; Pan et al., 2019), suggesting that MGEs responsible for macrolide and tetracycline resistance might be different from other streptococci (Huang et al., 2016b,c). *S. suis* is a key antibiotic resistance gene reservoir and a major zoonotic pathogen responsible for severe economic loss to the swine industry. This bacterium causes specific diseases in humans after contact with infected animals or derived food products. It caused human infection outbreaks in China in 1998 and 2005, respectively, and sporadic cases of *S. suis* infections in humans have occurred occasionally worldwide (Hui et al., 2005; Mazokopakis et al., 2005; Yu et al., 2006; Mai et al., 2008; CDC, 2013; Huang et al., 2019). Recent studies have demonstrated that the *erm*(B) and *tet*(O) genes co-existed on different ICEs in *S. suis* isolates of both pig and human origins (Holden et al., 2009; Zhang et al., 2011; Huang et al., 2016a,c). Previous results from our laboratory and other investigators have confirmed the intra-species transfer of the *erm*(B)- and *tet*(O)-carrying ICEs by conjugation (Huang et al., 2016a,c; Zhou et al., 2017; Pan et al., 2019). However, knowledge about types of *erm*(B) elements responsible for widespread macrolide resistance remains rare. In the present study, we identified three *erm*(B)-carrying transferable elements, including two *erm*(B)- and *tet*(O)-harboring putative ICEs, belonging to the ICE*Sa*2603 family, and a novel *erm*(B)-carrying GI, which can be horizontally transferred among *S. suis* strains with different serotypes.

MATERIALS AND METHODS

Bacterial Strains and Culture Condition

In this study, a total of 320 *S. suis* isolates obtained from humans and pigs in China from 2005 to 2018 were included. All *S. suis* strains were routinely cultivated on Todd–Hewitt broth (THB) or Todd–Hewitt agar (THA) plates supplemented with 5% calf serum at 37°C.

Genomic DNA Extraction and PCR Amplification

The crude genomic DNA was prepared using boiling extraction. The bacterial cultures were centrifuged (6,000*g* for 5 min at room temperature), and the pellets were harvested and resuspended in TE buffer (10 mM Tris–HCl, 1 mM EDTA, pH = 8.0). The mixtures were boiled for 10 min and incubated with ice for 10 min, then the mixtures were centrifuged, and the supernatants were collected. The extracted DNA was used as the template for PCR. All *S. suis* isolates were subjected to screen for the resistance genes of *erm*(B) and *tet*(O) in PCR analysis. The ICE*Sa*2603 family conserved genes of Int_{tyr} and *virB4* were characterized by a PCR mapping assay. To investigate the presence of circular/integrate forms of ICE and GI, two specific primer pairs (P1–P4 for ICE*Ssu*YSB17_*rplL* and P5–P8 for GI*Ssu*JHJ17_*rpsI*) were designed and used in PCR experiments. All the PCR primers were listed in **Supplementary Table S1**. Amplification reactions were performed in a total volume of 25 μl containing 12.5 μl 2 × Taq Plus Master Mix II (Vazyme, China), 1 μl of each primer (10 μM), 1 μl genomic DNA, and 9.5 μl water. The PCR assay was carried out in a thermocycler, comprising 5 min of pre-incubation at 94°C, followed by 35 cycles of 30 s at 94°C, 30 s at 50–60°C (determined by primers), and 1 min at 72°C. The final extension was performed for 10 min at 72°C.

Antibiotic Susceptibility Testing

Antimicrobial susceptibility testing was performed for determining the minimum inhibitory concentrations (MICs) to the corresponding antimicrobial agents according to the CLSI M100-ED28 guideline (CLSI, 2018). *Staphylococcus aureus* ATCC 29213 was used for quality control.

Transfer and Retransfer Experiments

We randomly selected six non-serotype 2 *S. suis* strains carrying the *erm*(B), *tet*(O), *virB4*, and Int_{tyr} genes that were used as donors (rifampicin and fusidic acid susceptibility and erythromycin resistance) (**Supplementary Table S2**). *S. suis* P1/7RF (rifampicin and fusidic acid resistance and erythromycin susceptibility) described in a previous study (Huang et al.,

2016a) was utilized as recipients, which was considered to be not competent until the *com*RS system was activated. For a long time, *S. suis* was thought to be a bacterium unable to transformation. However, recently, the natural competence of *S. suis* under laboratory conditions was demonstrated with the addition of a *com*X-inducing peptide (Zaccaria et al., 2014). *S. suis* SH28CIP and NP4CIP (ciprofloxacin resistance and erythromycin susceptibility) were used as recipients in retransfer experiments. Transfer and retransfer experiments were performed by filter mating as described previously (Li et al., 2011; Huang et al., 2016b), with minor modifications. In brief, donor and recipient strains were grown separately at 37°C. The bacterial cultures were centrifuged to harvest at the end of the exponential growth phase and then mixed at a ratio of 1:10 (donor to recipient). The mixtures were placed on sterile nitrocellulose filters on THA plates and incubated at 37°C for 4 h. Bacteria were removed from the filters by washing with 2 ml THB medium. Transconjugants were selected by THA plates containing appropriate antibiotics (50 mg/l erythromycin with 100 mg/l rifampicin and 100 mg/l fusidic acid in transfer assays or 100 mg/l ciprofloxacin in retransfer assays) and further confirmed the presence of the *erm*(B), *tet*(O), and type IV secretion system (T4SS) core genes by PCR. To rule out spontaneous mutation and the contribution of transformation to the genetic exchange during transfer, filter mating experiments were carried out in the presence of 10 μg/ml DNase I in transfer and retransfer assays, with donor and recipient control plates included. The residual DNA with the treatment of DNase I was quantified by quantitative PCR (qPCR) using primers targeting the *virB4* gene in wash buffer. The conjugation experiments were done in triplicate. The transfer frequency was calculated based on the number of observed transconjugants divided by the donors' initial number.

PFGE and DNA Hybridization

To determine the location of the *erm*(B) or *tet*(O) genes, genomic DNA from each of the donor strains, the recipient strains, and the corresponding transconjugants was digested with *Sma*I endonuclease and subjected to pulsed-field gel electrophoresis (PFGE) as previously described (Vela et al., 2003; Huang et al., 2015), followed by southern blotting and DNA hybridization analysis using *erm*(B)- or *tet*(O) probes with specific primers (**Supplementary Table S1**).

Whole-Genome Sequencing and Analysis

Bacterial cells were centrifuged, and the pellets were harvested and resuspended in TE buffer (10 mM Tris–HCl, 1 mM EDTA, pH 8.0). Total genomic DNA was extracted using an Omega Bacteria DNA Kit (OMEGA, China) according to the manufacturer's instructions. Purified genomic DNAs were submitted for 150 bp paired-end whole-genome sequencing (WGS) on the Illumina Hiseq 2000 platform (Biozeron, Shanghai, China). ABySS v2.0.2 was used for genome assembly with multiple-Kmer parameters (Jackman et al., 2017). The genomes were annotated using the Rapid Annotation of microbial genomes using Subsystems Technology (RAST) annotation server[2] (Overbeek et al., 2014), and the genetic elements were predicted using the ICEfinder[3]. ICEs and GI were identified by comparison with other MGEs from GenBank and were visualized using Mauve and Easyfig 2.2.2 (Sullivan et al., 2011).

Growth Curve and Fitness Measurements

The fitness difference between transconjugants and the recipient strains was calculated by *in vitro* growth and competition assays as described previously (Gagneux et al., 2006; Kodio et al., 2019). In *in vitro* growth assay, a single colony of each strain was picked from the agar plate and incubated overnight at 37°C Cultures were adjusted into the same optical density (OD), diluted 1:100 in fresh THB medium, and aliquoted to 1 ml at an interval of every hour, and the OD600 of bacterial cultures was measured for 24 h.

In *in vitro* competition assay, cultures of each competitor were adjusted to OD_{600} = 0.1, mixed in a 1:1 ratio, and diluted to 1:100 in 10 ml at 37°C for 24 h. The mixtures at both startpoint (0 h) and endpoint (24 h) were plated on THA plates without or with 50 mg/l erythromycin and incubated at 37°C for 48 h. The relative competitive fitness W was calculated using the formula $W = \ln(R_f/R_i)/\ln(S_f/S_i)$. R_i and S_i indicate the number of transconjugant and recipient cells at 0 h, respectively, and R_f and S_f indicate the number of transconjugant and recipient cells at 24 h, respectively.

GenBank Accession Numbers

The complete nucleotide sequences of ICE*Ssu*YSB17_*rplL* and GI*Ssu*JHJ17_*rpsI* have been deposited in the GenBank database under accession numbers MN876247 and MN876248, respectively.

RESULTS

Co-transfer of *erm*(B) With Other Antimicrobial Resistance Genes

There is a strong association between *erm*(B) and *tet*(O) in *S. suis* isolated from China and worldwide (Huang et al., 2015, 2016c). In this study, 221 *S. suis* strains (86.33%, 221/320) were co-existed of *erm*(B) and *tet*(O). In order to test the co-transfer frequency of *erm*(B) with *tet*(O), we randomly selected six *S. suis* strains as donors for conjugative transfer, which were all co-harboring the *erm*(B), *tet*(O), and T4SS core genes (**Supplementary Table S2**). Transconjugants were observed from strains YSB17, YSJ15, and JHJ17 under erythromycin selection with or without DNase I treatment. The residual DNA with the treatment of DNase I in mating experiments was detected by qPCR using primers targeting the *virB4* gene in wash buffer but with a negative result. For each strain, about 30–50 transconjugant clones were

[2]http://rast.nmpdr.org

[3]https://db-mml.sjtu.edu.cn/ICEfinder/ICEfinder.html

picked and detected to be positive for the *erm*(B), *tet*(O), and T4SS core genes by PCR. Retransfer assays using *S. suis* SH28CIP and NP4CIP as recipients were performed, but no transconjugant was obtained.

Two strains, YSB17 and YSJ15, successfully transferred the erythromycin and tetracycline resistance to recipient *S. suis* P1/7RF, with a calculated transfer frequency of $(5.75 \pm 1.18) \times 10^{-8}$ and $(3.84 \pm 1.29) \times 10^{-8}$. The two transconjugants, designated as SScYSB17 and SScYSJ15, respectively, exhibited macrolide and tetracycline resistance phenotypes and were tested positive for *erm*(B) and *tet*(O) (**Table 1**). The transfer frequency of the third transconjugant SScJHJ17 was $(4.31 \pm 1.53) \times 10^{-8}$, and SScJHJ17 showed erythromycin, streptomycin, and spectinomycin resistance but tetracycline-sensitive phenotype. It acquired not only *erm*(B) and *aadE* from donor strain JHJ17, which is responsible for erythromycin and high-level streptomycin resistance, respectively, but also *spw-like*, which exhibited 96.58% identity to *spw* in *E. faecalis* strain E211 (MK784777) (Wendlandt et al., 2013), and might mediate resistance to spectinomycin in *S. suis* SScJHJ17. However, *tet*(O) carried by donor strain JHJ17 was not detected in this conjugant strain.

Following PFGE separation, southern blotting, and hybridization with the *erm*(B) or *tet*(O) probes, the sizes of the transferable DNA fragments were deduced by comparing the profiles of the donor strains, the recipient strains, and the transconjugants. *Sma*I-PFGE analysis of YSB17 and JHJ17 conjugation pairs showed differences in three bands between the recipients and transconjugants (**Supplementary Figure S1**). An ∼460 kb fragment existed in the recipient P1/7RF but could not be detected in transconjugant SScYSB17. Instead, two fragments, with the sizes of ∼390 and ∼140 kb, were present in SScYSB17. These results suggested the successfully transferred element with an estimated size of approximately 70 kb, most probably ICE that carried *erm*(B) and *tet*(O), into the recipient's genome. Subsequent DNA hybridization revealed that the genes *erm*(B) and *tet*(O) were located on the different fragments, indicating the presence of *Sma*I restriction sites within this element (**Supplementary Figure S1A**). Similarly, the maximal fragment of recipient P1/7RF was replaced with two smaller fragments of transconjugant SScJHJ17, and the *erm*(B) gene was located in one of the fragments that differed from the recipient P1/7RF (**Supplementary Figure S1B**).

Characterization of Two *erm*(B)- and *tet*(O)-Carrying ICEs and an *erm*(B)-Carrying GI

To better understand the genetic context of the *erm*(B)-carrying elements, we determined the whole genomes of the donors YSB17, YSJ15, and JHJ17 and their respective transconjugants by WGS. In both YSB17 and its transconjugant SScYSB17, a single putative ICE carrying *erm*(B) and *tet*(O) in the chromosome was identified using the ICEfinder and designated as ICE*Ssu*YSB17_*rplL*. The *hyd* and *rplL* (**Figure 1A**, black color) are located at the terminals of ICE*Ssu*YSB17_*rplL*, which encoded a predicted hydrolase and 50S ribosomal protein L7/L12, respectively. ICE*Ssu*YSB17_*rplL* is 69,442 bp in length, with an average G + C content of 38%, and consists of 68 putative open reading frames (ORFs). A *Sma*I site existed at 4,843 bp downstream of the *erm*(B) gene and 9,665 bp upstream of the *tet*(O) gene, which is consistent with the *Sma*I-PFGE and hybridization results (**Figure 1A**). A 15-bp conserved sequence (5'-TTATTTAAGAGTAAC-3') was presented at both the left (L) and right (R) ends of the integrated ICE*Ssu*YSB17_*rplL* element. ICE*Ssu*YSB17_*rplL* was integrated into the 3'-end of the *rplL* gene and contained all 30 conserved core genes compared with ICE*Sa*2603. In addition, three intergenic hotspots (HS-1, HS-2, and HS-3) and three additional insertion sites were presented (**Figure 1A**). Among the three insertions, one was inserted in a previously identified site I-1, and one reverse transcriptase gene and one integrase gene were integrated within the SNF2 protein gene sequence, whereas *erm*(B) and *tet*(O) are located in HS-2 and I-1 variable regions, respectively. To trace the derivation of the resistance genes, comparative genome analyses were performed for HS-2 and I-1. The 5,869 bp HS-2 segment shared higher similarity with the corresponding sequences of the *S. suis* 9401240 (LR738724), ICE*Ssu*32457 (FR823304), ICE*Ssu*YS108 (MK211815), and *S. suis* (MN437484) (**Supplementary Figure S2A**). The content of 10,906-bp I-1

TABLE 1 | Characteristics of strains included in the filter mating conjugation experiments performed in this study.

Strains	Conjugation frequency[a]	MIC (mg/l)					
		RIF	FUS	ERY	TET	STR	SPC
P1/7RF		**256**	**256**	0.125	0.25	1,024	32
YSB17		≤0.0625	32	**>256**	**64**	>2,048	8
SScYSB17	$(5.75 \pm 1.18) \times 10^{-8}$	**256**	**256**	**>256**	**64**	1,024	32
YSJ15		≤0.0625	32	**>256**	**256**	>2,048	8
SScYSJ15	$(3.84 \pm 1.29) \times 10^{-8}$	**256**	**256**	**>256**	**128**	1,024	32
JHJ17		≤0.0625	32	**>256**	32	**>2,048**	**256**
SScJHJ17	$(4.31 \pm 1.53) \times 10^{-8}$	**256**	**256**	**>256**	0.25	**>2,048**	**256**

RIF, rifampin; FUS, fusidic acid; ERY, erythromycin; TET, tetracycline; STR, streptomycin; SPC, spectinomycin. [a]*The frequency is calculated by CFUs of transconjugants/donors. Resistance-related phenotypes related to transfer in conjugation assays were shown in bold.*

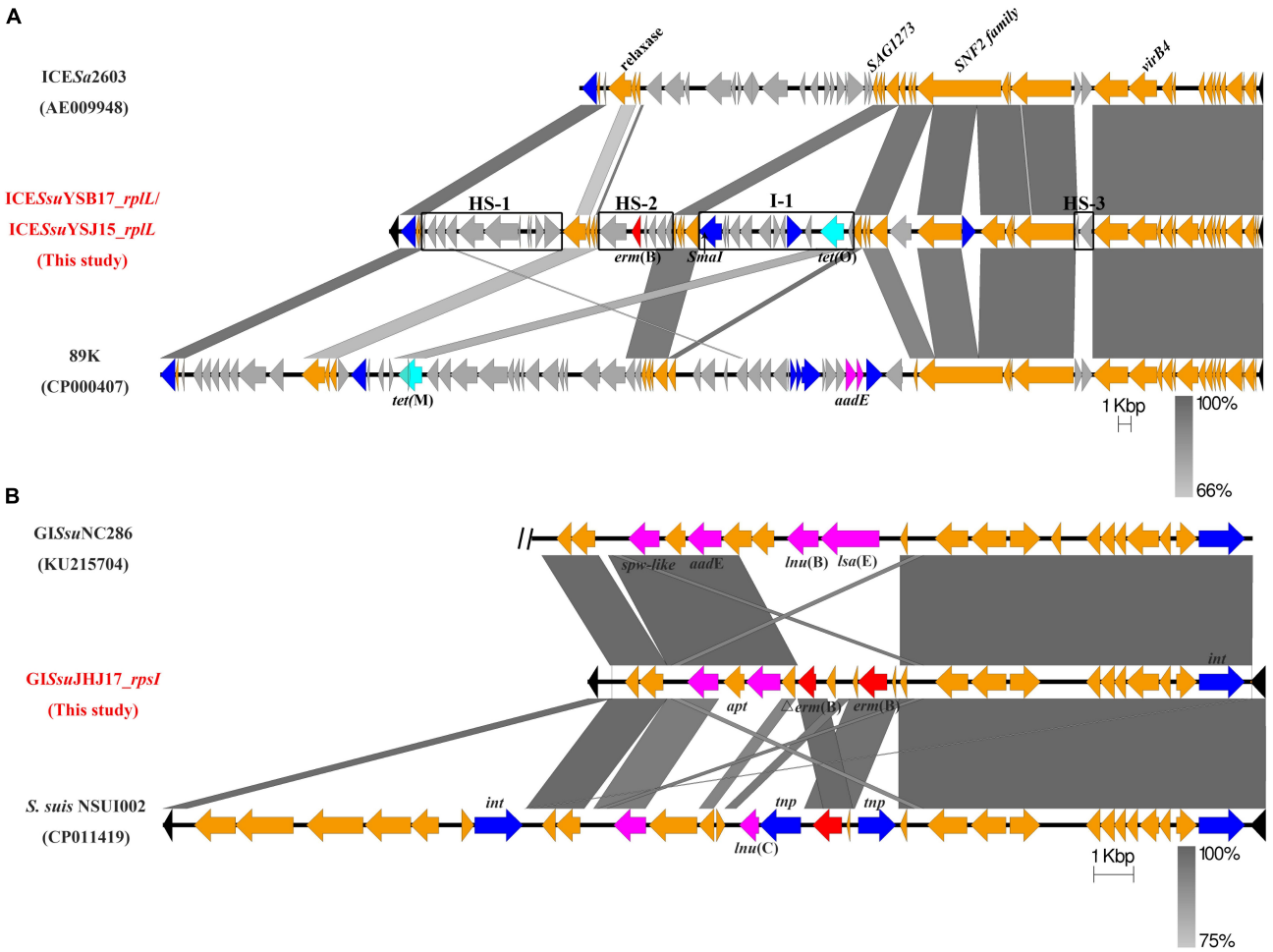


FIGURE 1 | Genetic features of the mobile genetic elements ICE*Ssu*YSB17_*rplL* and GI*Ssu*JHJ17_*rpsI*. The direction of the arrow indicates the direction of transcription. Homologous regions are shaded in gray. Genes are shown in different colors: the ICE and GI flanking chromosomal genes were shown in black, the 30 conserved core genes of the ICE*Sa*2603 family backbone and GI*Ssu*JHJ17_*rpsI* backbone are in orange, and variable genes are in light gray. Integrase/transposase/recombinase genes were highlighted in blue, *erm*(B) is marked in red, tet-resistant genes are in pale blue, and other resistant genes are labeled in pink. **(A)** Comparison of ICE*Ssu*YSB17_*rplL*/ICE*Ssu*YSJ15_*rplL* with ICE*Sa*2603 and 89K. Intergenic hotspots HS-1, HS-2, and HS-3 and insertion site I-2 were indicated. *Sma*I restriction site is marked by a black arrow. **(B)** GI*Ssu*JHJ17_*rpsI* from *S. suis* JHJ17 and linear DNA comparison against part of ICE*Ssu*NC286 and fragments of *S. suis* NSUI002. The two vertical black diagonal lines on the left of NC286 indicate that there is still a part of the sequence that is not shown.

segment showed identical nucleotide sequence with the *S. suis* NSUI060 (CP012911), *Blautia hansenii* DSM 20583 (CP022413), *Enterocloster clostridioformis* FDAARGOS_739 (CP050964), *Enterococcus cecorum* NCTC12421 (LS483306), *Streptococcus pyogenes* NCTC12057 (LS483331), and *Eubacterium hallii* EH1 (LT907978). The only difference is that the I-1 region contained an additional ORF (1,503 bp) encoding IS4 family transposase (**Supplementary Figure S2B**). Comparison of ICE*Ssu*YSB17_*rplL* with some other *erm*(B)- and *tet*(O)-carrying ICEs revealed highly conserved core genes but differed greatly in non-conserved regions (**Supplementary Figure S3**). In YSJ15 and its transconjugant SScYSJ15, we also detected a putative ICE, designated as ICE*Ssu*YSJ15_*rplL*, neatly identical to ICE*Ssu*YSB17_*rplL* with only five nucleotide differences. In JHJ17 but not the transconjugant SScJHJ17, a *tet*(O)-carrying putative ICE with all conserved modules was integrated into the 3'-end of the *rplL* gene (data not shown).

In both JHJ17 and its transconjugant SScJHJ17, a 16,195-bp sequence with 34% GC content was considered a putative GI and designated as GI*Ssu*JHJ17_*rpsI*. GI*Ssu*JHJ17_*rpsI* carried the *erm*(B) gene and was found to be integrated into a locus *rpsI*, the 3'-end of the gene encoding the ribosomal protein S9. Apart from a gene coding an integrase, no other putative conjugative elements, such as coupling proteins or elements participating in T4SS, were observed in GI*Ssu*JHJ17_*rpsI*. GI*Ssu*JHJ17_*rpsI* encodes 22 putative ORFs, 19 of them with the same direction of transcription as that of *erm*(B). An 8-bp conserved direct repeat sequence (5'-CCTGGTTT-3') was detected at both flanking of the GI*Ssu*JHJ17. BLAST analysis of GI*Ssu*JHJ17_*rpsI* showed that it had the highest similarity to GI*Ssu*NC286 (KU215704) and the genomic sequence of *S. suis* NSUI002 (CP011419) (**Figure 1B**).

In addition to the *erm*(B) gene (two copies), GI*Ssu*JHJ17_*rpsI* also contained the high-level streptomycin resistance gene *aadE* and the spectinomycin resistance gene *spw-like*. These genes are in agreement with the resistance profile of JHJ17 (**Table 1**).

Detection of the Extrachromosomal Circular Intermediate Forms of ICE*Ssu*YSB17_*rplL* and GI*Ssu*JHJ17_*rpsI*

ICEs and GIs can be excised from the chromosome with the aid of the integrase to generate the extrachromosomal circular form, which is the first step of its transfer lifecycle. In this study, two specific primer pairs (P1–P4 for ICE*Ssu*YSB17_*rplL* and P5–P8 for GI*Ssu*JHJ17_*rpsI*, the location of the primers were shown in **Supplementary Figures S4A,B**, respectively), were designed to detect the integrated and the extrachromosomal circular forms of ICE*Ssu*YSB17_*rplL* and GI*Ssu*JHJ17_*rpsI* (**Supplementary Table S1**). More specifically, P1/P2 and P3/P4 amplify the integrated form of ICE*Ssu*YSB17_*rplL* left and right terminals, respectively. P2/P3 detects whether there is a circular form of ICE*Ssu*YSB17_*rplL*. After ICE*Ssu*YSB17_*rplL* excision, P1/P4 detects an empty *att* site. For GI*Ssu*JHJ17_*rpsI* identification, the pairs used for P5–P8 are analogous to P1–P4. The results confirmed the presence of both the integrated and the extrachromosomal circular forms of ICE*Ssu*YSB17_*rplL* and GI*Ssu*JHJ17_*rpsI* in the original donors and the transconjugants, but absent in the recipient strain *S. suis* P1/7RF (**Supplementary Figures S4C,D**). The relatively low probability of occurrence of an excised form of ICE*Ssu*YSB17_*rplL* and GI*Ssu*JHJ17_*rpsI*, as reflected by the shallow bands of P2/P3 and P6/P7 PCR amplification, might be one of the causes of low frequency for transfer of these genetic elements. Analysis of the *att*ICE/*att*B and *att*L/*att*R amplicon sequences identified the 15-bp identical sequence (5'-TTATTTAAGAGTAAC-3'). Both the circular and excised forms of GI*Ssu*JHJ17_*rpsI* contained a copy of the 8-bp conserved sequence corresponding to the direct repeat sequence (5'-CCTGGTTT-3') site (data not shown).

Fitness of SScYSB17 and SScJHJ17

The biological cost of the horizontal acquisition of ICE*Ssu*YSB17_*rplL* or GI*Ssu*JHJ17_*rpsI* was investigated by *in vitro* growth and competition assays. During the *in vitro* growth assays, no significant differences were observed between the recipient *S. suis* P1/7RF and the two transconjugants SScYSB17 and SScJHJ17, suggesting that the acquisition of ICE*Ssu*YSB17_*rplL* (**Figure 2A**) and GI*Ssu*JHJ17_*rpsI* (**Figure 2B**) did not affect bacterial growth in THB medium, although both *S. suis* P1/7RF and the transconjugants showed growth delay compared with the donors and the original *S. suis* P1/7 strain.

In vitro competition assays showed that the transconjugants SScYSB17 and SScJHJ17 had relative fitness values W of 0.977 ± 0.085 and 0.982 ± 0.108, respectively, when compared with the recipient strain P1/7RF (**Figure 2C**). These results further suggest that there was no visible fitness cost when recipient strain acquired ICE*Ssu*YSB17_*rplL* or GI*Ssu*JHJ17_*rpsI*.

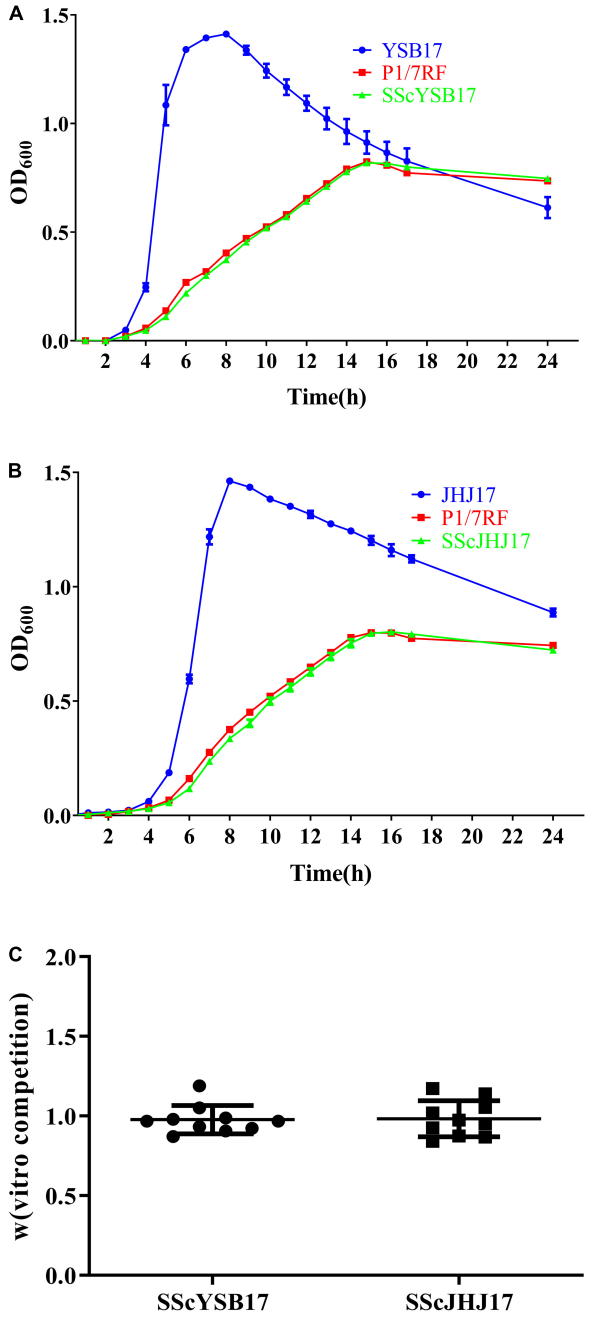


FIGURE 2 | The biological cost of the horizontal acquisition of ICE*Ssu*YSB17_*rplL* or GI*Ssu*JHJ17_*rpsI*. **(A)** Growth curves of the YSB17, P1/7RF, and SScYSB17 under the same conditions *in vitro*. **(B)** Growth curves of the JHJ17, P1/7RF, and SScJHJ17 under the same conditions *in vitro*. **(C)** The relative competitive fitness W of recipient P1/7RF and two transconjugants SScYSB17 and SScJHJ17 strain. The values are represented as mean ± *SD* of 10 independent experiments.

DISCUSSION

Macrolide-resistant *Streptococcus pneumoniae*, *S. pyogenes*, and *Streptococcus agalactiae* are 3 of the top 18 drug-resistant

threats as declared by the Centers for Disease Control and Prevention (CDC) in the United States in 2013 (CDC, 2013). Previous studies have suggested that *S. suis* is a reservoir of antimicrobial resistance (AMR) genes for other streptococcal pathogens (Palmieri et al., 2011; Huang et al., 2016a). The *erm*(B) gene is the most prevalent determinant conferring resistance to macrolide in streptococci clinical isolates (Chu et al., 2009; Haenni et al., 2018; Ichikawa et al., 2020). However, knowledge about the transfer of *erm*(B) as well as the related MGEs in *S. suis* remains unclear. In this study, we reported the co-transfer of *erm*(B) with other AMR genes among *S. suis* strains mediated by ICEs or GI, which could reveal the reason for the fast spread of macrolide-resistant *S. suis* in recent years in China.

Co-transfer of *erm*(B) and *tet*(O) was confirmed in two strains of *S. suis* serotype 21, which is co-located on ICEs of the ICE*Sa*2603 family. ICE*Sa*2603 family is highly prevalent in major *Streptococcus* species (Davies et al., 2009; Ambroset et al., 2015; Huang et al., 2016b). A variety of resistance genes responsible for resistance to tetracyclines, macrolides, or phenicols have been shown to be transferred inter-strains or inter-species by this family of ICEs (Chen et al., 2007; Palmieri et al., 2012; Marini et al., 2015; Huang et al., 2016b,c; Libante et al., 2019; Pan et al., 2019). Since *erm*(B) and *tet*(O) are located on two different variable regions, namely, HS-2 and I-1, and these two segments showed nearly identical sequence similarity to the corresponding sequences in *S. suis* and other Gram-positive cocci (**Supplementary Figure S2**), it is reasonable to speculate that ICE*Ssu*YSB17_*rplL* was evolved from acquisition of *erm*(B)-carrying HS-2 and *tet*(O)-carrying I-1 elements through a multi-step process. These results revealed the important role of the acquisition of AMR genes in ICEs diversity and evolution.

Co-transfer of *erm*(B) and *aadE-spw*-like elements was mediated by a novel GI, GI*Ssu*JHJ17_*rpsI*, which is integrated at the *rpsI* site, a conserved hotspot in *Streptococcus* species that was commonly integrated by IMEs and ICEs (Ambroset et al., 2015; Coluzzi et al., 2017; Libante et al., 2019). GIs are usually detected integrated into the 3'-end of the *tRNA* gene. However, two GIs were found in the *rpsI* gene, one carrying the *ant*(9)–*lnu*(C)–*erm*(B) genes (Libante et al., 2019) and another carrying the *aadE-lnu*(B)–*lsa*(E)–*spw*-like genes (Huang et al., 2016c). Moreover, GIs integrated into *rpsI* could be mobilized by subverting the relaxase and mating apparatus of a co-resident ICE (Libante et al., 2019). In this study, we also confirmed that GI*Ssu*JHJ17_*rpsI* was able to transfer from a *S. suis* serotype 29 isolate to serotype 2. We speculated that the transfer of GI*Ssu*JHJ17_*rpsI* was mobilized by a *tet*(O)-carrying ICE that harbored a fully functional mobilization module. It needs to be further proven by the inactivation of the *tet*(O)-carrying ICE. Studies have shown that some GIs not only need conjugative elements to promote their own transfer but also influence the transfer or stability of the helper co-resident elements (Guedon et al., 2017).

ICEs could integrate into the chromosome of bacteria and are capable to transfer to a new host uponconjugative transfer (Johnson and Grossman, 2015; Santoro et al., 2018). Functional ICEs were shown to excise from chromosome by site-specific recombination between *att*L and *att*R recombination sites, thus producing a covalently closed circular form of the ICE and a chromosomal excised *att*B site (Puymege et al., 2013). Under normal growth conditions, ICE*Ssu*YSB17_*rplL* and GI*Ssu*JHJ17_*rpsI* are mainly integrated into the chromosome. To check this integration state, we used primer pairs P1 + P2 and P3 + P4 to detect ICE*Ssu*YSB17_*rplL* in the chromosome and primer pairs P5 + P6 and P7 + P8 for GI*Ssu*JHJ17_*rpsI*. However, both MGEs could be excised from the bacterial genome and generated the extrachromosomal circular forms of ICE*Ssu*YSB17_*rplL* and GI*Ssu*JHJ17_*rpsI*, which identified the product by primers of P2 + P3 and P6 + P7, respectively. Furthermore, the empty *rplL attB* or *rpsI attB*' sites were detected by primers P1 + P4 and P5 + P8, respectively (**Supplementary Table S4**). This suggests that ICE*Ssu*YSB17_*rplL* and GI*Ssu*JHJ17_*rpsI* are functional and thus have the potential to be transferred. Previous studies have shown that excision of ICEs could be induced under environmental stress, including antimicrobials, such as ciprofloxacin and tetracycline (Beaber et al., 2004; Liu et al., 2017; Scornec et al., 2017). Considering the extensive use of antimicrobials in livestock and poultry, it is of great significance to evaluate the selection stress, especially antimicrobials, which are involved in inducing the excision and thereafter the transfer of the ICEs/GIs.

The acquisition of MGEs was thought to impose an immediate biological cost (Leon-Sampedro et al., 2016). However, the acquisition of ICE*Ssu*YSB17_*rplL* or GI*Ssu*JHJ17_*rpsI* in this study showed negligible fitness cost (**Figure 2**), which is consistent with our previous study (Huang et al., 2016b). In addition, the AMR-carrying ICEs or GIs enhance their survival under the corresponding antimicrobials. Those may explain the observation that the AMR-carrying ICE*Ssu*YSB17_*rplL*, GI*Ssu*JHJ17_*rpsI*, and similar ICEs are widely distributed in streptococci (Ambroset et al., 2015; Libante et al., 2019).

In summary, we identified three *erm*(B)-carrying transferable elements, including two *erm*(B)- and *tet*(O)-harboring ICEs of the ICE*Sa*2603 family and a novel *erm*(B)-carrying GI, which can be transferred between *S. suis* of different serotypes. The intraspecific transfer of *erm*(B)-carrying MGEs among different serotypes of *S. suis* strains might have contributed to the worldwide spread of macrolide resistance. This reinforces the need for strategies that inhibit the horizontal gene transfer of AMR-carrying MGEs.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository and accession numbers can be found below: (Repository: Genbank) (Accessions: BankIt2297772 Seq1 MN876247; BankIt2297772 Seq2 MN876248).

AUTHOR CONTRIBUTIONS

LC, JH, and LW developed the concept and designed the experiments. LC, JS, XD, and XW performed the experiments and

collected the data. LC, XH, and YH conducted all bioinformatics analyses. LC, JH, MS, and LW prepared the manuscript. All authors have contributed to, seen, and approved the manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (Nos. 31872517, 31702292, 31572567), the Fundamental Research Funds for the Central Universities (KJQN201827), the Natural Science Foundation of Shanghai City (19ZR1417400), the National Key R&D Program of China (2018YFD0500300), the Innovation Project for Postgraduate Training in Jiangsu Province (KYCX18_0714), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.628740/full#supplementary-material

REFERENCES

Ambroset, C., Coluzzi, C., Guedon, G., Devignes, M. D., Loux, V., Lacroix, T., et al. (2015). New insights into the classification and integration specificity of *Streptococcus* integrative conjugative elements through extensive genome exploration. *Front. Microbiol.* 6:1483. doi: 10.3389/fmicb.2015.01483

Beaber, J. W., Hochhut, B., and Waldor, M. K. (2004). SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* 427, 72–74. doi: 10.1038/nature02241

Bellanger, X., Payot, S., Leblond-Bourget, N., and Guedon, G. (2014). Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol. Rev.* 38, 720–760. doi: 10.1111/1574-6976.12058

Bojarska, A., Molska, E., Janas, K., Skoczynska, A., Stefaniuk, E., Hryniewicz, W., et al. (2016). *Streptococcus suis* in invasive human infections in Poland: clonality and determinants of virulence and antimicrobial resistance. *Eur. J. Clin. Microbiol. Infect. Dis.* 35, 917–925. doi: 10.1007/s10096-016-2616-x

Brenciani, A., Bacciaglia, A., Vecchi, M., Vitali, L. A., Varaldo, P. E., and Giovanetti, E. (2007). Genetic elements carrying *erm*(B) in *Streptococcus pyogenes* and association with *tet*(M) tetracycline resistance gene. *Antimicrob. Agents Chemother.* 51, 1209–1216. doi: 10.1128/AAC.01484-06

CDC (2013). *Antibiotic resistance threats in the United States, 2013 [Online]*. Available: http://www.cdc.gov/drugresistance/threat-report-2013 (accessed January 12, 2016).

Chen, C., Tang, J., Dong, W., Wang, C., Feng, Y., Wang, J., et al. (2007). A glimpse of streptococcal toxic shock syndrome from comparative genomics of *S. suis* 2 Chinese isolates. *PLoS One* 2:e315. doi: 10.1371/journal.pone.0000315

Chu, Y. W., Cheung, T. K., Chu, M. Y., Tsang, V. Y., Fung, J. T., Kam, K. M., et al. (2009). Resistance to tetracycline, erythromycin and clindamycin in *Streptococcus suis* serotype 2 in Hong Kong. *Int. J. Antimicrob. Agents* 34, 181–182. doi: 10.1016/j.ijantimicag.2009.01.007

CLSI (2018). *Performance Standards for Antimicrobial Susceptibility Testing. 28th ed. CLSI supplement M100*. Wayne, PA: Clinical and Laboratory Standards Institute.

Cochetti, I., Tili, E., Mingoia, M., Varaldo, P. E., and Montanari, M. P. (2008). erm(B)-carrying elements in tetracycline-resistant pneumococci and correspondence between Tn*1545* and Tn6003. *Antimicrob. Agents Chemother.* 52, 1285–1290. doi: 10.1128/AAC.01457-07

Coluzzi, C., Guedon, G., Devignes, M. D., Ambroset, C., Loux, V., Lacroix, T., et al. (2017). A glimpse into the world of integrative and mobilizable elements in streptococci reveals an unexpected diversity and novel families of mobilization proteins. *Front. Microbiol.* 8:443. doi: 10.3389/fmicb.2017.00443

Daccord, A., Ceccarelli, D., and Burrus, V. (2010). Integrating conjugative elements of the SXT/R391 family trigger the excision and drive the mobilization of a new class of *Vibrio* genomic islands. *Mol. Microbiol.* 78, 576–588. doi: 10.1111/j.1365-2958.2010.07364.x

Davies, M. R., Shera, J., Van Domselaar, G. H., Sriprakash, K. S., and McMillan, D. J. (2009). A novel integrative conjugative element mediates genetic transfer from group G *Streptococcus* to other β-hemolytic Streptococci. *J. Bacteriol.* 191, 2257–2265. doi: 10.1128/JB.01624-08

Feßler, A. T., Wang, Y., Wu, C., and Schwarz, S. (2018). Mobile macrolide resistance genes in staphylococci. *Plasmid* 99, 2–10. doi: 10.1016/j.plasmid.2018.05.001

Fyfe, C., Grossman, T. H., Kerstein, K., and Sutcliffe, J. (2016). Resistance to macrolide antibiotics in public health pathogens. *Cold Spring Harb. Perspect. Med.* 6:a25395. doi: 10.1101/cshperspect.a025395

Gagneux, S., Long, C. D., Small, P. M., Van, T., Schoolnik, G. K., and Bohannan, B. J. (2006). The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. *Science* 312, 1944–1946. doi: 10.1126/science.1124410

Guedon, G., Libante, V., Coluzzi, C., Payot, S., and Leblond-Bourget, N. (2017). The obscure world of integrative and mobilizable elements, highly widespread elements that pirate bacterial conjugative systems. *Genes* 8:337. doi: 10.3390/genes8110337

Gurung, M., Tamang, M. D., Moon, D. C., Kim, S. R., Jeong, J. H., Jang, G. C., et al. (2015). Molecular basis of resistance to selected antimicrobial agents in the emerging zoonotic pathogen *Streptococcus suis*. *J. Clin. Microbiol.* 53, 2332–2336. doi: 10.1128/JCM.00123-15

Haenni, M., Lupo, A., and Madec, J. Y. (2018). Antimicrobial resistance in *Streptococcus* spp. *Microbiol. Spectr.* 6, 1–25. doi: 10.1128/microbiolspec.ARBA-0008-2017

Holden, M. T., Hauser, H., Sanders, M., Ngo, T. H., Cherevach, I., Cronin, A., et al. (2009). Rapid evolution of virulence and drug resistance in the emerging zoonotic pathogen *Streptococcus suis*. *PLoS One* 4:e6072. doi: 10.1371/journal.pone.0006072

Horaud, T., Bouguenec, C. L., and Pepper, K. (1985). Molecular genetics of resistance to macrolides, lincosamides and streptogramin B (MLS) in streptococci. *J. Antimicrob. Chemother.* 16, 111–135. doi: 10.1093/jac/16.suppl_a.111

Huang, J., Liang, Y., Guo, D., Shang, K., Ge, L., Kashif, J., et al. (2016a). Comparative genomic analysis of the ICESa2603 family ICEs and spread of *erm*(B)- and *tet*(O)-carrying transferable 89K-subtype ICEs in swine and bovine isolates in China. *Front. Microbiol.* 7:55. doi: 10.3389/fmicb.2016.00055

Huang, J., Ma, J., Shang, K., Hu, X., Liang, Y., Li, D., et al. (2016b). Evolution and diversity of the antimicrobial resistance associated mobilome in *Streptococcus suis*: a probable mobile genetic elements reservoir for other streptococci. *Front. Cell Infect. Microbiol.* 6:118. doi: 10.3389/fcimb.2016.00118

Huang, J., Shang, K., Kashif, J., and Wang, L. (2015). Genetic diversity of *Streptococcus suis* isolated from three pig farms of China obtained by acquiring antibiotic resistance genes. *J. Sci. Food Agric.* 95, 1454–1460. doi: 10.1002/jsfa.6841

Huang, K., Zhang, Q., Song, Y., Zhang, Z., Zhang, A., Xiao, J., et al. (2016c). Characterization of spectinomycin resistance in *Streptococcus suis* leads to two novel insights into drug resistance formation and dissemination mechanism. *Antimicrob. Agents Chemother.* 60, 6390–6392. doi: 10.1128/AAC.01157-16

Huang, W., Wang, M., Hao, H., Yang, R., Xie, J., Su, J., et al. (2019). Genomic epidemiological investigation of a *Streptococcus suis* outbreak in Guangxi, China, 2016. *Infect. Genet. Evol.* 68, 249–252. doi: 10.1016/j.meegid.2018.12.023

Hui, A. C., Ng, K. C., Tong, P. Y., Mok, V., Chow, K. M., Wu, A., et al. (2005). Bacterial meningitis in Hong Kong: 10-years' experience. *Clin. Neurol. Neurosurg.* 107, 366–370. doi: 10.1016/j.clineuro.2004.10.006

Ichikawa, T., Oshima, M., Yamagishi, J., Muramatsu, C., and Asai, T. (2020). Changes in antimicrobial resistance phenotypes and genotypes in *Streptococcus suis* strains isolated from pigs in the Tokai area of Japan. *J. Vet. Med. Sci.* 82, 9–13. doi: 10.1292/jvms.19-0449

Jackman, S. D., Vandervalk, B. P., Mohamadi, H., Chu, J., Yeo, S., Hammond, S. A., et al. (2017). ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res.* 27, 768–777. doi: 10.1101/gr.214346.116

Johnson, C. M., and Grossman, A. D. (2015). Integrative and conjugative elements (ICEs): what they do and how they work. *Annu. Rev. Genet.* 49, 577–601. doi: 10.1146/annurev-genet-112414-055018

Kodio, O., Georges Togo, A. C., Sadio Sarro, Y. D., Fane, B., Diallo, F., Somboro, A., et al. (2019). Competitive fitness of *Mycobacterium tuberculosis in vitro*. *Int. J. Mycobacteriol.* 8, 287–291. doi: 10.4103/ijmy.ijmy_97_19

Leon-Sampedro, R., Novais, C., Peixe, L., Baquero, F., and Coque, T. M. (2016). Diversity and evolution of the Tn*5801-tet*(M)-like integrative and conjugative elements among *Enterococcus*, *Streptococcus*, and *Staphylococcus*. *Antimicrob. Agents Chemother.* 60, 1736–1746. doi: 10.1128/AAC.01864-15

Li, M., Shen, X., Yan, J., Han, H., Zheng, B., Liu, D., et al. (2011). GI-type T4SS-mediated horizontal transfer of the 89K pathogenicity island in epidemic *Streptococcus suis* serotype 2. *Mol. Microbiol.* 79, 1670–1683. doi: 10.1111/j.1365-2958.2011.07553.x

Libante, V., Nombre, Y., Coluzzi, C., Staub, J., Guedon, G., Gottschalk, M., et al. (2019). Chromosomal conjugative and mobilizable elements in *Streptococcus suis*: major actors in the spreading of antimicrobial resistance and bacteriocin synthesis genes. *Pathogens* 9, 1–23. doi: 10.3390/pathogens9010022

Liu, P., Wu, Z., Xue, H., and Zhao, X. (2017). Antibiotics trigger initiation of SCC*mec* transfer by inducing SOS responses. *Nucleic Acids Res.* 45, 3944–3952. doi: 10.1093/nar/gkx153

Mai, N. T., Hoa, N. T., Nga, T. V., Linh le, D., Chau, T. T., Sinh, D. X., et al. (2008). *Streptococcus suis* meningitis in adults in Vietnam. *Clin. Infect. Dis.* 46, 659–667. doi: 10.1086/527385

Marini, E., Palmieri, C., Magi, G., and Facinelli, B. (2015). Recombination between *Streptococcus suis* ICE*Ssu*32457 and *Streptococcus agalactiae* ICE*Sa*2603 yields a hybrid ICE transferable to *Streptococcus pyogenes*. *Vet. Microbiol.* 178, 99–104. doi: 10.1016/j.vetmic.2015.04.013

Marosevic, D., Kaevska, M., and Jaglic, Z. (2017). Resistance to the tetracyclines and macrolide-lincosamide-streptogramin group of antibiotics and its genetic linkage - a review. *Ann. Agric. Environ. Med. AAEM* 24, 338–344. doi: 10.26444/aaem/74718

Martel, A., Decostere, A., Leener, E. D., Marien, M., Graef, E. D., Heyndrickx, M., et al. (2005). Comparison and transferability of the *erm*(B) genes between human and farm animal streptococci. *Microb. Drug Resist.* 11, 295–302. doi: 10.1089/mdr.2005.11.295

Mazokopakis, E. E., Kofteridis, D. P., Papadakis, J. A., Gikas, A. H., and Samonis, G. J. (2005). First case report of *Streptococcus suis* septicaemia and meningitis from Greece. *Eur. J. Neurol.* 12, 487–489. doi: 10.1111/j.1468-1331.2005.00998.x

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

Palmieri, C., Magi, G., Mingoia, M., Bagnarelli, P., Ripa, S., Varaldo, P. E., et al. (2012). Characterization of a *Streptococcus suis tet*(O/W/32/O)-carrying element transferable to major streptococcal pathogens. *Antimicrob. Agents Chemother.* 56, 4697–4702. doi: 10.1128/AAC.00629-12

Palmieri, C., Varaldo, P. E., and Facinelli, B. (2011). *Streptococcus suis*, an emerging drug-resistant animal and human pathogen. *Front. Microbiol.* 2:235. doi: 10.3389/fmicb.2011.00235

Pan, Z., Liu, J., Zhang, Y., Chen, S., Ma, J., Dong, W., et al. (2019). A novel integrative conjugative element mediates transfer of multi-drug resistance between *Streptococcus suis* strains of different serotypes. *Vet. Microbiol.* 229, 110–116. doi: 10.1016/j.vetmic.2018.11.028

Princivalli, M. S., Palmieri, C., Magi, G., Vignaroli, C., Manzin, A., Camporese, A., et al. (2009). Genetic diversity of *Streptococcus Suis* clinical isolates from pigs and humans in Italy (2003-2007). *Euro Surveill.* 14:19310. doi: 10.2807/ese.14.33.19310-en

Puymege, A., Bertin, S., Chuzeville, S., Guedon, G., and Payot, S. (2013). Conjugative transfer and *cis*-mobilization of a genomic island by an integrative and conjugative element of *Streptococcus agalactiae*. *J. Bacteriol.* 195, 1142–1151. doi: 10.1128/JB.02199-12

Roberts, M. C. (2008). Update on macrolide-lincosamide-streptogramin, ketolide, and oxazolidinone resistance genes. *FEMS Microbiol. Lett.* 282, 147–159. doi: 10.1111/j.1574-6968.2008.01145.x

Santoro, F., Romeo, A., Pozzi, G., and Iannelli, F. (2018). Excision and circularization of integrative conjugative element Tn*5253* of *Streptococcus pneumoniae*. *Front. Microbiol.* 9:1779. doi: 10.3389/fmicb.2018.01779

Scornec, H., Bellanger, X., Guilloteau, H., Groshenry, G., and Merlin, C. (2017). Inducibility of Tn*916* conjugative transfer in *Enterococcus faecalis* by subinhibitory concentrations of ribosome-targeting antibiotics. *J. Antimicrob. Chemother.* 72, 2722–2728. doi: 10.1093/jac/dkx202

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

Tomich, P. K., An, F. Y., and Clewell, D. B. (1979). A transposon (Tn*917*) in *Streptococcus faecalis* that exhibits enhanced transposition during induction of drug resistance. *Cold Spring Harb. Symp. Quant. Biol.* 43(Pt 2), 1217–1221. doi: 10.1101/sqb.1979.043.01.138

Varaldo, P. E., Montanari, M. P., and Giovanetti, E. (2009). Genetic elements responsible for erythromycin resistance in streptococci. *Antimicrob. Agents Chemother.* 53, 343–353. doi: 10.1128/AAC.00781-08

Vela, A. I., Goyache, J., Tarradas, C., Luque, I., Mateos, A., Moreno, M. A., et al. (2003). Analysis of genetic diversity of *Streptococcus suis* clinical isolates from pigs in Spain by pulsed-field gel electrophoresis. *J. Clin. Microbiol.* 41, 2498–2502. doi: 10.1128/jcm.41.6.2498-2502.2003

Vela, A. I., Villalon, P., Saez-Nieto, J. A., Chacon, G., Dominguez, L., and Fernandez-Garayzabal, J. F. (2017). Characterization of *Streptococcus pyogenes* from animal clinical specimens, Spain. *Emerg. Infect. Dis.* 23, 2013–2016. doi: 10.3201/eid2312.151146

Wendlandt, S., Li, B., Lozano, C., Ma, Z., Torres, C., and Schwarz, S. (2013). Identification of the novel spectinomycin resistance gene *spw* in methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* of human and animal origin. *J. Antimicrob. Chemother.* 68, 1679–1680. doi: 10.1093/jac/dkt081

Wilson, D. N. (2014). Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat. Rev. Microbiol.* 12, 35–48. doi: 10.1038/nrmicro3155

Woodbury, R. L., Klammer, K. A., Xiong, Y., Bailiff, T., Glennen, A., Bartkus, J. M., et al. (2008). Plasmid-Borne *erm*(T) from invasive, macrolide-resistant *Streptococcus pyogenes* strains. *Antimicrob. Agents Chemother.* 52, 1140–1143. doi: 10.1128/AAC.01352-07

Xu, X., Cai, L., Xiao, M., Kong, F., Oftadeh, S., Zhou, F., et al. (2010). Distribution of serotypes, genotypes, and resistance determinants among macrolide-resistant *Streptococcus pneumoniae* isolates. *Antimicrob. Agents Chemother.* 54, 1152–1159. doi: 10.1128/AAC.01268-09

Yu, H., Jing, H., Chen, Z., Zheng, H., Zhu, X., Wang, H., et al. (2006). Human *Streptococcus suis* outbreak, Sichuan, China. *Emerg. Infect. Dis.* 12, 914–920. doi: 10.3201/eid1206.051194

Zaccaria, E., van Baarlen, P., de Greeff, A., Morrison, D. A., Smith, H., and Wells, J. M. (2014). Control of competence for DNA transformation in *Streptococcus suis* by genetically transferable pherotypes. *PLoS One* 9:e99394. doi: 10.1371/journal.pone.0099394

Zhang, A., Yang, M., Hu, P., Wu, J., Chen, B., Hua, Y., et al. (2011). Comparative genomic analysis of *Streptococcus suis* reveals significant genomic diversity among different serotypes. *BMC Genomics* 12:523. doi: 10.1186/1471-2164-12-523

Zhou, K., Xie, L., Han, L., Guo, X., Wang, Y., and Sun, J. (2017). ICE*Sag*37, a novel integrative and conjugative element carrying antimicrobial resistance genes and potential virulence factors in *Streptococcus agalactiae*. *Front. Microbiol.* 8:1921. doi: 10.3389/fmicb.2017.01921

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

frontiers in Microbiology

ORIGINAL RESEARCH
published: 01 April 2021
doi: 10.3389/fmicb.2021.626774

Genomic Diversity and Virulence Potential of ESBL- and AmpC-β-Lactamase-Producing *Escherichia coli* Strains From Healthy Food Animals Across Europe

Christa Ewers[1], Anno de Jong[2], Ellen Prenger-Berninghoff[1], Farid El Garch[2], Ursula Leidner[1], Sumeet K. Tiwari[3] and Torsten Semmler[3]*

[1] Department of Veterinary Medicine, Institute of Hygiene and Infectious Diseases of Animals, Justus Liebig University Giessen, Giessen, Germany, [2] European Antimicrobial Susceptibility Surveillance in Animals (EASSA) Study Group, Executive Animal Health Study Center (CEESA), Brussels, Belgium, [3] NG1 Microbial Genomics, Robert Koch Institute, Berlin, Germany

OPEN ACCESS

Edited by:
Edward Feil, University of Bath, United Kingdom

Reviewed by:
Nilton Lincopan, University of São Paulo, Brazil
Mohamed Salah Abbassi, Tunis El Manar University, Tunisia
Jorge Blanco, University of Santiago de Compostela, Spain

***Correspondence:**
Christa Ewers
Christa.Ewers@vetmed.uni-giessen.de

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

Received: *06 November 2020*
Accepted: *02 March 2021*
Published: *01 April 2021*

Citation:
Ewers C, de Jong A, Prenger-Berninghoff E, El Garch F, Leidner U, Tiwari SK and Semmler T (2021) Genomic Diversity and Virulence Potential of ESBL- and AmpC-β-Lactamase-Producing Escherichia coli Strains From Healthy Food Animals Across Europe. Front. Microbiol. 12:626774. doi: 10.3389/fmicb.2021.626774

The role of livestock animals as a putative source of ESBL/pAmpC *E. coli* for humans is a central issue of research. In a large-scale pan-European surveillance, 2,993 commensal *Escherichia* spp. isolates were recovered from randomly collected fecal samples of healthy cattle, pigs and chickens in various abattoirs. One-hundred *Escherichia* spp. isolates (0.5% from cattle, 1.3% pigs, 8.0% chickens) fulfilled the criteria for cefotaxime and ceftazidime non-wildtype (EUCAST). *In silico* screening of WGS data of 99 isolates (98 *E. coli* and 1 *E. fergusonii*) revealed bla_{SHV-12} (32.3%), $bla_{CTX-M-1}$ (24.2%), and bla_{CMY-2} (22.2%) as predominant ESBL/pAmpC types. Other types were bla_{SHV-2} (1.0%), $bla_{CTX-M-2/-14/-15}$ (1.0/6.1/1.0%), and bla_{TEM-52} (5.1%). Six isolates revealed AmpC-promoter mutations (position −42 (C > T) and one carried *mcr-1*. The majority (91.3%) of ESBL/pAmpC genes were located on plasmids. SHV-12 was mainly (50%) encoded on IncI1α plasmids (pST-3/-26/-95), followed by IncX3 (12.5%) and IncK2 (3.1%). The bla_{TEM-52} genes were located on IncI1α-pST-36 (60%) and IncX1 plasmids (20%). The dominant plasmid lineage among CTX-M-1 isolates was IncI1α (pST-3/-295/-317) (87.5%), followed by IncN-pST-1 (8.3%). CMY-2 was mostly identified on IncI1α (pST-12/-2) (54.5%) and IncK2 (31.8%) plasmids. Several plasmids revealed high similarity to published plasmids from human and animal Enterobacteriaceae. The isolates were assigned to phylogroups A/C (34.7/7.1%), B1 (27.6%), B2 (3.1%), D/F (9.2/10.2%), E (5.1%), and to *E.* clades (3.0%). With 51 known and 2 novel MLST types, a wide variety of STs was found, including STs previously observed in human isolates (ST10/38/117/131/648). ESBL/AmpC types or STs were rarely correlated with the geographic origin of the isolates or animal species. Virulence gene typing identified extraintestinal pathogenic *E. coli* (ExPEC; 2.0%), avian pathogenic *E. coli* (APEC; 51.5%), and atypical enteropathogenic *E. coli* (EPEC; 6.1%). In conclusion, the high diversity of STs and phylogenetic groups provides hardly any hint for clonal spread of single lineages but hints toward the dissemination of cephalosporin

resistance genes in livestock via distinct, globally successful plasmid lineages. Even though a number of isolates could not be assigned to a distinct pathotype, our finding of combined multidrug-resistance and virulence in this facultative pathogen should be considered an additional threat to public health.

Keywords: *Escherichia coli*, livestock, ESBL, AmpC, virulence, sequence type, plasmid, pathotype

INTRODUCTION

Since the turn of the century, the prevalence of infections due to extended-spectrum cephalosporin-resistant (ESC-R) Enterobacteriaceae increased globally both in hospitals and in the community, entailing a major public health concern. At the same time, the prevalence of ESC-R *Escherichia coli* has been increasingly reported in livestock, in the food chain, and in companion animals (Ewers et al., 2012; Kaesbohrer et al., 2019). Consequently, there has been a controversial discussion about the role of animals as a putative source of ESC-R *E. coli* for humans either by direct contact or consumption of contaminated food for many years (Carattoli, 2008; Ewers et al., 2012).

Resistance to third and fourth generation cephalosporins, which are rated highest priority critically important antimicrobials as defined by the World Health Organisation (WHO, 2019), is frequently conferred by the hydrolytic activity of extended-spectrum β-lactamases (ESBL) and plasmid-mediated AmpC-β-lactamases (pAmpC). The most clinically significant ESBL variants belong to the CTX-M, TEM and SHV families, pAmpCs are mostly represented by the CMY family (Bush and Fisher, 2011; Ewers et al., 2012; Carattoli, 2013). The frequent localization of bla_{ESBL} and bla_{AmpC} genes on plasmids and their common association with mobile genetic elements, like transposons and insertions sequences, contributes to their successful spread. Horizontal transfer facilitates easy transmission of ESBL/pAmpCs between bacteria of the same or closely related species, including commensals and pathogens, and including the intestinal microbiota of animals and humans (Ewers et al., 2012; Carattoli, 2013; Liebana et al., 2013; Madec et al., 2017; Rozwandowicz et al., 2018). ESBL-/pAmpC-encoding plasmids determined in the European Union (EU) frequently belong to the incompatibility groups (Inc) F, A/C, N, HI2, I1, and K, which differ in their host range and self-transmissibility (EFSA Panel on Biological Hazards (BIOHAZ), 2011; Liakopoulos et al., 2016). To understand the spread of ESC-R *E. coli* it is not only essential to consider the genetic context of ESBL/pAmpC plasmids but also the phylogenetic epidemiology and pathogenic potential of the bacterium.

E. coli comprises non-pathogenic commensals and strains that cause a variety of diseases in humans and animals. *E. coli* strains capable of causing extraintestinal infections, such as urinary tract infection, blood stream infection and soft tissue damage, are termed extraintestinal pathogenic *E. coli* (ExPEC), whereas strains that lead to intestinal diseases are designated as intestinal pathogenic *E. coli* (InPEC). While the classification of InPEC into different pathovars follows defined genetic criteria, i.e., concrete virulence-associated genes (VAGs) and/or phenotypic features, a wide range of VAGs have been associated with ExPEC and its various pathovars. It is assumed that ExPEC reside in the normal gut microbiota of healthy mammals and birds (Ewers et al., 2009; Koehler and Dobrindt, 2011; Starcic Erjavec and Zgur-Bertok, 2015). The combination of antimicrobial resistance (AMR) and virulence in *E. coli* strains constitutes a significant public health risk as exemplified by the successful global dissemination of ESBL-ExPEC lineages ST131 und ST648 (Ewers et al., 2014a; Schaufler et al., 2019), or by the CTX-M-15 producing enteroaggregative and Shigatoxin producing *E. coli* O104:H4 strains that caused the large German EHEC outbreak in 2011 (Karch et al., 2012).

In this study, we determined the prevalence of ESC-non-susceptible (ESC-non-S) and ESC-R isolates in a European collection of presumptive ESBL/pAmpC-producing *E. coli* and characterized ESBL/pAmpC plasmids and AMR genes as well as the clonal diversity and serotype of the strains based on whole genome sequence (WGS) analysis. Virulence gene profiles were determined to facilitate classification into ExPEC and InPEC pathovars in order to assess the pathogenic potential of antimicrobial resistant strains.

MATERIALS AND METHODS

Bacterial Isolates

The European Antimicrobial Susceptibility Surveillance in Animals (EASSA) monitors the antimicrobial susceptibility of zoonotic and commensal bacteria in healthy food-producing animals at slaughter across Europe (de Jong et al., 2013). The EASSA project includes the major countries of production of beef cattle, slaughter pigs and broiler chickens in the EU. Five or six countries were selected per animal species with ≥4 slaughterhouses in each country providing samples (usually about 200 per animal species). Colon and cecal samples were randomly collected from each of the major food-producing animal species. At most one isolate of each bacterial species was retained from each animal, which was randomly selected as being representative of a whole herd or flock. Our study comprised 2,993 non-repetitive *E. coli* isolates from broiler chicken (n = 1016), cattle (n = 841), pig (n = 1136)

Abbreviations: AMR, antimicrobial resistance; APEC, avian pathogenic *E. coli*; CARD, Comprehensive Antibiotic Resistance Database; CGE, Center for Genomic Epidemiology; EHEC, enterohemorrhagic *E. coli*; EIEC, enteroinvasive *E. coli*; EPEC, enteropathogenic *E. coli*; ESC, extended-spectrum cephalosporin; ExPEC, extraintestinal pathogenic *E. coli*; InPEC, intestinal pathogenic *E. coli*; MALDI-TOF MS, matrix-assisted laser desorption/ionization- time of flight mass spectrometry; MLST, multilocus sequence typing; pMLST, plasmid multilocus sequence typing; ST, sequence type; UPEC, uropathogenic *E. coli*; VAG, virulence-associated gene; WGS, whole genome sequencing.

isolated in Belgium (n = 167), Denmark (n = 208), France (n = 588), Germany (n = 426), Hungary (n = 328), Poland (n = 182), Spain (n = 410), the Netherlands (n = 387), and the United Kingdom (n = 297) isolated from February 2013 to March 2015 (2013, 1,181 isolates; 2014, 1,588 isolates; 2015, 224 isolates).

Antimicrobial Susceptibility Testing and Verification of Bacterial Species

Of the total *Escherichia* spp. collection 100 isolates fulfilled the criteria for cefotaxime and ceftazidime non-wild type according to EUCAST criteria, i.e., reduced susceptible or resistant to one or both cephalosporins [isolates with MICs of cefotaxime and/or ceftazidime $\geq$ 1 mg/L; EUCAST (The European Committee on Antimicrobial Susceptibility Testing), 2015; **Table 1**]. MALDI-TOF MS analysis (Bruker Daltonics, Bremen, Germany) confirmed 99 of these isolates as *E. coli* and one isolate as *E. fergusonii*.

The 100 isolates were further investigated for their susceptibility to 16 antimicrobials/antimicrobial combinations comprising 12 antibiotic classes used in human and/or veterinary medicine including the β-lactam antimicrobials ampicillin, cefepime, cefotaxime, ceftazidime and cefoxitin by using agar dilution according to Clinical Laboratory Standard Institute standards (CLSI, 2013). Minimum inhibitory concentrations (MICs) were interpreted according to CLSI criteria (CLSI, 2018), except those for colistin and tigecycline which were interpreted according to EUCAST guidelines [EUCAST (The European Committee on Antimicrobial Susceptibility Testing), 2015]. The *E. coli* strain ATCC 25922 was used for quality control. Multi-drug resistance (MDR) of an isolate was defined as clinical resistance to at least one agent in three or more antimicrobial classes (Magiorakos et al., 2012). The antimicrobials (and classes) included were ampicillin (penicillins), cefoxitin (cephamycins), cefotaxime, ceftazidime, and cefepime (extended-spectrum cephalosporins), ciprofloxacin (fluoroquinolones), chloramphenicol (phenicols), colistin (polymyxins), gentamicin (aminoglycosides), meropenem (carbapenems), trimethoprim/sulfamethoxazole (folate pathway inhibitors), sulfisoxazole (sulfonamides), tetracycline (tetracyclines), and tigecycline (glycylcyclines).

Whole Genome Sequencing

Ninety-nine cefotaxime and ceftazidime non-wild type *Escherichia* spp. isolates (one isolate was excluded due to repeated DNA degradation) were whole genome sequenced. DNA was extracted from the isolates using the DNA Blood and Tissue Kit according to the manufacturer's instruction (Qiagen, Hilden, Germany), followed by library preparation, using Nextera XT library (Illumina, San Diego, United States). DNA was sequenced by using an Illumina NextSeq 550 with multiplexing of 70 samples per flow cell using 150 bp paired end reads and a minimum of 70-fold coverage. Raw reads were adapter-trimmed by Flexbar v.3.0.3 (Resource Identification Portal RRID: SCR_013001), corrected using BayesHammer and assembled *de novo* using SPAdes v3.12.1 (RRID: SCR_000131). Assembled draft genomes were annotated using Prodigal (Prodigal, RRID: SCR_011936).

Resistance Gene Screening and Integron Typing

The web-based tool ResFinder 3.2[1], hosted at the Center for Genomic Epidemiology (CGE), was used to identify resistance

[1]https://cge.cbs.dtu.dk/services/ResFinder/

TABLE 1 | ESBL/pAmpC types in ESC-R *E. coli* (n = 99) and *E. fergusonii* (n = 1) from the EASSA collection 2013–2015.

		Ceftazidime		Cefotaxime		Country[b] (number of isolates)									Number of isolates per animal species[c]		
ESBL/AmpC	**No. of isolates**	**MIC range (mg/L)**	**MIC_{50} (mg/L)[a]**	**MIC range (mg/L)**	**MIC_{50} (mg/L)[a]**	**BE**	**DE**	**DK**	**ES**	**FR**	**HU**	**NL**	**PL**	**UK**	**Chicken**	**Pig**	**Cattle**
SHV-12	31	8–128	64	0.5–32	2				25	1	2	3			28	3	
SHV-2	1	2	–	4	–				1						1		
TEM-52	5	8–32	–	4–16	–				2		2	1			4	1	
CTX-M-1	23	2–16	8	8– > 32	> 32		4		3	10	5			1	17	5	1
CTX-M-2	1	4	–	32	–	1											1
CTX-M-14	6	0.5–64	–	8– > 32	–				6						4	2	
CTX-M-15	1	16	–	> 32	–								1				1
CMY-2	20	16–128	32	4–16	8	1			6		11	2			18	1	1
SHV-12 & CMY-2	1	128	–	16	–						1				1		
CTX-M-1 & CMY-2	1	64	–	8	–					1					1		
none/n.d.	9/1	8–32	–	2–32	–		1		3		2	4			7	3	
Totals, if applicable	**100**	**0.5–128**	**32**	**0.5– > 32**	**8**	**2**	**5**	**0**	**46**	**12**	**23**	**10**	**1**	**1**	**81**	**15**	**4**

n.d.: not determined (no WGS data; the relative proportion of ESBL/AmpC types among isolates given in the text refers to 99 isolates).
[a]*MIC_{50} was calculated if number of isolates exceeded 10.*
[b]*Country abbreviations are Belgium (BE), Germany (DE), Denmark (DK), Spain (ES), France (FR), Hungary (HU), the Netherlands (NL), Poland (PL), United Kingdom (UK).*
[c]*Chicken specimens were collected in ES, FR, HU, NL, UK; pig specimens in DE, DK, ES, FR, NL, UK; cattle specimens in BE, DE, FR, PL, and UK.*

genes and chromosomal mutations related to β-lactam (*ampC* promoter mutation), fluoroquinolone (mutations in *gyrA*, *gyrB*, *parA*, and *parC*), and colistin resistance (*pmrAB*) based on WGS (Zankari et al., 2012). Resistance gene screening was carried out by BLASTn (90% identity and 90% query coverage) analysis against homologous genes present in the Comprehensive Antibiotic Resistance Database[2].

To investigate the distribution of class 1, class 2, and class 3 integrons with prevalent arrays in the genomes of our *Escherichia* spp. isolates, we performed BLAST search against the available databases GenBank and INTEGRALL 1.2, using filter parameters of more than 90% nucleotide identity and at least 80% query coverage.

Plasmid Analysis

PlasmidFinder 2.1 and pMLST 2.0 (Carattoli et al., 2014) were applied to determine plasmid replicon types and plasmid multilocus sequence types (pST). To further characterize the identified ESBL and pAmpC gene harboring contigs, all contigs were aligned using Geneious (v. 8.1.9, Biomatters Ltd., Auckland, New Zealand) (Geneious, RRID: SCR_010519) to the respective genes (ESBL genes: $bla_{CTX-M-1}$ gene, GenBank accession no. DQ915955.1; $bla_{CTX-M-2}$, EU622041.1; $bla_{CTX-M-14}$, AF252622.2; $bla_{CTX-M-15}$, DQ302097.1; bla_{SHV-2}, AF282921.1; bla_{SHV-12}, AF462395.1; bla_{TEM-52}, AF027199.1; pAmpC gene: bla_{CMY-2}, X91840.1). Contigs containing one of these genes were aligned to publicly available ESBL-/pAmpC plasmids of different incompatibility groups (Inc) from GenBank (GenBank, RRID: SCR_002760) (**Supplementary Table S1**). All contigs of a respective isolate were then aligned to the reference plasmid that revealed highest similarity. In addition, contigs were mapped to the selected reference plasmids using the Geneious Map to Reference (Geneious, RRID: SCR_010519). *In silico* constructed plasmids were further examined for mobile genetic elements, using ISfinder (RRID: SCR_003020).

To display circular comparisons between plasmids we used the blast ring image generator software BRIG Version 0.95 (RRID: SCR_007802).

To test whether ESBL and AmpC-β-lactamase genes were transferable, conjugation was performed by the broth filter mating method at 37°C on 40 isolates which represented different ESBL/pAmpC genes and plasmid replicon types using plasmid-free sodium azide resistant *E. coli* J53 (J53 Azi^R) as recipient. Transconjugants were selected on Mueller Hinton agar plates supplemented with 100 mg/L sodium azide and 4 mg/L cefotaxime (Sigma-Aldrich, Germany). To confirm successful plasmid transfer, antimicrobial susceptibility testing of transconjugants and PCR for ESBL/pAmpC genes were performed. In addition to WGS analysis, plasmid sizes and location of ESBL/pAmpC genes were determined for all 99 isolates by S1 nuclease-restriction followed by pulsed-field gel electrophoresis and subsequent southern blotting. To verify the genomic location of ESBL and pAmpC genes we further performed an *in silico* search of whole genome sequences with mlplasmids v. 1.0.0 (Arredondo-Alonso et al., 2018).

[2]https://card.mcmaster.ca

Phylogenetic Analysis

Phylogenetic groups were determined by using the ClermonTyping method and its associated web-interface ClermonTyper, that allows a given strain sequence to be assigned to *E. albertii, E. fergusonii*, *Escherichia* clades I-V, *E. coli sensu stricto* as well as to the seven main *E. coli* phylogroups A, B1, C, E, D, F, and B2 (Clermont et al., 2013; Beghain et al., 2018). MLST 2.0[3] (Larsen et al., 2012) was applied to identify the multilocus sequence type (RRID: SCR_010245) of *E. coli* isolates following the Achtman scheme[4], which represents a 7-gene-scheme including genes *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA*.

Detection of Serotype and Virulence-Associated Genes

The sero(geno)type of *E. coli* strains was determined using the web-based serotyping tool SerotypeFinder 2.0[5] (Joensen et al., 2015). Screening for VAGs was carried out by NCBI BLASTn (RRID: SCR_004870) analysis against homologous genes present in an in-house database of 800 VAGs, gene variants or genomic islands from a subset of the VirulenceFinder database and in-house created and manually curated VAG reference sequences. Coverage length and sequence identity thresholds were 80 and 90%. We searched for genes that were previously linked with intestinal and extraintestinal pathogenic *E. coli* pathovars and that belonged to different categories (adhesin, toxin, iron uptake system, capsule synthesis, auto-transporter, and secretions system). A VAG score was definded as the total number of VAGs within an isolate. In case a VAG was detected multiple times within a single isolate, it was only counted once. The *kpsM*, *afa/dra*, and *sfa/foc* operons were considered present if any of the corresponding genes or allelic variants were identified. All *E. coli* isolates were further analyzed for the presence of *fimH* gene and allele type by aligning to a FimH database using FimTyper 1.0[6].

Determination of Associations Between *E. coli* Host, Country of Origin, Phylogroups, VAGs, and AMR Genes

The clustering of binary matrix indicating the presence and absence of antibiotic and virulence classes was performed by using package Rtsne version 0.15 (RRID: SCR_016900) with 5000 iteration and perplexity 15 in R version 3.6.1 (R Project for Statistical Computing, RRID: SCR_001905). Microreact version 5.93.0 was applied to visualize the clustering pattern in association with the metadata of the strains (Argimon et al., 2016). Kruskal-Wallis test as well as χ^2 and Fisher's exact test (IBM SPSS Statistics 27) were used to determine whether proportions for one variable (AMR gene, VAGs) were different between phylogroups and isolate origin (host and country). P-values < 0.001 were considered as statistical significant.

[3]https://cge.cbs.dtu.dk/services/MLST/

[4]https://pubmlst.org/mlst/

[5]https://cge.cbs.dtu.dk/services/SerotypeFinder/

[6]https://cge.cbs.dtu.dk/services/FimTyper/

Core Genome Analysis

Phylogenetic and population genetic relationships were determined by applying a gene-by-gene approach on the dataset to generate a core genome alignment and subsequently a phylogenetic tree. The core genome alignment was assembled by a gene-wise alignment with Mafft v7.407 (RRID: SCR_011811), of 1,366 core genes that were present in at least 99% of the strains (sequence similarity min. 70%, sequence coverage min. 90%) and were concatenated afterward. The resulting alignment was used to infer a phylogeny with 100 bootstrap replicates using RAxML v.8.2.10 (RAxML, RRID: SCR_006086) with a General Time Reversible model and gamma correction for among site rate variation. iTOL v5 (RRID: SCR_018174) was used to visualize the population structure in the context of available metadata.

RESULTS

Antimicrobial Susceptibility

Of the total collection of 2,993 presumed *E. coli* isolates, 100 (3.3%) isolates fulfilled the criteria for cefotaxime and ceftazidime non-wild type, i.e., reduced susceptible (ESC-non-S) or resistant (ESC-R) to one or both cephalosporins [EUCAST (The European Committee on Antimicrobial Susceptibility Testing), 2015; CLSI, 2018]. Spain (46 out of 100 ESC-non-S/ESC-R isolates; 46/410 = 11.2% within-country rate) and Hungary (23/328 = 7.0%) were the main sources of ESC-non-S/ESC-R isolates followed by France (12/588 = 2.0%), the Netherlands (10/387 = 2.6%), Germany (5/426 = 1.2%), Belgium (2/167 = 1.2%), Poland (1/182 = 0.6%), and United Kingdom (1/297 = 0.3%) while no ESC-non-S/ESC-R isolate was retrieved from 208 *E. coli* strains from Denmark. Most of the ESC-non-S/ESC-R isolates originated from poultry (81/100 = 81%), followed by pigs (15/100 = 15%), and cattle (4/100 = 4%) (**Table 1**). Also with regard to the total number of isolates investigated, ESC-non-S/ESC-R isolates were more prevalent in chickens (n = 81/1016; 8.0%) compared with pigs (n = 15/1136; 1.3%) and cattle (n = 4/841; 0.5%).

Percentages of clinical resistance of 100 ESC-non-S/ESC-R *Escherichia* spp. isolates to compounds other than cefotaxime and ceftazidime were in decreasing number: ampicillin 100%, sulfisoxazole 78%, tetracycline 68%, nalidixic acid 67%, trimethoprim/sulfamethoxazole 37%, ciprofloxacin 32%, cefoxitin 27%, chloramphenicol 27%, gentamicin 4%, and colistin 1%. All strains were susceptible to meropenem and tigecycline.

Almost all isolates (97%) revealed MDR. The most frequent MDR phenotype was combined resistance to penicillins, cephalosporins, tetracycline, sulfisoxazole, and either phenicols or folate pathway inhibitors (n = 8 each). Different proportions of isolates showed resistance to two (3%), three (18%), four (27%), five (23%), six (19%), seven (7%), and eight (3%) antibiotic classes. On average, isolates from cattle, chickens and pigs revealed a mean number of resistance to 4.3, 3.5, and 3.0 antibiotic classes, respectively. Thirty-four phenotypic AMR patterns (based on resistance of isolates to 12 antibiotic classes) were distinguished (**Supplementary Table S2**). Eleven of these patterns were unique; the remaining patterns were dispersed among a number of 2–8 isolates. AMR patterns appeared to be unrelated to host and country origin of the *E. coli* isolates.

Antimicrobial Resistance Genes and Integrons

The genomes of 99 out of 100 ESC-non-S/ESC-R isolates were sequenced successfully. One isolate from a chicken in Spain turned out to belong to the species *E. fergusonii*. Ninety out of these 99 isolates (90.9%) possessed one or more ESBL/pAmpC genes (**Table 1**). Another six isolates that lacked ESBL/pAmpC genes revealed mutations in the promoter of the chromosomally encoded *ampC* gene when compared with the wildtype *E. coli* K12 sequence (ATCC 25922; NZ_CP009072.1). Mutations were located at positions −42 (alternate −35 box, C > T), −18 (alternate −10 box, G > A), + 81 (AmpC coding region, A > G), and outside functional promtoter elements and AmpC coding region (−1, C > T; + 58, C > T). In three ESBL- and AmpC negative isolates, namely IHIT31981, IHIT31985 (both from chicken in the Netherlands) and IHIT32069 (pig, Netherlands), the genetic background of the ESC-R phenotype remained unclear.

Eight different ESBL/pAmpC genes were detected among the 99 *Escherichia* spp. isolates (98 *E. coli*, one *E. fergusonii*) from chickens (n = 80), pigs (n = 15), and cattle (n = 4) (**Figure 1A**). $Bla_{\mathrm{SHV-12}}$ was the most prevalent ESBL gene (32.3%) with the highest frequency observed among isolates from chickens (29/80 = 36.3%), followed by swine (3/15 = 20.0%). With regard to the geographical origin, isolates from Spain (25/46 = 54.3%) and from the Netherlands (3/9 = 33.3%) most frequently harbored $bla_{\mathrm{SHV-12}}$ (**Figure 1B**).

The genes $bla_{\mathrm{CTX-M-1}}$ (24/99, 24.2%) and $bla_{\mathrm{CMY-2}}$ (22/99, 22.2%) were the second and third most identified ESBL/pAmpC genes among the 99 *Escherichia* spp. isolates. The genes were nearly equally present among isolates from chickens ($bla_{\mathrm{CTX-M-1}}$: 18/80 = 22.5%; $bla_{\mathrm{CMY-2}}$: 20/80 = 25.0%). The $bla_{\mathrm{CTX-M-1}}$ gene showed the highest frequency in ESC-non-S/ESC-R isolates from pigs (33.3%), while with 25.0% of the four cattle isolates harbored this gene. With respect to country origin, the CTX-M-1 encoding gene was predominant in isolates from France (11/12 = 91.7%) while it was the only ESBL gene detected among isolates from Germany and the United Kingdom. Two isolates revealed co-presence of ESBL/pAmpC genes: $bla_{\mathrm{CMY-2}}$ and $bla_{\mathrm{SHV-12}}$ (chicken, Hungary) and $bla_{\mathrm{CMY-2}}$ and $bla_{\mathrm{CTX-M-1}}$ (chicken, France). Other ESBL types observed were $bla_{\mathrm{SHV-2}}$ (1/99 = 1.0%), $bla_{\mathrm{CTX-M-2/-14/-15}}$ (1, 6, and 1 out of 99 = 1.0, 6.1, and 1.0%), and $bla_{\mathrm{TEM-52}}$ (5/99 = 5.1%) (**Figures 1A,B**).

Besides ESBL/pAmpC genes, we observed several other genes among the 99 *Escherichia* spp. isolates. Additional β-lactamase genes identified were $bla_{\mathrm{TEM-1}}$ (32.3%) and bla_{OXA-9} (1.0%). Different proportions of isolates harbored genes encoding for resistance to aminoglycosides (*aadA1*, 43.4%; *aadA2*, 27.3%; *aadA5*, 11.1%; *aadA15* (2.0%); *aadA24* (2.0%), *aac(6')-Ib*, 2.0%; *strA/strB*, 28.3%; *aac(3)-IId*, 1.0%; *aac(3)IVa*, 4.0%; *aph(3″)-Ia*, 1.0%; *aph(3″)-Ic*, 3.0%), tetracyclines [*tet*(A), 52.5%; *tet*(B), 15.2%], folate pathway inhibitors (*dfrA1*, 24.2%; *dfrA7*, 1.0%;

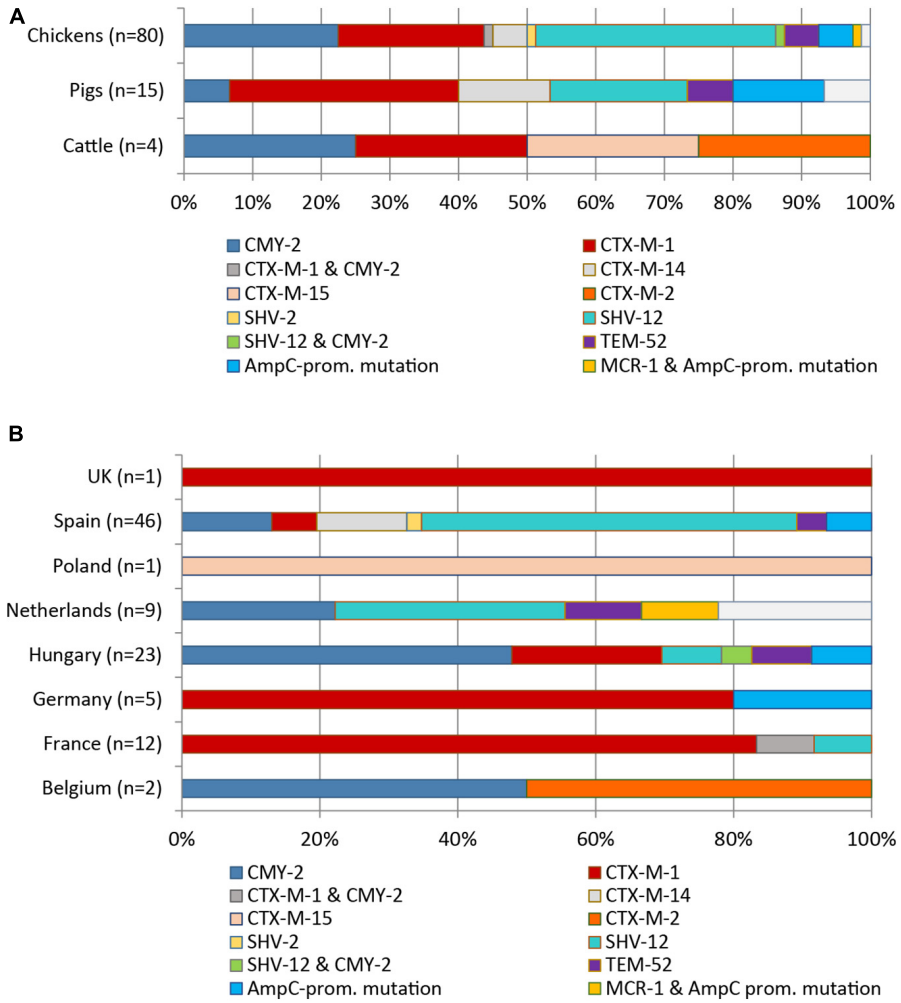


FIGURE 1 | Distribution of ESBL/pAmpC genes and AmpC promotor mutations among 99 ESC-non-S/ESC-R *E. coli* isolates with regard to livestock species **(A)** and country of origin **(B)**.

dfrA8, 1.0%; *dfrA12*, 4.0%; *dfrA16*, 1.0%; *dfrA17*, 12.1%; *sul1*, 24.2%; *sul2*, 44.4%; *sul3*, 25.3%), phenicols (*catA1*, 9.1%; *catB2* 1.0%; *catB3*, 2.2%; *cmlA*, 22.2%; *floR*, 2.0%), lincosamides [*Inu*(F), 2.0%], and macrolides [*mph*(A), 2.0%; *mph*(B), 9.1%, *mef*(B), 1.0%]. Apart from *tet*(A) and *mef*(B), the isolates commonly harbored other genes encoding efflux pumps and/or products involved in membrane permeability, such as *acrAB* (100%), *acrAD* (96.0%), *acrEF* (93.9%), *mdtAB* (98.0%), and *mdtEF* (98.0%). Almost 40% of the resistance plasmids revealed co-carriage of ESBL/pAmpC and AMR genes encoding tetracycline, aminoglycoside, phenicol, and fluoroquinolone resistance. Detailed results from resistance gene screening are provided in **Supplementary Document S1A**.

Aminoglycoside, trimethoprim and sulfonamide resistance genes were frequently located on integrons. More than half of the isolates harbored gene cassettes (GC) associated with integron class I (55.6%). Integron class II was less prevalent (3.0%) and integron class III was not present. The 58 integron-positive isolates revealed phenotypic resistance (mean value) to 5.5 antibiotic classes compared to 3.5 classes for isolates without integrons. In total, 17 and 2 different GCs were found on class I and class II integrons, respectively. The most common class I integron GC was *estX-psp-aadA2-cmlA1-aadA1-qacI* embedded in the genetic platform *tnp440-sul3-orf1-ygkA-yusZ-orf1-?mef(B)-tnp26* (7/55 integron class I isolates; 12.7%). It was almost identical with the GC present in SHV-12-IncI1α plasmid pCAZ590 (GenBank LT669764.1) from an *E. coli* isolate from chicken in Germany (Alonso et al., 2017). Other frequently identified class I integron GCs were *dfrA1-aadA1-qacEdelta1-sul1* (18.2%), and *dfrA17-aadA5-tnp-orf1-orf5-sul2* (14.6%). Eleven other GCs were distributed among the remaining 14 isolates, which is detailed in **Supplementary Document S2**.

10 isolates (10.1%) carried plasmid-mediated quinolone resistance (PMQR) gene *qnrS1*. Sixty-six strains (66.7%) had at least one mutation in GyrA [D87Y (n = 2), S83L (n = 33), S83L and D87H (n = 1), S83L and D87N (n = 29)], and 55.0% of these strains possessed at least one additional mutation in ParC [mostly S80I (n = 26)] (**Supplementary Document S2**). All

isolates with wildtype GyrA and ParC sequences (n = 33) were phenotypically susceptible to ciprofloxacin, even though three of these isolates harbored the *qnrS1* gene (MICs 0.5–1.0 mg/L). On the other hand, resistance to ciprofloxacin was usually (97.7%) associated with mutations in *gyrA* and *parC* genes, irrespective of the presence of *qnrS1* in the isolates.

Furthermore, we identified the plasmid-mediated colistin resistance gene *mcr-1* in one colistin resistant strain isolated from a broiler chicken sample in the Netherlands in 2013. Other *mcr* variants (*mcr-2–mcr-10*) or alleles were not present.

Genomic Location of ESBL, pAmpC and mcr-1 Genes and Plasmid Analysis

The most frequent replicon types identified were Col (MG828, RNAI, 156, pVC, 8282, BS512, MGD2, E10, MP18) (n = 161), FIB (n = 78), FII (n = 67), I1 (n = 65), FIC (n = 31), P0111 (n = 23), X1 (n = 20), and B/O/K/Z (n = 19), Q1 (n = 15), and FIA (n = 14). Among the replicon types that were found less than 10 times were I2 (n = 8), N (n = 6), X4 (n = 6), Y (n = 5), HI2 (n = 4), X3 (n = 4), and A/C2 (n = 2). Contig alignment to several completely assembled plasmids from GenBank database carrying the different ESBL-/AmpC/*mcr-1* genes identified in this study, together with mlplasmid analysis results and data from S1 nuclease digestion and Southern Blot hybridization enabled the identification of $bla_{ESBL/pAmpC}$ and *mcr-1* gene location and calculation of plasmid sizes.

The majority of strains (83/91; 91.2%) carried their ESBL/pAmpC/MCR-1 genes on plasmids. For 30 isolates, there was a link between ESBL/pAmpC genes and replicon sequences on the same *de novo* assembled contig. In the remaining isolates, WGS-based reconstruction, i.e., a mapping of contigs to previously published plasmid sequences together with plasmid prediction of contigs by using mlplasmids was performed to deduce the respective plasmid contigs. In some cases, we could not reconstruct the entire plasmid backbones. Consequently, Inc types of ESBL/pAmpC encoding plasmids could either not (n = 3 isolates Inc not typable) or only partially (n = 7 isolates non-IncI1-type) be assigned **Supplementary Document S2**. Transconjugation assays with 40 selected strains confirmed the transferability of ESBL/pAmpC/MCR-1 plasmids and phenotypic resistance (beta-lactam or colistin) for all strains and transconjugants, respectively (data not shown).

CTX-M Type β-Lactamases

The most common gene-plasmid combination was IncI1α-$bla_{CTX-M-1}$ (n = 21) (**Table 2**). Plasmid sizes ranged between 100 and 137 kb. The dominant plasmid lineage was pST-3 (CC3), followed by pST-295 (CC2) and the novel pST-317 (CC3) (*ardA*_19, *pilL*_2, *repI1*_2, *sogS*_1, *trbA*_4) that was identified in a chicken isolate from Spain. IncI1-pST-3 plasmids revealed three different resistance gene patterns: $bla_{CTX-M-1}$ and *sul2*, $bla_{CTX-M-1}$ and *sul2* and *tet* (A), and $bla_{CTX-M-1}$ and *sul2* and *aadA5* and *dfrA17*.

Twenty of these plasmids showed the genomic backbone of pC60-108, a 108.6 kb plasmid determined in an *E. coli* strain from chicken in Switzerland in 2013 (Wang et al., 2014; **Supplementary Document S3** and **Supplementary Figure S3**). Another two isolates from pigs in France and Germany revealed $bla_{CTX-M-1}$ on IncN/pST-1 plasmids that were 42.4 kb in size and showed a similarity of 98.7% to IncN/pST-1 plasmid pL2-43 (**Supplementary Document S3** and **Supplementary Figure S4**). Like pL2-43 (*E. coli*, lamb, Switzerland, 2013), our plasmids harbored macrolide resistance gene *mph*(A) in addition to $bla_{CTX-M-1}$ (Wang et al., 2014).

In four avian *E. coli* isolates from Spain, we located the $bla_{CTX-M-14}$ gene on a plasmid with unknown replicon type. The CTX-M-14 plasmids were about 110 kb in size and showed partial similarity with 55.1 kb CMY-2 plasmid 2016C-3936C1-unnamed2 (GenBank: CP018772) (**Supplementary Document S3** and **Supplementary Figure S5**).

The only ESBL-producing strain identified among *E. coli* isolates from Poland (cattle, ST617, phylogroup C) was at the same time the only strain that harbored CTX-M-15. The $bla_{CTX-M-15}$ gene was located on the chromosome and the genetic environment shows high similarity to that of human intra-abdominal infection ST4981 *E. coli* strain CH611_eco (CP017980.1) from China (data not shown). Although we could verify plasmid location of the $bla_{CTX-M-2}$ gene of a bovine isolate from Belgium, a comparison with published plasmid sequences revealed no significant results.

SHV-12

The bla_{SHV-12} genes (n = 32) were predominantly located on IncI1α plasmids (50.0%), followed by IncX3 (12.5%), and IncK2 (3.1%) (**Table 2**). For seven isolates, we could not reconstruct plasmid sequences. According to results from sequence and S1-nuclease restriction analysis we termed them "non-IncI1"plasmids. Three isolates carried the bla_{SHV-12} gene on the chromosome. IncI1α plasmids (95.4–119 kb) were assigned to pST-26/CC2 (n = 6; Spain and France) pST-3/CC3 (n = 7; Spain), and pST-95/CC9 (n = 3; Netherlands). ST-26 and ST-95 plasmids co-harbored resistance genes *sul3*, *tet*(A), *cmlA*, *aadA1,* and *aadA2b*, whereas pST-3 plasmids only possessed the bla_{SHV-12} gene. Plasmid backbones of SHV-12 IncI1α plasmids revealed similarity to that of bla_{SHV-12}-carrying plasmid pCAZ590 (GenBank: LT669764.1) that was identified in an *E. coli* isolate from chicken in Germany in 2011 (**Supplementary Document S3** and **Supplementary Figure S6**).

The SHV-12 IncX3 plasmids were 46.3–46.4 kb in size and were highly similar (99.9%) to each other and to the reference plasmid pEC-244 GenBank: KX618704) (**Supplementary Document S3** and **Supplementary Figure S7**). Like pEC-244 (*E. coli*, chicken feces, Netherlands), our plasmids harbored PMQR gene *qnrS1* in addition to bla_{SHV-12}. The IncX3 plasmids also showed significant similarity to the plasmid backbone of the SHV-12 plasmid pUHKPC33 (*K. pneumoniae*, human patient, United States) (GenBank: NZ_CP011992).

TEM-52

Of five bla_{TEM-52} genes observed, three were encoded on a 89.5 kb IncI1α pST-36/CC-3 plasmid. These plasmids showed > 99% identity to reference plasmid pESBL-117 (urine, human patient, the Netherlands) (GenBank: CP008734.1) (**Supplementary Document S3** and **Supplementary Figure S8**)

and carried no additional resistance genes. One chicken isolate from Spain harbored the bla_{TEM-52} gene on a 38.7 kb IncX1 plasmid with 99.7% identity to reference plasmid pDKX-TEM-52 (JQ269336.1) that was isolated from chicken meat in Denmark in 2006 (Johnson et al., 2012).

CMY-2

The bla_{CMY-2} gene that was present on plasmids with replicon types IncI1α (n = 12), IncK2 (n = 7), and IncA/C (n = 1) (**Table 2**). The 95–122 kb CMY-2 IncI α plasmids were classified into pMLST types pST-12 (n = 10) and pST-2 (n = 2), and carried not additional resistance gene.

Nine pST-12 plasmids revealed high similarity to the 99 kb-plasmid p11-004736-1-7_99 (*S.* Heidelberg, cattle, Canada, 2011) (**Supplementary Document S3** and **Supplementary Figure S9**; Labbe et al., 2016). IncI1α/pST-2 plasmids from chickens in Spain and Hungary showed > 99% similarity to plasmid pSA01AB09084001_92 (*S.* Heidelberg, chicken, cecal content, Canada, 2009) (**Supplementary Document S3** and **Supplementary Figure S10**; Labbe et al., 2016). The backbone of IncK2-CMY-2 plasmids (83–86 kb) was highly similar to the bla_{CMY-2}-carrying plasmid pDV45 (85.9 kb) from poultry retail meat (**Supplementary Document S3** and **Supplementary Figure S11**; Seiffert et al., 2017). While CMY-2 plasmids of the IncI1α and IncK2 replicon types encoded no further resistance genes, the IncA/C-pST-3 plasmid (99.9 kb) observed in a bovine *E. coli* isolate from Belgium carried resistance genes *dfrA12*, *tet*(A), *aadA2*, *aph(3″)-Ib*, *aph(6)-Id*, *floR*, *sul1*, and *sul2* in addition. It revealed a plasmid backbone of the 135.2 kb plasmid pSH163_135 (*S.* Heidelberg, turkey, United States, 2002) (**Supplementary Document S3** and **Supplementary Figure S12**; Han et al., 2012). It also showed high similarity to the 96 kb *K. pneumoniae* plasmid pKP_Goe_024-2, which was isolated from the abdominal drainage fluid from a human patient in Germany in 2014 and carried a bla_{OXA-48} instead of a bla_{CMY-2} gene. Two avian isolates from Spain and Hungary carrying CMY-2 plasmids without known replicon types revealed a genomic backbone of 55.1 kb-plasmid 2016C-3936C1 unnamed2 (*E. coli* O157, human, United States) (GenBank: CP018772) (**Supplementary Document S3** and **Supplementary Figure S13**).

MCR-1

The *mcr-1* gene in avian *E. coli* isolate IHIT31981 was encoded on a 33.3 kb IncX4 plasmid that was almost identical (99.9% nucleotide sequence identity) with other globally distributed MCR-1 IncX4 plasmids, such as pMCR-1-CT (GenBank: CP018773.2), determined in an *E. coli* isolate from a human patient in the United States.

Phylogenetic Grouping

Using the ClermonTyping method and its associated web-interface ClermonTyper, 98 isolates were verified as *E. coli* and

TABLE 2 | Genomic location of ESBL/AmpC and MCR-1 genes and assignment to Inc groups.

ESBL/AmpC/MCR-1	Location		Inc typing and pMLST					Plasmid size
	ch (host)*	pl	Inc type	ST	CC	No. (host)*	No. (country)*	
CTX-M-1	0	23	IncI1α	3	3	16 (C), 3 (P)	10 (F), H(5), G (3), E (1)	100–137 kb
				295	2	1 (C)	1 (E)	~105 kb
				317	3	1 (C)	1 (E)	~100 kb
			IncN	1	n.d.	2 (P)	1 (F), 1 (G)	~43 kb
CTX-M-14	2 (P)	4	n.d.	n.d.	n.d.	4 (C)	4 (E)	110 kb
CTX-M-15	1 (Ca)	0	n.d.	n.d.	n.d.	1 (Ca)	1 (PL)	none
CTX-M-2	0	1	n.d.	n.d.	n.d.	1 (Ca)	1 (B)	n.d.
SHV-12	3 (C), 1 (P)	28	IncI1α	26	2	6 (C)	5 (E), 1 (F)	113–119 kb
				3	3	6 (C), 1 (P)	7 (E)	95–100 kb
				95	9	3 (C)	3 (NL)	115–116 kb
			IncX3	n.d.	n.d.	4 (C)	3 (H), 1 (E)	46 kb
			IncK2	n.d.	n.d.	1 (C)	1 (E)	~80 kb
			non-IncI1α	n.d.	n.d.	6 (C), 1 (P)	7 (E)	80–200 kb
SHV-2	0	1	non-IncI1α	n.d.	n.d.	1 (C)	1 (E)	~130 kb
TEM-52	0	5	IncI1α	36	3	2 (C), 1 (P)	2 (H), 1 (NL)	89.5 kb
			IncX1	n.d.	n.d.	1 (C)	1 (E)	38.7 kb
			non-IncI1α	n.d.	n.d.	1 (C)	1 (E)	~95 kb
CMY-2	0	22	IncI1α	12	10	9 (C), 1 (P)	6 (H), 2 (NL), 2 (E)	96–99 kb
				2	2	2 (C)	1 (E), 1 (H)	93–94 kb
			IncK2	n.d.	n.d.	7 (C)	4 (H), 2 (E), 1 (F)	80–90 kb
			IncA/C	3	n.d.	1 (Ca)	1 (B)	100 kb
			n.d.	n.d.	n.d.	2 (C)	1 (E), 1 (H)	90–95 kb
MCR-1	0	1	IncX4	n.d.	n.d.	1 (C)	1 (NL)	33.3 kb

**ch, chromosomal; pl, plasmid; C, chicken; P, pig; Ca, cattle; B, Belgium; G, Germany; E, Spain; F, France; H, Hungary; NL, Netherlands; PL, Poland; n.d., not determined.*

one, as expected, as *E. fergusonii*. The majority of the 98 *E. coli* isolates were assigned to phylogenetic groups A (34.7%) and B1 (27.6%). Group C, a newly defined phylogroup within the formerly highly diverse group A (Clermont et al., 2013), was identified in 7.1% of the isolates. Phylogenetic groups B2 (3.1%), D (9.2%), and F (10.2%), the latter one representing a novel group D-related cluster, were less prevalent. Three isolates (3.0%) from chickens in Spain belonged to clade I. Five isolates (5.1%) were assigned to phylogenetic group E (**Figure 2A**). All phylogroups B2 and D and 90% of group F isolates, often referred to as extraintestinal pathogenic *E. coli* (ExPEC) groups, were from chickens. Those phylogroups regarded as commensals (A, B1, C) were detected in 62.5% of chicken, 100% of cattle, and in 93.3% of pig isolates.

The distribution of ESBL/pAmpC types revealed no significant association to the phylogenetic group of a strain (**Figure 2B**). Among 12 antibiotic classes tested, phylogroup B1 strains showed the highest mean number of resistance (resistant to 5.19 ± 1.36 classes), followed by isolates belonging to group E (4.8 ± 0.87), A (4.79 ± 1.28), clade I (4.67 ± 1.33), F (4.4 ± 1.05), and D (4.11 ± 1.04). Group C (3.71 ± 0.89) and B2 (3.67 ± 0.33) isolates revealed the lowest antimicrobial resistance with regard to different antibiotic classes.

Distribution of Serotypes

Overall, 48 different serotypes were identified among the 99 ESC-non-S/ESC-R isolates (**Supplementary Document S2**). Fifteen isolates, including *E. fergusonii*, were O not typable. Twenty-nine serotypes appeared as singletons. Twenty-one isolates revealed serotypes that are frequently associated with ExPEC including O1, O2, and O6 (n = 2 each), O7 (n = 6), O8 (n = 8), and O78 (n = 1) (Ewers et al., 2007). In addition, the CMY-2-ST131 isolates revealed serotype O25b:H4, which is highly linked with this ST. Moreover, one CMY-2 producing isolate (IHIT32055, ST533) from a chicken in Hungary revealed serotype O157:H10, which is frequently observed among enterohemorrhagic strains.

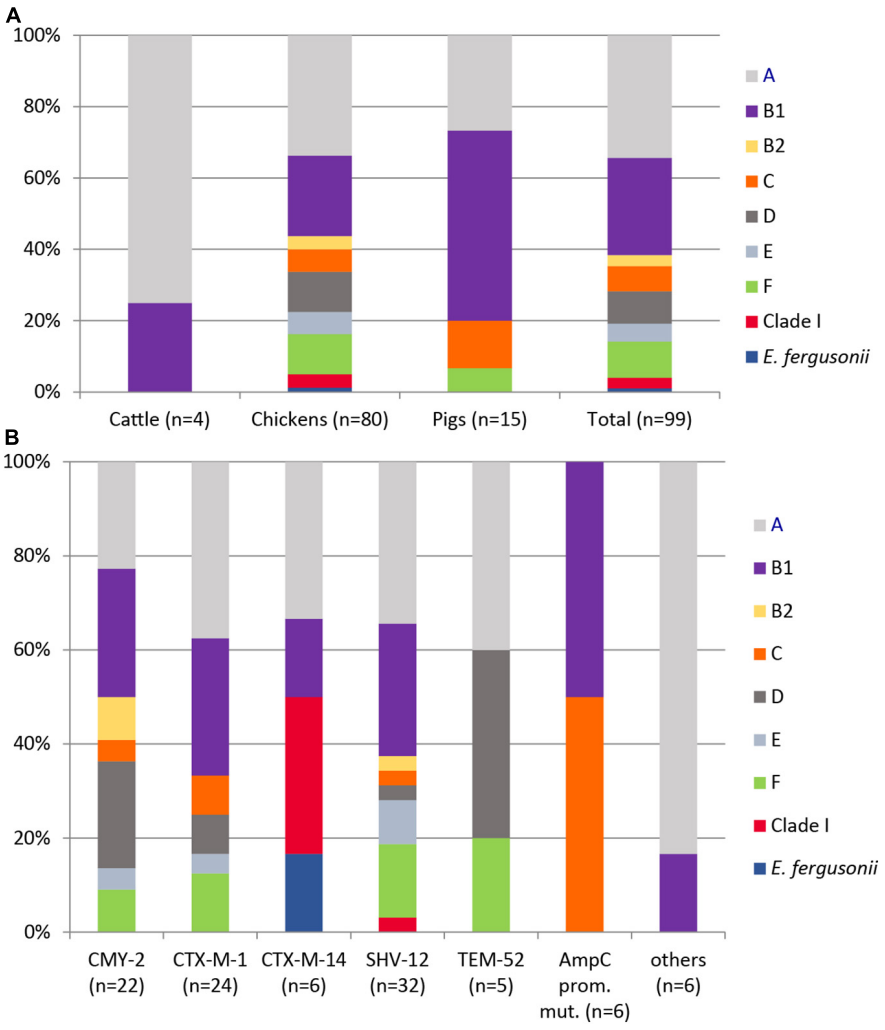


FIGURE 2 | Distribution of phylogenetic groups among 99 ESC-non-S/ESC-R *Escherichia* spp. isolates with respect to animal host **(A)** and ESBL/pAmpC genes and AmpC-promoter mutations **(B)**.

Other serotypes that are commonly recorded from EHEC and EPEC (O26, O103, O111, and O145) or from ETEC (O45, O138, O139, O141, O147, and O149) were only rarely detected (two TEM-51 O149:H20 isolates from chickens in Hungary; two avian O45 isolates from Spain (CMY-2) and France (CTX-M-1).

Distribution of Virulence-Associated Genes (VAGs) and Pathotyping

We tested our 99 *Escherichia* spp. strains for 800 VAGs associated with intestinal and extraintestinal pathogenic *E. coli* pathovars. More than 300 of these VAGs, belonging to the categories adhesion (n = 98), iron acquisition (n = 45), invasion and protection (n = 42), secretion system (n = 115), and autotransporter/toxin (n = 19) were identified in the isolates with different frequencies (**Supplementary Document S1**).

Mean Number of VAGs

The mean number (± SD) of VAGs detected was 113.04 ± 15.84. The VAG content was highest in phylogroup D isolates (134.4 ± 13.7), followed by groups E (130.8 ± 14.1), C (127.1 ± 14.9), B2 (119. 7 ± 2.7), F (115.6 ± 12.1), B1 (113.7 ± 13.3), clade I (113.0 ± 2.0), and group A (102.6 ± 14.0). Mean VAG numbers according to ESBL/pAmpC genes carried by the isolates were in decreasing order: 125.2 ± 11.6 (AmpC-promoter mutation); 119.5 ± 16.7 (CMY-2); 114.5 ± 22.0 (CTX-M-1); 113.0 (SHV-2, one isolate); 113.0 ± 6.0 (TEM-52); 111.4 ± 10.1 (SHV-12); 104.0 (CTX-M-15, one isolate); 99.2 ± 16.7 (CTX-M-14); 84.0 (CTX-M-2, one isolate) (**Supplementary Table S4**). Isolates from pigs (117.0 ± 17.5) and chickens (113.5 ± 14.9) harbored significantly higher numbers of VAGs than isolates from cattle (89.5 ± 9.6).

InPEC-Related VAGs

The majority of isolates carried none of the 43 InPEC-related VAGs (**Table 3**). One or two of these VAGs were present in 43.4% of the isolates, and only 6.1% of the isolates carried 8–12 genes. None of the isolates possessed invasion plasmid antigen gene *ipaH*, which would indicate the presence of enteroinvasive *E. coli* (EIEC). Most typical factors associated with the enteroaggregative *E. coli* (EAEC) pathotype, such as aggregative adherence fimbriae AAF/I to AAF/V (*aggA*, *aafABCD, agg3A*, *agg4A*, *agg5A*), cytotoxic autotransporter protease Pet (*pet*), antiaggregation protein dispersin (*aap*, *aatPABDC*), or a type IV secretion system (*aai*), as well as the global regulator of these genes, AggR (*aggR*, *agg* variants), were also not present. Only three factors that are commonly, though not exclusively, present in EAEC, namely enteroaggregative heat-stable enterotoxin EAST-1 (*EAST-1*, 31.3%), serine protease/mucinase Pic (*pic*, 6.1%) and the enteroaggregative immunoglobulin repeat protein Air (*air*, 21.2%), were detected in our isolates. Due to the absence of genes for cytotoxic necrotizing factors (*cnf1-cnf3*) and cytolethal distending toxins (*cdtA-cdtC*) and of shiga toxin genes (*stx1* and *stx2* variants) the presence of NTEC (necrotoxigenic *E. coli*) and STEC (Shigatoxin-producing *E. coli*) was excluded (**Supplementary Document S1**).

None of our isolates carried ETEC-related toxin genes (*esta, estb, eltA, eltB*). Few isolates harbored genes for ETEC-related fimbriae F4 (*fae*, 4.0%, 2 × chicken, 1 × cattle, 1 × pig), F5 (*fan*, 3.0%, 3 × chicken), and F17 (f17, 1.0%, 1 × cattle), while F6 and F18 fimbriae genes were not detected. Other VAGs often reported for ETEC were found in 31.0% (*EAST-1*—heat-stable enterotoxin 1 gene), 30.3% (*aidA*–AIDA-I-like adhesion gene), 5.1% (*iha*–bifunctional enterobactin receptor adhesin protein gene), and 4.0% (*tia*—toxigenic invasion locus gene) of the isolates.

Six isolates (6.1%) represented atypical enteropathogenic *E. coli* (aEPEC) as they harbored intimin gene *eae* and lacked bundle-forming pili adhesin genes *bfpA-L*, which together with

TABLE 3 | Distribution of InPEC- and ExPEC-related virulence-associated genes (VAGs) in ESC-non-S/ESC-R *E. coli* isolates.

Pathotype	Positive isolates	
	n	%
InPEC[a]		
0 genes	15	15.2
1–2 genes	43	43.4
3–4 genes	16	16.2
5–7 genes	0	0
8–12 genes	6	6.1
EAEC, EIEC, ETEC, NTEC	0	0
EPEC (typical/atypical)	0/6	0/6.1
STEC (including EHEC, EDEC)	0	0
ExPEC		
Group 1 (ExPEC)[b]		
0 genes	75	75.8
1 gene	22	22.2
≥2 genes	**2**	2.0
Group 2 (UPEC)[c]		
ExPEC genesand *cnf1/cnf2* and *hlyD*	0	0
Group 3 (APEC)[d]		
0 genes	18	18.2
1–2 genes	30	30.3
≥3 genes	**51**	**51.5**
Group 4 (ExPEC-like)[e]		
3–5 genes	9	9.1
6–10 genes	31	31.3
11–15 genes	35	35.4
16–20 genes	17	17.2
21–26 genes	7	7.1

[a] *Isolates carrying one or more of 43 InPEC-related VAGs among the categories (i) toxin: cnf1, cnf2, cnf3, cdtA, cdtB, cdtC, EAST-1, eltA, eltB, esta, estb, stx1, stx2; (ii) adhesion: aaf, aggA, aggR, aidA, bfpA, eae, f17, fae, fan, fas, fed, iha, stgA, tir; (iii) invasion, secrection, autotransporter: ces, eatA, escN, espD, etgA, ipaH, lifA, pet, pic, sepDL, tia, tibA; (iv) others: aap, air, ler, nleA (see* ***Supplementary Document S1*** *for detailed distribution).*

[b] *Isolates with genes (afa/dra, focG, iutA, kpsMTII, papA and/or papC, sfa/focDE) that have been suggested to define a strain as an ExPEC; ≥ 2 genes = ExPEC, according to Johnson et al. (2003).*

[c] *Isolates with ≥ 2 genes plus cnf1/cnf2 and hlyD, which are more specifically related to uropathogenic E. coli (UPEC).*

[d] *Isolates with genes (iroN, iss, iutA, ompT, hlyF) that have been suggested to define avian ExPEC; ≥ 3 genes = avian pathogenic E. coli (APEC) according to Johnson et al. (2008).*

[e] *Isolates with genes (n = 32) or gene clusters (n = 17) frequently, but not exclusively associated with ExPEC; isolates with > 10 genes/gene clusters = ExPEC-like (see* ***Supplementary Table S3*** *for list of VAGs).*

eae are indicative for typical EPEC (**Table 3**). Intimin genes revealed ≥ 99.8% nucleotide sequence identity to *eae* type β1 (GenBank AF200363.1). All strains additionally harbored the *tir* (translocated intimin receptor) gene, and most isolates additionally possessed several genes encoding for type III secretion proteins, effectors, and regulators (*esc, ces, epa, epr, etgA, mpc, sepDL*), for locus of enterocyte and effacement (LEE) regulator gene *ler*, and for non-LEE-encoded type III effector genes (*nle*). The ST10 chicken isolate IHIT32040 was the only aEPEC strain that carried genes for the AIDA-I-like adhesin protein (*aidA*) and for Stg fimbriae (*stgABCD*). Three of the aEPEC strains additionaly harbored the porcine attaching and effacing-associated gene *paa* that promotes adherence to intestinal epithelial cells in a characteristic A/E pattern (Batisson et al., 2003).

ExPEC-Related Genes

More than half of the 99 isolates fulfilled the criteria to be classified as avian ExPEC, also termed avian pathogenic *E. coli* (APEC). In detail, 51.5% of our isolates harbored at least three of the five VAGs *iroN, iss, iutA, ompT,* and *hlyF* that have been suggested to define APEC (**Table 3**; Johnson et al., 2008). A few years earlier, Johnson and Russo (2005) defined the group of ExPEC on the basis of the presence of ≥ 2 VAGs or gene combinations, including genes for P fimbrial genes (*papAH* and/or *papC*), S and FIC fimbriae (*sfa/focDE*), Dr. antigen-binding adhesin (*afa/draBC*), group II capsule polysaccharides (*kpsMTII*), and ferric aerobactin receptor (*iutA*) (Johnson and Russo, 2005). This definition includes all pathotypes that were primarily "ill-defined," including uropathogenic *E. coli* (UPEC), neonatal meningitis-associated *E. coli* (NMEC), sepsis-causing *E. coli* (SEPEC), and APEC. Following this definition, only two of our isolates qualified as ExPEC, namely a CMY-2-positive avian isolate from Spain (phylogroup D, ST38) and a CTX-M-1-producing porcine isolate from Germany (phylogroup B1, ST453).

As current evidence suggests that no single virulence determinant renders an ExPEC isolate capable of causing site-specific disease, we additionally screened our isolates for an extensive set of 49 VAGs (**Supplementary Table S3**) that have previously been linked with different ExPEC pathovars in order to categorize them as ExPEC-like pathovar (Dale and Woodford, 2015). Seven isolates, all from chicken, possessed 21–26 ExPEC-related VAGs and belonged to phylogroups F (n = 4), B2 (n = 3), and C (n = 1). Another 17 isolates, including A, B1, C, and D strains, harbored 16–20 genes. The majority of strains harbored 6–10 (31.3%) or 11–15 (35.4%) of the VAGs tested. All strains were negative for afimbrial adhesin Afa (*afaA-G*), Dr. fimbriae (*draA-E, draP*) and S-fimbriae (*sfaA-H*, *sfaS*), as well as for toxin genes *cnf1-3* and *sat*, i.e., for virulence determinants that are frequently observed among UPEC. Invasion-related gene *gimB*, which has been reported for APEC and NMEC, was also not detected. Genes detected only once were α-hemolysin gene *hlyA and hlyD*, salmochelin receptor gene *iutA*, and Expec adhesin gene *yqi*. Other ExPEC-related genes, encoding P fimbriae (*pap*; 4.0%), K1 capsule synthesis (*neu*; 4.0%), uropathogenic specific factor (*usp*; 8.1%), serine protease Pic (*pic*; 6.1%), vacuolating autotransporter toxin Vat (*vat*; 3.0%), adherence protein Iha (*iha*; 5.1%), invasion factor IbeA (*ibeA*; 6.1%), or antigen 43 (*agn43*; 6.1%) occurred in 2–8 of the isolates. Between 91.0 and 99.0% of the isolated possessed Curli fiber (*csg*) and type I fimbriae (*fim*) genes. The most often detected iron acquisition genes were *ent* (98.0%), *sit* (66.7%), *iuc* (54.5%), and *iro* (46.5%), coding for enterobactin siderophore, iron transport system, aerobactin siderophore, and a salmochelin siderophore system, respectively. Among the protectins, increased serum survival protein *iss* (77.8%) and outer membrane exclusion protein *traT* (75.8%) were most prevalent (**Supplementary Table S3**).

Clonal Diversity

With 51 known and two novel types, we found a wide variety of multilocus sequence types (STs) among our strains. For isolates from broilers (n = 80), pigs (n = 15) and cattle (n = 4) we determined 42, 14, and 3 different STs, respectively (**Table 4**). Thirty-two of these STs have been assigned to 18 clonal complexes (CCs), while 21 STs did not cluster within one of the CCs defined in the Enterobase *E. coli* database (**Supplementary Document S2**)[7]. The most often identified STs were ST117 and ST10 (n = 6 each), ST665 (n = 5) ST23, ST38, ST155 (n = 4 each), ST101, ST354, ST4980, ST752, ST770, and ST88 (n = 3 each). ST648 and ST131, two globally distributed lineages that are frequently associated with ESBL/AmpC production were identified in a CTX-M-1-producing chicken isolate from Spain (ST648) and in two CMY-2 positive chicken ST131 isolates from different farms in Hungary. A correlation between STs and animal species or country of origin was not evident. In addition, ESBL and AmpC β-lactamase types, as well as plasmid replicon types showed a random distribution among the different STs.

We further determined the relatedness of strains from different animal species and countries by WGS comparison. Although the application of comparative core genome analysis of 98 *E. coli* isolates, based on 1,330 orthologous genes, achieved a much higher resolution regarding the genetic relatedness of the ESBL/pAmpC/MCR-1 strains (SNP-based comparison of 1,330 genes vs. allele-based comparison of seven MLST genes), it basically reflected the high genetic diversity observed by MLST analysis (**Figure 3**). As expected, isolates forming one ST and/or CC by based on MLST data as well as isolates that shared the same phylogenetic groups almost always clustered together in the cgMLST scheme. In contrast, neither ESBL/AmpC-types nor plasmid Inc types followed the phylogenetic distribution of the isolates but were broadly distributed among the different cgMLST clusters.

Correlation of Antimicrobial Resistance and VAG Pattern With Metadata of the Isolates

We first visualized the clustering pattern, i.e., AMR genes categorized according to their mode of action with the metadata of isolates (**Supplementary Figure S1**). As expected, there was

[7]https://enterobase.warwick.ac.uk/

an association between the category "antibiotic inactivation" and ESBL/pAmpC types ($p < 0.001$). In addition, single AMR genes showed associations with either host (e.g., *floR*, *aph(3)-Ia*, *aph(3)-Ib*, *aph(6)-Id*, and *mphA* to cattle; *aph(3)-Ib*, *aph(6)-Id*, *dfrA12*, and *pmrE* to chicken), country, ESBL/AmpC-type, and phylogenetic group.

The same analysis was done using VAGs grouped into different categories as the clustering pattern (**Supplementary Figure S2**). Significant associations were predominantly observed between VAGs and the phylogenetic group. In total 36 adhesion-related, 11 autotransporter and toxin genes, 22 invasion and protectin genes, 14 iron acquisition genes and 33 secretion systems genes revealed several *P*-values < 0.001. In contrast, only 14, 21, and 7 VAGs showed significant associations with the variables host, country, and ESBL/AmpC type.

TABLE 4 | Multilocus sequence types and ESBL/AmpC types of 99 ESC-non-S/ESC-R *Escherichia* spp. from livestock animals.

Chicken		Chicken (continued)	
ST (n)	**ESBL/pAmpC***	**ST (n)**	**ESBL/pAmpC**
ST10 (5)	SHV-12 (3), CTX-M-1, CTX-M-14	ST1594 (2)	CTX-M-1 (1)
ST23 (4)	CTX-M-1, CMY-2, none (2)	ST1621 (1)	SHV-12
ST38 (4)	CMY-2 (3), CTX-M-1 (2)	ST1730 (1)	none
ST48 (2)	SHV-12 (2)	ST4980 (3)	TEM-52 (1), SHV-12, CTX-M-1
ST68 (1)	CMY-2	ST3406 (1)	CTX-M-1
ST88 (1)	CTX-M-1	ST3994 (1)	none
ST93 (1)	CMY-2	ST4118 (2)	TEM-52 (2)
ST101 (2)	SHV-12, CMY-2	ST4243 (1)	CMY-2
ST117 (6)	CTX-M-1 (2), CMY-2, TEM-52, SHV-12/CMY-2	ST7104 (1)	CTX-M-1
		ST7852** (1)	CTX-M-14
ST131 (2)	CMY-2 (2)	ST10807 (1)	CTX-M-1
ST135 (1)	SHV-12		
ST154 (1)	SHV-12	**Swine**	
ST155 (4)	SHV-12 (3), CTX-M-1	**ST (n)**	**ESBL/pAmpC**
ST156 (2)	none (2)	ST10 (1)	CTX-M-1
ST162 (1)	CTX-M-1	ST58 (1)	CTX-M-1
ST189 (1)	TEM-52	ST56 (1)	CTX-M-1
ST354 (3)	SHV-12 (3)	ST88 (2)	SHV-12 (1), none (1)
ST371 (1)	CMY-2	ST101 (1)	SHV-12, TEM-52
ST398 (2)	SHV-12 (2)	ST345 (1)	SHV-12
ST453 (1)	CMY-2	ST398 (1)	CTX-M-1
ST533 (1)	CMY-2	ST453 (1)	CTX-M-1
ST648 (1)	CTX-M-1	ST847 (1)	none
ST665 (5)	CMY-2 (2), SHV-2, SHV-12, CTX-M-1	ST1147 (1)	CTX-M-14
		ST1304 (1)	CMY-2
ST752 (3)	SHV-12 (2), CMY-2	ST2197 (1)	CTX-M-14
ST770 (3)	CTX-M-14 (2), SHV-12	ST10813 (1)	none
ST997 (2)	SHV-12 (2)		
ST1137 (1)	CTX-M-1	**Cattle**	
ST1158 (1)	SHV-12	**ST**	**ESBL/pAmpC**
ST1246 (1)	SHV-12	ST515 (1)	CMY-2
ST1304 (1)	CMY-2	ST617 (1)	CTX-M-15
ST1431 (1)	SHV-12	ST744 (2)	CTX-M-1, CTX-M-2
ST1551 (1)	CTX-M-1		

**In case an ESBL/AmpC type appears more than once, the number is indicated in brackets.*

***Escherichia fergusonii (IHIT32041).*

DISCUSSION

This study focused on the identification and investigation of ESBL and pAmpC genes in ESC-non-susceptible/ESC-resistant commensal *E. coli* isolates. Special attention was paid to the antimicrobial resistance phenotype of the isolates and their plasmids with regard to replicon type and genetic plasmid structure. We further investigated the molecular epidemiology of *E. coli* isolates with respect to their genome-based phylogeny and assignment to intestinal and extraintestinal *E. coli* pathotypes. The presence of potentially virulent and antimicrobial resistant ESBL/pAmpC-producing *E. coli* originating from healthy production animals could be a significant issue for human and animal health.

One hundred ESC-non-susceptible/ESC-resistant isolates were identified among 2,993 commensal *E. coli* isolates collected over a time period of 25 months (February 2013–March 2015) from different livestock species and European countries. The highest rate of ESC-non-susceptible/ESC-resistant isolates was detected in poultry (8.0%), whereas pigs (1.3%) and cattle (0.5%) revealed comparable low numbers of such isolates. These findings are fairly in agreement with the numbers reported by the European Food Safety Authority (EFSA) and the European Centre for Disease Prevention and Control (ECDC). Here the number of presumptive ESBL and AmpC phenotypes identified in indicator *E. coli* isolates rated to 5.4% for broiler flocks in 2014 [EFSA (European Food Safety Authority) and ECDC (European Centre for Disease Prevention and Control), 2016], and to 1.5 and 2.0% for fattening pigs and calves under 1 year of age, respectively, in 2015 [EFSA (European Food Safety Authority) and ECDC (European Centre for Disease Prevention and Control), 2017]. The lower number observed for cattle in our study might be due to the mixed origin of isolates according to animal age. Different studies could show that veal calves harbor higher proportions of ESBL/AmpC-producing bacteria than beef cattle [Haenni et al., 2014; EFSA (European Food Safety Authority) and ECDC (European Centre for Disease Prevention and Control), 2017]. Recent data from 2017 to 2018 indicated that the occurrence of presumptive ESBL- and/or AmpC-producing indicator *E. coli* isolates was still generally low in reporting European countries, ranging from 0.6 to 6.3% in isolates from pigs, from 1.2 to 5.3% in isolates from calves and from 0.6 to 7.9% in isolates from broilers [EFSA (European Food Safety Authority) and ECDC (European Centre for Disease Prevention and Control), 2020]. Carbapenemase producing (CP) *E. coli* were not detected in our study, which may be due to the absence of such isolates and/or the

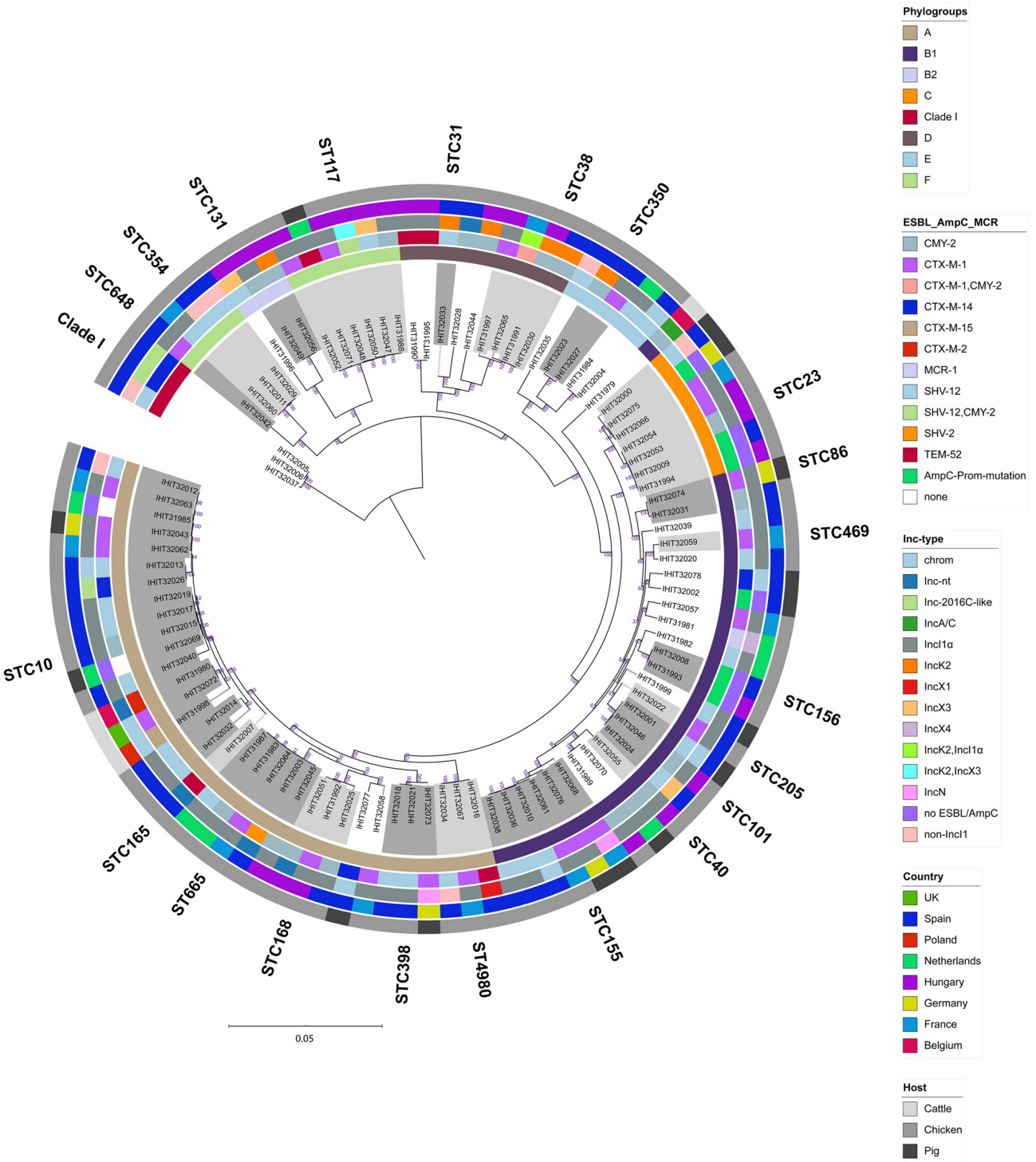


FIGURE 3 | Maximum-likelihood tree of 98 *E. coli* genomes (*E. fergusonii* isolate IHIT32041 was not included). The phylogeny is based on 1336 orthologous genes and was calculated by RAxML. Graphical visualization was achieved by iTOL v5. Colored rings represent (from inner to outer ring the (i) phylogroup, (ii) ESC-non-S/ESC-R resistance determinant, (iii) ESBL/AmpC/MCR-1 plasmid replicon type, (iv) country of isolation, and (v) host.

cultivation of the samples on non-selective media. The presence of CP bacteria on livestock farms was first reported in 2012 in Germany (Fischer et al., 2013). Since then, increasingly emerged in different livestock species and countries (Anderson and Boerlin, 2020). Despite of these findings, the overall prevalence of CPs in livestock is still low as shown by the results from

the specific carbapenemase producers monitoring in European countries between 2016 and 2018 [EFSA (European Food Safety Authority) and ECDC (European Centre for Disease Prevention and Control), 2020].

Almost all isolates (97.0%) from our study revealed an MDR phenotype, i.e., resistance to antimicrobial agents of at least three classes, and 29.0% of the isolates showed resistance to ≥ six classes of antimicrobial agents. This is consistent with previous studies from different countries. Ninety percent of cefotaxime and/or ciprofloxacin resistant, commensal *E. coli* isolated from food-producing animals in Belgium were resistant to several other antibiotics (Lambrecht et al., 2018). MDR was also observed in 88.1% of ESBL-producing isolates obtained between 2008 and 2014 from diseased food-producing animals in Germany (Michael et al., 2017). In this study, 39.4% of the isolates, predominantly from cattle, were resistant to six antibiotic classes. Several authors assumed a correlation between antimicrobial usage in food-producing animals and the prevalence of AMR bacteria in those animals. According to publicly available national or international reports from seven European countries, the level of veterinary use of specific antimicrobials strongly correlates to the level of resistance toward these agents in commensal *E. coli* isolates in pigs, poultry and cattle (Chantziaras et al., 2014). A significant correlation between antimicrobial use and resistance in commensal *E. coli* was also demonstrated for livestock animals in Belgium (Callens et al., 2018). In our study, isolates were frequently (32–100%) resistant to ampicillin, sulfisoxazole, tetracycline, trimethoprim/sulfamethoxazole, and ciprofloxacin. Antimicrobial agents of such classes are among the most sold agents to veterinarians in different European countries (Lekagul et al., 2018; Roth et al., 2019), indicating that AMR and frequent antibiotic usage could coincide to an extent that remains to be determined.

The identification of eight different ESBL/pAmpC types indicated a high diversity of β-lactamases among our isolates, which is consistent with previous studies (Valentin et al., 2014; Ceccarelli et al., 2019). The two major ESBL genes were $bla_{\mathrm{SHV-12}}$ (35.6%) and $bla_{\mathrm{CTX-M-1}}$ (24.2%), while $bla_{\mathrm{CMY-2}}$ (22.2%) was the only pAmpC gene detected. The CTX-M-1 β-lactamase is the most common ESBL in livestock animals and food, and also SHV-12 ranks among the most predominant ESBL types within Enterobacteriaceae of diverse origins (Ewers et al., 2012; Valentin et al., 2014; Alonso et al., 2017; Liakopoulos et al., 2018). Regarding $bla_{\mathrm{CMY-2}}$, an increasing number of *E. coli* carrying this gene was reported in the European livestock production in recent years (Ewers et al., 2012; Larsen et al., 2012; Irrgang et al., 2018; Liakopoulos et al., 2018; Pietsch et al., 2018; Ceccarelli et al., 2019; EFSA (European Food Safety Authority) and ECDC (European Centre for Disease Prevention and Control), 2019). Consistent with our findings (25.0% of ESC-R/ESC-non-resistant poultry isolates carried $bla_{\mathrm{CMY-2}}$), a prevalence of more than 30% among global ESC-R *E. coli* expressed CMY-2, particularly in poultry (Ewers et al., 2012).

ESBL genes $bla_{\mathrm{CTX-M-14}}$ (6.1%, chicken and pig) and $bla_{\mathrm{TEM-52}}$ (5.1%, chicken and pig) were present in lower numbers among our isolates, while $bla_{\mathrm{SHV-2}}$ (chicken), $bla_{\mathrm{CTX-M-2}}$ (cattle) and $bla_{\mathrm{CTX-M-15}}$ (cattle) occurred only once. Together with CTX-M-1 and CTX-M-15, CTX-M-14 represents one of the most common ESBL type in *E. coli* isolates from cattle and its largely moderate occurrence in *E. coli* isolates derived from chickens and pigs has also been demonstrated (Ewers et al., 2012; Palmeira et al., 2020).

AmpC activity due to chromosomal *ampC* promoter mutations was found in 6.1% (6/99) of our isolates. All seven isolates revealed a mutation pattern previously described as *ampC* type 03 (Mulvey et al., 2005). This type was reported as one of the most prevalent variants among cefoxitin-resistant clinical *E. coli* isolates in different studies (Mulvey et al., 2005; Peter-Getzlaff et al., 2011). The *ampC* promoter alterations created an alternate displaced promoter whose mutation at position −42 (C > T) is thought to have strong effect on promoter strength and consequently on hyperproduction of *ampC* (Caroff et al., 2000).

The majority of the 70 ESBL and 22 pAmpC genes determined in this study were plasmid located (84/92; 91.3%). IncI1α was the most represented plasmid family (52/92; 56.5%) that occurred in 57.5% (46/80) of chicken and in 40.0% (6/15) of pig isolates, followed by IncK2 (8/92; 8.7%; all from chicken), IncX3 (4/92; 4.3%, chicken), IncN (2/92; 2.2%, pigs), IncA/C (1/92; 1.2%, cattle), and IncX1 (1/92; 1.2%, chicken). For 17.4% (16/92) of our ESBL/pAmpC plasmids, we could not assign a definite Inc type. Previous studies highlighted the challenge in reconstructing plasmids encoding resistance genes from WGS, mostly due to the lack of long reads (Orlek et al., 2017; Arredondo-Alonso et al., 2018; Ludden et al., 2019).

CTX-M-1-producing isolates belonged largely (75.0%) to IncI1α/pST-3 plasmids, which represent a globally distributed, genetically highly similar plasmid lineage that is disseminated in livestock animals and humans in Europe (Carattoli, 2008; Zurfluh et al., 2014; Irrgang et al., 2018; Touzain et al., 2018).[8] IncI1 plasmids are highly conjugative, and transmission from commensal to pathogenic bacteria poses a risk for animal and human health (Carattoli, 2013). One of our CTX-M-1 plasmids was assigned as pST-295, a plasmid type that was first identified in a commensal CTX-M-1 ST1604 *E. coli* isolate from chicken in Denmark in 2015 (Liu et al., 2019). In that study, four novel CTX-M-1 IncIα plasmid lineages (pST-293–pST-296) were determined among livestock isolates. Together with our finding of novel type pST-317, this could reflect a possible diversifying evolutionary process in these otherwise highly conserved plasmids. Noteworthy, the plasmid transfer of several plasmid types, including IncI1-pST3-CTX-M-1 and IncI1-pST295-CTX-M-1 could be significantly increased *in vitro* in a strain-dependent manner following exposure to the antibiotics cefotaxime, ampicillin, and ciprofloxacin (Liu et al., 2019).

The diverse population of *E. coli* isolates carrying CTX-M-1 IncI1 plasmids (21 IncI1 plasmids among 18 different MLST types) suggests that the dissemination of this ESBL gene is not due to the spread of single clonal lineages but is more likely the result of vertical transmission of plasmids by horizontal transfer. Also Irrgang and co-workers (2018) reported 51 different STs among 89 CTX-M-1 producing *E. coli* isolates from German food samples.

[8]https://pubmlst.org/plasmid/

In agreement with recent reports, SHV-12 was mainly encoded on IncI1α plasmids (50%) in our isolates, followed by IncX3 (12.5%) and IncK2 (3.1%) (Huijbers et al., 2014; Alonso et al., 2017; Apostolakos et al., 2020). When typing 23 SHV-12 positive *E. coli* from human, animal and food sources, Alonso et al. (2017) identified IncI1 as predominant replicon type (73.9%), while IncK2 (13.0%) and IncX3 (4.3%) were less prevalent (Alonso et al., 2017). IncI1-pST-3-SHV-12 was suggested as a poultry-associated plasmid lineage, whereas IncI1-pST26-SHV-12 plasmids were associated with a wide host range contributing to the spread of bla_{SHV-12} genes among different environments (Accogli et al., 2013; Alonso et al., 2017). We cannot conclude such an association from our data, as all pST-26 plasmids were from chicken isolates and pST-3 plasmids from chicken and pig isolates. However, this lack of correlation may be due to the predominance of chicken isolates among our collection.

Our SHV-12 IncX3 plasmids were almost identical in plasmid backbone and bla_{SHV-12} flanking region to pEC-244, which was the first completely sequenced IncX3 plasmid of animal origin. In this study, IncX3 plasmids from humans and animals exhibited remarkable synteny in their backbone and differed only in their bla_{SHV-12}-flanking region (Liakopoulos et al., 2018). IncX3 plasmids are conjugative and highly stable, while they exert no fitness cost on their bacterial host, thus highlighting the epidemic potential of these plasmids.

In eight *E. coli* isolates from chickens, bla_{CMY-2} (n = 7) and bla_{SHV-12} (n = 1) has been linked to plasmids of the recently defined incompatibility group IncK2 (Seiffert et al., 2017). IncK2 plasmids are mainly associated with the spread of bla_{CMY-2} and $bla_{CTX-M-14}$ genes in Europe, particularly in Spain and the United Kingdom, and are frequently reported from *E. coli* from livestock sources (Dhanji et al., 2012; Rozwandowicz et al., 2018; Apostolakos et al., 2020). de Been et al. (2014) identified bla_{CMY-2} carrying IncK2 plasmids in human and poultry isolates belonging to evolutionary distinct backgrounds, suggesting that these plasmids efficiently spread through *E. coli* populations of different reservoirs. We also identified IncK2 plasmids in isolates belonging to different phylogenetic groups (B2, D, E, and F) and to six different MLST types, including ST38, ST117, and ST131. The most frequent replicon type identified among bla_{CMY-2} carrying plasmids in our study was IncIα (n = 12 isolates from Netherlands, Hungary, and Spain). IncI1α and IncK2 were also found as major replicon types of bla_{CMY-2} carrying plasmids from *E. coli* isolates from humans, animals and food in Germany (Pietsch et al., 2018). Due to high sequence identity of plasmids, also shown for several plasmids in our study, the authors suggested that plasmid-mediated, rather than clonal spread likely plays an important role for emergence and transmission of bla_{CMY-2} between aminals and humans.

Some of the clones (ST10/A, ST23/C, ST38/D, ST117/F, and ST155/B1) detected in our study more than once have been previously associated with ESBL/pAmpC phenotypes and are widely spread among different environments, including clinical settings (Ewers et al., 2012; de Been et al., 2014; Alonso et al., 2017; Hussain et al., 2019; Ludden et al., 2019). Only two CMY-2 positive isolates belonged to phylogroup B2 and to ST131. The isolates from chickens in Hungary represented the B2-O25b-H4-ST131-*fimH*22-clade B lineage that is mostly composed of fluoroquinolone susceptible isolates, and has been associated with poultry and human bloodstream infections (Pitout and DeVinney, 2017). Several other STs found in our study, including ST88/C, ST93/A, ST101/B1, ST135/B2, ST617/A, ST648/F, and ST1431/B1, have frequently been determined in ESBL- and AmpC-producing *E. coli* from human and animal sources (Ewers et al., 2012; de Been et al., 2014; Pietsch et al., 2018; Schaufler et al., 2019; Apostolakos et al., 2020).

Thirty-one isolates revealed 10 of the sequence types (ST10, ST117, ST131, ST23, ST354, ST38, ST58, ST617, ST648, and ST88) that belong to the top 20 ExPEC STs, as previously determined based on a meta-analysis including 169 global studies (Manges et al., 2019). However, the classic ExPEC STs ST73 and ST95 that are successful extraintestinal pathogens but have also be reported as persistent intestinal colonizers in humans and animals were not present among our collection (Riley, 2014). It is commonly accepted that ExPEC lineages, such as ST95, ST73, ST12, and ST127 exhibit lower multidrug resistance levels, while only few are capable of combining MDR and virulence, like ST648, ST131, and ST410 (Ewers et al., 2014a, 2016; Schaufler et al., 2016; Manges et al., 2019).

The virulence potential of ESBL- and AmpC-producing *E. coli* originating from healthy livestock animals could be a significant issue for public health. Nevertheless, data on extensive virulence gene typing of AMR *E. coli* from livestock animals, including a rigorous screening for a set of > 800 VAGs related with intestinal and extraintestinal pathotypes, are scarce (Hussain et al., 2019; Apostolakos et al., 2020). We identified the intimin gene *eae*, which is a surrogate for atypical EPEC and several other InPEC-related genes, including the translocated intimin receptor gene *tir* in five ESBL- (SHV-12 and TEM-52) and one CMY-2-producing chicken *E. coli* isolates belonging to four different STs. Mueller et al. determined the *eae* gene in 3.6 and 2.9% of ESBL-producing *E. coli* isolates from livestock (n = 28) and healthy humans (n = 34), respectively. Consistent with our findings, they did not identify genes indicating the presence of EIEC (*ipaH*), ETEC (*sta, stb*, *est*), and STEC (*stx1, stx2*), suggesting a low prevalence of intestinal pathotypes among MDR bacteria from livestock (Mueller et al., 2016). Although bovines are the primary reservoir of STEC, previous reports confirm a low prevalence of ESBL-producing STEC among cattle and other livestock animals (Ewers et al., 2014b; Pietsch et al., 2018). Apostolakos et al. (2020) could assign the majority of 100 ESBL/pAmpC-*E. coli* isolated from the broiler production pyramid in Italy to defined ExPEC or InPEC pathotypes by virulence gene analysis. They identified 56% of their isolates as atypical EAEC, based on the presence of *aatA* and the absence of *aggR*. Both typical and atypical EAEC, that have been predominantly associated with pediatric diarrhea in developing countries, were not present in our isolates as they all lacked the pathotype-specific genes *aatA*, *aggR*, and *aaiC* (Bamidele et al., 2019).

Regarding ExPEC, distinct sequence types, particularly ST131, ST648, and ST410 have been reported as successful pandemic lineages that combine multidrug-resistance and virulence (Ewers et al., 2014a, 2016; Schaufler et al., 2016; Manges et al., 2019).

Whether an *E. coli* isolate falls into the category of ExPEC, likely depends on the definition proposed by different authors. More than half of our 99 isolates fulfilled the criteria to be classified as avian ExPEC (APEC), including 19.6% isolates from non-avian sources. Apostolakos et al. (2020) identified 39 ESBL-*E. coli* broiler isolates (39%) belonging to 13 different STs as APEC. While they APEC group mainly (38.5%) consisted of ST744 and ST429/ST9298, our APEC strains were even more diverse (31 STs among 51 isolates) and only few STs were not unique, including ST117 (n = 6), ST23 (n = 4), ST38 (n = 3), ST88 (n = 3), and ST101 (n = 3). While we had no indication for a co-location of APEC-VAGs and ESBL-/AmpC genes on plasmids, others identified several virulence genes, including *sitA-D*, *iucA-D*, *iutA*, *hlyF*, *ompT*, *iss*, *iroN*, *cvaA-C*, and *cvi* on bla_{CMY-2} IncF plasmid obtained from a diseased French broiler, suggesting a putative threat for the easy dissemination of ExPEC virulence factors and resistance determinants (Touzain et al., 2018).

Only two of our isolates (CMY-2-ST38 and CTX-M-1-ST453) could be classified as ExPEC, whereas 67.7% of the isolates were considered as ExPEC-like (>10 ExPEC genes). Using the same criteria we applied, Apostolakos et al. (2020) identified 51% of their isolates belonging to 12 STs from different production chains of an integrated broiler company in Italy as ExPEC. In a study from Brazil, 58% of ESBL *E. coli* isolates from chicken carcasses harbored 3–5 of the ExPEC VAGs *iutA*, *hlyF*, *iss*, *iroN*, and *ompT* (Cyoia et al., 2018). The authors suggested that chicken meat is a potential reservoir of MDR *E. coli* strains harboring resistance and virulence genes that could pose a serious threat to human public health. This is further corroborated by a study dealing with ESBL-*E. coli* from poultry and human in India. Among 15 ExPEC-associated genes (*papA/C/E/F/G*, *fimH*, *pic*, *sat*, *tsh*, *vat*, *iutA*, *ireA*, *iroN*, *fyuA*, and *usp*) Hussain et al. (2019) identified comparable numbers in broiler (median 4) and human isolates (median 3) and overlaps in the AMR and VAG profile and phylogenetic background of a subset of strains.

Interestingly, several isolates harbored both InPEC- and ExPEC-related genes (**Supplementary document S1**), resembling what has previously been reported as hybrid pathotype, which is probably best exemplified by EHEC as a long-standing EPEC/STEC hetero-pathogen (Santos et al., 2020). Our aEPEC strains additionally harbored 9–12 InPEC related genes, such as ColV plasmid gene *cvi*, iron acquisition gene *fyuA*, *iucA-D*, *irp2*, and *sitA-D*, increased serum survival protein gene *iss*, and haem acquisition protein gene *hma*, which plays a critical role in the colonization of the urinary tract. The ExPEC and APEC strains carried up to four InPEC-related genes, including those encoding for AidA-I adhesin-like protein invasion-associated protein Tia, long polar fimbriae Stg, serine protease Pic, iron-related haem receptor Iha, and ETEC F4 and F5 fimbriae. Apostolakos et al. (2020) identified 30 ESBL/AmpC broiler isolates that displayed an aEAEC/ExPEC pathotype. They suggested that the virulence gene repertoire of ESBL/pAmpC *E. coli* may explain their adaptation to and persistence in different niches. As the capacity for genome interrogations is constantly rising, it seems obvious that various VAGs, previously either linked with InPEC or ExPEC pathotypes, will be found among different *E. coli* pathotypes to which they have not traditionally been associated (Santos et al., 2020). This, together with the availability of extensive AMR data based on an ever growing number of genome sequences may help to predict the emergence of novel multidrug resistant and virulent strains.

CONCLUSION

Our data suggest that cephalosporin resistance genes are mainly disseminated in livestock animals via distinct plasmids. Plasmid backbones within different plasmid lineages were often almost identical and were shared by phylogenetically unrelated isolates from chickens, cattle, and swine. In addition, they revealed significant similarity to plasmids from human isolates. We could demonstrate that *E. coli* of various phylogenetic groups and STs can combine antimicrobial resistance and virulence, even though a number of isolates could not be assigned to a distinct pathotype. In summary, this work significantly contributes to the understanding of the epidemiology and virulence potential of cephalosporin-resistant *E. coli* from livestock animals.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

Ethical review and approval was not required for the study in accordance with the local legislation and institutional requirements.

AUTHOR CONTRIBUTIONS

CE supervised the entire project. CE and AJ drafted the manuscript. CE and FE designed the study. EP-B, CE, and AJ have provided raw data and analyzed the data. UL conducted the laboratory experiments. ST, CE, and TS conducted the analyses of the sequencing data. All authors critically reviewed the manuscript.

FUNDING

This study was supported in part by the Executive Animal Health Study Center (CEESA). ST and TS were funded by the Federal Ministry of Education and Research (BMBF) under project number 01KI1703B as part of the ERA-NET COFUND antimicrobial resistances (JPI-EC-AMR) "HECTOR." CE was funded by the Federal Ministry of Education and Research (BMBF) under project number 01KI1908B as part of the JPI-AMR project "OASIS."

ACKNOWLEDGMENTS

We thank the veterinarians, the different slaughterhouses and the national microbiological laboratories involved in the sampling and isolation procedures.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.626774/full#supplementary-material

REFERENCES

Accogli, M., Fortini, D., Giufre, M., Graziani, C., Dolejska, M., Carattoli, A., et al. (2013). IncI1 plasmids associated with the spread of CMY-2, CTX-M-1 and SHV-12 in *Escherichia coli* of animal and human origin. *Clin. Microbiol. Infect.* 19, E238–E240. doi: 10.1111/1469-0691.12128

Alonso, C. A., Michael, G. B., Li, J., Somalo, S., Simon, C., Wang, Y., et al. (2017). Analysis of bla_{SHV-12}-carrying *Escherichia coli* clones and plasmids from human, animal and food sources. *J. Antimicrob. Chemother.* 72, 1589–1596. doi: 10.1093/jac/dkx024

Anderson, R., and Boerlin, P. (2020). Carbapenemase-producing *Enterobacteriaceae* in animals and methodologies for their detection. *Can. J. Vet. Res.* 84, 3–17.

Apostolakos, I., Feudi, C., Eichhorn, I., Palmieri, N., Fasolato, L., Schwarz, S., et al. (2020). High-resolution characterisation of ESBL/pAmpC-producing *Escherichia coli* isolated from the broiler production pyramid. *Sci. Rep.* 10:11123. doi: 10.1038/s41598-020-68036-9

Argimon, S., Abudahab, K., Goater, R. J. E., Fedosejev, A., Bhai, J., Glasner, C., et al. (2016). Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb. Genom.* 2:e000093. doi: 10.1099/mgen.0.000093

Arredondo-Alonso, S., Rogers, M. R. C., Braat, J. C., Verschuuren, T. D., Top, J., Corander, J., et al. (2018). mlplasmids: a user-friendly tool to predict plasmid- and chromosome-derived sequences for single species. *Microb. Genom.* 4:e000224. doi: 10.1099/mgen.0.000224

Bamidele, O., Jiang, Z. D., and Dupont, H. (2019). Occurrence of putative virulence-related genes, *aatA*, *aggR* and *aaiC*, of enteroaggregative *Escherichia coli* (EAEC) among adults with travelers' diarrhea acquired in Guatemala and Mexico. *Microb. Pathog.* 128, 97–99. doi: 10.1016/j.micpath.2018.12.030

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

Beghain, J., Bridier-Nahmias, A., Le Nagard, H., Denamur, E., and Clermont, O. (2018). ClermonTyping: an easy-to-use and accurate in silico method for *Escherichia* genus strain phylotyping. *Microb. Genom.* 4:e000192. doi: 10.1099/mgen.0.000192

Bush, K., and Fisher, J. F. (2011). Epidemiological expansion, structural studies, and clinical challenges of new beta-lactamases from gram-negative bacteria. *Annu. Rev. Microbiol.* 65, 455–478. doi: 10.1146/annurev-micro-090110-102911

Callens, B., Cargnel, M., Sarrazin, S., Dewulf, J., Hoet, B., Vermeersch, K., et al. (2018). Associations between a decreased veterinary antimicrobial use and resistance in commensal *Escherichia coli* from Belgian livestock species (2011-2015). *Prev. Vet. Med.* 157, 50–58. doi: 10.1016/j.prevetmed.2017.10.013

Carattoli, A. (2008). Animal reservoirs for extended spectrum beta-lactamase producers. *Clin. Microbiol. Infect.* 14(Suppl. 1), 117–123. doi: 10.1111/j.1469-0691.2007.01851.x

Carattoli, A. (2013). Plasmids and the spread of resistance. *Int. J. Med. Microbiol.* 303, 298–304. doi: 10.1016/j.ijmm.2013.02.001

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

Caroff, N., Espaze, E., Gautreau, D., Richet, H., and Reynaud, A. (2000). Analysis of the effects of -42 and -32 *ampC* promoter mutations in clinical isolates of *Escherichia coli* hyperproducing *ampC*. *J. Antimicrob. Chemother.* 45, 783–788. doi: 10.1093/jac/45.6.783

Ceccarelli, D., Kant, A., van Essen-Zandbergen, A., Dierikx, C., Hordijk, J., Wit, B., et al. (2019). Diversity of plasmids and genes encoding resistance to extended spectrum cephalosporins in commensal *Escherichia coli* from Dutch livestock in 2007-2017. *Front. Microbiol.* 10:76. doi: 10.3389/fmicb.2019.00076

Chantziaras, I., Boyen, F., Callens, B., and Dewulf, J. (2014). Correlation between veterinary antimicrobial use and antimicrobial resistance in food-producing animals: a report on seven countries. *J. Antimicrob. Chemother.* 69, 827–834. doi: 10.1093/jac/dkt443

Clermont, O., Christenson, J. K., Denamur, E., and Gordon, D. M. (2013). The Clermont *Escherichia coli* phylotyping method revisited: improvement of specificity and detection of new phylogroups. *Environ. Microbiol. Rep.* 5, 58–65. doi: 10.1111/1758-2229.12019

CLSI (2013). *Performance Standards for Antimicrobial Disk and Dilution Susceptibility Tests for Bacteria isolated from Animals*. Wayne, PA: CLSI.

CLSI (2018). *Performance Standards for Antimicrobial Susceptibility Testing: 28th Informational Supplement M100-S28*. Wayne, PA: CLSI.

Cyoia, P. S., Koga, V. L., Nishio, E. K., Houle, S., Dozois, C. M., de Brito, K. C. T., et al. (2018). Distribution of ExPEC virulence factors, bla_{CTX-M}, *fosA3*, and *mcr-1* in *Escherichia coli* isolated from commercialized chicken carcasses. *Front. Microbiol.* 9:3254. doi: 10.3389/fmicb.2018.03254

Dale, A. P., and Woodford, N. (2015). Extra-intestinal pathogenic *Escherichia coli* (ExPEC): disease, carriage and clones. *J. Infect.* 71, 615–626. doi: 10.1016/j.jinf.2015.09.009

de Been, M., Lanza, V. F., de Toro, M., Scharringa, J., Dohmen, W., Du, Y., et al. (2014). Dissemination of cephalosporin resistance genes between *Escherichia coli* strains from farm animals and humans by specific plasmid lineages. *PLoS Genet.* 10:e1004776. doi: 10.1371/journal.pgen.1004776

de Jong, A., Thomas, V., Klein, U., Marion, H., Moyaert, H., Simjee, S., et al. (2013). Pan-European resistance monitoring programmes encompassing food-borne bacteria and target pathogens of food-producing and companion animals. *Int. J. Antimicrob. Agents* 41, 403–409. doi: 10.1016/j.ijantimicag.2012.11.004

Dhanji, H., Khan, P., Cottell, J. L., Piddock, L. J., Zhang, J., Livermore, D. M., et al. (2012). Dissemination of pCT-like IncK plasmids harboring CTX-M-14 extended-spectrum beta-lactamase among clinical *Escherichia coli* isolates in the United Kingdom. *Antimicrob. Agents Chemother.* 56, 3376–3377. doi: 10.1128/AAC.00313-12

EFSA (European Food Safety Authority) and ECDC (European Centre for Disease Prevention and Control) (2016). The European Union summary report on antimicrobial resistance in zoonotic and indicator bacteria from humans, animals and food in 2014. *EFSA J.* 14:4380. doi: 10.2903/j.efsa.2016.4380

EFSA (European Food Safety Authority) and ECDC (European Centre for Disease Prevention and Control) (2017). The European Union summary report on antimicrobial resistance in zoonotic and indicator bacteria from humans, animals and food in 2015. *EFSA J.* 15:4694. doi: 10.2903/j.efsa.2017.4694

EFSA (European Food Safety Authority) and ECDC (European Centre for Disease Prevention and Control) (2019). The European Union summary report on antimicrobial resistance in zoonotic and indicator bacteria from humans, animals and food in 2017. *EFSA J.* 17:5598. doi: 10.2903/j.efsa.2019.5598

EFSA (European Food Safety Authority) and ECDC (European Centre for Disease Prevention and Control) (2020). The European Union summary report on antimicrobial resistance in zoonotic and indicator bacteria from humans, animals and food in 2017/2018. *EFSA J.* 18:6007. doi: 10.2903/j.efsa.2020.6007

EFSA Panel on Biological Hazards (BIOHAZ) (2011). Scientific Opinion on the public health risks of bacterial strains producing extended-spectrum β-lactamases and/or AmpC β-lactamases in food and food-producing animals. *EFSA J.* 9:2322. doi: 10.2903/j.efsa.2011.2322

EUCAST (The European Committee on Antimicrobial Susceptibility Testing) (2015). *Breakpoint Tables for Interpretation of MICs and Zone Diameters. Version 5.0, 2015*. Available online at: http://www.eucast.org (accessed December 16, 2020).

Ewers, C., Antao, E. M., Diehl, I., Philipp, H. C., and Wieler, L. H. (2009). Intestine and environment of the chicken as reservoirs for extraintestinal pathogenic

Escherichia coli strains with zoonotic potential. *Appl. Environ. Microbiol.* 75, 184–192. doi: 10.1128/AEM.01324-08

Ewers, C., Bethe, A., Semmler, T., Guenther, S., and Wieler, L. H. (2012). Extended-spectrum beta-lactamase-producing and AmpC-producing *Escherichia coli* from livestock and companion animals, and their putative impact on public health: a global perspective. *Clin. Microbiol. Infect.* 18, 646–655. doi: 10.1111/j.1469-0691.2012.03850.x

Ewers, C., Bethe, A., Stamm, I., Grobbel, M., Kopp, P. A., Guerra, B., et al. (2014a). CTX-M-15-D-ST648 *Escherichia coli* from companion animals and horses: another pandemic clone combining multiresistance and extraintestinal virulence? *J. Antimicrob. Chemother.* 69, 1224–1230. doi: 10.1093/jac/dkt516

Ewers, C., Goettig, S., Buelte, M., Fiedler, S., Tietgen, M., Leidner, U., et al. (2016). Genome sequence of avian *Escherichia coli* strain IHIT25637, an extraintestinal pathogenic *E. coli* strain of ST131 encoding colistin resistance determinant MCR-1. *Genome Announc.* 4:e00863-16. doi: 10.1128/genomeA.00863-16

Ewers, C., Li, G., Wilking, H., Kiessling, S., Alt, K., Antao, E. M., et al. (2007). Avian pathogenic, uropathogenic, and newborn meningitis-causing *Escherichia coli*: how closely related are they? *Int. J. Med. Microbiol.* 297, 163–176. doi: 10.1016/j.ijmm.2007.01.003

Ewers, C., Stamm, I., Stolle, I., Guenther, S., Kopp, P. A., Fruth, A., et al. (2014b). Detection of Shiga toxin- and extended-spectrum beta-lactamase-producing *Escherichia coli* O145:NM and Ont:NM from calves with diarrhoea. *J. Antimicrob. Chemother.* 69, 2005–2007. doi: 10.1093/jac/dku042

Fischer, J., Rodríguez, I., Schmoger, S., Friese, A., Roesler, U., Helmuth, R., et al. (2013). *Salmonella enterica* subsp. *enterica* producing VIM-1 carbapenemase isolated from livestock farms. *J. Antimicrob. Chemother.* 68, 478–480. doi: 10.1093/jac/dks393

Haenni, M., Chatre, P., Metayer, V., Bour, M., Signol, E., Madec, J. Y., et al. (2014). Comparative prevalence and characterization of ESBL-producing *Enterobacteriaceae* in dominant versus subdominant enteric flora in veal calves at slaughterhouse, France. *Vet. Microbiol.* 171, 321–327. doi: 10.1016/j.vetmic.2014.02.023

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

Huijbers, P. M., Graat, E. A., Haenen, A. P., van Santen, M. G., van Essen-Zandbergen, A., Mevius, D. J., et al. (2014). Extended-spectrum and AmpC beta-lactamase-producing *Escherichia coli* in broilers and people living and/or working on broiler farms: prevalence, risk factors and molecular characteristics. *J. Antimicrob. Chemother.* 69, 2669–2675. doi: 10.1093/jac/dku178

Hussain, A., Shaik, S., Ranjan, A., Suresh, A., Sarker, N., Semmler, T., et al. (2019). Genomic and functional characterization of poultry *Escherichia coli* from India revealed diverse extended-spectrum beta-lactamase-producing lineages with shared virulence profiles. *Front. Microbiol.* 10:2766. doi: 10.3389/fmicb.2019.02766

Irrgang, A., Hammerl, J. A., Falgenhauer, L., Guiral, E., Schmoger, S., Imirzalioglu, C., et al. (2018). Diversity of CTX-M-1-producing *E. coli* from German food samples and genetic diversity of the $bla_{\mathrm{CTX-M-1}}$ region on IncI1 ST3 plasmids. *Vet. Microbiol.* 221, 98–104. doi: 10.1016/j.vetmic.2018.06.003

Joensen, K. G., Tetzschner, A. M., Iguchi, A., Aarestrup, F. M., and Scheutz, F. (2015). Rapid and easy in silico serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *J. Clin. Microbiol.* 53, 2410–2426. doi: 10.1128/JCM.00008-15

Johnson, J. R., and Russo, T. A. (2005). Molecular epidemiology of extraintestinal pathogenic (uropathogenic) *Escherichia coli*. *Int. J. Med. Microbiol.* 295, 383–404. doi: 10.1016/j.ijmm.2005.07.005

Johnson, J. R., Murray, A. C., Gajewski, A., Sullivan, M., Snippes, P., Kuskowski, M. A., et al. (2003). Isolation and molecular characterization of nalidixic acid-resistant extraintestinal pathogenic *Escherichia coli* from retail chicken products. *Antimicrob. Agents Chemother.* 47, 2161–2168. doi: 10.1128/aac.47.7.2161-2168.2003

Johnson, T. J., Bielak, E. M., Fortini, D., Hansen, L. H., Hasman, H., Debroy, C., et al. (2012). Expansion of the IncX plasmid family for improved identification and typing of novel plasmids in drug-resistant *Enterobacteriaceae*. *Plasmid* 68, 43–50. doi: 10.1016/j.plasmid.2012.03.001

Johnson, T. J., Wannemuehler, Y., Doetkott, C., Johnson, S. J., Rosenberger, S. C., and Nolan, L. K. (2008). Identification of minimal predictors of avian pathogenic *Escherichia coli* virulence for use as a rapid diagnostic tool. *J. Clin. Microbiol.* 46, 3987–3996. doi: 10.1128/JCM.00816-08

Kaesbohrer, A., Bakran-Lebl, K., Irrgang, A., Fischer, J., Kampf, P., Schiffmann, A., et al. (2019). Diversity in prevalence and characteristics of ESBL/pAmpC producing *E. coli* in food in Germany. *Vet. Microbiol.* 233, 52–60. doi: 10.1016/j.vetmic.2019.03.025

Karch, H., Denamur, E., Dobrindt, U., Finlay, B. B., Hengge, R., Johannes, L., et al. (2012). The enemy within us: lessons from the 2011 European *Escherichia coli* O104:H4 outbreak. *EMBO Mol. Med.* 4, 841–848. doi: 10.1002/emmm.201201662

Koehler, C. D., and Dobrindt, U. (2011). What defines extraintestinal pathogenic *Escherichia coli*? *Int. J. Med. Microbiol.* 301, 642–647. doi: 10.1016/j.ijmm.2011.09.006

Labbe, G., Edirmanasinghe, R., Ziebell, K., Nash, J. H., Bekal, S., Parmley, E. J., et al. (2016). Complete genome and plasmid sequences of three Canadian isolates of *Salmonella enterica* subsp. *enterica* Serovar Heidelberg from human and food sources. *Genome Announc.* 4:e01526-15. doi: 10.1128/genomeA.01526-15

Lambrecht, E., Van Meervenne, E., Boon, N., Van de Wiele, T., Wattiau, P., Herman, L., et al. (2018). Characterization of cefotaxime- and ciprofloxacin-resistant commensal *Escherichia coli* originating from Belgian farm animals indicates high antibiotic resistance transfer rates. *Microb. Drug Resist.* 24, 707–717. doi: 10.1089/mdr.2017.0226

Larsen, M. V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R. L., et al. (2012). Multilocus sequence typing of total-genome-sequenced bacteria. *J. Clin. Microbiol.* 50, 1355–1361. doi: 10.1128/JCM.06094-11

Lekagul, A., Tangcharoensathien, V., and Yeung, S. (2018). The use of antimicrobials in global pig production: a systematic review of methods for quantification. *Prev. Vet. Med.* 160, 85–98. doi: 10.1016/j.prevetmed.2018.09.016

Liakopoulos, A., Mevius, D., and Ceccarelli, D. (2016). A review of SHV extended-spectrum beta-lactamases: neglected yet ubiquitous. *Front. Microbiol.* 7:1374. doi: 10.3389/fmicb.2016.01374

Liakopoulos, A., van der Goot, J., Bossers, A., Betts, J., Brouwer, M. S. M., Kant, A., et al. (2018). Genomic and functional characterisation of IncX3 plasmids encoding bla_{SHV-12} in *Escherichia coli* from human and animal origin. *Sci. Rep.* 8:7674. doi: 10.1038/s41598-018-26073-5

Liebana, E., Carattoli, A., Coque, T. M., Hasman, H., Magiorakos, A. P., Mevius, D., et al. (2013). Public health risks of enterobacterial isolates producing extended-spectrum beta-lactamases or AmpC beta-lactamases in food and food-producing animals: an EU perspective of epidemiology, analytical methods, risk factors, and control options. *Clin. Infect. Dis.* 56, 1030–1037. doi: 10.1093/cid/cis1043

Liu, G., Bogaj, K., Bortolaia, V., Olsen, J. E., and Thomsen, L. E. (2019). Antibiotic-induced, increased conjugative transfer is common to diverse naturally occurring ESBL plasmids in *Escherichia coli*. *Front. Microbiol.* 10:2119. doi: 10.3389/fmicb.2019.02119

Ludden, C., Raven, K. E., Jamrozy, D., Gouliouris, T., Blane, B., Coll, F., et al. (2019). One Health genomic surveillance of *Escherichia coli* demonstrates distinct lineages and mobile genetic elements in isolates from humans versus livestock. *mBio* 10:e02693-18. doi: 10.1128/mBio.02693-18

Madec, J. Y., Haenni, M., Nordmann, P., and Poirel, L. (2017). Extended-spectrum beta-lactamase / AmpC- and carbapenemase-producing *Enterobacteriaceae* in animals: a threat for humans? *Clin. Microbiol. Infect.* 23, 826–833. doi: 10.1016/j.cmi.2017.01.013

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

Manges, A. R., Geum, H. M., Guo, A., Edens, T. J., Fibke, C. D., and Pitout, J. D. D. (2019). Global extraintestinal pathogenic *Escherichia coli* (ExPEC) lineages. *Clin. Microbiol. Rev.* 32:e00135-18. doi: 10.1128/CMR.00135-18

Michael, G. B., Kaspar, H., Siqueira, A. K., de Freitas Costa, E., Corbellini, L. G., Kadlec, K., et al. (2017). Extended-spectrum beta-lactamase (ESBL)-producing *Escherichia coli* isolates collected from diseased food-producing animals in the GERM-Vet monitoring program 2008-2014. *Vet. Microbiol.* 200, 142–150. doi: 10.1016/j.vetmic.2016.08.023

Mueller, A., Stephan, R., and Nuesch-Inderbinen, M. (2016). Distribution of virulence factors in ESBL-producing *Escherichia coli* isolated from the environment, livestock, food and humans. *Sci. Total Environ.* 541, 667–672. doi: 10.1016/j.scitotenv.2015.09.135

Mulvey, M. R., Bryce, E., Boyd, D. A., Ofner-Agostini, M., Land, A. M., Simor, A. E., et al. (2005). Molecular characterization of cefoxitin-resistant *Escherichia coli* from Canadian hospitals. *Antimicrob. Agents Chemother.* 49, 358–365. doi: 10.1128/AAC.49.1.358-365.2005

Orlek, A., Stoesser, N., Anjum, M. F., Doumith, M., Ellington, M. J., Peto, T., et al. (2017). Plasmid classification in an era of whole-genome sequencing: application in studies of antibiotic resistance epidemiology. *Front. Microbiol.* 8:182. doi: 10.3389/fmicb.2017.00182

Palmeira, J. D., Haenni, M., Metayer, V., Madec, J. Y., and Ferreira, H. M. N. (2020). Epidemic spread of IncI1/pST113 plasmid carrying the extended-spectrum beta-lactamase (ESBL) $bla_{\mathrm{CTX-M-8}}$ gene in *Escherichia coli* of Brazilian cattle. *Vet. Microbiol.* 243:108629. doi: 10.1016/j.vetmic.2020.108629

Peter-Getzlaff, S., Polsfuss, S., Poledica, M., Hombach, M., Giger, J., Bottger, E. C., et al. (2011). Detection of AmpC beta-lactamase in *Escherichia coli*: comparison of three phenotypic confirmation assays and genetic analysis. *J. Clin. Microbiol.* 49, 2924–2932. doi: 10.1128/JCM.00091-11

Pietsch, M., Irrgang, A., Roschanski, N., Brenner Michael, G., Hamprecht, A., Rieber, H., et al. (2018). Whole genome analyses of CMY-2-producing *Escherichia coli* isolates from humans, animals and food in Germany. *BMC Genomics* 19:601. doi: 10.1186/s12864-018-4976-3

Pitout, J. D., and DeVinney, R. (2017). *Escherichia coli* ST131: a multidrug-resistant clone primed for global domination. *F1000Res.* 6:F1000FacultyRev–1195. doi: 10.12688/f1000research.10609.1

Riley, L. W. (2014). Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. *Clin. Microbiol. Infect.* 20, 380–390. doi: 10.1111/1469-0691.12646

Roth, N., Kaesbohrer, A., Mayrhofer, S., Zitz, U., Hofacre, C., and Domig, K. J. (2019). The application of antibiotics in broiler production and the resulting antibiotic resistance in *Escherichia coli:* a global overview. *Poult. Sci.* 98, 1791–1804. doi: 10.3382/ps/pey539

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

Santos, A. C. M., Santos, F. F., Silva, R. M., and Gomes, T. A. T. (2020). Diversity of hybrid- and hetero-pathogenic *Escherichia coli* and their potential implication in more severe diseases. *Front. Cell. Infect. Microbiol.* 10:339. doi: 10.3389/fcimb.2020.00339

Schaufler, K., Semmler, T., Wieler, L. H., Trott, D. J., Pitout, J., Peirano, G., et al. (2019). Genomic and functional analysis of emerging virulent and multidrug-resistant *Escherichia coli* lineage sequence type 648. *Antimicrob. Agents Chemother.* 63:e00243-19. doi: 10.1128/AAC.00243-19

Schaufler, K., Semmler, T., Wieler, L. H., Wohrmann, M., Baddam, R., Ahmed, N., et al. (2016). Clonal spread and interspecies transmission of clinically relevant ESBL-producing *Escherichia coli* of ST410–another successful pandemic clone? *FEMS Microbiol. Ecol.* 92:fiv155. doi: 10.1093/femsec/fiv155

Seiffert, S. N., Carattoli, A., Schwendener, S., Collaud, A., Endimiani, A., and Perreten, V. (2017). Plasmids carrying $bla_{CMY-2/4}$ in *Escherichia coli* from poultry, poultry meat, and humans belong to a novel IncK subgroup designated IncK2. *Front. Microbiol.* 8:407. doi: 10.3389/fmicb.2017.00407

Starcic Erjavec, M., and Zgur-Bertok, D. (2015). Virulence potential for extraintestinal infections among commensal *Escherichia coli* isolated from healthy humans–the Trojan horse within our gut. *FEMS Microbiol. Lett.* 362:fnu061. doi: 10.1093/femsle/fnu061

Touzain, F., Le Devendec, L., de Boisseson, C., Baron, S., Jouy, E., Perrin-Guyomard, A., et al. (2018). Characterization of plasmids harboring $bla_{\mathrm{CTX-M}}$ and bla_{CMY} genes in *E. coli* from French broilers. *PLoS One* 13:e0188768. doi: 10.1371/journal.pone.0188768

Valentin, L., Sharp, H., Hille, K., Seibt, U., Fischer, J., Pfeifer, Y., et al. (2014). Subgrouping of ESBL-producing *Escherichia coli* from animal and human sources: an approach to quantify the distribution of ESBL types between different reservoirs. *Int. J. Med. Microbiol.* 304, 805–816. doi: 10.1016/j.ijmm.2014.07.015

Wang, J., Stephan, R., Power, K., Yan, Q., Hachler, H., and Fanning, S. (2014). Nucleotide sequences of 16 transmissible plasmids identified in nine multidrug-resistant *Escherichia coli* isolates expressing an ESBL phenotype isolated from food-producing animals and healthy humans. *J. Antimicrob. Chemother.* 69, 2658–2668. doi: 10.1093/jac/dku206

WHO (2019). *Critically Important Antimicrobials for Human Medicine, 6th Revision*. Geneva: World Health Organization.

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

Zurfluh, K., Wang, J., Klumpp, J., Nuesch-Inderbinen, M., Fanning, S., and Stephan, R. (2014). Vertical transmission of highly similar $bla_{\mathrm{CTX-M-1}}$-harboring IncI1 plasmids in *Escherichia coli* with different MLST types in the poultry production pyramid. *Front. Microbiol.* 5:519. doi: 10.3389/fmicb.2014.00519

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

frontiers in Microbiology

ORIGINAL RESEARCH
published: 12 April 2021
doi: 10.3389/fmicb.2021.617349

Clonal CTX-M-15-Producing *Escherichia coli* ST-949 Are Present in German Surface Water

Linda Falgenhauer[1,2], Anja zur Nieden[1], Susanne Harpel[1], Jane Falgenhauer[2,3] and Eugen Domann[1,2,3]*

[1] *Institute of Hygiene and Environmental Medicine, Justus Liebig University Giessen, Giessen, Germany,* [2] *German Center for Infection Research (DZIF), Partner Site Giessen-Marburg-Langen, Justus Liebig University Giessen, Giessen, Germany,* [3] *Institute of Medical Microbiology, Justus Liebig University Giessen, Giessen, Germany*

Extended-spectrum beta-lactamase (ESBL)-producing bacterial isolates are emerging within the last years. To understand this emergence, a thorough genome-based analysis of ESBL isolates from different sources (One Health approach) is needed. Among these, analysis of surface water is underrepresented. Therefore, we performed a genome-based analysis of ESBL-producing *Escherichia coli* isolates from surface water samples. Water samples were collected from eleven different surface water sites (lakes, river). ESBL-producing *E. coli* were recovered from these samples using filters and chromogenic media. Whole-genome sequencing of ESBL-producing *E. coli* was performed followed by determination of the multilocus sequence type (ST), ESBL-type, and virulence genes. Phylogenetic analysis was done using single nucleotide analysis. From all water samples taken, nineteen ESBL-producing *E. coli* were recovered. All of them harbored an ESBL gene. Nine different multilocus STs were determined, among which ST-949 was the ST detected most frequently. Phylogenetic analysis of ST-949 isolates revealed that all those isolates were closely related. In addition, they harbored an identical chromosomal insertion of $bla_{CTX-M-15}$, indicating a clonal relationship among these isolates. Genetic comparison with isolates from all over the world revealed that these isolates were closely related to human clinical isolates derived from New Zealand and Sweden. An ESBL-producing *E. coli* ST-949 clone was detected in German surface waters. Its close relationship to human clinical isolates suggests its ability to colonize or even infect humans. Our findings reveal that water sources indeed may play a hitherto underreported role in spread of ESBL-producing isolates.

Keywords: CTX-M-15, ST-949, ESBL-*E. coli*, water samples, WGS

OPEN ACCESS

Edited by:
Edward Feil,
University of Bath, United Kingdom

Reviewed by:
Jean-Yves Madec,
Agence Nationale de Sécurité Sanitaire de l'Alimentation, de l'Environnement et du Travail (ANSES), France
Ákos Tóth,
National Public Health Center, Hungary

****Correspondence:***
Linda Falgenhauer
Linda.falgenhauer@mikrobio.med.uni-giessen.de

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

Received: *14 October 2020*
Accepted: *24 March 2021*
Published: *12 April 2021*

Citation:
Falgenhauer L, zur Nieden A, Harpel S, Falgenhauer J and Domann E (2021) Clonal CTX-M-15-Producing Escherichia coli ST-949 Are Present in German Surface Water. Front. Microbiol. 12:617349. doi: 10.3389/fmicb.2021.617349

INTRODUCTION

Extended-spectrum beta-lactamase (ESBL)-producing bacterial isolates are emerging in the last years (Peirano and Pitout, 2019). The spread of ESBL-producers is a clear One Health issue, as they have been found to be present in different sources, animals, humans, and environment (Hooban et al., 2020). This is true for Germany as well. In Germany, 6.3% of humans are colonized with ESBL-producing *Escherichia coli* isolates (Valenza et al., 2014). In diseased food-producing animals,

the prevalence of ESBL-producing *E. coli* ranges between 0.8 and 11.2% depending on the animal species (Michael et al., 2017).

The commonly accepted opinion is that all different sources play a role in the spread of ESBL-producing bacteria. To be able to track the transfer routes of ESBL-producers among different sources, a thorough understanding of the epidemiology of these bacteria is needed. The method of choice to perform an in-depth epidemiological analysis is to use whole genome sequence-based methods. They have been used a lot in human and veterinary medicine (in particular to track outbreaks), but genome-based data from water sources are still very rare. Few epidemiological studies have been performed to analyze the genomes of ESBL-producing isolates from water samples. These studies showed a high identity between ESBL producers from water samples and clinical samples indicating a spread from either clinical to water sources or vice versa (Fagerström et al., 2019).

In order to gain more insight into this topic, an investigation was performed that included water samples from official and unofficial bathing sites at lakes and a river in Hesse, Germany.

MATERIALS AND METHODS

Sampling Procedure

During the bathing season 2018, samples were taken from swimming lakes in Hesse (n = 10). According to the European Bathing Water Directive (BWD; EG 2006/7) the sites were checked at least monthly for the presence of coliform bacteria. Additionally, samples from unofficial bathing sites of the Hessian river Lahn around Marburg and Giessen were taken (n = 9). Procedures for sampling as well as preparation, filtration, and enumeration were performed conforming with the DIN EN ISO 9308-2 (K6-1) 07-2014, DIN EN ISO 19458 (K19), DIN EN ISO 8199 (K20) 01-2008, and DIN EN ISO 9380-1: 2014 (K12) regulations within 24 h.

Characterization of ESBL-Producing Isolates

For detection of ESBL-producing isolates, water samples were filtered and the filters put onto Brilliance^{TM} ESBL chromogenic medium (OXOID, Wesel, Germany). For isolates growing on the chromogenic agar, species confirmation was performed using MALDI-TOF-MS (Biomérieux, Nürtingen, Germany). Antibiotic susceptibility testing and ESBL phenotype confirmation was performed using the VITEK 2 System (AST-N263 cards, Biomérieux, Nürtingen, Germany). Classification of the antibiotic resistance/susceptibility was performed according to EUCAST criteria[1].

Whole-Genome Sequencing

Short-read whole genome sequencing was performed for all *E. coli* isolates growing on the chromogenic medium (n = 21). DNA from overnight cultures was isolated using the Purelink genomic DNA kit (ThermoFisher, Dreieich, Germany). Short read sequencing was performed on a NextSeq 500 machine (Illumina, Eindhoven, Netherlands) using a Nextera XT sequencing library with an average read length of 115 nt and an average coverage of 33.5 x. Raw reads were processed using the ASA^{3}P pipeline using default parameters (Schwengers et al., 2020).

Long-read sequencing of a representative *E. coli* ST-949 isolate (EDCC5518) was performed using the Nanopore technology. The library was prepared using the native barcoding kit (EXP-NBD103, Oxford Nanopore Technologies Ltd., Oxford, United Kingdom) and 1D chemistry (SQK-LSK108). Sequencing was performed using the SpotON Flow Cell Mk I R9 Version (FLO-MIN106) on a MinION/MinIT machine with an average read length of 4,137 nt. Basecalling was performed directly on the MinIT machine. Demultiplexing was performed using Porechop (v. 0.2.3[2]). Hybrid assembly was performed using Unicycler (v. 0.4.7) (Wick et al., 2017) and the short and long reads with default parameters.

Genome-Based Analyses

In silico multilocus sequence typing of *E. coli* isolates was performed using the scheme presented by Wirth et al. (2006). Antibiotic resistance genes, plasmid incompatibility groups and *fimH* types were determined using the bacterial analysis pipeline of the Center for Genomic Epidemiology[3]. Insertion elements were determined using ISFinder (Siguier et al., 2006). Virulence gene determination was performed using ASA^{3}P (Schwengers et al., 2020). Comparative genome analysis was performed using the HarvestSuite package (Treangen et al., 2014). Publicly available assembled *E. coli* genomes of the multilocus sequence type (ST) ST-949 were downloaded using Enterobase (as of 8th June 2020, **Supplementary Table 1**) (Zhou et al., 2020). Geographical representation of sampling sites was visualized using MicroReact (Argimón et al., 2016).

RESULTS AND DISCUSSION

Detection and Phenotypic Characterization of ESBL-Producing *Escherichia coli* Samples

During the bathing season of 2018 (June–August), fifty-five samples from nineteen sampling sites were collected. Of these samples, forty-four did not show growth of isolates on ESBL chromogenic agar. Notably, the 2018 summer was a comparatively hot summer[4] resulting in low water levels. From the remaining water samples (n = 11, **Figure 1**), nineteen ESBL-producing bacterial isolates were detected (**Table 1**). Only *E. coli* isolates were detected. Environmental data and characterization of the sampling sites are shown in **Table 1**.

For all *E. coli* isolates growing on the chromogenic plates, the ESBL phenotype was confirmed. Phenotypic

[1]https://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_11.0_Breakpoint_Tables.pdf

[2]https://github.com/rrwick/Porechop

[3]http://www.genomicepidemiology.org/

[4]http://www.dwd.de and http://www.wetter.de

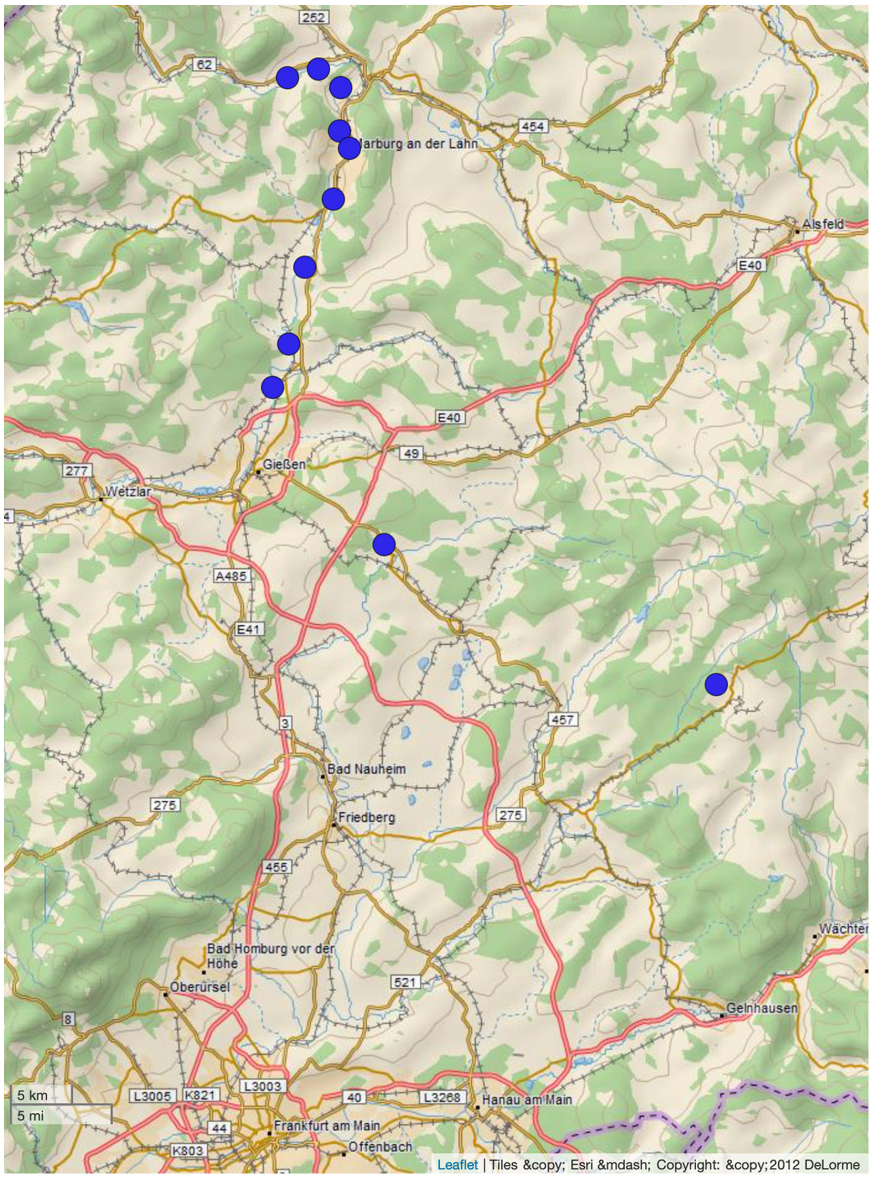


FIGURE 1 | Distribution of sampling sites. The figure was generated using microreact (Argimón et al., 2016).

resistance to antibiotics other than beta-lactams was detected very seldom and included resistance to fluoroquinolones (4/21) and trimethoprim/sulfamethoxazole (5/21) (**Supplementary Table 2**). The isolates were not resistant to carbapenems. According to the classification proposed by Magiorakos et al. (2012), all isolates were multidrug-resistant (resistant to ≥3 different antibiotic classes; **Supplementary Table 2**).

Genome-Based Analysis of *Escherichia coli* Isolates

All *E. coli* isolates harbored an ESBL gene (**Table 2**). The most common ESBL gene detected was $bla_{CTX-M-15}$ (n = 17) followed by $bla_{CTX-M-1}$ (n = 2) and $bla_{CTX-M-27}$ (n = 2). The predominance of $bla_{CTX-M-15}$ in our study is concordant with the results in other studies performed in Europe (Kittinger et al., 2016; Jorgensen et al., 2017). The *E. coli* isolates encoded other antibiotic resistance genes conferring resistance to aminoglycosides (7/21), fluoroquinolones (15/21, *qnrS1*), sulfonamide (5/21), trimethoprim (5/21), and tetracycline (2/21) (**Table 2**). In concordance with previous reports (Rodríguez-Martínez et al., 2011), the presence of *qnrS1* did not lead to high-level fluoroquinolone resistance (MIC > 0.5 mg/L) in our isolates.

Multilocus sequence typing revealed that nine different STs were present (**Table 2** and **Figure 2**). Of these, three were detected

more than once: ST-949 (n = 11), ST-131 (n = 2), and ST-1431 (n = 2). *E. coli* ST-949 and ST-1431 isolates harbored $bla_{CTX-M-15}$, while *E. coli* ST-131 isolates harbored $bla_{CTX-M-15}$ or $bla_{CTX-M-27}$.

To our knowledge, *E. coli* ST-949 have been reported in only five publications worldwide, indicating that this ST is less frequent and might represent an emerging clone (Oh et al., 2014; Potron et al., 2017; Potel et al., 2018; Fagerström et al., 2019; Sedrati et al., 2020). The total number of publicly available *E. coli* ST-949 isolates in the Enterobase database is 41 [as of 11th August 2020 (Zhou et al., 2020)], a very low number compared with frequent multilocus STs as e.g., ST-131 (n = 9202, as of 11th August 2020).

ESBL-producing ST-1431 *E. coli* isolates have been detected more often in animal sources (livestock, pets, wild animals) than in humans (Rocha-Gracia et al., 2015; Bachiri et al., 2017; Seiffert et al., 2017).

In this study, we detected two ST131 isolates. *E. coli* ST-131 are frequently associated with human clinical infections (Nicolas-Chanoine et al., 2014), in particular those depicting the *fim*H type H30 and harboring CTX-M-15 or CTX-M-27 (Nicolas-Chanoine et al., 2014; Stoesser et al., 2016). EDCC5529 depicted the fimH41 *fim*H-type and harbored $bla_{CTX-M-15}$. EDCC5535 depicted a fimH30 *fim*H type and characteristic properties of the ST-131 C1-M27 clade (Matsumura et al., 2016): $bla_{CTX-M-27}$, the GyrA S83L/D87N and ParC S80I/E84V mutations leading to fluoroquinolone resistance and the M27PP1 phage. Therefore, it is a member of the C1-M27 clade usually associated with human isolates (Matsumura et al., 2016; Ghosh et al., 2017). Thus, EDCC5535 might have originated from human sources.

Deeper Analysis of *Escherichia coli* ST-949 Isolates

The most common ST within the ESBL *E. coli* was ST-949 (**Table 2** and **Figure 2**). Therefore, we analyzed these isolates in more detail. *E. coli* ST-949 is known to be associated with carbapenem-resistance (Potron et al., 2017; Potel et al., 2018) or ETEC pathotypes (Oh et al., 2014). They have been isolated from environmental samples (water samples) collected in Sweden, where the authors could show that the water isolates were highly related with isolates derived from a hospital that was adjoining the water source (Fagerström et al., 2019).

Because *E. coli* ST-949 are known to be pathogenic (Oh et al., 2014), we analyzed all available *E. coli* ST-949 isolates for the presence of virulence genes. The ST-949 isolates from this study harbored only ExPEC virulence genes (e.g., iron acquisition genes, Enterobactin, **Supplementary Figure 2**). The isolates detected in New Zealand and Sweden harbored the same sets of virulence genes. Other ST-949 harbored also toxins (Shigatoxin) and hemolysins indicating that ST-949 isolates differ widely in their virulence capabilities.

A whole-genome-based analysis of the *E. coli* ST-949 isolates from this study and those from Enterobase revealed two different findings (**Figure 3** and **Supplementary Table 1**): Firstly, ST-949 isolates are divided into two different clusters. Cluster A (including our isolates) consists of isolates found in water,

TABLE 1 | Environmental data and enumeration results of sampling sites with ESBL-positive samples.

Sampling site #	Isolate #	Species	Site	*E. coli** [CFU/100 ml]	*Enterobacter** [CFU/100 ml]	Temperature air [°C]	Temperature water [°C]	Sampling date	Sampling time
1	EDCC5518	*Escherichia coli*	Bathing lake	77	15	21	20	19.06.18	09:20
	EDCC5519	*Escherichia coli*							
	EDCC5520	*Escherichia coli*							
	EDCC5522	*Escherichia coli*		15	46	23	24	17.07.18	09:00
2	EDCC5521	*Escherichia coli*	Bathing lake	<15	15	23	24	16.07.18	09:30
6	EDCC5523	*Escherichia coli*	Bathing lake	109	161	28	27	07.08.18	10:00
	EDCC5524	*Escherichia coli*							
10	EDCC5525	*Escherichia coli*	River	<15	<15	27	21	08.08.18	9:41
11	EDCC5526	*Escherichia coli*	River	30	<15	28	22	08.08.18	10:21
13	EDCC5527	*Escherichia coli*	River	1,509	144	27	19	08.08.18	11:38
14	EDCC5528	*Escherichia coli*	River	1,749	94	25	19	08.08.18	12:09
	EDCC5529	*Escherichia coli*							
	EDCC5530	*Escherichia coli*							
15	EDCC5531	*Escherichia coli*	River	5,352	640	28	23	08.08.18	12:31
16	EDCC5532	*Escherichia coli*	River	1,931	197	32	23	08.08.18	12:58
	EDCC5533	*Escherichia coli*		1,931	197	32	23	08.08.18	12:58
17	EDCC5534	*Escherichia coli*	River	110	<15	32	24	08.08.18	13:36
	EDCC5535	*Escherichia coli*		110	<15	32	24	08.08.18	13:36
19	EDCC5536	*Escherichia coli*	River	77	<15	32	24	08.08.18	15:53
	EDCC5537	*Escherichia coli*							
	EDCC5538	*Escherichia coli*							

**Enumeration performed based on DIN EN ISO 9308-2 (K6-1) 07-2014 regulation; EDCC, ED culture collection.*

TABLE 2 | Results of the genome-based analysis of the ESBL-producing *E. coli* isolates.

Isolate	ST	Aminoglycoside	Beta-lactam	Macrolide	Phenicol	Quinolone	Sulfonamide	Tetracycline	Trimethoprim	*fimH* type	Plasmid incompatibility groups
EDCC5518	949		$bla_{CTX-M-15}$			*qnrS1*				H121	IncI1 and p0111
EDCC5519	949		$bla_{CTX-M-15}$			*qnrS1*				H121	IncI1
EDCC5520	949		$bla_{CTX-M-15}$			*qnrS1*				H121	IncI1 and p0111
EDCC5521	3,246		$bla_{CTX-M-1}$							H65	IncFIA, IncI1, IncFIB (AP001918), IncFII(29), and ColRNAI
EDCC5522	949		$bla_{CTX-M-15}$			*qnrS1*				H121	IncI1 and IncA/C2
EDCC5523	949		$bla_{CTX-M-15}$			*qnrS1*				H121	IncI1
EDCC5524	949		$bla_{CTX-M-15}$			*qnrS1*				H121	IncI1
EDCC5525	949		$bla_{CTX-M-15}$			*qnrS1*				H121	IncI1
EDCC5526	949		$bla_{CTX-M-15}$			*qnrS1*				H121	IncI1
EDCC5527	1,490	*aadA1*	$bla_{CTX-M-15}$			*qnrS1*			*dfrA1*	H25	IncFII, IncFIB (AP001918), IncFII (pCoo), and IncB/O/K/Z
EDCC5528	2,797		$bla_{CTX-M-15}$			*qnrS1*				H54	IncFII (pHN7A8) and IncB/O/K/Z
EDCC5529	131	*strA, strB*	$bla_{CTX-M-15}$, bla_{TEM-1B}	*mph(A)*			*sul2*	*tet(A)*		H41	IncFII(29), IncFIB (AP001918), and Col156
EDCC5530	155	*aadA5*	$bla_{CTX-M-1}$				*sul2*		*dfrA17*	N.D.	IncI1 and IncFII(pCoo)

(Continued)

TABLE 2 | Continued

Isolate	ST	Aminoglycoside	Beta-lactam	Macrolide	Phenicol	Quinolone	Sulfonamide	Tetracycline	Trimethoprim	*fimH* type	Plasmid incompatibility groups
EDCC5531	226	*aadA1*	$bla_{CTX-M-15}$, bla_{OXA-1}		*catA1*			*tet(B)*		H41	IncFII, ColRNAI, and Col(MG828)
EDCC5532	949		$bla_{CTX-M-15}$			*qnrS1*				H121	IncI1
EDCC5533	949		$bla_{CTX-M-15}$			*qnrS1*				H121	IncI1, Col8282, ColRNAI, Col156, and Col(MG828)
EDCC5534	949		$bla_{CTX-M-15}$			*qnrS1*				H121	IncI1
EDCC5535	131		$bla_{CTX-M-27}$							H30	IncFII (pRSB107), IncFIA, IncFIB(AP001918), Col8282, Col156, and Col(MG828)
EDCC5536	295	*aadA5*	$bla_{CTX-M-27}$, bla_{TEM-1B}	*mph(A)*			*sul1*		*dfrA17*	H54	IncFII (pRSB107), IncFIB (AP001918), IncFII (pCoo), IncY, and ColRNAI
EDCC5537	1,431	*aadA1, aadA2, strA, strB*	$bla_{CTX-M-15}$, bla_{TEM-1B}		*cmlA1*	*qnrS1*	*sul2, sul3*		*dfrA12*	H32	IncI1, IncX1, IncY, and Col156
EDCC5538	1,431	*aadA1,aadA2, strA, strB*	$bla_{CTX-M-15}$, bla_{TEM-1B}		*cmlA1*	*qnrS1*	*sul2, sul3*		*dfrA12*	H32	IncI1, IncX1, IncY, and Col156

EDCC, ED culture collection; N.D., not detected.

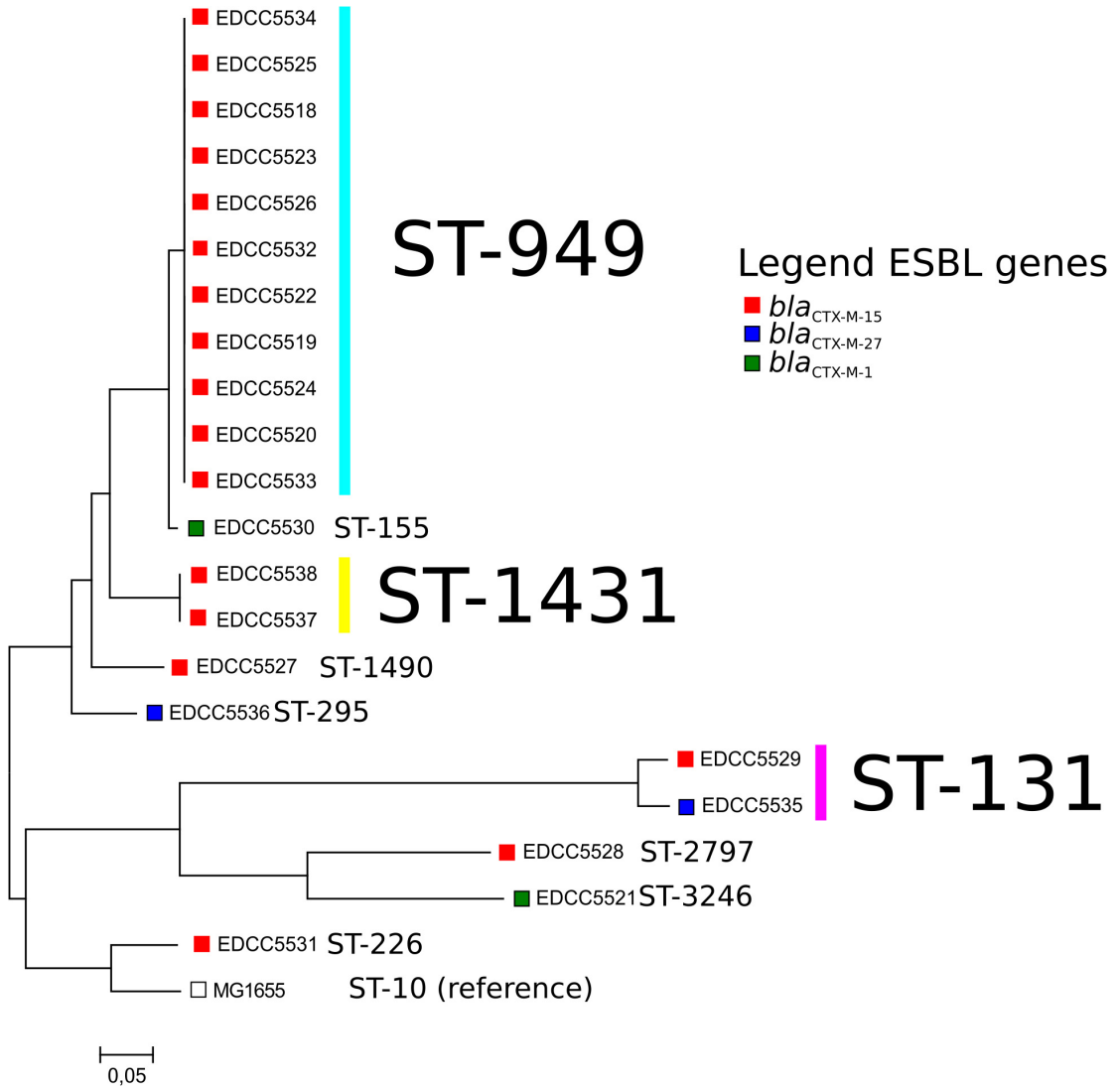


FIGURE 2 | Core-genome-based comparison of the *E. coli* isolates analyzed in this study. Boxes depict the ESBL gene present in these isolates. For a representation of the isolate geographical location and the phylogeny, please see the following link: https://microreact.org/project/bELapwXL1Fi8ZdznzZHF1H.

human, and livestock samples, while cluster B includes also isolates from companion animals. Secondly, only cluster A isolates contain CTX-M-15, while Cluster B isolates harbor either CTX-M alleles other than CTX-M-15, or AmpC beta-lactamases. Cluster B harbored two ST-949 isolates from Germany from livestock and an unknown source.

The *E. coli* ST-949 isolates from this study (Cluster A, n = 11) were highly related to ST-949 isolates found in New Zealand (ESC_FB9670AA_AS, human isolate) and Sweden (n = 4, water and human isolates, **Figure 3**). All these 16 isolates harbored a complex antibiotic resistance gene region including not only the $bla_{CTX-M-15}$ gene, but also the fluoroquinolone resistance gene *qnrS1* and several different insertion sequences (**Supplementary Figure 3**). The antibiotic resistance region was inserted in the chromosome at an identical location in the isolates from this study and the isolates from New Zealand, while the isolates from Sweden harbored the identical region inserted at a different location of the chromosome. This finding indicates that the acquisition of $bla_{CTX-M-15}$ in the two different clones was presumably from a different source and was independent in both clones.

The epidemiological link between Germany and New Zealand is not clear. It may indicate that the ST-949 clone found in Germany is present worldwide, but this is only an assumption as the total number of ST-949 isolates throughout the world is very low. It remains to be clarified whether ST-949 is an emerging ST and whether it is present in other sources.

The epidemiological link between ST-949 isolates from our study is partly explainable. All ST-949 river isolates originate from the same river (sampling sites 10, 11, 16, and 17), indicating a common source of contamination. Possible sources

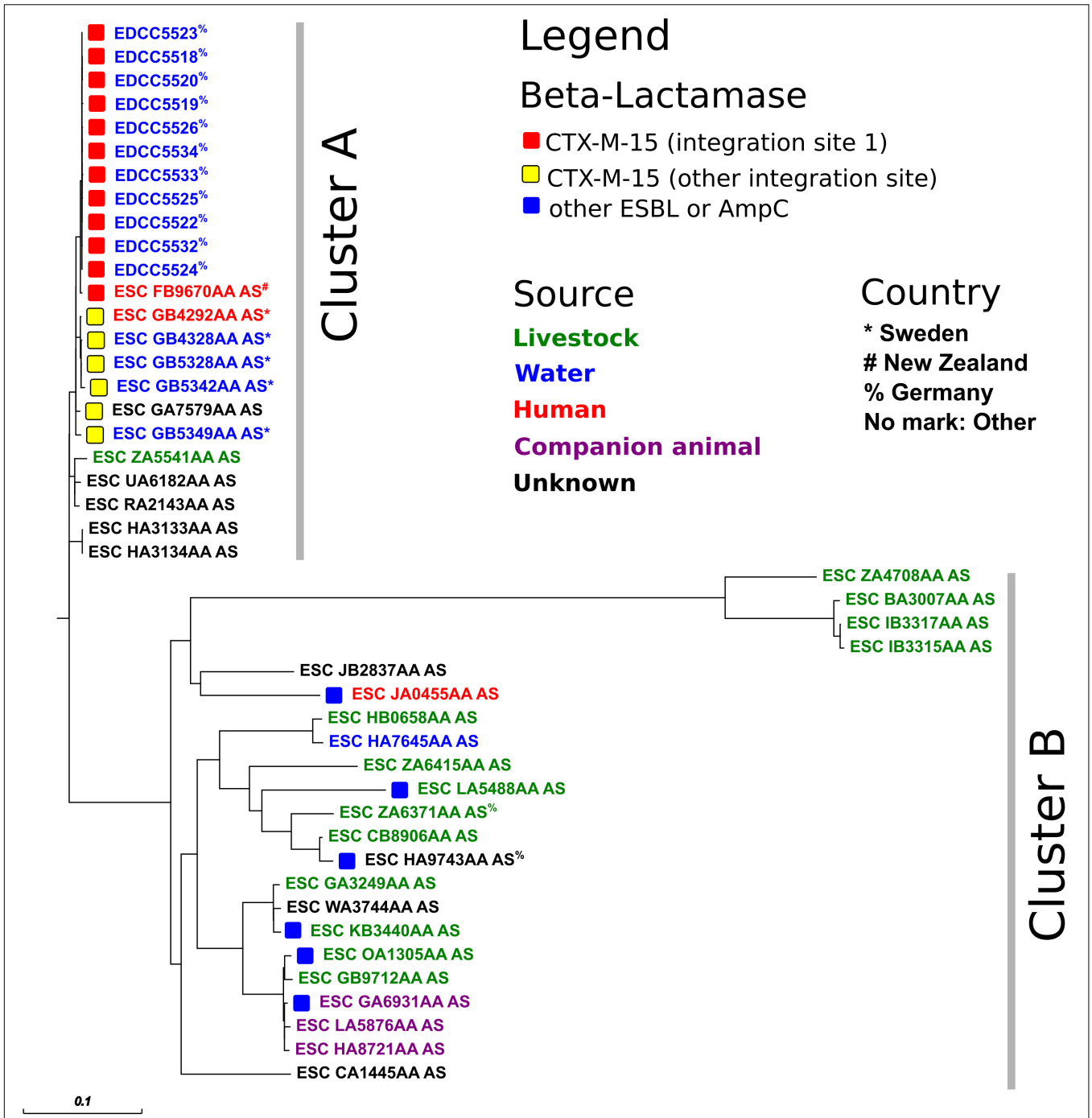


FIGURE 3 | Core-genome-based phylogenetic tree of *E. coli* ST-949 isolates from this study and Enterobase (as of 8th June 2020). For a representation of the isolate sampling site, please see **Supplementary Figure 1**.

of contamination along the river might be either agriculture, two large university hospitals whose cleared wastewater end up in the river itself or human influence through tourism, as the river is frequently used for recreational purposes. Sampling site 6 is located close to the sampled river, indicating a possible contamination through the river by flooding. What is not completely clear, is the epidemiological link between sampling site 1 and the other sites. They are not interconnected by any water flows. A possible connection between those might have been movement of humans or animals (in particular birds).

ST-949 isolates have never been reported in Germany. Therefore, this is the first study detecting ST-949 *E. coli* in Germany. Its predominance in our study indicates that either ST-949 *E. coli* might resemble *E. coli* isolates only present in water

sources or a new emerging multilocus ST in Germany. To prove this hypothesis, more studies are required.

CONCLUSION

In this study, we characterized ESBL-producing *E. coli* isolates from water samples. Our results show that the main MLST type is ST-949, reported in only a few number of very recent publications. In addition, it has been associated with human disease. This indicates that it might be an emerging ST with human pathogenic potential that could spread through water sources.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA656216.

AUTHOR CONTRIBUTIONS

ED designed the study. AN and SH collected samples and data. ED performed antibiotic resistance determination. LF, AN, JF, and ED analyzed the data. LF, AN, and ED wrote the manuscript that was critically reviewed and approved by all authors.

FUNDING

This study has been funded through the Hessian Competence Center for Clinical Hygiene (HuKKH) funded by the Hessian Ministry for Education and Research and the German Center for Infection Research (Grant # 8032808811) funded by the Federal Ministry of Education and Research.

ACKNOWLEDGMENTS

We would like to thank Silke Zechel-Gran and Christina Gerstmann for excellent technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.617349/full#supplementary-material

REFERENCES

Argimón, S., Abudahab, K., Goater, R. J. E., Fedosejev, A., Bhai, J., Glasner, C., et al. (2016). Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb. Genomics* 2:e000093. doi: 10.1099/mgen.0.000093

Bachiri, T., Bakour, S., Ladjouzi, R., Thongpan, L., Rolain, J. M., and Touati, A. (2017). High rates of CTX-M-15-producing *Escherichia coli* and *Klebsiella pneumoniae* in wild boars and barbary macaques in Algeria. *J. Glob. Antimicrob. Resist.* 8, 35–40. doi: 10.1016/j.jgar.2016.10.005

Fagerström, A., Mölling, P., Khan, F. A., Sundqvist, M., Jass, J., and Söderquist, B. (2019). Comparative distribution of extended-spectrum beta-lactamase–producing *Escherichia coli* from urine infections and environmental waters. *PLoS One* 14:e0224861. doi: 10.1371/journal.pone.0224861

Ghosh, H., Doijad, S., Falgenhauer, L., Fritzenwanker, M., Imirzalioglu, C., and Chakraborty, T. (2017). bla$_{CTX-M-27}$–Encoding *Escherichia coli* sequence type 131 lineage C1-m27 clone in clinical isolates, Germany. *Emerg. Infect. Dis.* 23, 1754–1756. doi: 10.3201/eid2310.170938

Hooban, B., Joyce, A., Fitzhenry, K., Chique, C., and Morris, D. (2020). The role of the natural aquatic environment in the dissemination of extended spectrum beta-lactamase and carbapenemase encoding genes: a scoping review. *Water Res.* 180, 1–12. doi: 10.1016/j.watres.2020.115880

Jorgensen, S. B., Soraas, A. V., Arnesen, L. S., Leegaard, T. M., Sundsfjord, A., and Jenum, P. A. (2017). A comparison of extended spectrum beta-lactamase producing *Escherichia coli* from clinical, recreational water and wastewater samples associated in time and location. *PLoS One* 12:e0186576. doi: 10.1371/journal.pone.0186576

Kittinger, C., Lipp, M., Folli, B., Kirschner, A., Baumert, R., Galler, H., et al. (2016). *Enterobacteriaceae* isolated from the river danube: antibiotic resistances, with a focus on the presence of ESBL and carbapenemases. *PLoS One* 11:e0165820. doi: 10.1371/journal.pone.0165820

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

Matsumura, Y., Pitout, J. D. D., Gomi, R., Matsuda, T., Noguchi, T., Yamamoto, M., et al. (2016). Global *Escherichia coli* sequence type 131 clade with bla$_{CTX-M-27}$ gene. *Emerg. Infect. Dis.* 22, 1900–1907. doi: 10.3201/eid2211.160519

Michael, G. B., Kaspar, H., Siqueira, A. K., de Freitas Costa, E., Corbellini, L. G., Kadlec, K., et al. (2017). Extended-spectrum β-lactamase (ESBL)-producing *Escherichia coli* isolates collected from diseased food-producing animals in the GERM-Vet monitoring program 2008–2014. *Vet. Microbiol.* 200, 142–150. doi: 10.1016/j.vetmic.2016.08.023

Nicolas-Chanoine, M.-H., Bertrand, X., and Madec, J.-Y. (2014). *Escherichia coli* ST131, an intriguing clonal group. *Clin. Microbiol. Rev.* 27, 543–574. doi: 10.1128/CMR.00125-13

Oh, K. H., Kim, D. W., Jung, S. M., and Cho, S. H. (2014). Molecular characterization of enterotoxigenic *Escherichia coli* strains isolated from diarrheal patients in Korea during 2003-2011. *PLoS One* 9:e96896. doi: 10.1371/journal.pone.0096896

Peirano, G., and Pitout, J. D. D. (2019). Extended-spectrum β−Lactamase-producing *Enterobacteriaceae*: update on molecular epidemiology and treatment options. *Drugs* 79, 1529–1541. doi: 10.1007/s40265-019-01180-3

Potel, C., Ortega, A., Martinez-Lamas, L., Bautist, V., Regueiro, B., and Oteo, J. (2018). Interspecies transmission of the bla$_{OXA-48}$ gene from a *Klebsiella pneumoniae* high-risk clone of sequence type 147 to different *Escherichia coli* clones in the gut microbiota. *Antimicrob. Agents Chemother.* 62, 2017–2019. doi: 10.1128/AAC.01699-17

Potron, A., Bernabeu, S., Cuzon, G., Pontiès, V., Blanchard, H., Seringe, E., et al. (2017). Analysis of OXA-204 carbapenemase-producing *Enterobacteriaceae* reveals possible endoscopyassociated transmission, France, 2012 to 2014. *Eurosurveillance* 22, 1–8. doi: 10.2807/1560-7917.ES.2017.22.49.17-00048

Rocha-Gracia, R. C., Cortés-Cortés, G., Lozano-Zarain, P., Bello, F., Martínez-Laguna, Y., and Torres, C. (2015). Faecal *Escherichia coli* isolates from healthy dogs harbour CTX-M-15 and CMY-2 β-lactamases. *Vet. J.* 203, 315–319. doi: 10.1016/j.tvjl.2014.12.026

Rodríguez-Martínez, J. M., Cano, M. E., Velasco, C., Martínez-Martínez, L., and Pascual, Á (2011). Plasmid-mediated quinolone resistance: an update. *J. Infect. Chemother.* 17, 149–182. doi: 10.1007/s10156-010-0120-2

Schwengers, O., Hoek, A., Fritzenwanker, M., Falgenhauer, L., Hain, T., Chakraborty, T., et al. (2020). ASA^{3}P: an automatic and scalable pipeline for the assembly, annotation and higher level analysis of closely related bacterial isolates. *PLoS Comput. Biol.* 16:e1007134. doi: 10.1371/journal.pcbi.1007134

Sedrati, T., Menoueri, M. N., Tennah, S., Ngaiganam, E. P., Azzi, O., Chadi, H., et al. (2020). Molecular characterization of multidrug resistant *Escherichia coli* isolated from milk of dairy cows with clinical mastitis in Algeria. *J. Food Prot.* 83, 2173–2178. doi: 10.4315/JFP-20-198

Seiffert, S. N., Carattoli, A., Schwendener, S., Collaud, A., Endimiani, A., and Perreten, V. (2017). Plasmids Carrying bla$_{(CMY-2/4)}$ in *Escherichia coli* from poultry, poultry meat, and humans belong to a novel IncK subgroup designated IncK2. *Front. Microbiol.* 8:407. doi: 10.3389/fmicb.2017.00407

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

Stoesser, N., Sheppard, A. E., Pankhurst, L., De Maio, N., Moore, C. E., Sebra, R., et al. (2016). Evolutionary history of the global emergence of the *Escherichia coli* epidemic clone ST131. *MBio* 7:e2162. doi: 10.1128/mBio.02162-15

Treangen, T. J., Ondov, B. D., Koren, S., and Phillippy, A. M. (2014). The harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* 15:524. doi: 10.1186/s13059-014-0524-x

Valenza, G., Nickel, S., Pfeifer, Y., Eller, C., Krupa, E., Lehner-Reindl, V., et al. (2014). Extended-spectrum-β-lactamase-producing *Escherichia coli* as intestinal colonizers in the German community. *Antimicrob. Agents Chemother.* 58, 1228–1230. doi: 10.1128/AAC.01993-13

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

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L. H., et al. (2006). Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol. Microbiol.* 60, 1136–1151. doi: 10.1111/j.1365-2958.2006.05172.x

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

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

frontiers in Microbiology

ORIGINAL RESEARCH
published: 14 May 2021
doi: 10.3389/fmicb.2021.662575

Circulation of Extended-Spectrum Beta-Lactamase-Producing *Escherichia coli* of Pandemic Sequence Types 131, 648, and 410 Among Hospitalized Patients, Caregivers, and the Community in Rwanda

*Elias Eger[1], Stefan E. Heiden[1], Katja Korolew[2], Claude Bayingana[3], Jules M. Ndoli[3,4], Augustin Sendegeya[3,4], Jean Bosco Gahutu[3,4†], Mathis S. E. Kurz[5], Frank P. Mockenhaupt[5], Julia Müller[1], Stefan Simm[2] and Katharina Schaufler[1]**

[1] Pharmaceutical Microbiology, Institute of Pharmacy, University of Greifswald, Greifswald, Germany, [2] Institute of Bioinformatics, University Medicine Greifswald, Greifswald, Germany, [3] College of Medicine and Health Sciences, University of Rwanda, Kigali, Rwanda, [4] University Teaching Hospital of Butare, Butare, Rwanda, [5] Institute of Tropical Medicine and International Health, Charité Medical University of Berlin, Berlin, Germany

OPEN ACCESS

Edited by:
Guido Werner,
Robert Koch Institute (RKI), Germany

Reviewed by:
Tim Downing,
Dublin City University, Ireland
Torsten Semmler,
Robert Koch Institute (RKI), Germany

***Correspondence:**
Katharina Schaufler
katharina.schaufler@uni-greifswald.de

[†] Deceased

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

Received: 01 February 2021
Accepted: 20 April 2021
Published: 14 May 2021

Citation:
Eger E, Heiden SE, Korolew K, Bayingana C, Ndoli JM, Sendegeya A, Gahutu JB, Kurz MSE, Mockenhaupt FP, Müller J, Simm S and Schaufler K (2021) Circulation of Extended-Spectrum Beta-Lactamase-Producing Escherichia coli of Pandemic Sequence Types 131, 648, and 410 Among Hospitalized Patients, Caregivers, and the Community in Rwanda. Front. Microbiol. 12:662575. doi: 10.3389/fmicb.2021.662575

Multi-drug resistant (MDR), gram-negative *Enterobacteriaceae*, such as *Escherichia coli* (*E. coli*) limit therapeutic options and increase morbidity, mortality, and treatment costs worldwide. They pose a serious burden on healthcare systems, especially in developing countries like Rwanda. Several studies have shown the effects caused by the global spread of extended-spectrum beta-lactamase (ESBL)-producing *E. coli*. However, limited data is available on transmission dynamics of these pathogens and the mobile elements they carry in the context of clinical and community locations in Sub-Saharan Africa. Here, we examined 120 ESBL-producing *E. coli* strains from patients hospitalized in the University Teaching Hospital of Butare (Rwanda), their attending caregivers as well as associated community members and livestock. Based on whole-genome analysis, the genetic diversification and phylogenetics were assessed. Moreover, the content of carried plasmids was characterized and investigated for putative transmission among strains, and for their potential role as drivers for the spread of antibiotic resistance. We show that among the 30 different sequence types (ST) detected were the pandemic clonal lineages ST131, ST648 and ST410, which combine high-level antimicrobial resistance with virulence. In addition to the frequently found resistance genes $bla_{CTX-M-15}$, *tet*(34), and *aph(6)-Id*, we identified *csg* genes, which are required for curli fiber synthesis and thus biofilm formation. Numerous strains harbored multiple virulence-associated genes (VAGs) including *pap* (P fimbriae adhesion cluster), *fim* (type I fimbriae) and *chu* (Chu heme uptake system). Furthermore, we found phylogenetic relationships among strains from patients and their caregivers or related community members and animals, which indicates transmission of pathogens. Also, we

demonstrated the presence and potential transfer of identical/similar ESBL-plasmids in different strains from the Rwandan setting and when compared to an external plasmid. This study highlights the circulation of clinically relevant, pathogenic ESBL-producing *E. coli* among patients, caregivers and the community in Rwanda. Combining antimicrobial resistance with virulence in addition to the putative exchange of mobile genetic elements among bacterial pathogens poses a significant risk around the world.

Keywords: ESBL—*E. coli*, whole-genome sequencing, Rwanda, virulence factors, phylogenetic analysis

INTRODUCTION

The versatility of *Escherichia coli* (*E. coli*) is based on the diversity of genetic substructures within this species (Whittam et al., 1983). In addition to commensal strains, which are an essential part of the non-anaerobic intestinal microflora of humans, other mammals and birds, pathogenic variants occur. The dissimilarity of these pathotypes depends also on their virulence attributes, resulting in a wide range of pathologies in both humans and animals. The intestinal pathogenic *E. coli* (InPEC) express characteristic virulence factors that allow to adhere and invade intestinal cells, causing specific enteric and diarrheal diseases. While InPEC are obligate pathogens, extraintestinal pathogenic *E. coli* (ExPEC) are part of the intestinal microbiome but exhibit a heterogeneous composition of virulence factors to colonize niches such as the urinary tract (Kaper et al., 2004). They can thus cause infections in almost any organ or non-intestinal site, regardless of the state of the host's immune system (Russo and Johnson, 2000). However, a strict differentiation of pathogenic and commensal *E. coli* is difficult, provided by their rapid geno- and phenotypic adaptation to changing environmental conditions, for example through horizontal gene transfer (Pallen and Wren, 2007). Despite the plasticity of the genome, phylogenetic studies have shown some clonality within the population structure of *E. coli*, from which seven distinct phylogenetic groups were derived (Jaureguy et al., 2008; Touchon et al., 2009; Clermont et al., 2013). Usually, commensal strains and obligatory pathogens belong to the phylogroups A and B1, whereas strains with extended virulent attributes (mainly ExPEC) are part of the phylogroups B2, D, and F, with the latter as a sister group of B2 (Escobar-Páramo et al., 2004; Clermont et al., 2013). Multi-locus sequence typing (MLST) allows additional classification and several phylogenetic studies suggest the spread of pandemic high-risk clonal lineages including primarily sequence type (ST) 131 (Nicolas-Chanoine et al., 2008; Ewers et al., 2010; Hussain et al., 2012), ST648 (Ewers et al., 2014; Schaufler et al., 2019), ST410 (Schaufler et al., 2016b; Zurita et al., 2020), putatively ST405 (Manges et al., 2019), and others.

The management of zoonotic infections caused by antibiotic-resistant bacteria has become a multidisciplinary challenge for all modern healthcare systems and is nowadays often approached in a holistic One Health context. Bacterial pathogens spread through direct contact among humans and animals, indirectly by (environmental) pollution and also through non-living and living vectors (Rahman et al., 2020). One example for the latter are houseflies, which have been demonstrated to carry antibiotic-resistant pathogens including extended-spectrum beta-lactamases-(ESBL)-producing *E. coli* (Rahuma et al., 2005; Heiden et al., 2020b; Tufa et al., 2020) non-susceptible to third-generation cephalosporins (e.g., cefotaxime) and monobactams (e.g., aztreonam) (Bevan et al., 2017). Notably, ESBL enzyme production is often accompanied by cross- and co-resistances (Cantón and Coque, 2006; Hidron et al., 2008; Pitout, 2012) resulting in multi-drug resistant (MDR) representatives (Beceiro et al., 2013).

Both pandemic clonal lineages including the aforementioned ST131, ST648 and others, as well as mobile genetic elements (i.e., ESBL-plasmids) drive the spread of antibiotic resistance and virulence-associated genes (VAGs) (Cantón and Coque, 2006). Interestingly, previous studies suggest that ESBL-plasmid carriage does not ineluctably reduce bacterial fitness, which seems particularly true for specific clonal lineages (McNally et al., 2016; Schaufler et al., 2016a; Ranjan et al., 2018; Monárrez et al., 2019; Schaufler et al., 2019). The combination of MDR with high-level bacterial virulence and fitness leads to the emergence of these pandemic, high-risk clonal lineages and subsequently contributes to treatment failures, increasing morbidity, and mortality (Melzer and Petersen, 2007; Schwaber and Carmeli, 2007; Tumbarello et al., 2007; Beceiro et al., 2013; Schaufler et al., 2019; Heiden et al., 2020a).

The One Health concept—addressing human, animal and environmental health—encounters some challenges, especially in low-income countries like Sub-Saharan Africa/Rwanda. On the one hand, the lack of surveillance systems may result in inadequate establishment and implementation of hygienic strategies and therapy guidelines (Muvunyi et al., 2011; Ntirenganya et al., 2015; Carroll et al., 2016). On the other hand, uncontrolled over-the-counter sale of partially counterfeit and substandard antibiotic drugs (Kayumba et al., 2004; Carroll et al., 2016) as well as close human-livestock contact and household crowding might contribute to the broad occurrence and interspecies transmission of MDR bacteria in Sub-Saharan Africa.

This study aimed to investigate whether (i) ESBL-producing *E. coli* circulate among patients, caregivers, the community, and animals in Rwanda, (ii) some of these belong to pandemic high-risk clonal lineages and how they are phylogenetically related, (iii) they demonstrate virulence-associated features, (iv) their mobile genetic elements contribute to the spread of antibiotic resistance.

MATERIALS AND METHODS

Bacterial Strains

The *E. coli* strains investigated in this study were sampled over a time period of 8 weeks at the University Teaching Hospital of Butare (Rwanda) in 2014 (previously described by Kurz et al., 2017). Rectal swabs (Sarstedt AG & Co. KG, Nümbrecht, Germany) were collected from patients and caregivers at admission and discharge as well as from several community members and animals. Each patient had their own caregiver, who were usually relatives accompanying the patient upon admission. They stayed in the patient's room and were involved in personal care of the patient and food preparation. This is a common practice in African hospitals (Hoffman et al., 2012; Ugochukwu, 2013). Sample groups consisting of a patient and related caregiver, and associated family members, neighbors and/or pets were included in the same study-ID. The samples were plated onto chromogenic agar (CHROMagar-ESBL, Mast Diagnostica GmbH, Reinfeld, Germany) supplemented with 2 μg/mL cefotaxime (Cayman Chemical Company, Ann Arbor, United States) and incubated at 37°C. For putative ESBL-positive colonies, the production of ESBL and/or ampicillinase (AmpC) was verified (ESBL-AmpC-Detection Test, Mast Diagnostica GmbH, Reinfeld, Germany) and all strains positive for AmpC only were excluded. The strains were stored at −80°C in LB broth (Carl Roth GmbH & Co. KG, Karlsruhe, Germany) supplemented with 20% (V/V) glycerol (Merck KGaA, Darmstadt, Germany). Originally, we have obtained overall 289 ESBL-producing *E. coli* strains (from patients, caregivers, community members, and animals), with 120 selected strains (based on related study-IDs) that were whole-genome sequenced. Additionally, flies caught with fly traps at different wards of the hospital examined in a previous study (Heiden et al., 2020b) were partly included in this study (**Supplementary Table 1**).

Whole-Genome Sequencing

One single *E. coli* colony was cultured in LB broth supplemented with 2 μg/mL cefotaxime overnight and the total DNA was extracted using the MasterPure™ DNA Purification Kit for Blood, Version II (Lucigen, Middleton, United States) according to the manufacturer's instructions. DNA was purity-controlled and quantified using NanoDrop™ 2000 (Thermo Fisher Scientific Inc., Waltham, United States). WGS was performed in collaboration with LGC (LGC Genomics GmbH, Berlin, Germany) with 150 bp paired-end-reads using Illumina NextSeq 500/550 V2.

Genomic Analysis

Raw reads were quality-trimmed, adapter-trimmed and contaminant-filtered using BBDuk from BBTools v. 38.86[1]. After *de novo* assembly (at a maximum coverage of 100×) using shovill v. 1.1.0[2] in combination with SPAdes v. 3.14.1 (Bankevich et al., 2012), draft genomes were polished by mapping all trimmed reads back to the contigs with bwa v. 0.7.17 (Li and Durbin, 2009), processing SAM/BAM files marking optical duplicates with Samtools v. 1.10 (Li et al., 2009) and calling variants with Pilon v. 1.23 (Walker et al., 2014) (**Supplementary Table 2**).

Plasmid sequences of all strains were manually extracted using similarity searches (BLASTN Megablast) against the NCBI nucleotide collection for visualization in BRIG v. 0.95-dev.0004 (Alikhan et al., 2011). The *in silico* MLST, antibiotic resistance/virulence gene and single-nucleotide polymorphism (SNP) detection was carried out using mlst v. 2.19.0[3], ABRicate v. 1.0.0[4], and snippy v. 4.6.0[5]. We inferred core SNP phylogenies for pandemic sequence types. Alignments were filtered for recombinations using Gubbins v. 2.4.1 (Croucher et al., 2015) and core SNPs extracted using snp-sites v. 2.5.1 (Page et al., 2016) [core SNP sites filtered out: 1,219 (ST131); 0 (ST354); 5,370 (ST405); 2,940 (ST410); 3,564 (ST648)]. The final core SNP alignments contained 208 (ST131; reference: PBIO440), 20 (ST354; reference: PBIO388), 177 (ST405; reference: PBIO397), 154 (ST410; reference: PBIO289), and 135 (ST648; reference: PBIO368) sites (**Supplementary Table 3**). Maximum likelihood trees were inferred with RAxML-NG v. 1.0.0 (Kozlov et al., 2019) using GTR + G (discrete GAMMA model of rate heterogeneity with 4 categories) and searching from 500 random and 500 parsimony-based starting trees. The best-scoring maximum likelihood tree supplemented with support values from 1,000 non-parametric bootstrap replicates was midpoint-rooted and visualized with iTOL v. 5.7 (Letunic and Bork, 2019). To assess the population structure of all available genomes, we inferred an additional phylogeny constructed with JolyTree v 2.0.19092ac (Criscuolo, 2019) of all 120 strains from this study and 13 housefly isolates previously published (Heiden et al., 2020b). A synteny plot comparing ST38 chromosome- and plasmid-derived contigs was created with genoPlotR v 0.8.9 (Luan and Li, 2004).

Minimum Inhibitory Concentration of Colistin

When the genotype was positive for colistin resistance (presence of mcr genes), we evaluated the resistance phenotype by determining the minimum inhibitory concentration (MIC) using MICRONAUT MIC-Strip Colistin (Merlin Diagnostika GmbH, Bornheim, Germany) according to the manufacturer's instructions and interpreted the results according to the published breakpoints of EUCAST (The European Committee on Antimicrobial Susceptibility Testing, 2021). Experiments were performed thrice.

RESULTS

Phylogenetic Grouping and Multi-Locus Sequence Typing

The largest fraction of the 120 ESBL-producing *E. coli* originated from rectal swabs of patients (50.8%; 61/120), followed by caregivers (38.3%; 46/120), neighbors (5.0%; 6/120), family members (4.2%; 5/120), and animals (1.7%; 2/120) (**Figure 1**

[1]https://sourceforge.net/projects/bbmap/

[2]https://github.com/tseemann/shovill/

[3]https://github.com/tseemann/mlst/

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

[5]https://github.com/tseemann/snippy/

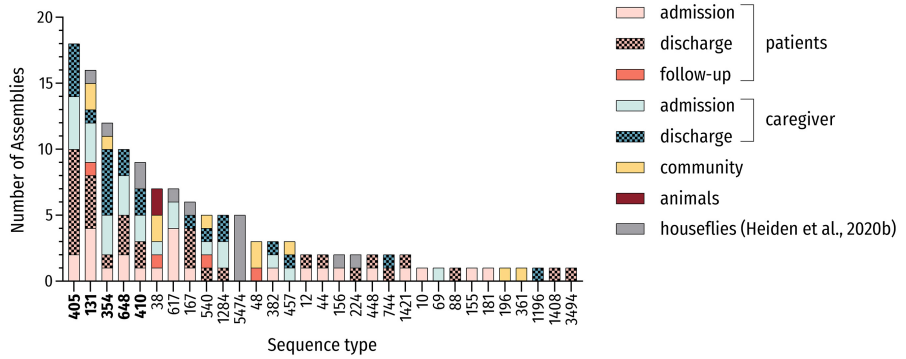


FIGURE 1 | MLST-based distribution of investigated strains. Distribution of all investigated strains including genomes of ESBL-producing *E. coli* carried by houseflies (Heiden et al., 2020b; *n* = 133) and associated sequence types (STs). The five most prevalent STs are in bold.

and **Supplementary Table 1**). The majority of strains belonged to phylogroup A (30.8%; 37/120), which usually comprises commensal strains. The remaining strains were distributed among the phylogroups D (21.7%; 26/120), F (20.0%; 24/120), and B2 (14.2%; 17/120) as well as phylogroups B1 (6.7%; 8/120) and C (6.7%; 8/120), based on Clermont's revised *E. coli* phylotyping method (Clermont et al., 2013).

Multi-locus sequence typing (MLST) from genomic data revealed 30 different STs. In total, sixty percent (18/30) of identified STs were present more than once and 10 or more strains belonged to 4 main STs (13.3%; 4/30) including the high-risk ST131 and ST648 *E. coli* lineages. These most frequently encountered STs were ST405 (15.0%; 18/120), ST131 (12.5%, 15/120), ST354 (9.2%; 11/120) and ST648 (8.3%; 10/120), which accounted for 45.0% (54/120) of all strains. Including the pandemic ST410 lineage, more than half of all strains (50.8%; 61/120) belonged to one of the five most prevalent STs in this study (**Figure 1**). These sequence types belonged to phylogroups B2 (ST131), D (ST405), and F (ST354, ST648) as well as phylogroup C (ST410), thus underlining the heterogeneity of phylogenetic backgrounds among ESBL-producing *E. coli*.

Phylogenetic Relationships

To elucidate phylogenetic relationships, we constructed a tree in an alignment-free manner for all investigated genomes (**Supplementary Figure 1**). Additionally, we inferred phylogenies, which are based on SNPs in the core genome of strains belonging to the five predominant STs of this study to assess potential transmission scenarios (**Figure 2**). For comparative reasons, we included 13 genomes of ESBL-producing *E. coli* isolated from houseflies (Heiden et al., 2020b) originating from the same Rwandan hospital.

The phylogenetic analysis (**Supplementary Figure 1**) shows that the *E. coli* strains were distributed among six distinct phylogroups and grouped into several clades according to their sequence type. Within these ST-associated phylogenies, several sub-clades were defined with genomes interspersed in patients, caregivers, related community members as well as animals and flies, which suggests common phylogenetic backgrounds (Clermont et al., 2011) potentially based on interspecies transmission. For example, PBIO458 (study-ID 60) and PBIO459 [(study-ID 133)—both ST38 isolated from two animals—clustered with four different strains from patients, family members and neighbors, PBIO272 (patient admission, study-ID 60), PBIO451 (neighbor, study-ID 60), PBIO455 (family member, study-ID 60), and PBIO467 (patient follow-up, study-ID 60)]. These findings are corroborated by results of two of our previous studies, where we demonstrated the likely transmission of ESBL-producing *E. coli* ST38 among humans and animals (Guenther et al., 2017; Schaufler et al., 2018). Note, however, that in this current study, only one representative of ST38 [PBIO302 (caregiver admission, study-ID 131)] carried a chromosomally encoded $bla_{CTX-M-15}$ gene and the before mentioned strains carried plasmid-encoded ESBLs (**Supplementary Figure 1**), which is contrary to our previous findings. We then compared the $bla_{CTX-M-15}$ gene-carrying chromosomal contig of PBIO302 to two of the plasmid-encoded $bla_{CTX-M-15}$ sequences of ST38 (PBIO272 and PBIO459; **Supplementary Figure 2**). The chromosomal sections of PBIO302, PBIO272, and PBIO459 were highly similar, except the chromosomal insertion of $bla_{CTX-M-15}$ in PBIO302. This resistance gene was flanked by transposable elements, as described below.

In addition, strains with the numbers PBIO1939, PBIO1942, PBIO1945, PBIO1946, and PBIO1947 from houseflies were in the same sub-clade as strains isolated from different human sources indicating the potential role of living vectors in the spread of pathogenic bacteria.

For the phylogenetic trees of the five predominant STs of this study (**Figure 2**) it is interesting to notice that some genomes stemming from different sources were more closely related than genomes from the same source. For example, PBIO283 [(study-ID 92) **Figure 2A**, ST131, sub-clade 1], which originates from a caregiver at admission differed in 0.2 ± 0.0003 SNPs/Mbp with strains isolated from the related patient at admission [PBIO286 (study-ID 92)] as well as discharge [PBIO285 (study-ID 92)] and unrelated patients at discharge [PBIO293 (study-ID 114) and PBIO296 (study-ID 117)]. Moreover, PBIO440 [(study-ID 133) **Figure 2A**, ST131, sub-clade 2], isolated from a community member, only varied in 0.2 SNPs/Mbp compared to a follow-up strain of an already discharged patient [PBIO405 (study-ID

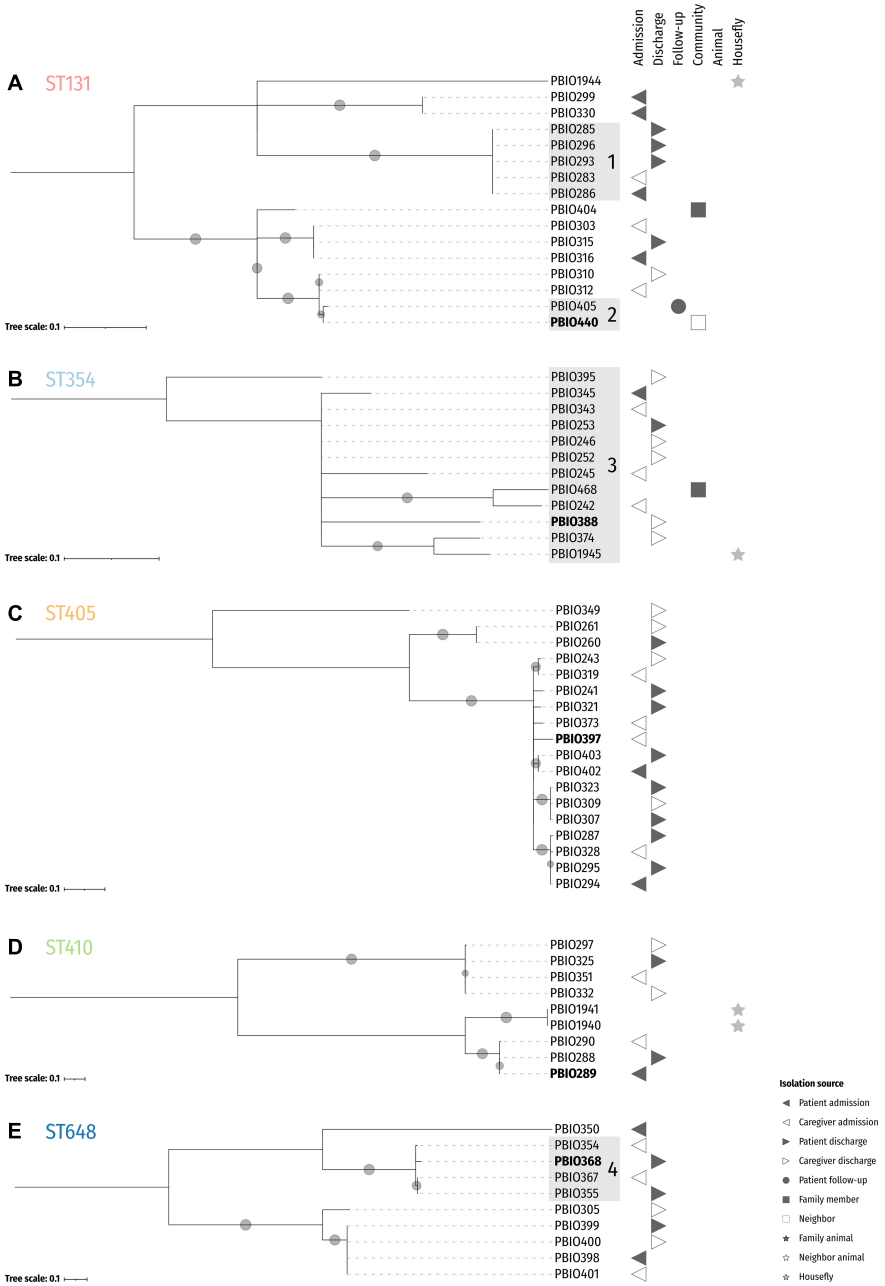


FIGURE 2 | Midpoint-rooted maximum likelihood core SNP phylogenies of the five dominating sequence types. Reference isolates for ST131 (PBIO440; **A**), ST354 (PBIO388; **B**), ST405 (PBIO397; **C**), ST410 (PBIO289; **D**), and ST648 (PBIO368; **E**) are highlighted in bold. The trees are based on alignments with 208 **(A)**, 20 **(B)**, 177 **(C)**, 154 **(D)**, and 135 **(E)** sites. Sub-clades are highlighted in gray. Circles at branches display bootstrap proportions ≥50% (1,000 replicates). Symbols besides the tree depict isolation sources as given in the legend.

434)]. Notably, all strains belonging to ST354 (**Figure 2B**, sub-clade 3) only differed in 1.0 ± 0.4 SNPs/Mbp including one strain carried by a housefly (PBIO1945). Also interesting was the difference between PBIO368 [(study-ID 335) **Figure 2E**, ST648, sub-clade 4], originating from a patient at discharge, and strains of distinct sources [PBIO354 (caregiver admission, study-ID 265), PBIO355 (patient discharge, study-ID 265), and PBIO367 (caregiver admission, study-ID 288)], differing in 0.9 ± 0.1 SNPs/Mbp. These numbers of SNPs were up to 10-fold lower than described for clonal EHEC strains during an outbreak in Germany (1.8 SNPs/Mbp) (Grad et al., 2012; Been et al., 2014), suggesting the circulation of only a handful of sequence types in this African setting, which interestingly happen to mostly be international high-risk clonal lineages. In addition, some strains from identical STs were carried by both flies and humans, for example PBIO1945 and PBIO374 (caregiver discharge, study-ID

401; **Figure 2B**), again indicating the role of flies in the spread of antibiotic-resistant pathogens.

Antimicrobial Resistance Determinants

The predominant ESBL-gene was $bla_{CTX-M-15}$ (87.5%; 105/120). Furthermore, 76 strains (63.3%; 76/120) carried bla_{OXA-1} and eight (6.7%; 8/120) $bla_{CTX-M-27}$. Previous studies reported the co-occurrence of $bla_{CTX-M-15}$ and bla_{OXA-1}, whereas other members of the CTX-M family (e.g., $bla_{CTX-M-27}$) and bla_{OXA-1} appear to be mutually exclusive (Schaufler et al., 2016b; Livermore et al., 2019; Bodendoerfer et al., 2020). Here, almost all strains carrying the bla_{OXA-1} gene (98.7%; 75/76) carried the $bla_{CTX-M-15}$ gene in addition and none showed the combination of bla_{OXA-1} and another CTX-M gene. The co-occurence of distinct CTX-M genes was not detected. For the majority of bla_{CTX-M}-positive strains (71.1%; 81/114), the corresponding gene was encoded on a plasmid (**Supplementary Figure 1**). Interestingly, all strains of phylogroup F (20.0%; 24/120) were carriers of $bla_{CTX-M-15}$ genes, of which 87.5% (21/24) were located on the chromosome (all strains belonging to ST648 and ST354). These chromosomally encoded genes were flanked by an insertion sequence IS*Ecp1* upstream and a Tn*2* downstream. Also note that the strains PBIO242 and PBIO245 (ST354) showed two of these motifs consecutively. The $bla_{CTX-M-15}$ gene of the ST410 strains PBIO288, PBIO289, and PBIO290 were found in proximity to an upstream-located IS*Ecp1* only. Previous studies have already demonstrated the diversity and global distribution of these complex transposable units in enterobacteria (Poirel et al., 2005; Lartigue et al., 2006; Decano and Downing, 2019; Ludden et al., 2020; Yoon et al., 2020).

Genes encoding for carbapenem-hydrolyzing enzymes, like bla_{OXA-48} or bla_{NDM-1}, were not present.

A growing body of ESBL-producers is MDR (Cantón and Coque, 2006; Hidron et al., 2008; Pitout, 2012; Beceiro et al., 2013). In our study, all investigated strains (100.0%; 120/120) were carriers of genes conferring resistance to three or more different classes of antimicrobial agents, thus exhibiting a MDR genotype (Magiorakos et al., 2012). In total, 120 strains (100.0%; 120/120) carried resistance genes to tetracyclines (*tet*), followed by genes encoding for aminoglycoside (*aac, aad*, and *aph* [98.3%; 118/120]), sulfonamide [*sul* (90.0%; 108/120)], trimethoprim [*dfr* (89.2%; 107/120)], chloramphenicol [*cat* (72.5%; 87/120)], and fluoroquinolone [*aac(6′)-Ib-cr* and *qnr* (71.7%; 86/120)] resistances. Two strains (1.7%; 2/120) belonging to ST181 and ST540 were carriers of *mcr-9* but showed phenotypic susceptibility to colistin, which is a last-resort antibiotic (MIC of both strains: 0.5 μg/mL). Interestingly, previous studies suggest that this latest member of the *mcr* gene family does not always confer phenotypic resistance to colistin in clinical isolates, although overexpression in *E. coli* TOP10 cells leads to increased MICs (Carroll et al., 2019; Kieffer et al., 2019; Tyson et al., 2020).

In addition to the resistances described, genes encoding for efflux pumps were found frequently, with *mdfA* in all (120/120) and *acrB* in 85.8% (103/120) of all genomes.

Virulence-Associated Genes

As previously demonstrated by us and others, the combination of MDR and high-level bacterial virulence seems to be a hallmark of pandemic high-risk clonal lineages (Hussain et al., 2012; Ewers et al., 2014; Shaik et al., 2017; Schaufler et al., 2019). To investigate the strains' genetic virulome, we analyzed the genomes of the predominant ST131 (n = 15), ST648 (n = 10), ST354 (n = 11), ST405 (n = 18), and ST410 (n = 7) for these features (**Figure 3**).

The strains belonging to the five predominant STs carried several VAGs, mainly associated with adherence, antiphagocytosis, biofilm formation, invasion, iron uptake and bacterial secretion. The ability to attach to surfaces/cells and form biofilms is a common strategy used by bacterial populations to resist antibiotic treatment and host defense mechanisms as well as cause infection (Moser et al., 2017; Amanatidou et al., 2019). In particular, genes for the P fimbriae adhesion cluster [*pap* operon (70.5%; 43/61)], Dr family of adhesins (4.9%; 3/61) and type I fimbriae [*fim* (100%; 61/61)], which are necessary for uroepithelia cell adhesion and invasion, and, thus, for causing urinary tract infection (Mulvey, 2002), were frequently found. Notably, strains belonging to ST354 and ST410 showed a lack of *pap* genes, which is consistent with previous findings (Vangchhia et al., 2016; Zogg et al., 2018; Schaufler et al., 2019). Furthermore, we detected several members of the *csg* gene family in all genomes (100%; 120/120). These genes encode curli fibers, which are essential components of bacterial biofilms (Hammar et al., 1995; Evans and Chapman, 2014).

The ability to acquire intracellular heme and hemolysin, which is based on the expression of iron uptake-associated genes [e.g., Chu heme uptake system (88.5%; 54/61), yersiniabactin (86.9%; 53/61), and aerobactin (65.6%; 40/61)], is an effective strategy for iron utilization during infection (Fischbach et al., 2006) and is another important virulence-associated feature in the repertoire of these bacteria.

Mobile Genetic Elements

We next investigated the occurrence and circulation/transfer of ESBL-plasmids among strains and thus their contribution to the spread of antibiotic resistance in the Rwandan setting.

In total, 107 strains (89.5%; 107/120) carried plasmids with incompatibly (Inc) group FIB, followed by IncFIA (75.8%; 91/120) and IncFII (70.8%; 85/120). In particular IncF plasmids are frequently associated with genes encoding ESBL-genes, other resistances as well as virulence features important for iron acquisition (Han et al., 2012), serum resistance (Ranjan et al., 2018), and biofilm formation (Schaufler et al., 2016a).

To better assess similar plasmid backgrounds, we compared the plasmid sequences of all strains that carried a plasmid-borne $bla_{CTX-M-15}$ gene against the plasmid sequence of PBIO241 (ST405; patient discharge; study-ID 159; **Figure 4**) as a representative for the most prevalent ST. Keep in mind, however, that the selection of the reference biases the visual representation when a large number of plasmid sequences is absent in the query sequences. On the other hand, plasmid sequences, not present in the reference but in queries, are missing in this approach. The plasmid backbones of other ST405

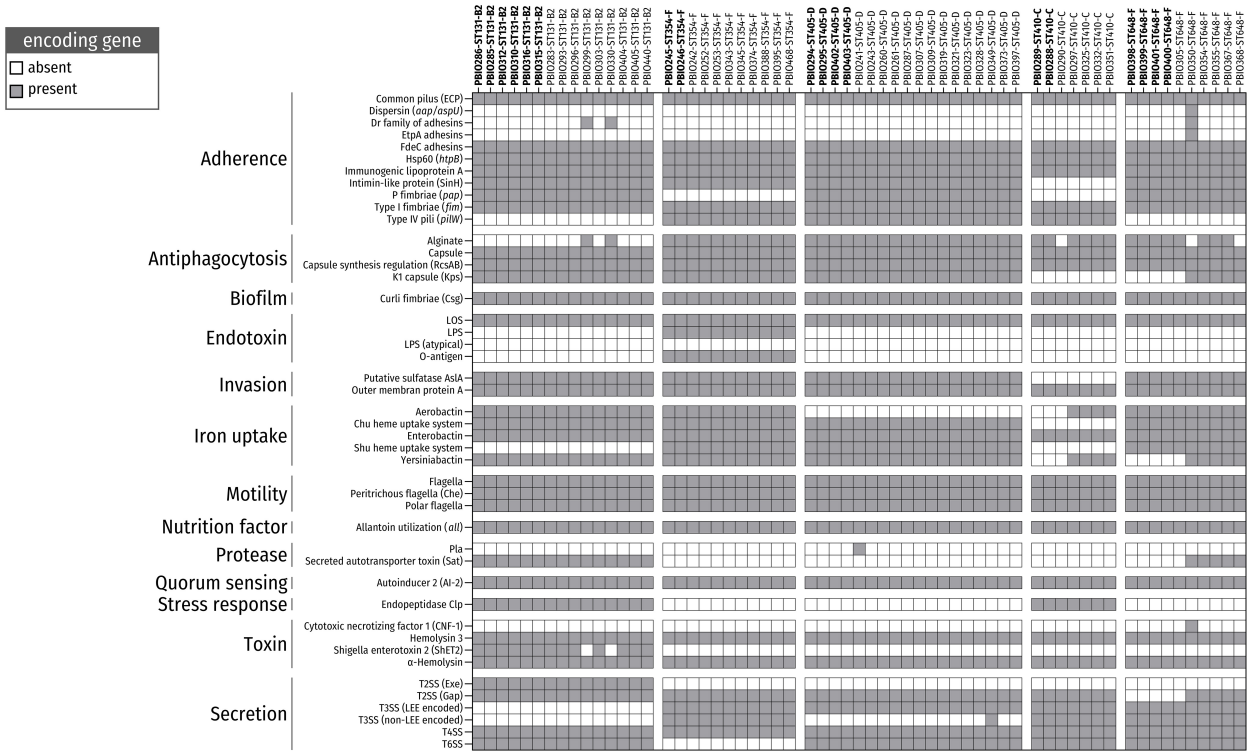


FIGURE 3 | Presence/absence of virulence factors (VAGs) among ESBL-producing *E. coli* belonging to the five most prevalent sequence types (STs). Color-filled boxes show the presence of genes (coverage and identity ≥ 65%) encoding for different VAGs.

strains were highly similar to the reference. In contrast, plasmids from strains of other pandemic STs (ST131 and ST410) only showed few similarities when compared to the reference plasmid underlining their genetic diversity. Interestingly, in addition, the plasmid sequences of four ST167 strains [PBIO276 (patient discharge, study-ID 78), PBIO277 (patient admission, study-ID 78), PBIO301 (caregiver discharge, study-ID 131), and PBIO306 (patient discharge, study-ID 131)] were highly similar to the reference. Five other strains [PBIO256 (ST12, patient admission, study-ID 40), PBIO257 (ST12, patient discharge, study-ID 40), PBIO258 (ST1421, patient admission, study-ID 40), PBIO1948 (ST5474, housefly), and PBIO1949 (ST5474, housefly)] showed identical plasmid sequences in parts of the BRIG visualization that were absent in most of the other sequences.

We also compared some of our plasmid sequences of representative strains (criteria: top BLAST-hits) against a publicly available reference plasmid pIV_IncHI2_CTX_M_15 (Marchetti et al., 2020; **Supplementary Figure 3**). This $bla_{\mathrm{CTX-M-15}}$-containing plasmid has been originally obtained from an *E. coli* ST58 strain causing a deadly puppy infection in Italy in 2019. Interestingly, the plasmid backbone of PBIO251 (ST155, patient admission, study-ID 153) was nearly identical when compared to this external Italian plasmid. The plasmids of PBIO275 (ST540, caregiver admission, study-ID 78), PBIO385 (ST3494, patient discharge, study-ID 434), and PBIO1943 (ST5474, housefly) also showed high similarities to the reference and thus among each other.

The presence of similar resistance plasmids in strains of distinct sequence types and sample groups indicates the potential transmission of mobile genetic elements and the growing prevalence of successful plasmid families (Carattoli, 2011).

DISCUSSION

In contrast to Europe, the United States of America and Australia, only little information is available concerning the exact characteristics and distribution of ESBL-producing *E. coli* in Africa. This study reveals the broad occurrence and potential circulation of several international high-risk clonal lineages in a hospital and associated locations in Rwanda, following up on a publication by Kurz et al. (2017). Despite the potential bias due to the analysis of both admission and discharge strains of the same patient and/or caregiver, it is remarkable that our sample set was dominated by ST131, ST648, and ST410, which have been frequently reported from humans, animals and the environment (Nicolas-Chanoine et al., 2008; Ewers et al., 2010, 2014; Hussain et al., 2012; Schaufler et al., 2016b, 2019) and which are "classic" pandemic, high-risk clonal lineages. In addition to these, we found ST354 to be the third most detected sequence type in this study. Its emergence has been previously described globally except for the African continent (Manges et al., 2019). This locally restricted accumulation of one single ST in combination with a small number of other STs indicates re-entering and circulation of

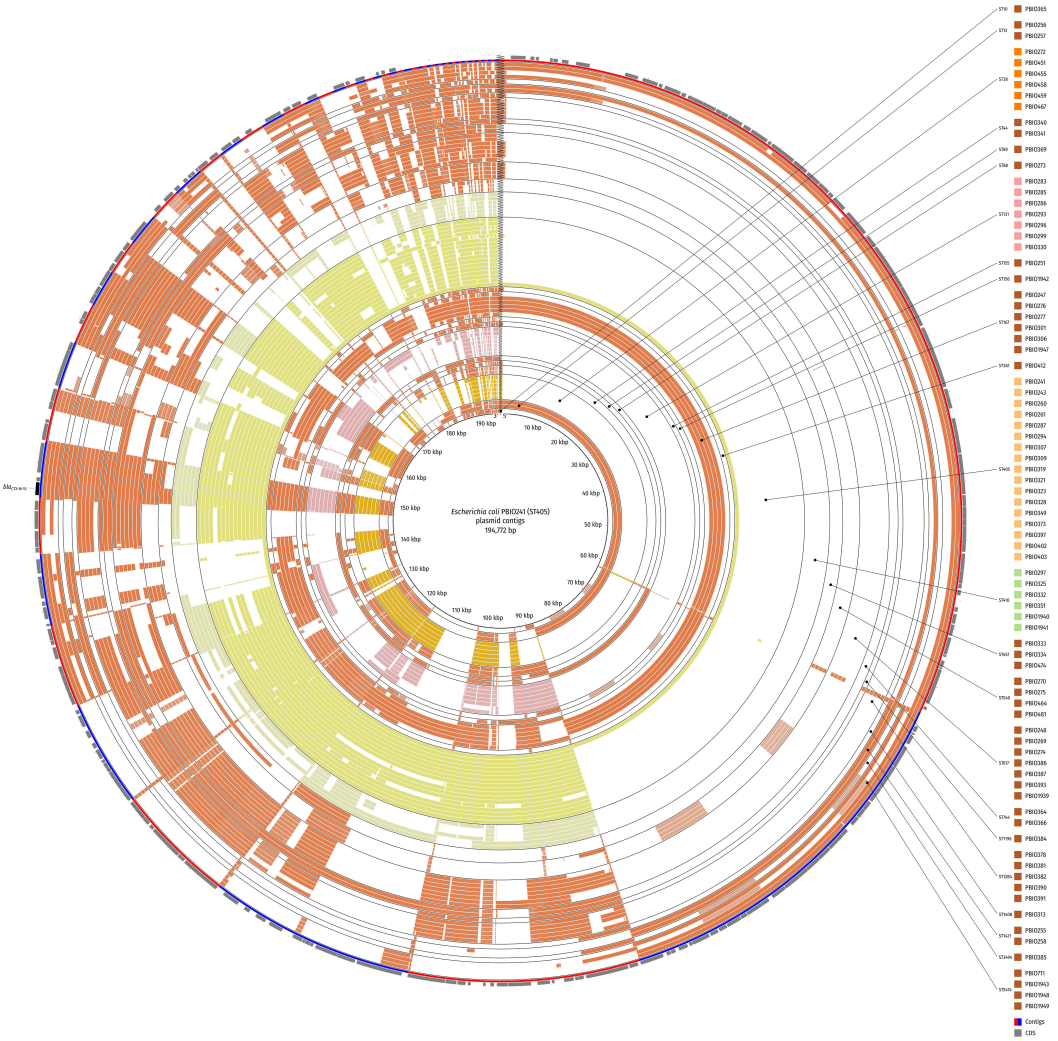


FIGURE 4 | Circular comparison of the plasmid background of strains with plasmid-borne $bla_{CTX-M-15}$ gene. The BLAST-classified plasmid contigs were searched against plasmid contigs of PBIO241 (ST405) as a reference. Strains are ordered by ST and name in ascending order and depicted in concentric rings from inner to outer. Red and blue arcs show contig boundaries and the outermost ring contains coding sequences (CDS; gray arrows) with the $bla_{CTX-M-15}$ gene highlighted in black. The comparison was created with BLAST Ring Image Generator (BRIG).

dominant bacterial lineages in the Rwandan hospital and among family members/neighbors and animals.

Interestingly, within the ST-associated clades, some genomes without genetic differences were interspersed in humans (hospitalized patients and caregivers as well as community members) and animals. This putative lack of host adaptation and the close phylogenetic relationships indicate the colonization and rapid transmission of several clones within the community and the potential transmission into the clinical setting and vice versa, underlined by the high acquisition rates of ESBL-producing *E. coli* during hospitalization as described previously (Kurz et al., 2017).

The resistance genes found in this study confer resistances to antibiotics frequently used in veterinary medicine and/or in sub-therapeutic doses as food supplements and growth promotors in Africa (Eagar et al., 2012; Adesokan et al., 2015; Mainda et al., 2015; Manishimwe et al., 2017). When also considering the zoonotic character of ESBL-producing *E. coli*, it is not surprising that we found clonal strains with similar patterns of resistance features in the different sample groups. Transmission likely occurred among patients and caregivers/family members and was also influenced by livestock animals due to close human-animal contact (Klous et al., 2016).

Two strains of this study carried the *mcr-9* gene but were phenotypically susceptible to colistin. This phenomenon was first reported in 2019 (Carroll et al., 2019; Kieffer et al., 2019). Due to the structural heterogeneity compared to other *mcr* genes (65% amino acid identity with the most closely related *mcr-3* gene) and the weak inactivation of colistin, the clinical importance of mcr-9 is unknown (Tyson et al., 2020).

Notably, some of the strains showed extensive, chromosomally encoded virulence-associated features. The CNF-1-encoding

gene (*cnf1*) detected in strain PBIO350 (ST648), for example, is associated with causing neonatal meningitis (Khan et al., 2002). In addition to the major virulence factors of meningitis-associated and uropathogenic *E. coli* (like P fimbriae adhesion cluster, K1 capsule, heme utilization systems, and the secreted autotransporter toxin), the strains showed VAGs typical for InPEC (especially the various bacterial secretion systems) underlining the clinical relevance of these pathogens.

Finally, we demonstrate that similar plasmid sequences were present in strains from different sample groups, thus likely indicating mobile genetic element transmission, and underlining the importance of plasmid-driven spread of antimicrobial resistance independent of the host's phylogenetic background (Schaufler et al., 2016a; Ranjan et al., 2018). Interesting in addition to similarities among strains from the Rwandan setting is in particular the close relationship to an external plasmid, which has been obtained only recently (Marchetti et al., 2020). This highlights the sometimes global spread of such mobile genetic elements and their bacterial hosts.

CONCLUSION

In this study, we investigated and identified the presence of clinically relevant ESBL-producing *E. coli* that circulate among patients, caregivers, the community and animals in a Rwandan setting. The findings contribute to the understanding of the global dissemination of bacterial high-risk clonal lineages, their virulence features as well as plasmid transmissions. They also underline the potential role of houseflies in this harmful dynamic.

DATA AVAILABILITY

The data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB42795 (https://www.ebi.ac.uk/ena/browser/view/PRJEB42795).

AUTHOR CONTRIBUTIONS

KS and EE designed the study. EE and JM performed the laboratory and phenotypic experiments. SH, SS, and KK performed the bioinformatics analyses. KS, EE, SH, KK, CB, JN, AS, JG, MK, FM, and SS analyzed the data. KS, EE, and SH wrote the manuscript and prepared the tables and figures. All authors read and approved the final version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.662575/full#supplementary-material

REFERENCES

Adesokan, H. K., Akanbi, I. O., Akanbi, I. M., and Obaweda, R. A. (2015). Pattern of antimicrobial usage in livestock animals in south-western Nigeria: the need for alternative plans. *Onderstepoort J. Vet. Res.* 82:816. doi: 10.4102/OJVR.V82I1.816

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

Amanatidou, E., Matthews, A. C., Kuhlicke, U., Neu, T. R., McEvoy, J. P., and Raymond, B. (2019). Biofilms facilitate cheating and social exploitation of β-lactam resistance in *Escherichia coli*. *NPJ Biofilms Microb.* 5:36. doi: 10.1038/s41522-019-0109-2

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

Beceiro, A., Tomás, M., and Bou, G. (2013). Antimicrobial resistance and virulence: a successful or deleterious association in the bacterial world? *Clin. Microbiol. Rev.* 26, 185–230. doi: 10.1128/CMR.00059-12

Been, M., de Lanza, V. F., Toro, M., de, Scharringa, J., Dohmen, W., et al. (2014). Dissemination of cephalosporin resistance genes between *Escherichia coli* strains from farm animals and humans by specific plasmid lineages. *PLoS Genet.* 10:e1004776. doi: 10.1371/journal.pgen.1004776

Bevan, E. R., Jones, A. M., and Hawkey, P. M. (2017). Global epidemiology of CTX-M β-lactamases: temporal and geographical shifts in genotype. *J. Antimicrob. Chemother.* 72, 2145–2155. doi: 10.1093/jac/dkx146

Bodendoerfer, E., Marchesi, M., Imkamp, F., Courvalin, P., Böttger, E. C., and Mancini, S. (2020). Co-occurrence of aminoglycoside and β-lactam resistance mechanisms in aminoglycoside- non-susceptible *Escherichia coli* isolated in the zurich area, switzerland. *Int. J. Antimicrob. Agents* 56:106019. doi: 10.1016/j.ijantimicag.2020.106019

Cantón, R., and Coque, T. M. (2006). The CTX-M beta-lactamase pandemic. *Curr. Opin. Microbiol.* 9, 466–475. doi: 10.1016/j.mib.2006.08.011

Carattoli, A. (2011). Plasmids in gram negatives: molecular typing of resistance plasmids. *Int. J. Med. Microbiol.* 301, 654–658. doi: 10.1016/j.ijmm.2011.09.003

Carroll, L. M., Gaballa, A., Guldimann, C., Sullivan, G., Henderson, L. O., and Wiedmann, M. (2019). Identification of novel mobilized colistin resistance gene mcr-9 in a multidrug-resistant, colistin-susceptible *Salmonella enterica* serotype typhimurium isolate. *mBio* 10, e819–e853. doi: 10.1128/mBio.00853-19

Carroll, M., Rangaiahagari, A., Musabeyezu, E., Singer, D., and Ogbuagu, O. (2016). Five-year antimicrobial susceptibility trends among bacterial isolates from a tertiary health-care facility in kigali, Rwanda. *Am. J. Trop Med. Hyg.* 95, 1277–1283. doi: 10.4269/ajtmh.16-0392

Clermont, O., Christenson, J. K., Denamur, E., and Gordon, D. M. (2013). The clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ. Microbiol. Rep.* 5, 58–65.

Clermont, O., Olier, M., Hoede, C., Diancourt, L., Brisse, S., Keroudean, M., et al. (2011). Animal and human pathogenic *Escherichia coli* strains share common genetic backgrounds. *Infect. Genet. Evol.* 11, 654–662.

Criscuolo, A. (2019). A fast alignment-free bioinformatics procedure to infer accurate distance-based phylogenetic trees from genome assemblies. *Res. Ideas Outcomes* 5:e36178. doi: 10.3897/rio.5.e36178

Croucher, N. J., Page, A. J., Connor, T. R., Delaney, A. J., Keane, J. A., Bentley, S. D., et al. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using gubbins. *Nucleic Acids Res.* 43:e15. doi: 10.1093/nar/gku1196

Decano, A. G., and Downing, T. (2019). An *Escherichia coli* ST131 pangenome atlas reveals population structure and evolution across 4,071 isolates. *Sci. Rep.* 9:17394. doi: 10.1038/s41598-019-54004-5

Eagar, H., Swan, G., and van Vuuren, M. (2012). A survey of antimicrobial usage in animals in south africa with specific reference to food animals. *J. S. Afr. Vet. Assoc.* 83:16. doi: 10.4102/jsava.v83i1.16

Escobar-Páramo, P., Clermont, O., Blanc-Potard, A.-B., Bui, H., Le Bouguénec, C., and Denamur, E. (2004). A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. *Mol. Biol. Evol.* 21, 1085–1094. doi: 10.1093/molbev/msh118

Evans, M. L., and Chapman, M. R. (2014). Curli biogenesis: order out of disorder. *Biochim. Biophys. Acta* 1843, 1551–1558. doi: 10.1016/j.bbamcr.2013.09.010

Ewers, C., Bethe, A., Stamm, I., Grobbel, M., Kopp, P. A., Guerra, B., et al. (2014). CTX-M-15-D-ST648 *Escherichia coli* from companion animals and horses: another pandemic clone combining multiresistance and extraintestinal virulence? *J. Antimicrob. Chemother.* 69, 1224–1230. doi: 10.1093/jac/dkt516

Ewers, C., Grobbel, M., Stamm, I., Kopp, P. A., Diehl, I., Semmler, T., et al. (2010). Emergence of human pandemic O25:H4-ST131 CTX-M-15 extended-spectrum-beta-lactamase-producing *Escherichia coli* among companion animals. *J. Antimicrob. Chemother.* 65, 651–660. doi: 10.1093/jac/dkq004

Fischbach, M. A., Lin, H., Liu, D. R., and Walsh, C. T. (2006). How pathogenic bacteria evade mammalian sabotage in the battle for iron. *Nat. Chem. Biol.* 2, 132–138. doi: 10.1038/nchembio771

Grad, Y. H., Lipsitch, M., Feldgarden, M., Arachchi, H. M., Cerqueira, G. C., Fitzgerald, M., et al. (2012). Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in europe, 2011. *Proc. Natl. Acad. Sci. U S A* 109, 3065–3070. doi: 10.1073/pnas.1121491109

Guenther, S., Semmler, T., Stubbe, A., Stubbe, M., Wieler, L. H., and Schaufler, K. (2017). Chromosomally encoded ESBL genes in *Escherichia coli* of ST38 from mongolian wild birds. *J. Antimicrob. Chemother.* 72, 1310–1313. doi: 10.1093/jac/dkx006

Hammar, M. R., Arnqvist, A., Bian, Z., Olsén, A., and Normark, S. (1995). Expression of two csg operons is required for production of fibronectin- and congo red-binding curli polymers in *Escherichia coli* K-12. *Mol. Microbiol.* 18, 661–670. doi: 10.1111/j.1365-2958.1995.mmi_18040661.x

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

Heiden, S. E., Hübner, N.-O., Bohnert, J. A., Heidecke, C.-D., Kramer, A., Balau, V., et al. (2020a). A *Klebsiella pneumoniae* ST307 outbreak clone from germany demonstrates features of extensive drug resistance, hypermucoviscosity, and enhanced iron acquisition. *Genome Med.* 12:113. doi: 10.1186/s13073-020-00814-6

Heiden, S. E., Kurz, M. S. E., Bohnert, J., Bayingana, C., Ndoli, J. M., Sendegeya, A., et al. (2020b). Flies from a tertiary hospital in Rwanda carry multidrug-resistant gram-negative pathogens including extended-spectrum beta-lactamase-producing E. coli sequence type 131. *Antimicrob. Resist. Infect. Control* 9:34. doi: 10.1186/s13756-020-0696-y

Hidron, A. I., Edwards, J. R., Patel, J., Horan, T. C., Sievert, D. M., Pollock, D. A., et al. (2008). NHSN annual update: antimicrobial-resistant pathogens associated with healthcare-associated infections: annual summary of data reported to the national healthcare safety network at the centers for disease control and prevention, 2006-2007. *Infect. Control Hosp. Epidemiol.* 29, 996–1011. doi: 10.1086/591861

Hoffman, M., Mofolo, I., Salima, C., Hoffman, I., Zadrozny, S., Martinson, F., et al. (2012). Utilization of family members to provide hospital care in malawi: the role of hospital guardians. *Malawi Med. J.* 24, 74–78.

Hussain, A., Ewers, C., Nandanwar, N., Guenther, S., Jadhav, S., Wieler, L. H., et al. (2012). Multiresistant uropathogenic *Escherichia coli* from a region in India where urinary tract infections are endemic: genotypic and phenotypic characteristics of sequence type 131 isolates of the CTX-M-15 extended-spectrum-β-lactamase-producing lineage. *Antimicrob. Agents Chemother.* 56, 6358–6365. doi: 10.1128/AAC.01099-12

Jaureguy, F., Landraud, L., Passet, V., Diancourt, L., Frapy, E., Guigon, G., et al. (2008). Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. *BMC Genom.* 9:560. doi: 10.1186/1471-2164-9-560

Kaper, J. B., Nataro, J. P., and Mobley, H. L. (2004). Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* 2, 123–140. doi: 10.1038/nrmicro818

Kayumba, P. C., Risha, P. G., Shewiyo, D., Msami, A., Masuki, G., Ameye, D., et al. (2004). The quality of essential antimicrobial and antimalarial drugs marketed in rwanda and tanzania: influence of tropical storage conditions on in vitro dissolution. *J. Clin. Pharm. Ther.* 29, 331–338. doi: 10.1111/j.1365-2710.2004.00568.x

Khan, N. A., Wang, Y., Kim, K. J., Chung, J. W., Wass, C. A., and Kim, K. S. (2002). Cytotoxic necrotizing factor-1 contributes to *Escherichia coli* K1 invasion of the central nervous system. *J. Biol. Chem.* 277, 15607–15612.

Kieffer, N., Royer, G., Decousser, J.-W., Bourrel, A.-S., Palmieri, M., La Ortiz De Rosa, J.-M., et al. (2019). mcr-9, an inducible gene encoding an acquired phosphoethanolamine transferase in *Escherichia coli*, and its origin. *Antimicrob. Agents Chemother.* 63, e1819–e1866. doi: 10.1128/AAC.00965-19

Klous, G., Huss, A., Heederik, D. J., and Coutinho, R. A. (2016). Human–livestock contacts and their relationship to transmission of zoonotic pathogens, a systematic review of literature. *One Health* 2, 65–76. doi: 10.1016/j.onehlt.2016.03.001

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

Kurz, M. S. E., Bayingana, C., Ndoli, J. M., Sendegeya, A., Durst, A., Pfüller, R., et al. (2017). Intense pre-admission carriage and further acquisition of ESBL-producing *Enterobacteriaceae* among patients and their caregivers in a tertiary hospital in Rwanda. *Trop Med Int Health* 22, 210–220. doi: 10.1111/tmi.12824

Lartigue, M.-F., Poirel, L., Aubert, D., and Nordmann, P. (2006). In vitro analysis of ISEcp1B-mediated mobilization of naturally occurring beta-lactamase gene blaCTX-M of *Kluyvera ascorbata*. *Antimicrob Agents Chemother.* 50, 1282–1286. doi: 10.1128/AAC.50.4.1282-1286.2006

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

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

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

Livermore, D. M., Day, M., Cleary, P., Hopkins, K. L., Toleman, M. A., Wareham, D. W., et al. (2019). OXA-1 β-lactamase and non-susceptibility to penicillin/β-lactamase inhibitor combinations among ESBL-producing *Escherichia coli*. *J. Antimicrob. Chemother.* 74, 326–333. doi: 10.1093/jac/dky453

Luan, Y., and Li, H. (2004). Model-based methods for identifying periodically expressed genes based on time course microarray gene expression data. *Bioinformatics* 20, 332–339. doi: 10.1093/bioinformatics/btg413

Ludden, C., Decano, A. G., Jamrozy, D., Pickard, D., Morris, D., Parkhill, J., et al. (2020). Genomic surveillance of *Escherichia coli* ST131 identifies local expansion and serial replacement of subclones. *Microb. Genom.* 6, e000352. doi: 10.1099/mgen.0.000352

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

Mainda, G., Bessell, P. R., Muma, J. B., McAteer, S. P., Chase-Topping, M. E., Gibbons, J., et al. (2015). Prevalence and patterns of antimicrobial resistance among *Escherichia coli* isolated from zambian dairy cattle across different production systems. *Sci. Rep.* 5:12439. doi: 10.1038/srep12439

Manges, A. R., Geum, H. M., Guo, A., Edens, T. J., Fibke, C. D., and Pitout, J. D. D. (2019). Global extraintestinal pathogenic *Escherichia coli* (ExPEC) lineages. *Clin. Microbiol. Rev.* 32, e118–e135. doi: 10.1128/CMR.00135-18

Manishimwe, R., Nishimwe, K., and Ojok, L. (2017). Assessment of antibiotic use in farm animals in Rwanda. *Trop Anim. Health Prod.* 49, 1101–1106. doi: 10.1007/s11250-017-1290-z

Marchetti, M. V., Bitar, I., Mercato, A., Nucleo, E., Marchesini, F., et al. (2020). Deadly puppy infection caused by an MDR *Escherichia coli* O39 $bla_{\mathrm{CTX-M-15}}$, blaCMY-2, blaDHA-1, and aac(6)-Ib-cr - positive in a breeding kennel in central italy. *Front. Microbiol.* 11:584. doi: 10.3389/fmicb.2020.00584

McNally, A., Oren, Y., Kelly, D., Pascoe, B., Dunn, S., Sreecharan, T., et al. (2016). Combined analysis of variation in core, accessory and regulatory genome regions provides a super-resolution view into the evolution of bacterial populations. *PLoS Genet* 12:e1006280. doi: 10.1371/journal.pgen.1006280

Melzer, M., and Petersen, I. (2007). Mortality following bacteraemic infection caused by extended spectrum beta-lactamase (ESBL) producing E. coli

compared to non-ESBL producing E. coli. *J. Infect.* 55, 254–259. doi: 10.1016/j.jinf.2007.04.007

Monárrez, R., Braun, M., Coburn-Flynn, O., Botelho, J., Odetoyin, B. W., Otero-Vera, J. I., et al. (2019). A large self-transmissible resistance plasmid from nigeria contains genes that ameliorate a carrying cost. *Sci. Rep.* 9:19624. doi: 10.1038/s41598-019-56064-z

Moser, C., Pedersen, H. T., Lerche, C. J., Kolpen, M., Line, L., Thomsen, K., et al. (2017). Biofilms and host response - helpful or harmful. *APMIS* 125, 320–338. doi: 10.1111/apm.12674

Mulvey, M. A. (2002). Adhesion and entry of uropathogenic *Escherichia coli*. *Cell Microbiol.* 4, 257–271. doi: 10.1046/j.1462-5822.2002.00193.x

Muvunyi, C. M., Masaisa, F., Bayingana, C., Mutesa, L., Musemakweri, A., Muhirwa, G., et al. (2011). Decreased susceptibility to commonly used antimicrobial agents in bacterial pathogens isolated from urinary tract infections in Rwanda: need for new antimicrobial guidelines. *Am. J. Trop Med. Hyg.* 84, 923–928. doi: 10.4269/ajtmh.2011.11-0057

Nicolas-Chanoine, M.-H., Blanco, J., Leflon-Guibout, V., Demarty, R., Alonso, M. P., Caniça, M. M., et al. (2008). Intercontinental emergence of *Escherichia coli* clone O25:H4-ST131 producing CTX-M-15. *J. Antimicrob. Chemother.* 61, 273–281. doi: 10.1093/jac/dkm464

Ntirenganya, C., Manzi, O., Muvunyi, C. M., and Ogbuagu, O. (2015). High prevalence of antimicrobial resistance among common bacterial isolates in a tertiary healthcare facility in Rwanda. *Am. J. Trop. Med. Hyg.* 92, 865–870. doi: 10.4269/ajtmh.14-0607

Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, J. A., et al. (2016). SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genom.* 2:e000056. doi: 10.1099/mgen.0.000056

Pallen, M. J., and Wren, B. W. (2007). Bacterial pathogenomics. *Nature* 449, 835–842. doi: 10.1038/nature06248

Pitout, J. D. D. (2012). Extraintestinal pathogenic *Escherichia coli*: a combination of virulence with antibiotic resistance. *Front. Microbiol.* 3:9. doi: 10.3389/fmicb.2012.00009

Poirel, L., Lartigue, M.-F., Decousser, J.-W., and Nordmann, P. (2005). ISEcp1B-mediated transposition of blaCTX-M in *Escherichia coli*. *Antimicrob. Agents Chemother.* 49, 447–450. doi: 10.1128/AAC.49.1.447-450.2005

Rahman, M. T., Sobur, M. A., Islam, M. S., Ievy, S., Hossain, M. J., El Zowalaty, M. E., et al. (2020). Zoonotic diseases: etiology, impact, and control. *Microorganisms* 8:1405. doi: 10.3390/microorganisms8091405

Rahuma, N., Ghenghesh, K. S., Ben Aissa, R., and Elamaari, A. (2005). Carriage by the housefly (*Musca domestica*) of multiple-antibiotic-resistant bacteria that are potentially pathogenic to humans, in hospital and other urban environments in misurata, libya. *Ann. Trop Med. Parasitol.* 99, 795–802.

Ranjan, A., Scholz, J., Semmler, T., Wieler, L. H., Ewers, C., Müller, S., et al. (2018). ESBL-plasmid carriage in E. coli enhances in vitro bacterial competition fitness and serum resistance in some strains of pandemic sequence types without overall fitness cost. *Gut. Pathog.* 10:24. doi: 10.1186/s13099-018-0243-z

Russo, T. A., and Johnson, J. R. (2000). Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli*: ExPEC. *J. Infect. Dis.* 181, 1753–1754. doi: 10.1086/315418

Schaufler, K., Nowak, K., Düx, A., Semmler, T., Villa, L., Kourouma, L., et al. (2018). Clinically relevant ESBL-producing K. pneumoniae ST307 and E. coli ST38 in an urban west african rat population. *Front. Microbiol.* 9:150. doi: 10.3389/fmicb.2018.00150

Schaufler, K., Semmler, T., Pickard, D. J., Toro, M., de La Cruz, F., de, et al. (2016a). Carriage of extended-spectrum beta-lactamase-plasmids does not reduce fitness but enhances virulence in some strains of pandemic E. coli lineages. *Front. Microbiol.* 7:336. doi: 10.3389/fmicb.2016.00336

Schaufler, K., Semmler, T., Wieler, L. H., Wöhrmann, M., Baddam, R., Ahmed, N., et al. (2016b). Clonal spread and interspecies transmission of clinically relevant ESBL-producing *Escherichia coli* of ST410–another successful pandemic clone? *FEMS Microbiol. Ecol.* 92:fiv155. doi: 10.1093/femsec/fiv155

Schaufler, K., Semmler, T., Wieler, L. H., Trott, D. J., Pitout, J., Peirano, G., et al. (2019). Genomic and functional analysis of emerging virulent and multidrug-resistant *Escherichia coli* lineage sequence type 648. *Antimicrob. Agents Chemother.* 63, e219–e243. doi: 10.1128/AAC.00243-19

Schwaber, M. J., and Carmeli, Y. (2007). Mortality and delay in effective therapy associated with extended-spectrum beta-lactamase production in *Enterobacteriaceae* bacteraemia: a systematic review and meta-analysis. *J. Antimicrob. Chemother.* 60, 913–920. doi: 10.1093/jac/dkm318

Shaik, S., Ranjan, A., Tiwari, S. K., Hussain, A., Nandanwar, N., Kumar, N., et al. (2017). Comparative genomic analysis of globally dominant ST131 clone with other epidemiologically successful extraintestinal pathogenic *Escherichia coli* (ExPEC) lineages. *mBio* 8, e1517–e1596. doi: 10.1128/mBio.01596-17

The European Committee on Antimicrobial Susceptibility Testing. (2021). *Breakpoint tables for interpretation of MICs and zone diameters: Version 11.0*. Available online at: http://www.eucast.org

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., et al. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.* 5:e1000344. doi: 10.1371/journal.pgen.1000344

Tufa, T. B., Fuchs, A., Wienemann, T., Eggers, Y., Abdissa, S., Schneider, M., et al. (2020). Carriage of ESBL-producing Gram-negative bacteria by flies captured in a hospital and its suburban surroundings in ethiopia. *Antimicrob. Resist. Infect. Control* 9:175. doi: 10.1186/s13756-020-00836-0

Tumbarello, M., Sanguinetti, M., Montuori, E., Trecarichi, E. M., Posteraro, B., Fiori, B., et al. (2007). Predictors of mortality in patients with bloodstream infections caused by extended-spectrum-beta-lactamase-producing *Enterobacteriaceae*: importance of inadequate initial antimicrobial treatment. *Antimicrob Agents Chemother.* 51, 1987–1994. doi: 10.1128/AAC.01509-06

Tyson, G. H., Li, C., Hsu, C.-H., Ayers, S., Borenstein, S., Mukherjee, S., et al. (2020). The mcr-9 gene of *Salmonella* and *Escherichia coli* is not associated with colistin resistance in the united states. *Antimicrob. Agents Chemother.* 64, e520–e573. doi: 10.1128/AAC.00573-20

Ugochukwu, G. (2013). Roles of nurses in sub-saharan african region. *Int. J. Nurs. Midwifery* 5, 117–131. doi: 10.5897/IJNM2013.0104

Vangchhia, B., Abraham, S., Bell, J. M., Collignon, P., Gibson, J. S., Ingram, P. R., et al. (2016). Phylogenetic diversity, antimicrobial susceptibility and virulence characteristics of phylogroup F *Escherichia coli* in australia. *Microbiology* 162, 1904–1912.

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

Whittam, T. S., Ochman, H., and Selander, R. K. (1983). Geographic components of linkage disequilibrium in natural populations of *Escherichia coli*. *Mol. Biol. Evol.* 1, 67–83. doi: 10.1093/oxfordjournals.molbev.a040302

Yoon, E.-J., Gwon, B., Liu, C., Kim, D., Won, D., Park, S. G., et al. (2020). Beneficial chromosomal integration of the genes for CTX-M extended-spectrum β-lactamase in *Klebsiella pneumoniae* for stable propagation. *mSystems* 5, e420–e459. doi: 10.1128/mSystems.00459-20

Zogg, A. L., Simmen, S., Zurfluh, K., Stephan, R., Schmitt, S. N., and Nüesch-Inderbinen, M. (2018). High prevalence of extended-spectrum β-lactamase producing *Enterobacteriaceae* among clinical isolates from cats and dogs admitted to a veterinary hospital in switzerland. *Front. Vet. Sci.* 5:62. doi: 10.3389/fvets.2018.00062

Zurita, J., Yánez, F., Sevillano, G., Ortega-Paredes, D., Paz, Y., and Miño, A. (2020). Ready-to-eat street food: a potential source for dissemination of multidrug-resistant *Escherichia coli* epidemic clones in quito, ecuador. *Lett. Appl. Microbiol.* 70, 203–209.

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

frontiers in Cellular and Infection Microbiology

MINI REVIEW
published: 18 May 2021
doi: 10.3389/fcimb.2021.636290

Centralised or Localised Pathogen Whole Genome Sequencing: Lessons Learnt From Implementation in a Clinical Diagnostic Laboratory

Alicia G. Beukers[1], *Frances Jenkins*[1] *and Sebastiaan J. van Hal*[1,2]*

[1] Department of Microbiology and Infectious Diseases, Royal Prince Alfred Hospital, Sydney, NSW, Australia, [2] Faculty of Medicine, University of Sydney, Sydney, NSW, Australia

OPEN ACCESS

Edited by:
Guido Werner,
Robert Koch Institute (RKI),
Germany

Reviewed by:
Artur J. Sabat,
University Medical Center
Groningen, Netherlands
Bryan Schmitt,
Indiana University Bloomington,
United States

***Correspondence:**
Sebastiaan J. van Hal
Sebastiaan.vanhal@health.nsw.gov.au

Specialty section:
This article was submitted to
Clinical Microbiology,
a section of the journal
Frontiers in Cellular
and Infection Microbiology

Received: 04 February 2021
Accepted: 14 April 2021
Published: 18 May 2021

Citation:
Beukers AG, Jenkins F
and van Hal SJ (2021)
Centralised or Localised Pathogen
Whole Genome Sequencing:
Lessons Learnt From
Implementation in a Clinical
Diagnostic Laboratory.
Front. Cell. Infect. Microbiol. 11:636290.
doi: 10.3389/fcimb.2021.636290

Whole genome sequencing (WGS) has had widespread use in the management of microbial outbreaks in a public health setting. Current models encompass sending isolates to a central laboratory for WGS who then produce a report for various levels of government. This model, although beneficial, has multiple shortcomings especially for localised infection control interventions and patient care. One reason for the slow rollout of WGS in clinical diagnostic laboratories has been the requirement for professionally trained personal in both wet lab techniques and in the analysis and interpretation of data, otherwise known as bioinformatics. A further bottleneck has been establishment of regulations in order to certify clinical and technical validity and demonstrate WGS as a verified diagnostic test. Nevertheless, this technology is far superior providing information that would normally require several diagnostic tests to achieve. An obvious barrier to informed outbreak tracking is turnaround time and requires isolates to be sequenced in real-time to rapidly identify chains of transmission. One way this can be achieved is through onsite hospital sequencing with a cumulative analysis approach employed. Onsite, as opposed to centralised sequencing, has added benefits including the increased agility to combine with local infection control staff to iterate through the data, finding links that aide in understanding transmission chains and inform infection control strategies. Our laboratory has recently instituted a pathogen WGS service within a diagnostic laboratory, separate to a public health laboratory. We describe our experience, address the challenges faced and demonstrate the advantages of de-centralised sequencing through real-life scenarios.

Keywords: centralised, localised, whole genome sequencing, clinical microbiology, pathogen genomics, public health

INTRODUCTION – PATHOGEN WHOLE GENOME SEQUENCING: CURRENT SITUATION AND FUTURE DIRECTIONS

Outbreaks of multi-drug resistant (MDR) pathogens represent a constant threat to healthcare systems worldwide. Infections with MDR pathogens are associated with higher morbidity and mortality rates than susceptible pathogens due to limited available treatment options (Scott, 2009; Vincent et al., 2009; Gandra et al., 2019). Effective infection control is therefore important for mitigating potential outbreaks.

Pathogen whole genome sequencing (WGS) has been demonstrated as a useful tool for identifying outbreaks and indicating points of intervention (Arnold, 2015; Gilchrist et al., 2015; Brown et al., 2019). WGS has become progressively available due to its increased affordability, which has led to its improved uptake for pathogen WGS in a clinical diagnostic space (Deurenberg et al., 2017; Berry et al., 2020). This technology is constantly evolving with some of the more recent advancements including a more rapid turn-around time of results, especially in relation to technologies such as Oxford Nanopore Technologies (Tyler et al., 2018).

Currently, pathogen WGS follows a centralised sequencing approach modelled off a public health process where isolates are sequenced at a centralised laboratory, analysed and a generic analysis report produced (Grant et al., 2018). Although this model is beneficial, by reducing the costs associated with WGS, the turnaround time can be quite long, effectively reducing the practicality of using WGS data in outbreak investigations. Further, there is disconnect between those requesting the data and those analysing and interpreting it, limiting its useability.

A transition to WGS at a local level would therefore be more suitable for use in a diagnostic laboratory, as data can be generated and analysed in near real-time. Uptake of WGS at a local level, however, has mostly been hindered by the lack of staff with the necessary bioinformatics skills needed for the analysis of data generated. Furthermore, establishment of necessary regulations are still under development to try and standardise this technology across multiple sites.

Here we review our experience with implementing a pathogen WGS service at a localised level in a clinical diagnostic space and demonstrate the benefits of establishing such a service at a local vs a centralised level.

CENTRALISED VS LOCALISED SEQUENCING – SHORTCOMINGS OF A PUBLIC HEALTH MODEL IN A CLINICAL DIAGNOSTIC SETTING

Improvements in next generation sequencing (NGS) technologies, including development of benchtop sequencing and remote cloud computing and data storage, are helping to support a move to a decentralised model. Our laboratory currently houses an Illumina MiSeq, an Illumina iSeq and more recently an Oxford Nanopore MinION platform. These instruments take up a relatively small benchtop area of between 3.8, 1.0 m^2 and the size of the computer used to run the MinION platform respectively. Our laboratory started sequencing bacterial isolates in 2017 and since then has undergone a successful accreditation review by the Australian regulatory body; National Association of Testing Authorities, Australia (NATA) in 2019.

Our sequencing service currently serves our hospital, a quaternary 900-bed referral hospital that provides several state-wide services, including liver transplantation. Bacterial isolates are sequenced in our laboratory for a number of purposes including outbreak investigations, confirmation of species identification and investigation and confirmation of antibiotic resistance and virulence genes. Sequencing requests are made by our infectious disease specialists with both the wet- and dry-lab procedures undertaken by our microbiologists on site. Our laboratory performs pathogen WGS on approximately 1000 bacterial isolates per year. Pathogen WGS is incorporated into our routine laboratory workflow. In brief, bacterial cultures are first identified by species through matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry (MS). Susceptibility testing is then performed using the Vitek and complimented with disc diffusion and E-tests. Further confirmatory testing is performed by our molecular department using PCR for confirmation of vancomycin resistance genes, methicillin-resistant *Staphylococcus aureus* (MRSA), and confirmation of carbapenem resistance genes in Enterobacteriaceae isolates with a raised meropenem MIC. Following WGS, bioinformatic analysis of the samples are performed on site using an *in-house* bioinformatics pipeline by a qualified microbiologist and a customised report generated. A schematic overview of our laboratory process is presented in **Figure 1**. There are several major benefits that are experienced by sequencing at a local level compared to centralisation. First, reduced turnaround time is the most significant improvement as samples do not have to be transported to external sites for sequencing. Moreover, current clinical microbiology practice involves several processes which can take several days for culturing, species identification and susceptibility testing, and weeks for molecular typing. For slow growing bacteria, results can take months to be obtained. WGS can produce all this information in as little as a week (Didelot et al., 2012). Prioritisation of samples at a centralised laboratory is often based on a first-in first-served basis due to the large quantities of samples received which also adds to the turnaround time. At a localised level, samples can be selected based on priority. Fast turnaround times, so the data can be assessed and interpreted in real-time, are especially important in assessing outbreaks, maximising the benefits of utilising WGS technology.

In a centralised laboratory, there is a major disconnect between the samples sequenced and their associated metadata, limiting connections that can be made between isolates. In our laboratory, we undertake a cumulative analysis approach whereby each successive isolate of the same bacterial species is analysed as a group in addition to individual case studies. This allows connections between isolates to be established which may

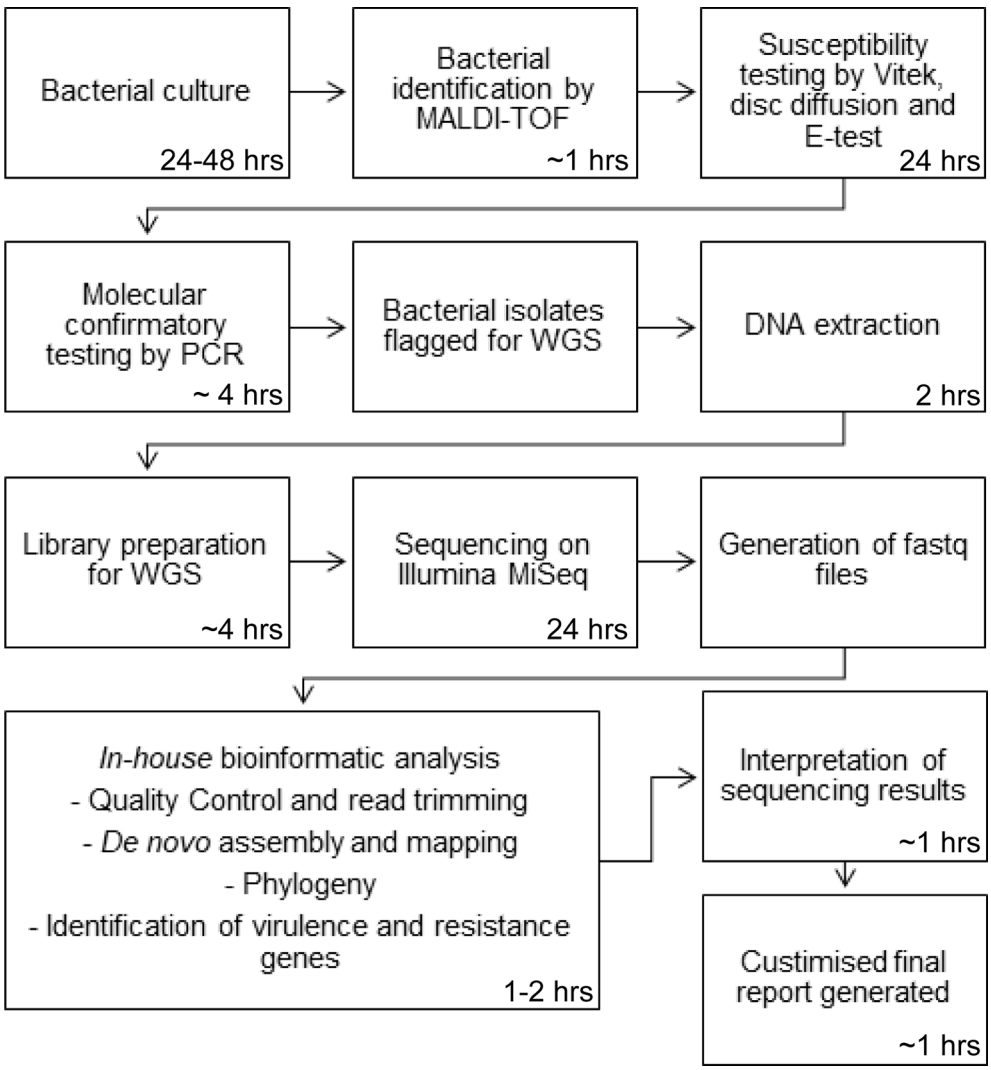


FIGURE 1 | Schematic overview of the laboratory process for WGS.

not necessarily be obvious due to large timespans between collection and sequencing. Further, with centralised sequencing and generation of a generic report there is limited ability for interpretation. At a localised level there is a greater degree of communication between clinicians, microbiologist and bioinformaticians allowing shared knowledge and greater understanding of the data generated.

An additional strength of sequencing at a local level is the small number of samples that are considered important. In our setting, a single MRSA from NICU is considered worthy of sequencing with detection of an extended outbreak over time. In a centralised setting, such as in public health, the importance of a single isolate is not recognised and would not warrant sequencing unless a cluster of isolates with suggested links are identified.

The largest bottleneck presented to decentralisation of WGS is the bioinformatics. This is becoming less of an issue as more graphical user-friendly interfaces are being developed (e.g., Galaxy, Illumina's Basespace). Development of effective semi-automatic pipelines have also allowed those with basic bioinformatics skills to analyse data. However, for a more in-depth or customised analysis, some knowledge of Unix-systems is required in order to utilise the large number of free bioinformatics software packages available. Furthermore, scientific knowledge of the biology of various micro-organisms that are investigated as well as their important genomic features are essential for a comprehensive analysis.

The final obstacle for a decentralised model is development of an accreditation process to standardise testing across multiple laboratory sites. This is still in the development phase, but as more and more laboratories decide to uptake this technology, the accreditation process will become more streamlined and robust. Some of the caveats to accreditation include both wet- and dry-lab parts requiring validation, the methods of WGS are applicable to all microbial species so standardisation needs to be broad and

validation of the WGS workflow may have to be performed using a selection of species as opposed to every species. As WGS is far superior to many of the technologies currently used in the laboratory, it can also provide multiple pieces of information such as phylogeny, antibiotic and virulence genes and MLST, with each of these aspects requiring verification.

REQUIREMENTS AND REGULATION OF PATHOGEN WHOLE GENOME SEQUENCING

NATA is the regulatory body responsible for certifying laboratories across Australia that they comply with regulatory standards and give confidence in the data and reports produced to those seeking their services. A number of criteria are required to be met for a laboratory to receive certification and are based on relevant international standards (e.g., ISO/IEC 17025, ISO 15189, ISO/IEC 17020). The process of accreditation involves a number of steps including first an enquiry followed by an advisory visit, application, document review, assessment and finally accreditation. After the accreditation process is complete, scheduled reassessments and surveillance visits become incorporated into the laboratory process to ensure continued compliance. To help laboratories ensure they comply with NATA, the accreditation process is administered jointly with the RCPA (Royal College of Pathologist Australasia).

WGS has been used routinely in human medical genomics, with standards well established within this field (ISO 15189 Medical Laboratories – Requirements for quality and competence, Requirements for human medical genome testing utilising massively parallel sequencing technologies). Currently, there is no specific guidance regarding quality issues, validation and requirements of supervision using massively parallel sequencing technology in relation to microbial sequencing.

One of the major differences when applying frameworks used in medical genomics is the focus of reporting and sequencing directed at target regions or genes with strict guidelines for the interpretation of mutations. For pathogen WGS, this is less straight forward as the whole genome of micro-organisms are being assessed. The data is

TABLE 1 | Recent publications demonstrating applicability of bacterial WGS in clinical diagnostic microbiology.

Bacteria	Title of study	Investigation type	Methodology	Major findings	Ref
Staphylococcus aureus	A multicentre outbreak of ST45 MRSA containing deletions in the spa gene in New South Wales, Australia	Epidemiology, assessment of diagnostic tests	WGS (Illumina MiSeq), read mapping (BWA), SNP calling (freebayes), assembly (SPAdes), phylogeny (treeAnnotator program)	Identified deletion in *spa* gene of *S. aureus* by WGS that inhibited Cepheid Xpert® MRSA/SA BC test therefore informing company of corrective modifications to be made in real time	Beukers et al., 2020
Escherichia coli, Enterococcus faecalis, Enterococcus faecium, Mycobacterium tuberculosis, S. aureus, Streptococcus pneumoniae, Staphylococcus epidermidis	Recommendations to address the difficulties encountered when determining linezolid resistance from whole-genome sequencing data	Antibiotic resistance	Bioinformatics, visualisation (CLC genomics Workbench)	Identification of linezolid resistance site (G2576T) in various organisms	Beukers et al., 2018
Enterococcus faecium	Relentless spread and adaptation of non-typeable vanA vancomycin-resistant *Enterococcus faecium*: a genome-wide investigation	Epidemiology, outbreak investigation	WGS (NextSeq 500), mapping (BWA; Stampy), SNP calling (freebayes), assembly (SPAdes), MLST, phylogeny (RaxML), mutation rate (BEAST)	Emergence of vanA pstS negative *E. faecium* in NSW with evidence of outbreaks and inter-hospital transmission	van Hal et al., 2018
Cronobacter sakazakii	*Cronobacter sakazakii* infection from expressed breast milk, Australia	Epidemiology, outbreak investigation	WGS (Illumina MiSeq), mapping (BWA), SNP calling (freebayes), phylogeny (FastTree)	Epidemiological link established between contaminated expressed breast milk and infant infection with *C. sakazakii*	McMullan et al., 2018
Enterococcus faecium	Failure of daptomycin β-lactam combination therapy to prevent resistance emergence in *Enterococcus faecium*	Antibiotic resistance	WGS (Ion Torrent PGM), mapping (CLC Genomics Workbench)	Variable daptomycin MICs associated with the same mutation in *liaF* and *cls* genes	Menon et al., 2018
Salmonella enterica, *Klebsiella pneumoniae*	Hospital acquisition of New Delhi Metallo β-Lactamase type-1 (NDM-1) *Salmonella enterica* through inter-species plasmid transmission	Epidemiology, outbreak investigation, horizontal gene transfer	WGS (Illumina MiSeq, Oxford Nanopore Technologies), mapping (BWA), SNP calling (freebayes), assemblies (Skesa, Unicycler), antimicrobial resistance (AMRFinder)	Confirmation of two separate outbreaks; one of *K. pneumoniae* and one of *Salmonella* enterica. Epidemiological links established between two outbreaks with inter-species plasmid transfer of NDM-1 established	Beukers et al.,
Haemophilus influenza	Whole genome sequencing identifies opportunistic non-typeable *Haemophilus influenza* rather than a hypervirulent clone	Epidemiology, virulence	WGS (Illumina MiSeq), assembly (SKESA), MLST, mapping (BWA), SNP calling (freebayes), phylogeny (FastTree)	Four cases of invasive non-typeable *H. influenza* infection were demonstrated to be unrelated and with no novel virulence factors	John et al., 2020

somewhat open to interpretation and requires educated judgement for interpreting the results and for making decisions.

As with any other microbiology test demonstrating ongoing quality proficiency is necessary. Although programs exist, these are likewise directed at public health laboratories. A pilot program has been established by RCPA for assessing WGS of infectious agents, with ambitions to continue this program into the future.

Overall, for pathogen WGS to transition to a localised level, it is important that these requirements and regulations become well established. This will assist diagnostic laboratories in the set up and formation of pathogen WGS in their own facilities. It will also ensure the services provided at different locations become standardised, providing more confidence in a decentralised model over a centralised one.

BENEFITS OF ONSITE SEQUENCING – CASE STUDIES

Pathogen WGS has been routinely performed in our hospital for the last two years. This has been beneficial in a clinical diagnostic space. The reasons for testing can be categorised into three broad categories: patient-centred, hospital-centred and/or investigational.

Patient-centred sequencing directly impacts patient care or outcomes for example, we sequence all *Burkholderia cepacia* complex isolates from respiratory tract samples of cystic fibrosis patients to confirm speciation. These results influence decisions around suitability for lung transplantation.

Similarly, in hotel quarantine, decision about care and release from quarantine are based on SARS-CoV-2 genomics. Our laboratory thus began using WGS sequencing to identify variants of concern (VOC) of SARS-CoV-2 in real-time. To enable a faster throughput and ability to sequence a single isolate, SARS-CoV-2 WGS long-read sequencing is performed using Oxford Nanopore Technologies (ONT) (Bull et al., 2020). One of the major advantages of introducing this technology has been the real time monitoring of genome coverage allowing us to generate a lineage report within 8 hours.

Other patient-centred applications included antimicrobial resistance testing, antiviral resistance testing and metagenomics, although the later remains under investigation. Additional examples of the uses of pathogen WGS are outlined in some of our recent publications listed in **Table 1**.

To highlight how pathogen WGS at a local level has benefited our infection control team (hospital-centred sequencing) we outline an outbreak investigation that spanned over a 17-month period (**Figure 2**). Our laboratory routinely sequences Enterobacteriaceae isolates with a raised meropenem MIC, indicative of carbapenem resistance. From January 2018 until July 2018, our hospital detected several patients colonised with NDM-1 producing *Klebsiella pneumoniae*. Analysis of the sequencing data in conjunction with

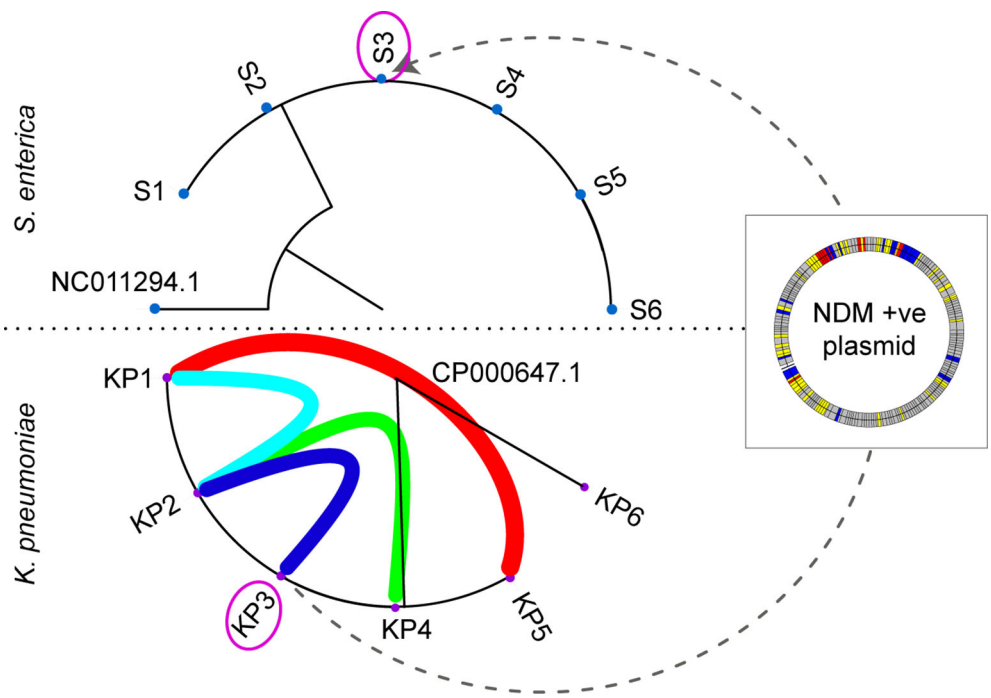


FIGURE 2 | Maximum likelihood phylogeny of the Salmonella enterica (S1-S6) and Klebsiella pneumoniae (KP1-KP6) outbreaks. The purple circled isolate (KP3 and S3) originate from the same patient. The coloured arcs in the K. pneumoniae outbreak demonstrate the epidemiological links between patients, with KP6 confirmed as unrelated to the outbreak. The dotted line between KP3 and S3 indicates the movement of the NDM-1 plasmid from K. pneumoniae and S. enterica within patient 3. The NDM-1 plasmid is depicted in the overlayed box with coloured bands representative of different genes (grey: hypothetical, yellow: mobile genetic elements, red: antibiotic resistance genes, and blue: other).

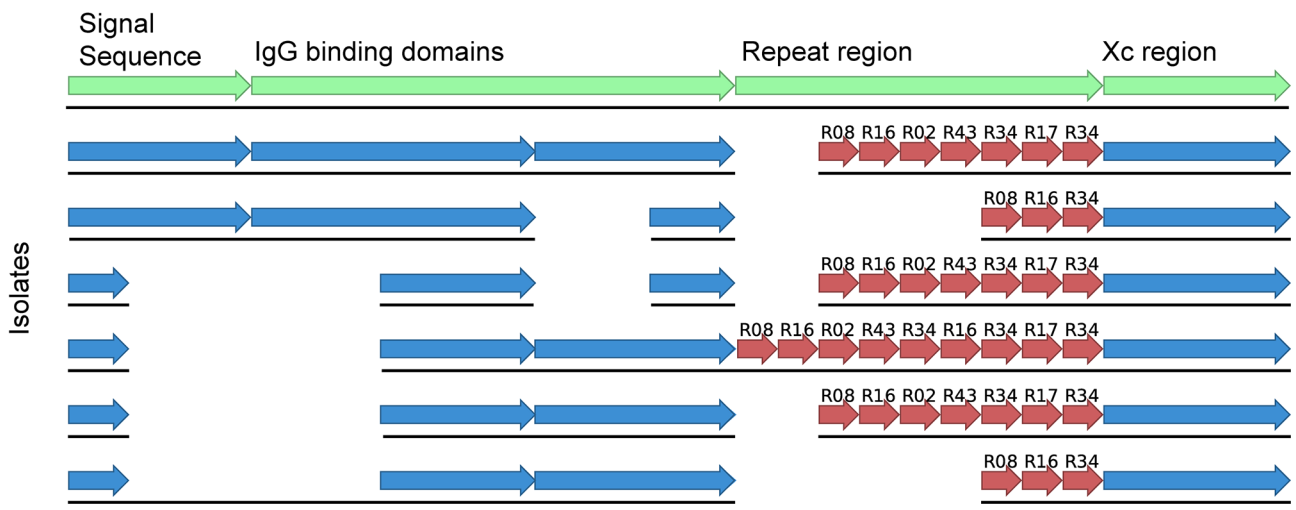


FIGURE 3 | Alignment of the *spa* gene from ST45 MRSA of various *spa* types and indicating the detected deletion event contained within the signal sequence and the IgG binding domains. The first sequence shows the various regions present in the *spa* gene in green whilst the sequences below indicate isolates of different *spa* types with or without the deletion.

the epidemiological data provided by our infection control staff we were able to confirm an outbreak. Genomic linkages were also fed back leading to additional reviews of bed movements and exploration of possible "missed" transmission events. With near real-time analysis we were able to direct interventions and identify areas for environmental testing.

Around 7 months after this outbreak was detected, we had increased incidence of *Salmonella enterica* in several patients. WGS combined with epidemiological data, similarly, confirmed *S. enterica* outbreak. Interestingly, one of the patients carried an NDM-1 producing *S. enterica* and had co-located on the same ward as one of the patients involved in the NDM-1 *K. pneumoniae* outbreak. We investigated this further and established inter-species transmission of a plasmid carrying the NDM-1 resistance gene had occurred between *K. pneumoniae* and *S. enterica*.

The connections determined between these two separate outbreaks would unlikely to have been recognised if these isolates were sequenced at a centralised level, considering the timespan that occurred between these two events. As our infectious disease specialists and scientists responsible for performing WGS communicate closely, links between these two outbreaks became more readily apparent and the analysis was able to be modified accordingly.

Finally, WGS has allowed us to investigate unexpected testing features encountered in the diagnostic laboratory. We noticed an increase in MRSA isolates that were mis-identified as coagulase-negative *S. aureus*. WGS was able to identify a deletion event in the *spa* gene that occurred during the emergence of ST45 MRSA in NSW (**Figure 3**). This deletion made the assay unreliable as a diagnostic test. The company was informed of our results and they were able to modify their diagnostic assay accordingly (Beukers et al., 2020).

CONCLUSION

The benefits of onsite pathogen WGS have been clearly demonstrated, with the most advantageous factor of localised compared centralised pathogen WGS an increased turnaround time. As WGS costs continue to decrease and the technology continues to advance, it is becoming more and more feasible for smaller laboratories to implement this technology.

AUTHOR CONTRIBUTIONS

AB and SH wrote the initial manuscript. FJ and SH revised the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

Arnold, C. (2015). Outbreak Breakthrough: Using Whole-Genome Sequencing to Control Hospital Infection. *Environ. Health Perspect.* 123 (11), A281–A286. doi: 10.1289/ehp.123-A281

Berry, I. M., Melendrez, M. C., Bishop-Lilly, K. A., Rutvisuttinunt, W., Pollett, S., Talundzic, E., et al. (2020). Next Generation Sequencing and Bioinformatics Methodologies for Infectious Disease Research and Public Health: Approaches, Applications, and Considerations for Development of Laboratory Capacity. *J. Infect. Dis.* 221 (Supplement_3), S292–S307. doi: 10.1093/infdis/jiz286

Beukers, A. G., Hasam, H., Hegstad, K., and van Hal, S. J. (2018). Recommendations to Address the Difficulties Encountered When Determining Linezolid Resistance From Whole-Genome Sequencing Data. *Antimicrob. Agents Chemother.* 62 (8), e00613–e00618. doi: 10.1128/AAC.00613-18

Beukers, A. G., Newton, P., Hudson, B., Ross, K., Gottlieb, T., O'Sullivan, M., et al. (2020). A Multicentre Outbreak of ST45 MRSA Containing Deletions in the

Spa Gene in New South Wales, Australia. *J. Antimicrob. Chemother.* 75 (5), 1112–1116. doi: 10.1093/jac/dkz560

Brown, E., Uday, D., McGarry, S., and Gerner-Smidt, P. (2019). Use of Whole-Genome Sequencing for Food Safety and Public Health in the United States. *Foodborne Pathog. Dis.* 16 (7), 441–450. doi: 10.1089/fpd.2019.2662

Bull, R. A., Adikari, T. N., Ferguson, J. M., Hammond, J. M., Stevanovski, I., Beukers, A. G., et al. (2020). Analytical Validity of Nanopore Sequencing for Rapid SARS-CoV-2 Genome Analysis. *Nat. Commun.* 11 (1), 6272. doi: 10.1038/s41467-020-20075-6

Deurenberg, R., Bathoorn, E., Chlebowicz, M. A., Couto, N., Ferdous, M., García-Cobos, S., et al. (2017). Application of Next Generation Sequencing in Clinical Microbiology and Infection Prevention. *J. Biotechnol.* 243, 16–24. doi: 10.1016/j.jbiotec.2016.12.022

Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E., and Crook, D. W. (2012). Transforming Clinical Microbiology With Bacterial Genome Sequencing. *Nat. Rev. Genet.* 13 (9), 601–612. doi: 10.1038/nrg3226

Gandra, S., Tseng, K. K., Arora, A., Bhowmik, B., Robinson, M. L., Panigrahi, B., et al. (2019). The Mortality Burden of Multidrug-Resistant Pathogens in India: A Retrospective, Observational Study. *Clin. Infect. Dis.* 69 (4), 563–570. doi: 10.1093/cid/ciy955

Gilchrist, C. A., Turner, S. D., Riley, M. F., Petri, W. A., and Hewlett, E. L. (2015). Whole-genome Sequencing in Outbreak Analysis. *Clin. Microbiol. Rev.* 28 (3), 541–563. doi: 10.1128/CMR.00075-13

Grant, K., Jenkins, C., Arnold, C., Green, J., and Zambon, M. (2018). *Implementing Pathogen Genomics. A Case Study* (Accessed 3 September 2020).

John, M. A., Beukers, A. G., Chan, R., and van Hal, S. J. (2020). Whole Genome Sequencing Identifies Opportunistic non-Typeable *Haemophilus Influenza* Rather Than a Hypervirulent Clone. *Pathology* S0031–3025(20)30951-X. doi: 10.1016/j.pathol.2020.08.011

McMullan, R., Menon, V., Beukers, A. G., Jenson, S. O., van Hal, S. J., and Davis, R. (2018). *Cronobacter Sakazakii* Infection From Expressed Breast Milk, Australia. *Emerg. Infect. Dis.* 24 (2), 393–394. doi: 10.3201/eid2402.171411

Menon, V., Davis, R., Shackel, N., Espedido, B. A., Beukers, A. G., Jensen, S. O., et al. (2018). Failure of Daptomycin β-Lactam Combination Therapy to Prevent Resistance Emergence in *Enterococcus Faecium*. *Diagn. Microbiol. Infect. Dis.* 90 (2), 120–122. doi: 10.1016/j.diagmicrobio.2017.10.017

Scott, R. D. (2009) *The Direct Medical Costs of Healthcare-Associated Infections in US Hospitals and the Benefits of Prevention*. Available at: https://www.cdc.gov/hai/pdfs/hai/scott_costpaper.pdf (Accessed 3 September 2020).

Tyler, A. D., Mataseje, L., Urfano, C. J., Schmidt, L., Antonation, K. S., Mulvey, M. R., et al. (2018). Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. *Sci. Rep.* 8 (1), 10931. doi: 10.1038/s41598-018-29334-5

van Hal, S. J., Beukers, A. G., Timms, V. J., Ellem, J. A., Taylor, P., Maley, M. W., et al. (2018). Relentless Spread and Adaptation of non-Typeable Vana Vancomycin-Resistant *Enterococcus Faecium*: A Genome-Wide Investigation. *J. Antimicrob. Chemother.* 73 (6), 1487–1491. doi: 10.1093/jac/dky074

Vincent, J. L., Rello, J., Marshall, J., Silva, E., Anzueto, A., Martin, C. D., et al. (2009). International Study of the Prevalence and Outcomes of Infection in Intensive Care Units. *JAMA* 302 (21), 2323–2329. doi: 10.1001/jama.2009.1754

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

frontiers in Microbiology

METHODS
published: 28 May 2021
doi: 10.3389/fmicb.2021.649517

Decentralized Investigation of Bacterial Outbreaks Based on Hashed cgMLST

Carlus Deneke, Laura Uelze, Holger Brendebach, Simon H. Tausch and Burkhard Malorny*

Department Biological Safety, German Federal Institute for Risk Assessment, Berlin, Germany

Whole-genome sequencing (WGS)-based outbreak investigation has proven to be a valuable method for the surveillance of bacterial pathogens. Its utility has been successfully demonstrated using both gene-by-gene (cgMLST or wgMLST) and single-nucleotide polymorphism (SNP)-based approaches. Among the obstacles of implementing a WGS-based routine surveillance is the need for an exchange of large volumes of sequencing data, as well as a widespread reluctance to share sequence and metadata in public repositories, together with a lacking standardization of suitable bioinformatic tools and workflows. To address these issues, we present *chewieSnake*, an intuitive and simple-to-use cgMLST workflow. *ChewieSnake* builds on the allele calling software chewBBACA and extends it by the concept of allele hashing. The resulting hashed allele profiles can be readily compared between laboratories without the need of a central allele nomenclature. The workflow fully automates the computation of the allele distance matrix, cluster membership, and phylogeny and summarizes all important findings in an interactive HTML report. Furthermore, *chewieSnake* can join allele profiles generated at different laboratories and identify shared clusters, including a stable and intercommunicable cluster nomenclature, thus facilitating a joint outbreak investigation. We demonstrate the feasibility of the proposed approach with a thorough method comparison using publically available sequencing data for *Salmonella enterica*. However, *chewieSnake* is readily applicable to all bacterial taxa, provided that a suitable cgMLST scheme is available. The workflow is freely available as an open-source tool and can be easily installed via conda or docker.

Keywords: cgMLST, WGS typing, molecular surveillance, comparative microbial genomics, *Salmonella*

OPEN ACCESS

Edited by:
Kristin Hegstad, University Hospital of North Norway, Norway

Reviewed by:
Magaly Toro, University of Chile, Chile
Shabarinath Srikumar, United Arab Emirates University, United Arab Emirates

****Correspondence:***
Carlus Deneke
carlus.deneke@bfr.bund.de

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *19 January 2021*
Accepted: *25 March 2021*
Published: *28 May 2021*

Citation:
Deneke C, Uelze L, Brendebach H, Tausch SH and Malorny B (2021) Decentralized Investigation of Bacterial Outbreaks Based on Hashed cgMLST. Front. Microbiol. 12:649517. doi: 10.3389/fmicb.2021.649517

INTRODUCTION

Whole-genome sequencing (WGS)-based typing approaches allow the highly discriminatory comparison of the similarity of bacterial genomes. The results are used for the investigation of disease outbreaks, source attribution, contamination control, and surveillance of bacterial pathogens (Franz et al., 2016; Ronholm et al., 2016; Jagadeesan et al., 2019). A number of sequence-based typing approaches exist, each with their own advantages and drawbacks (Uelze et al., 2020b).

Abbreviations: AD, allele distance; API, application programming interface; BLAST, basic local alignment search tool; CDS, coding sequence; cgMLST, core genome MLST; DAG, directed acyclic graph; DNA, deoxyribonucleic acid; HTML, hypertext markup language; INDELs, insertion and deletions; ISO, International Organization for Standardization; NCBI, National Center for Biotechnology Information; MLST, multilocus sequence typing; NGS, next-generation sequencing; SNP, single-nucleotide polymorphism; SRA, sequence read archive; ST, sequence type; wgMLST, whole-genome MLST; WGS, whole-genome sequencing.

Generally, the analysis of single-nucleotide polymorphisms (SNPs) is considered the method with the highest resolution. In particular, the National Center for Biotechnology Information (NCBI) pathogen detection pipeline is an SNP-based molecular typing system for global microbial surveillance based on publically available sequencing data[1]. Another high-resolution typing approach is the core genome/whole-genome multilocus sequence typing (cg/wgMLST), which was derived from the concept of classic multilocus sequence typing (MLST), with the distinction that the initial seven-gene multi-locus scheme was expanded to hundreds or thousands of gene loci (Maiden et al., 2013). CgMLST is a gene-by-gene approach, which functions by aligning complete or draft genome assemblies to a scheme consisting of a set of loci and a collection of associated numbered allele sequences. One of the advantages of cgMLST is that no outbreak specific reference is required, and therefore, it is a suitable and unbiased method to identify possible clusters from samples from an entire species. During the allele calling step, each locus is searched in the assembly, and if it matches an existing allele sequence, the number of that allele is assigned. In the case that an allele sequence is not yet contained in the scheme, a new allele number is created, and the allele sequence is added for future inquiries. From the set of allele numbers for each locus, a so-called allele profile is derived. The similarity between two or several genomes is estimated by comparing their respective allele profiles and calculating the total number of different alleles. Allele differences are first determined pairwise, before a distance matrix is derived by cross-comparison for all samples. Finally, a phylogenetic tree can be computed from the distance matrix through various clustering techniques such as neighbor-joining, minimum-spanning trees, or hierarchical clustering. Large minimum-spanning trees can be visualized, e.g., with grapeTree[2] (Zhou et al., 2018) or PHLYOViZ[3] (Francisco et al., 2012).

Core genome multilocus sequence typing clustering results are dependent on the choice of cgMLST scheme, with the number and type of loci being of great importance. The loci for each scheme should be chosen carefully to account for the unique genetic background of different species. Generally, the chosen loci should be part of the core genome (hence the name) of a taxonomic group and as such be present in the majority of all isolates of this group. A number of cgMLST schemes for major species are curated by various (at times commercial) organizations and research groups, such as Enterobase[4], Institut Pasteur[5], Ridom SeqSphere+[6], or chewBACCA[7], and can often be obtained freely from their websites.

One major disadvantage of a classic numbered cgMLST allele sequence scheme is that results from different laboratories, even when using the same scheme, are not directly comparable. This is caused by the fact that different local instances of an (initially identical) cgMLST scheme quickly diverge from each other, as new allele sequences are added, leading to the assignment of the same allele number for different allele sequences, or of the same allele sequence to be attributed to two or more different allele numbers. This can only be prevented by real-time synchronization of the local allele database with a (centrally) curated cgMLST allele nomenclature server, as implemented in Ridom SeqSphere and chewieNS (Mamede et al., 2020).

One possible solution is the replacement of chronologically numbered allele numbers with numbers or strings that are directly related to the underlying allele sequence. Using the allele sequence itself is impractical, as it hugely increases the size of an allele profile. Alternatively, *allele hashes* can be derived directly and unambiguously from the allele sequence. Allele hashes uniquely map the nucleotide sequence into a fixed-size hash value. Thus, independently discovered, identical allele sequences always result in the identical allele hash. This has the major advantage of allowing a decentralized nomenclature-free allocation of sequencing types, with no need for harmonization with a central unit. The practicability of the hashing approach has been demonstrated by the bioinformatic tool SISTR (Yoshida et al., 2016) for the purpose of serotyping *Salmonella* spp. and for *Clostridium difficile* in (Eyre et al., 2019).

Existing bioinformatic cgMLST standalone tools differ not only in their implemented cgMLST scheme but also in their specific allele calling algorithm and whether the source code is freely available, i.e., open- or closed-source. Two popular commercial solutions (source code not freely available) with a graphical user interface (GUI), which can be run on Windows systems, are Ridom SeqSphere+[8] and Bionumerics[9]. Both employ a closed system with a central nomenclature server. In comparison, two open-source, command-line tools, MentaLiST (Feijao et al., 2018) and chewBACCA (Silva et al., 2018), allow users to utilize their own cgMLST scheme. ChewBBACA is a comprehensive pipeline for cgMLST calling, as well as for the creation of new cgMLST schemes. The allele calling algorithm of chewBACCA considers the Blast Score Ratio in order to determine the allele sequences. Among the advantages of chewBBACA is that it automatically incorporates novel alleles into a scheme. Its usability has furthermore been demonstrated in numerous studies (Macedo et al., 2019; Lüth et al., 2020; Pinto et al., 2020; Uelze et al., 2020a).

Here, we present *chewieSnake*, an automated analysis pipeline that encompasses the whole analysis process from reads or draft assemblies to a final user-friendly cgMLST report. The pipeline implements chewBBACA for allele calling and performs allele hashing, computation of an allele distance (AD) matrix, and a minimum-spanning tree, as well as a clustering analysis by AD. The analysis results are summarized in an interactive HTML report.

We demonstrate the workflow's usability by analyzing a large public database for *Salmonella enterica* and validate

[1]https://www.ncbi.nlm.nih.gov/pathogens/

[2]https://achtman-lab.github.io/GrapeTree/MSTree_holder.html

[3]https://online.phyloviz.net/index

[4]https://enterobase.warwick.ac.uk/

[5]https://bigsdb.pasteur.fr/

[6]https://www.cgmlst.org/ncs

[7]https://chewbbaca.online/

[8]https://www.ridom.de/seqsphere/

[9]https://www.applied-maths.com/applications/wgmlst

it by comparing it to two publically available, state-of-the-art methods—Enterobase and NCBI pathogen detection. Additionally, we simulate a scenario of decentralized allele calling and central cluster analysis.

Finally, we want to emphasize that the presented method is not restricted to *Salmonella* spp. but is readily applicable to all bacterial taxa, provided a suitable cgMLST scheme is available.

METHODS

Implementation and Availability

The provided software consists of two workflows—*chewieSnake* and *chewieSnake_join*. Additionally, several modules of the software are available on their own and can be applied in conjunction with other analysis workflows. The software is available open-source[10] and can be easily installed using bioconda (Grüning et al., 2018) or docker (Merkel, 2014).

ChewieSnake Workflow

ChewieSnake implements a workflow for allele calling, computation of the ADs, and a minimum-spanning tree, with analysis results summarized in an interactive HTML report. Central to the workflow is the concept of allele hashes that allow a nomenclature free comparison of allele profiles (see **Figure 1**).

Snakemake workflow

The workflow consists of a Snakemake pipeline (Koster and Rahmann, 2012) with a set of assemblies (or sequencing reads) as input listed in a sample sheet. Snakemake resolves all dependencies in a directed acyclic graph (DAG), runs all necessary components for each sample, merges all results, and renders the final HTML report using Rmarkdown (rmarkdown, 2020) (see also **Supplementary Figure 1**). The design of the workflow is such that samples may be added continuously without the need to re-compute analysis results for previously analyzed samples, thus optimizing the computational time/resources for maintaining an allele profile database containing a large number of samples. Nevertheless, the clustering, phylogeny, and report are updated whenever new samples are provided.

Allele calling

The basis for the central allele calling step is the assembly based allele caller chewBBACA (Silva et al., 2018). Its main concept lies in the identification of coding sequences (CDSs) using prodigal (Hyatt et al., 2010). While existing alleles are detected via exact sequence matching, novel alleles are quality controlled via a BLASTP routine. For the usage of chewBBACA in a Snakemake workflow—and in particular to allow the continuous addition of samples—a fixed version (2.12.0) was included in the *chewieSnake* repository with minor adaptations to their output: chewBBACA is called for one sample at a time, and the output is stored in a folder with the sample's name. Possibly conflicting parallel writing to the allele database is blocked.

[10]https://gitlab.com/bfr_bioinformatics/chewieSnake

Allele hashing

The allele profiles as provided from chewBBACA are reformatted into a GrapeTree compatible format; and for each sample, the allele numbers—as provided in the applied cgMLST scheme—are converted into allele hashes. This step guarantees that the same allele hash is assigned for any identical allele sequence. Thus, independent identification of alleles for different samples and on different computing facilities always leads to the same allele hash and therefore clustering result.

The provided function *alleleprofile_hasher.py* looks up the allele sequence for each allele number in an allele profile in the scheme and computes a CRC32 hash using python's zlib package, yielding a unique integer.

The function *alleleprofile_hasher.py* can be used as a standalone program, e.g., if a user prefers to use their own analysis workflow (based on chewBBACA) and thus allows the nomenclature free sharing of allele profiles regardless of the presented *chewieSnake* workflow.

Unique hashed sequence types

To generate unique sequence types for each sample (hashIDs), the hashing approach is also applied. For each allele profile, the allele hashes (including possibly missing loci) are combined into a single string, and a hash value of the string provides the unique sequence type. It should be noted that this approach is stricter than clustering samples at 0 AD, since missing alleles are explicitly accounted for.

The provided function *hashID.py* can be used standalone, thus enabling users to infer hashed sequence type for any kind of allele profile.

Allele distance matrix and minimum-spanning tree

Based on the allele profiles, the workflow uses GrapeTree (Zhou et al., 2018) for the computation of the AD matrix. GrapeTree allows the computation of the distance in various ways, which differ in how missing alleles are treated. In addition, GrapeTree is also used to infer the phylogeny of all samples with minimum-spanning trees. The resulting Newick file can be readily rendered with GrapeTree's graphic visualization software or other phylogenetic visualization tools. This step and all subsequent steps are repeated whenever new sample data are added to the workflow.

Sample clustering

ChewieSnake hierarchically clusters the distance matrix of all samples (using, e.g., single linkage or average linkage hierarchical clustering) using the R-function *hclust* (R Core Team, 2013). Subsequently, given the clustering, a set of pre-defined thresholds assign cluster numbers to all samples at different thresholds—using dendextend (Galili, 2015). This provides so-called *cluster addresses* or *cluster zip-codes*, which allow an additional description of the relatedness of a set of samples. This is a similar approach to SNP addressed in SnapperDB (Dallman et al., 2018) or HierCC in Enterobase (Zhou et al., 2020b).

The function *Clustering_DistanceMatrix.R* is provided in the scripts directory and can be used standalone for any kind of distance matrix, including SNP distance matrices.

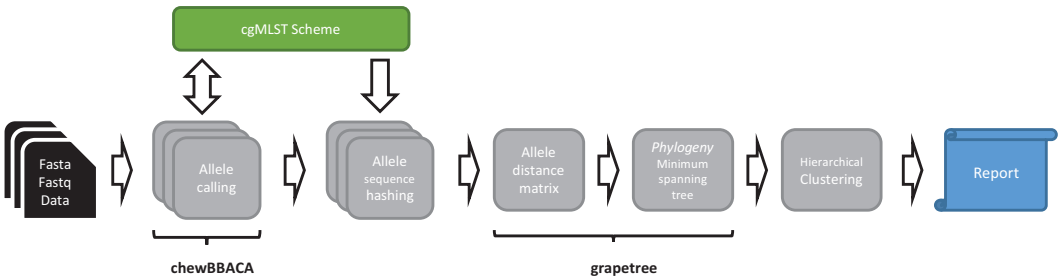


FIGURE 1 | Schematic representation of the *chewieSnake* workflow. The workflow starts from a set of assemblies (fasta data) or a set of raw reads (fastq data), which are then assembled into fasta files. On these, cgMLST allele calling is performed using the tool chewBBACA, utilizing a designated cgMLST scheme, to which newly found alleles are added in the process. The resulting allele profiles are converted to hashed allele profiles. All profiles are subsequently combined; and an allele distance matrix and a minimum-spanning tree phylogeny are computed with the tool GrapeTree. Then, cluster types are generated by hierarchical clustering at different allele distance thresholds. Finally, the results are summarized in an interactive HTML report.

HTML report

After processing and collection of all sample data, the entire dataset is summarized in an interactive user-friendly HTML report. The report is organized into different tabs and designed with the aim to give all scientists (regardless of bioinformatic background) access to the most important results:

In the *Allele statistics* tab, key quality assessment parameters of the allele calling step are shown for each sample, such as the number of alleles found—as well as the reason for missing alleles. The *Allele distance table* tab features a searchable table with all pairwise ADs. The *Allele distance matrix* tab provides a colored and zoomable visualization of the distance matrix. The *Clustering* tab presents a searchable table of the clustering address at the pre-defined thresholds. Furthermore, single-linkage trees are printed for all identified clusters at the pre-defined cluster threshold. The *Minimum-spanning tree* tab displays a simple (static) visualization of the phylogenetic tree. Lastly, the tabs *Links to files*, *Config and parameters*, and *Help* give further directions for more detailed analysis and enable optimal reproducibility.

An example report for a small test set can be found at https://bfr_bioinformatics.gitlab.io/chewieSnake/report_chewiesnake.html.

User input and parameter choice

The user has control over a large set of parameters, which are available with the command *chewieSnake.py –help*.

Importantly, if only reads are available for analysis, *chewieSnake* can also be given a set of reads as input, which it then trims using fastp (Chen et al., 2018) and subsequently assembles into draft genomes using shovill[11], before proceeding with the Snakemake workflow. Further important parameter choices are the GrapeTree distance method (see above), the clustering method, and thresholds, as well as chewBBACA specific parameters. All parameters are transparently saved and included in the report for reproducibility.

[11]https://github.com/tseemann/shovill

Possibilities in chewieSnake for inter-lab –comparison

The *chewieSnake* workflow contains the –comparison option to compare a set of query data with a pre-computed allele profile. This allows the identification of clusters between query data and comparison data. The workflow processes the query data in the same fashion as described above. When the (hashed) allele profiles are computed, these are joined with the comparison allele profiles. Next, all samples from the comparison allele profiles that match to any of the query data within a predefined *joining_threshold* are extracted; and a distance matrix, distance table (query vs. comparison data), and a minimum-spanning tree are computed. Again, all findings are summarized in an HTML report, highlighting the matches between the two datasets. In conclusion, this approach allows a quick screening of a potentially large comparison allele database for matches to queries of interest, e.g., in an outbreak situation. This strategy might also be useful when a comparison of newly sequenced data to existing allele data is desired.

An example report is provided at https://bfr_bioinformatics.gitlab.io/chewieSnake/report_chewiesnake_comparison.html.

ChewieSnake_Join Workflow

The central allele hashing concept within the *chewieSnake* workflow allows the nomenclature-free comparison of allele profiles generated by different laboratories. To facilitate the comparison and establish an inter-laboratory outbreak clustering nomenclature, the *chewieSnake_join* workflow conducts all necessary steps from individual *chewieSnake* results to a joined clustering report. The conceptual model is depicted in **Figure 2A**, whereas the essential steps are shown in **Figure 2B**. In the following, we describe the core elements of the workflow in more detail.

Compiling pre-computed allele data from a set of laboratories

All laboratories setup and run the *chewieSnake* workflow independently on their set of data. Laboratories can validate their results by inspecting the allele quality and the local clustering results.

The central results (provided in the files allele_profiles.tsv, allele_statistics.tsv, timestamps.tsv) need to be collected from

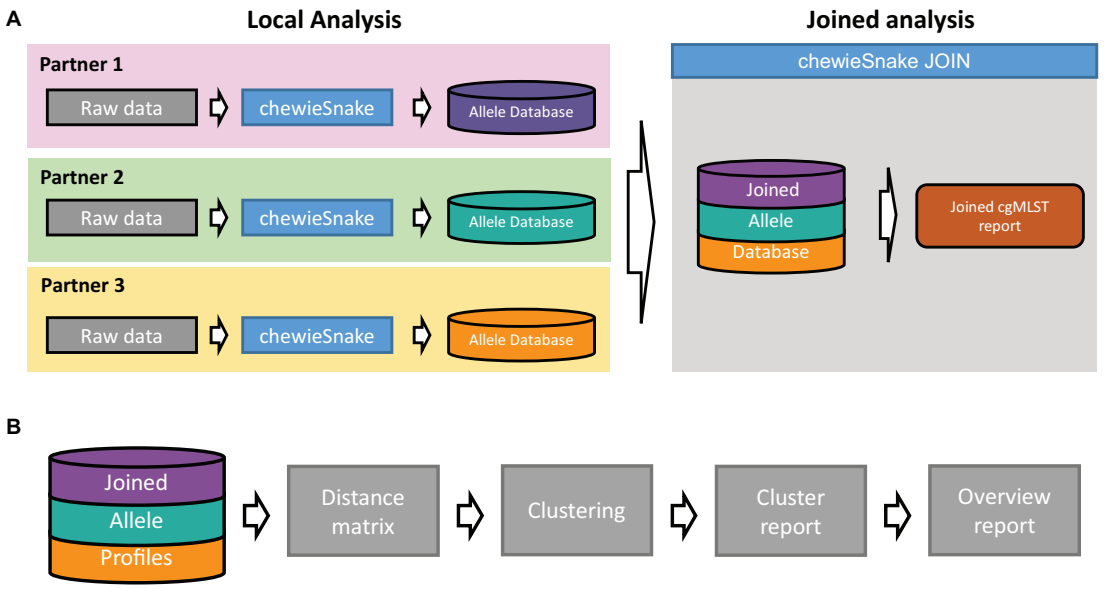


FIGURE 2 | Schematic representation of the *chewieSnake_join* workflow. **(A)** A set of laboratories locally employ the *chewieSnake* workflow to process incoming raw sequencing data. The resulting allele profiles are stored in individual allele databases, which are subsequently joined with the *chewieSnake_join* workflow. In addition, the *chewieSnake_join* workflow computes a joined cgMLST report. **(B)** Detailed description of the *chewieSnake_join* workflow. From the joined allele profiles, the allele distance matrix is computed, followed by cluster typing of the entire dataset. Detailed reports for each cluster as well as an overall report summarize the results in an interactive HTML format.

each laboratory. The location of all files is provided in an *allele sheet* (a tsv file containing each laboratories' acronym, as well as paths to the allele profiles, allele statistics, and timestamps). This is the input to the *chewieSnake_join* workflow.

The contributed data can be updated continuously. On each execution of *chewieSnake_join*, the workflow identifies new samples and repeats subsequent steps automatically where necessary.

Joining allele and associated data

The individual allele profiles (and other files) are joined into single files, and a file tying the laboratory of origin to the sample name is generated. Subsequently, in analogy to the *chewieSnake* workflow, a distance matrix and minimum-spanning tree are computed on the basis of the full dataset.

Cluster and subcluster identification

The joined AD matrix is input to the clustering module of *chewieSnake_join*. Again, hierarchical clustering together with a predefined threshold divides the dataset into individual clusters (and unrelated orphan samples). Since the clustering may assign new cluster numbers each time it is invoked, special emphasis is put on a stable clustering nomenclature: The cluster numbers are matched to a list of cluster names (which can be numbers such as *CT_0001* or predefined names such as *Cluster_Kairo*), whenever the clustering is repeated, e.g., after addition of more samples. Hereby, it is guaranteed that the same cluster name is assigned to the same set of samples. Exceptions such as the merging of two or more clusters are considered and handled adequately. In a similar fashion, for the samples in each cluster, a subclustering at a lower subclustering threshold is performed—again using stable names such as alpha and beta—thus allowing an intercommunicable cluster nomenclature at a finer level.

Additional options and parameters

Apart from setting clustering methods and thresholds, a number of additional options can be invoked in *chewieSnake_join*.

In particular, individual representative samples can be matched to external cluster names (using the *–external_cluster_names* flag). For instance, the latter can describe epidemiologically defined clusters—independent from the molecular data. *ChewieSnake_join* links these external cluster names to the internal cluster names, which may aid the tracking of established clusters in the report (see below). Additionally, samples can be assigned to e.g., serovars or larger clades using the *–serovar_info* flag. Again, this might facilitate the cluster analysis provided in the report. The matching of these additional data proceeds by providing a list of representative samples and their association to external clusters and serovar, respectively. *ChewieSnake_join* associates these representative samples to the identified clusters, and this association becomes accessible in the report.

The pool of names for the cluster naming can be modified by providing a list of available cluster names with the *–cluster_names* flag.

All available options and parameters can be inspected using *chewieSnake_join.py –help*.

Joined HTML report

Central to the facilitation of a joined outbreak analysis is the *chewiesnake_join* HTML report. This report summarizes all

findings from the data aggregation and clustering. It is organized into different tabs, as follows.

The *Overview tab* condenses the information about samples per laboratories, date of sample analysis, and number of clusters shared between laboratories. The *Cluster summary* tab provides the central information for the clustering analysis in a first (searchable) table that is organized by cluster. It provides readily accessible information such as cluster size, latest updates, duration (age) of cluster, cluster nomenclature, and the matching to external cluster names and serovar. Another table lists all samples and their respective cluster assignments. The *Orphans* tab displays all samples that, under the chosen threshold, do not belong to any existing cluster, together with their distance to the closest cluster. The *Inter-cluster relation* tab describes the more global relationship between clusters. The *Allele QC* tab allows an inspection of the allele quality for the entire dataset. The tabs *Links to files*, *Config and parameters*, and *Help* provide additional information.

An example report is provided at https://bfr_bioinformatics.gitlab.io/chewieSnake/report_chewiesnake_join.html.

For each identified cluster, a more detailed cluster report is available. This report summarizes the subclusters, the AD matrix of all samples in the cluster, the time evolution of the cluster (according to the timestamps of the analyses), and the cluster's and all subcluster's phylogeny.

An example report is provided at https://bfr_bioinformatics.gitlab.io/chewieSnake/clustering/CT_0004/clusterreport.html.

Dataset for Evaluation

Raw Data

In total, 1,263 WGS sequencing data were obtained from BioProject PRJEB31846. The dataset comprises diverse *S. enterica* serovars collected between the years 1999 and 2019 and sequenced by the National Reference Laboratory for *Salmonella* using the Nextera XT or DNA Flex kit (Illumina GmbH, München, Germany) on Illumina MiSeq and NextSeq instruments. The data are described in more detail in (Uelze et al., 2019).

Assembly

Data were trimmed [with fastp (Chen et al., 2018), version 0.19.5] and assembled [with shovill-spades (see text footnote 11), version 1.1.0] using the AQUAMIS pipeline[12] (version v1.2.0) (Deneke et al., 2021). All samples passed basic quality checks, such as sufficient base quality, coverage depth, genome length, and contig number. Furthermore, no evidence for sample contamination was detected. The assemblies are available under https://zenodo.org/record/4338293.

cgMLST Scheme

The cgMLST scheme for *Salmonella* was downloaded on June 11, 2018, from Enterobase (Zhou et al., 2020a) using the web API. As the sequencing data were published at a later date, this ensured that the downloaded cgMLST scheme did not already contain allele numbers from our dataset. Only this strategy allows an unbiased analysis of the allele calling process and in particular the assignment/identification of novel alleles, as would be the case for newly generated sequencing data unknown to a specific reference allele database. The scheme is available under https://zenodo.org/record/4724927.

Allele Calling

The assembled draft genomes were analyzed using the *chewieSnake* workflow. All parameters and software versions are specified in the **Supplementary Data Sheet 2**. The resulting *chewieSnake* report is available under https://bfr_bioinformatics.gitlab.io/chewiesnake_publicationdata/chewiesnake/cgmlst_report.html.

Simulation of Decentralized Analysis

The dataset was split randomly into three non-overlapping sub-datasets. The membership of each sample is listed within **Supplementary Data Sheet 1**. A cgMLST analysis with *chewieSnake* (with the same software version and parameters as for the full dataset) was performed separately on each sub-dataset. For each analysis, an identical but separate allele database was employed.

In order to simulate a decentralized surveillance system over a given time span, the timestamp files—which originally contain the date of the analyses—were modified to random dates from the period of January 2020 to October 2020. Other than for visualization purposes, this had no impact on the analyses.

Results were joined using the *chewieSnake* joining workflow (see above). All parameters are specified in the **Supplementary Data Sheet 2**. The resulting joined report is available under https://bfr_bioinformatics.gitlab.io/chewiesnake_publicationdata/chewiesnake_join/report.html.

Enterobase Data

Available cgMLST metadata for Bioproject PRJEB31846 were searched and downloaded from https://enterobase.warwick.ac.uk/species/index/senterica on November 2, 2020. The set of allele profiles was downloaded on November 2, 2020, and subsequently filtered for all entries of BioProject PRJEB31846. A total of 1,158 data were also found on Enterobase.

An AD matrix was computed from the allele profiles using GrapeTree in complete analogy to the *chewieSnake* workflow. Subsequently, the distance matrix was hierarchically clustered, and a cluster address was assigned using the provided script *Clustering_DistanceMatrix.R*. Thus, all analysis steps from the allele profiles were identical to the *chewieSnake* workflow.

National Center for Biotechnology Information Data

Salmonella SNP distances were obtained from NCBI Pathogen detection[13] on September 10, 2020, corresponding to version PDG000000002.1968. With the use of a custom script, the pairwise SNP distances were filtered such that all distances of pairs originating from Biosamples associated with BioProject PRJEB31846 were retained. Overall, 811 samples were part of an SNP cluster on NCBI pathogen, and the samples were found in 120 distinct SNP clusters. Note that the definition of an SNP

[12] https://gitlab.com/bfr_bioinformatics/AQUAMIS

[13] ftp://ftp.ncbi.nlm.nih.gov/pathogen/Results/Salmonella/

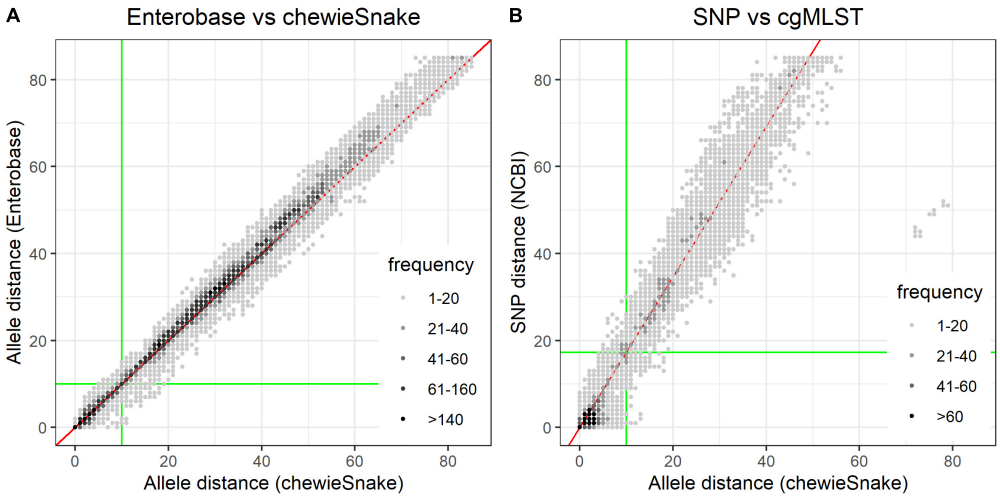


FIGURE 3 | Pairwise distances show very high method congruence. **(A)** Comparison of allele distance between Enterobase and *chewieSnake*. **(B)** Comparison of NCBI_pathogen SNP and *chewieSnake* allele distance. The gray shading indicates the frequency of observed pairwise distances. Both figures show the very high concordance between the different methods. This is also reflected in the Spearman correlation coefficients (0.984 and 0.973). The slope of a linear fit (red line) to the NCBI_SNP and *chewieSnake* comparison is ∼1.73; i.e., on average, 1 allele difference corresponds to 1.73 single-nucleotide polymorphisms. For the computation of sensitivity and specificity, the green lines mark clustering thresholds. All points in the lower left quadrant are counted as true positives, and all points in the upper right quadrant as true negatives, whereas the points in the top left and bottom right quadrants are false positives and false negatives, respectively. Using these thresholds also leads to the high adjusted Wallace and adjusted Rand coefficients (see **Tables 1**, **2**). Though overall the agreement is very high, some remaining differences at low distances may still influence interpretation of the very fine-grained phylogeny. Note that for the Enterobase comparison, only distances below 85 AD were displayed.

cluster on NCBI pathogen does not correspond to the hierarchical clustering studied in this contribution but rather to the set of samples with the same reference for SNP calling.

For this analysis, the SNP distance data were hierarchically clustered, and cluster numbers were assigned to each sample using a custom script in the same fashion as described above.

Methods for Comparison Analysis

Comparison of Pairwise Distances

Pairwise distances of each method were obtained as described in the individual sections above. Each resulting distance matrix was read into R and converted to a linearized table with the pair's sample names as keys. The linearized distance tables resulting from different methods were matched using the keys, followed by redundancy and self-hits removal. Based on these distance-method-comparison tables (provided in the **Supplementary Data sheet 1**), the Spearman correlation was computed, and the pairwise distance comparison plots were drawn (**Figures 3A,B**).

Computation of Sensitivity/Specificity Values

For each method combination, the pairwise distances of all sample pairs were obtained and a distance threshold was chosen, such that samples could be grouped into possible related clusters. Though no single fixed threshold is sufficient and advisable for the differentiation of outbreak to non-outbreak strains (Simon et al., 2018; Radomski et al., 2019), an allele threshold of 10 is considered suitable for an initial clustering (Besser et al., 2019; Uelze et al., 2021). For comparability with cgMLST clustering results, an SNP threshold of 17 was chosen (see section "Comparison With National Center for Biotechnology Information Pathogen Single-Nucleotide Polymorphism Results" for details), which is in agreement with Pightling et al. (2018). If the pairwise distance was below or above the threshold for both methods under comparison, it was considered as true positive (TP) or true negative (TN), respectively. If the pairwise distance was below the threshold for only one of the methods under comparison, it was considered as false positives (FPs) (method 1) or false negatives (FNs) (method 2) (see also **Figure 3**). From the TP, TN, FN, and FP values, the corresponding values for specificity and sensitivity were derived.

Comparison of Clustering

The distance matrices resulting from each method were hierarchically clustered and divided into clusters addresses using the thresholds described above. Direction-dependent concordance of clustering, expressed by the adjusted Wallace coefficient, was calculated with the Comparing Partitions online tool available at http://www.comparingpartitions.info (Carriço et al., 2006). The adjusted Wallace coefficient allows the direct interpretation of clustering results, ensuring that an agreement is not caused by chance (Pinto et al., 2008; Severiano et al., 2011). The same resource also enables the computation of the adjusted Rand index, which allows the estimation of the global congruence of two typing approaches (Hubert and Arabie, 1985).

RESULTS

ChewieSnake Workflow

We developed *chewieSnake*, an end-to-end analysis workflow for allele calling, allele profile clustering, and computation of a minimum-spanning tree (see **Figure 1**). Its core components rely

on a number of state-of-the-art bioinformatic tools including chewBBACA and GrapeTree. Central to the workflow is the concept of allele hashes that allow a nomenclature-free comparison of allele profiles.

From the user perspective, *chewieSnake* provides an intuitive and simple-to-use workflow that orchestrates all necessary steps from a set of reads or assemblies to a highly informative, interactive cgMLST report displaying sample relatedness and their phylogeny.

Thus, *chewieSnake* enables scientists without profound bioinformatic background to perform a set of complex tasks for a reproducible high-throughput analysis that can be readily applied in outbreak investigations.

Moreover, *chewieSnake*'s built-in hashing algorithm includes the possibility to directly compare allele data generated independently, for instance, on different sites. In particular, the *chewieSnake_join* workflow provides a straightforward and easy-to-implement method for a distributed outbreak investigation system. The workflow provides users with highly informative HTML reports, which allow the quick identification of shared clusters between laboratories, the definition of a common cluster nomenclature, and detailed reports for each identified cluster (see **Figure 2**).

The *chewieSnake* workflow and program codes are fully open-source and can be obtained from https://gitlab.com/bfr_bioinformatics/chewieSnake. Details on the workflow components are given in section "Implementation and Availability."

Workflow Validation

Summary of *ChewieSnake* Analysis

The dataset contains WGS data of 1,263 sequenced *S. enterica* isolates from BioProject PRJEB31846. All samples could be analyzed with *chewieSnake*, and sufficient loci for cgMLST analysis were found. On average, 97.8% of the loci were found, and an allele sequence could be identified. More than 95% of all loci were found for all but two samples. The sample with the least fraction featured 94.3% of all loci, corresponding to 170 missing loci. Therefore, sufficient targets/loci for cgMLST analysis could be found for all samples.

Overall, we detected 33,387 novel alleles not previously known to the cgMLST scheme. For each sample, we found 26.4 novel alleles on average, with only 17% of the samples containing no novel allele. Furthermore, novel alleles were evenly found at the beginning and end of the experiment, showing no signs of saturation (see **Supplementary Figure 3**). Thus, finding novel alleles is very common, and properly treating these alleles is key for a comparable data analysis (see also **Supplementary Figure 4**).

For a distance cutoff of 10, we found that the data separate into 170 different clusters containing at least two samples. A total of 439 samples did not cluster within the threshold (orphan samples) (see also **Supplementary Figure 2**).

Repeatability of Analysis

We repeated the entire analysis using the original, unaltered Enterobase scheme, i.e., by running *chewieSnake* on the same dataset again. Overall, we found large repeatability. The same novel alleles were identified in both datasets and assigned the identical allele hashes. Thus, an identical allele sequence consistently leads to an identical allele hash. Moreover, the same number of clusters was detected.

However, in some cases, minor allele differences were found. Among the epidemiologically relevant sample pairs with distances below 20 AD, 2% of the sample pairs differed by 1 AD. The reason for these differences lies in the different order of execution of the sample allele calling (which is not pre-determined by Snakemake but randomly assigned on each execution) and the treatment of novel alleles by chewBBACA. The strict BLASTP step in chewBBACA might discard novel alleles, while the allele might be counted as an exact match if the allele sequence is already known to the database. If a sample featured a missing locus in the first analysis, but in the second analysis an allele difference was found on this locus, the allele difference would be reduced by 1 in the first analysis compared with the second. Indeed, 720 and 562 loci were missing in only one of the analyses, but not both.

The **Supplementary File 1** provides a more detailed analysis of the dataset. The analysis of the dataset is summarized in the cgmlst report https://bfr_bioinformatics.gitlab.io/chewiesnake_publicationdata/chewiesnake/cgmlst_report.html.

Comparison With Enterobase

The comparison between all sample combinations reveals a very high concordance between Enterobase and *chewieSnake* ADs with a Spearman correlation of 0.984 (see **Figure 3A**). Most distances are found at or near a line with slope 1 (a linear regression reveals a slope of 1.02). This extends also to high ADs (see **Supplementary Figure 5**). A histogram of the method differences of epidemiologically relevant sample pairs (i.e., within 30 AD) reveals that distances mostly differ by one or two alleles only, with a mean of non-zero method differences of 1.8 AD. We also found a tendency of higher distances in Enterobase (see also **Supplementary Figure 6**).

When performing a threshold analysis at 10 AD, sample pairs are found with 0.96 sensitivity and 0.99 specificity within or outside the same distance threshold (amounting to an accuracy of 0.99). Moreover, a closer look at the FPs and FNs reveals that the misclassified samples differ by a few (mostly one or two) allele differences only. Thus, these apparent misclassifications result from the assumption of a fixed threshold. The maximum AD in Enterobase for a sample pair within 10 AD in *chewieSnake* cluster was 15, and the average of all putatively misclassified pairs in Enterobase was 11.8.

When hierarchically clustering the distance matrix obtained by both methods, the (direction-dependent) adjusted Wallace coefficient was 0.955 and 0.975 (see **Table 1**). The adjusted Rand index also reveals a high method congruence (see **Table 2**).

It is important to note that very similar conclusions can be drawn from different allele thresholds. Evidently, method differences by one or two alleles have a larger impact on clusters determined at lower compared with larger distance thresholds. Thus, a very low threshold may lead to a decrease in the cluster congruence. However, fixed thresholds are most relevant for

linking samples to outbreak clusters, while for a closer analysis of the phylogeny, the exact tree topology is more useful.

Comparison With National Center for Biotechnology Information Pathogen Single-Nucleotide Polymorphism Results

Overall, there is a linear relationship between SNP and AD (see **Figure 3B**). A linear model fit reveals that statistically 1 AD corresponds to 1.73 SNP differences. The Spearman correlation reads 0.973. This suggests that samples within an AD of 10 should have an SNP distance of smaller or equal to 17 (definition of TPs). A threshold analysis shows a concordant clustering decision of 95% between cgMLST and SNP analysis (sensitivity—all points in the lower left quadrant in **Figure 3B**). Similarly, both methods agree on assigning a sample pair in 96% of all cases not to be part of a cluster (see **Figure 3B**).

We hierarchically clustered the distance matrices from both methods (cgMLST and SNP.) The resulting (direction-dependent) adjusted Wallace coefficient was 0.922 and 0.968, respectively (see **Table 1**). In addition, the adjusted Rand index reveals a high method congruence (see **Table 2**). Again, similar conclusions can also be drawn from different allele and SNP thresholds.

At small distances, i.e., well below the applied threshold, SNP and ADs are more scattered than between the cgMLST approaches. Thus, although samples can be concordantly associated with outbreak clusters, the detailed phylogeny may differ. This is, however, not surprising given the very different nature of SNP and cgMLST approaches.

Table 1 also yields the congruence between the reference methods Enterobase and NCBI pathogen. It shows that the adjusted Wallace coefficient between *chewieSnake* and the reference methods was even higher than the congruence among the reference methods.

TABLE 1 | (Direction-dependent) adjusted Wallace coefficient for all method comparisons.

	chewieSnake	*chewieSnake_join*	Enterobase	NCBI pathogen
chewieSnake	1	0.992	0.955	0.922
chewieSnake_join	1.000	1	0.964	0.931
Enterobase	0.975	0.975	1	0.935
NCBI pathogen	0.968	0.968	0.964	1

For each method, the distance matrix was hierarchically clustered using single-linkage trees. The clustering is based on a cluster threshold of 10 for the cgMLST methods and of 17 for the SNP method (NCBI). SNP, single-nucleotide polymorphism; NCBI, National Center for Biotechnology Information.

TABLE 2 | Discriminatory power according to adjusted Rand index.

	chewieSnake	*chewieSnake_join*	Enterobase	NCBI pathogen
chewieSnake				
chewieSnake_join	0.996			
Enterobase	0.965	0.969		
NCBI pathogen	0.944	0.949	0.949	

All methods show very high discriminatory power, above >0.94. Noteworthy is that chewieSnake and chewieSnake_join have a nearly perfect accordance. NCBI, National Center for Biotechnology Information.

Simulated Decentralized cgMLST Analysis

The concordance between the *chewieSnake* analysis on the entire dataset (centralized approach, see above) and a simulated decentralized *chewieSnake_join* analysis has been evaluated. For the latter, the original dataset was randomly divided into three equally sized, non-overlapping subsets, analyzed individually with *chewieSnake* and joined subsequently with *chewieSnake_join* (see section "Methods").

The ADs obtained from centrally and decentrally analyzed samples are nearly identical—the Spearman correlation is 1 (0.9999997). **Table 3** shows that at almost all thresholds with the same number of clusters were identified (the only exception was 160 vs. 159 distinct clusters at 5 AD). Furthermore, almost all sample pairs fall into the same 10 AD cluster, amounting to over 99.9% sensitivity and specificity. In the three exceptions, the sample pairs had 10 AD and 11 AD, respectively, not leading to a different epidemiological interpretation. The (direction-dependent) adjusted Wallace coefficient was 0.992 and 1.000, respectively (see **Table 1**). The adjusted Rand index reads 0.996 as well, demonstrating that centralized and decentralized allele calling is fully compatible (see **Table 2**).

The origin of the detected small differences was found to be the same, as discussed in the repeatability analysis of the *chewieSnake* analysis (see above). Importantly, the allele calling always leads to the same allele sequence and hence the same allele hash. Instead, the differences are due to the different execution order of the samples' allele calling, which impacts the discovery of novel alleles and exclusion of loci in chewBBACA. Potential allele differences between a sample pair may be masked when in one sample the locus was classified for instance as a *non-informative paralogous hit*. This classification is, however, subject to whether the allele was already known to the scheme prior to the allele calling. For the present dataset, 246 loci in samples in the centralized dataset and 191 loci in samples in the decentralized dataset were discarded as paralogues. Additionally, eight loci were discarded as being too close to a contig border. Hence, subtle but epidemiologically negligible differences may occur due to the different order of samples during the allele calling step of chewBBACA.

TABLE 3 | Cluster counts for *chewieSnake* and simulated *chewieSnake_join* analysis.

	1,000	200	100	50	20	10	5	1
chewieSnake	50	71	83	97	150	170	159	102
chewieSnake_join	50	71	83	97	150	170	160	102

The number of identified clusters agrees at all different thresholds with a single difference for CT_5.

DISCUSSION

In this study, we demonstrated the broad utility and validity of our newly developed end-to-end cgMLST analysis workflow. The method successfully determines clusters fully unsupervised without the need of user intervention. The proposed approach works both for a centralized data analysis, as well for a decentralized analysis followed by central clustering. This is a unique feature of the implemented allele hashing algorithm. Thus, the research community benefits from a versatile, easy-to-use tool that can be readily applied for molecular surveillance. It shall be noted that the presented method is readily applicable to all taxa where a suitable cgMLST scheme is available.

In comparison with existing approaches, for example, Ridom SeqSphere+ (with cgMLST.org) or the very recent contribution chewieNS (Mamede et al., 2020), *chewieSnake* does not rely on a central allele nomenclature. In the case of chewieNS, chewBBACA is also implemented as the underlying allele calling software and thus is prone to the same distinction between exactly determined and newly inferred alleles. The chewieNS approach functions without an allele hashing step on the user side, but as a downside, it introduces the requirement to synchronize the scheme prior to every allele calling. The synchronization is critical for ensuring comparable allele numbers and might be challenging in outbreak situations, when suspected samples may be sequenced and analyzed at different sites at the same time. Nonetheless, if a nomenclature-based solution has been used to create allele profiles at different sites, the *chewieSnake_join* workflow can be used for merging of the allele results and a joint cluster interpretation. Also, the concept of allele hashing could be directly incorporated in the allele calling software, e.g., in future upgrades of chewBBACA.

As pointed out above, the discovered remaining differences between central and decentralized approaches are not related to the hashing algorithm but rather to unique features of the underlying chewBBACA software. These issues could therefore only be remedied by substantial changes to the chewBBACA allele calling algorithms. However, as discussed previously, the effect on cluster detection is negligible.

Results of *chewieSnake* are in strong agreement with Enterobase, although both approaches differ substantially regarding their implementation. This might be somewhat expected given that both methods rely on the same cgMLST scheme. Our analyses revealed that distances in Enterobase tend to be larger, which are mainly caused by the greater number of loci available in Enterobase. Due to the CDS prediction step in chewBBACA, fewer loci are accessible in *chewieSnake*, leading to a smaller effective scheme size in the latter.

The accuracy for correct cluster detection was furthermore confirmed by comparison with an independent complementary method—SNP calling from NCBI pathogen detection. Overall, we found a very high consistency. As anticipated, ADs and SNP distances are not fully comparable. On the one hand, an allele pair might differ by more than one SNP—thus, the AD underestimates the SNP distance. On the other hand, allele differences might originate from other mutation events such as indels (insertions and deletions) and are therefore not characterized as an SNP. In the latter, allele difference might also overestimate the SNP distance. Nonetheless, the large concordance between SNP and cgMLST results also boosts confidence for the epidemiological assessment.

Alternatives to self-hosted and open-source cgMLST analysis workflows such as *chewieSnake* are central analysis systems such as Enterobase (Zhou et al., 2020a) (or also NCBI pathogen detection (see text footnote 1), INNUENDO (Llarena et al., 2018), etc.]. Among the advantages are greater standardization, more rigid control of parameters, and a larger data pool [e.g., NCBI pathogen and Enterobase incorporate all data from NCBI sequence read archive (SRA)]. Conversely, these approaches limit flexibility in terms of changing parameters and selecting a scheme and a species of choice. Also, *chewieSnake* and in particular its decentralized feature might render it more easy to share data and results when data confidentiality concerns apply. Furthermore, given that the amount of data present in public repositories (and consequently their computational needs) grow exponentially, *chewieSnake* is our approach to envision future tractability in terms of lean computing and storage resources by decentralizing and outsourcing the core analysis steps to the data providers.

In addition to the presented novel methods, the detailed analysis and contributed data provide an excellent dataset for the validation of further methods (and parameters/schemes). For this aim, one simply needs to compute a distance matrix and either compare the pairwise distances or cluster the distance matrix to a desired threshold for the imputation of relevant metrics such as the adjusted Wallace coefficient (see a more detailed sketch in the **Supplementary Data sheet 2** and the provided scripts). It is worth noting that the validation approach of this contribution is also in agreement with the forthcoming ISO 23418 titled "Whole genome sequencing for typing and genomic characterization of foodborne bacteria—General requirements and guidance (ISO/DIS 23418, 2020)" for validation of bioinformatic software, in particular by validating with publicly accessible data and methods. However, the present analysis also indicates that further harmonization of bioinformatic approaches is needed (Jagadeesan et al., 2019; Coipan et al., 2020). In the cases where this is not possible (e.g., cgMLST vs. SNP), the expected differences in the epidemiological interpretation from different methods need to be further discussed within the research community.

DATA AVAILABILITY STATEMENT

The developed software can be found in https://gitlab.com/bfr_bioinformatics/chewieSnake, as well as in bioconda and Docker Hub (https://hub.docker.com/r/bfrbioinformatics/chewiesnake). The raw data analyzed for this study can be found in BioProject PRJEB31846. The assembled genome data can be found under https://zenodo.org/record/4338293 and the used cgMLST scheme under https://zenodo.org/record/4724927. The workflows' analyses can be found in https://bfr_bioinformatics.gitlab.io/chewiesnake_publicationdata. All further relevant data for reproducing the results can be found in the **Supplementary Data Sheets 1, 2**.

AUTHOR CONTRIBUTIONS

CD, LU, ST, and BM designed the project. CD wrote the software with support from HB and ST. CD, LU, and ST conducted the analysis. All authors wrote and approved the manuscript.

FUNDING

The BfR has received financial support from the Federal Government for LU on the basis of a resolution of the German Bundestag and funded by the Ministry of Health within the framework of the project Integrated Genome-Based Surveillance of *Salmonella* (GenoSalmSurv), decision ZMVI1-2518FSB709 of November 26, 2018. HB was funded by the BeONE project within the One Health European Joint Programme (OHEJP) (JRP27-R2-FBZ-BeONE).

ACKNOWLEDGMENTS

We are grateful for the continuous collaboration with the German Federal State laboratories, which provided us with the *Salmonella* isolates for sequencing. We are thankful to the Binfis from the 4th Microbial Bioinformatics Hackathon where the hashing concept was discussed. We also thank the project partners from the GenoSalmSurv project for continuous testing and feedback.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.649517/full#supplementary-material

REFERENCES

Besser, J. M., Carleton, H. A., Trees, E., Stroika, S. G., Hise, K., Wise, M., et al. (2019). Interpretation of whole-genome sequencing for enteric disease surveillance and outbreak investigation. *Foodborne Pathog. Dis.* 16, 504–512. doi: 10.1089/fpd.2019.2650

Carriço, J. A., Silva-Costa, C., Melo-Cristino, J., Pinto, F. R., de Lencastre, H., Almeida, J. S., et al. (2006). Illustration of a common framework for relating multiple typing methods by application to macrolide-resistant *Streptococcus pyogenes*. *J. Clin. Microbiol.* 44, 2524–2532. doi: 10.1128/jcm.02536-05

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

Coipan, C. E., Dallman, T. J., Brown, D., Hartman, H., van der Voort, M., van den Berg, R. R., et al. (2020). Concordance of SNP- and allele-based typing workflows in the context of a large-scale international *Salmonella* Enteritidis outbreak investigation. *Microb. Genom.* 26:e000318. doi: 10.1099/mgen.0.000318

Dallman, T., Ashton, P., Schafer, U., Jironkin, A., Painset, A., Shaaban, S., et al. (2018). SnapperDB: a database solution for routine sequencing analysis of bacterial isolates. *Bioinformatics* 34, 3028–3029. doi: 10.1093/bioinformatics/bty212

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

Eyre, D. W., Peto, T. E. A., Crook, D. W., Walker, A. S., and Wilcox, M. H. (2019). Hash-based core genome multilocus sequence typing for *Clostridium difficile*. *J. Clin. Microbiol.* 58:e001037-19.

Feijao, P., Yao, H.-T., Fornika, D., Gardy, J., Hsiao, W., Chauve, C., et al. (2018). MentaLiST - a fast MLST caller for large MLST schemes. *Microb. Genom.* 4:e000146.

Francisco, A. P., Vaz, C., Monteiro, P. T., Melo-Cristino, J., Ramirez, M., and Carriço, J. A. (2012). PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinform.* 13:87. doi: 10.1186/1471-2105-13-87

Franz, E., Gras, L. M., and Dallman, T. (2016). Significance of whole genome sequencing for surveillance, source attribution and microbial risk assessment of foodborne pathogens. *Curr. Opin. Food Sci.* 8, 74–79. doi: 10.1016/j.cofs.2016.04.004

Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* 31, 3718–3720. doi: 10.1093/bioinformatics/btv428

Grüning, B., Dale, R., Sjödin, A., Chapman, B. A., Rowe, J., Tomkins-Tinch, C. H., et al. (2018). Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nat. Methods* 15, 475–476. doi: 10.1038/s41592-018-0046-7

Hubert, L., and Arabie, P. (1985). Comparing partitions. *J. Classif.* 2, 193–218. doi: 10.1007/bf01908075

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* 11:119. doi: 10.1186/1471-2105-11-119

ISO/DIS 23418 (2020). *Microbiology of the Food Chain—Whole Genome Sequencing for Typing and Genomic Characterization of Foodborne Bacteria—General Requirements and Guidance, (German and English Version).* Geneva: International Organization for Standardization.

Jagadeesan, B., Gerner-Smidt, P., Allard, M. W., Leuillet, S., Winkler, A., Xiao, Y., et al. (2019). The use of next generation sequencing for improving food safety: translation into practice. *Food Microbiol.* 79, 96–115. doi: 10.1016/j.fm.2018.11.005

Koster, J., and Rahmann, S. (2012). Snakemake - a scalable bioinformatics workflow engine. *Bioinformatics* 28, 2520–2522. doi: 10.1093/bioinformatics/bts480

Llarena, A., Ribeiro-Gonçalves, B. F., Nuno Silva, D., Halkilahti, J., Machado, M. P., Da Silva, M. S., et al. (2018). INNUENDO: a cross-sectoral platform for the integration of genomics in the surveillance of food-borne pathogens. *EFSA Support Publ.* 15:1498E. doi: 10.2903/sp.efsa.2018.EN-1498

Lüth, S., Deneke, C., Kleta, S., and Al Dahouk, S. (2020). Translatability of WGS typing results can simplify data exchange for surveillance and control of *Listeria monocytogenes*. *Microb. Genom.* 7:491. doi: 10.1099/mgen.0.000491

Macedo, R., Pinto, M., Borges, V., Nunes, A., Oliveira, O., Portugal, I., et al. (2019). Evaluation of a gene-by-gene approach for prospective whole-genome sequencing-based surveillance of multidrug resistant *Mycobacterium tuberculosis*. *Tuberculosis* 115, 81–88. doi: 10.1016/j.tube.2019.02.006

Maiden, M., van Rensburg, M., Bray, J., Earle, S. G., Ford, S. A., Jolley, K. A., et al. (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat. Rev. Microbiol.* 11, 728–736. doi: 10.1038/nrmicro3093

Mamede, R., Vila-Cerqueira, P., Silva, M., Carriço, J. A., and Ramirez, M. (2020). Chewie Nomenclature Server (chewie-NS): a deployable nomenclature server for easy sharing of core and whole genome MLST schemas. *Nucleic Acids Res.* 49, D660–D666. doi: 10.1093/nar/gkaa889

Merkel, D. (2014). Docker: lightweight Linux containers for consistent development and deployment. *Linux J.* 2:2.

Pightling, A. W., Pettengill, J. B., Luo, Y., Baugher, J. D., Rand, H., and Strain, E. (2018). Interpreting whole-genome sequence analyses of foodborne bacteria for regulatory applications and outbreak investigations. *Front. Microbiol.* 9:1482. doi: 10.3389/fmicb.2018.01482

Pinto, F. R., Melo-Cristino, J., and Ramirez, M. A. (2008). Confidence interval for the wallace coefficient of concordance and its application to microbial typing methods. *PLoS One* 3:e3696. doi: 10.1371/journal.pone.0003696

Pinto, M., Borges, V., Isidro, J., Rodrigues, J. C., Vieira, L., Borrego, M. J., et al. (2020). *Neisseria gonorrhoeae* clustering to reveal major European whole-genome-sequencing-based genogroups in association with antimicrobial resistance. *Microb. Genom.* 7:481. doi: 10.1099/mgen.0.000481

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Radomski, N., Cadel-Six, S., Cherchame, E., Felten, A., Barbet, P., Palma, F., et al. (2019). A simple and robust statistical method to define genetic relatedness of samples related to outbreaks at the genomic scale - application to retrospective *Salmonella* foodborne outbreak investigations. *Front. Microbiol.* 10:2413. doi: 10.3389/fmicb.2019.02413

rmarkdown (2020). *Dynamic rmarkdown: Dynamic Documents for R. R package version 2.5.*. Available online at: https://rmarkdown.rstudio.com

Ronholm, J., Nasheri, N., Petronella, N., and Pagotto, F. (2016). Navigating microbiological food safety in the era of whole-genome sequencing. *Clin. Microbiol. Rev.* 29, 837–857. doi: 10.1128/cmr.00056-16

Severiano, A., Pinto, F. R., Ramirez, M., and Carriço, J. A. (2011). Adjusted wallace coefficient as a measure of congruence between typing methods. *J. Clin. Microbiol.* 49, 3997–4000.

Silva, M., Machado, M. P., Silva, D. N., Rossi, M., Moran-Gilad, J., Santos, S., et al. (2018). chewBBACA: a complete suite for gene-by-gene schema creation and strain identification. *Microb. Genom.* 4:e000166.

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

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

Uelze, L., Borowiak, M., Brinks, E., Deneke, C., Stingl, K., Kleta, S., et al. (2020a). German-wide interlaboratory study compares consistency, accuracy and reproducibility of whole-genome short read sequencing. *Front. Microbiol.* 11:e0573972. doi: 10.3389/fmicb.2020.573972

Uelze, L., Grützke, J., Borowiak, M., Hammerl, J. A., Juraschek, K., Deneke, C., et al. (2020b). Typing methods based on whole genome sequencing data. *One Health Outlook* 2:3.

Uelze, L., Borowiak, M., Deneke, C., Szabó, I., Fischer, J., Tausch, S. H., et al. (2019). Performance and accuracy of four open-source tools for *in silico* serotyping of *Salmonella* spp. based on whole-genome short-read sequencing data. *Appl. Environ. Microbiol.* 86:e002265-19.

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

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

Zhou, Z., Charlesworth, J., and Achtman, M. (2020b). HierCC: a multi-level clustering scheme for population assignments based on core genome MLST. *bioRxiv* [Preprint]. doi: 10.1101/2020.11.25.397539v1

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

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

frontiers in Microbiology

ORIGINAL RESEARCH
published: 02 July 2021
doi: 10.3389/fmicb.2021.692453

Variation in Accessory Genes Within the *Klebsiella oxytoca* Species Complex Delineates Monophyletic Members and Simplifies Coherent Genotyping

*Amar Cosic[1,2†], Eva Leitner[3†], Christian Petternel[3‡], Herbert Galler[3], Franz F. Reinthaler[3], Kathrin A. Herzog-Obereder[4], Elisabeth Tatscher[4], Sandra Raffl[1], Gebhard Feierl[3], Christoph Högenauer[2,4], Ellen L. Zechner[1,2,5] and Sabine Kienesberger[1,2,5]**

[1] Institute of Molecular Biosciences, University of Graz, Graz, Austria, [2] BioTechMed-Graz, Graz, Austria, [3] Diagnostic and Research Institute of Hygiene, Microbiology and Environmental Medicine, Medical University of Graz, Graz, Austria, [4] Division of Gastroenterology and Hepatology, Department of Internal Medicine, Medical University of Graz, Graz, Austria, [5] Field of Excellence BioHealth, University of Graz, Graz, Austria

Members of the *Klebsiella oxytoca* species complex (KoSC) are emerging human pathogens causing infections of increasing significance especially in healthcare settings. KoSC strains are affiliated with distinct phylogroups based on genetic variation at the beta-lactamase gene (bla_{OXY}) and it has been proposed that each major phylogroup represents a unique species. However, since the typing methods applied in clinical settings cannot differentiate every species within the complex, existing clinical, epidemiological and DNA sequence data is frequently misclassified. Here we systematically examined the phylogenetic relationship of KoSC strains to evaluate robustness of existing typing methods and to provide a simple typing strategy for KoSC members that cannot be differentiated biochemically. Initial analysis of a collection of *K. oxytoca*, *K. michiganensis, K. pasteurii*, and *K. grimontii* strains of environmental origin showed robust correlation of core phylogeny and blaOXY grouping. Moreover, we identified species-specific accessory gene loci for these strains. Extension of species correlation using database entries initially failed. However, assessment of average nucleotide identities (ANI) and phylogenetic validations showed that nearly one third of isolates in public databases have been misidentified. Reclassification resulted in a robust reference strain set for reliable species identification of new isolates or for retyping of strains previously analyzed by multi-locus sequence typing (MLST). Finally, we show convergence of ANI, core gene phylogeny, and accessory gene content for available KoSC genomes. We conclude that also the monophyletic members *K. oxytoca*, *K. michiganensis*, *K. pasteurii* and *K. grimontii* can be simply differentiated by a PCR strategy targeting bla_{OXY} and accessory genes defined here.

Keywords: bacterial phylogeny, *Klebsiella oxytoca* species complex, taxonomic classification, necrotizing enterocolitis, bacterial cytotoxicity, intestinal disease

OPEN ACCESS

Edited by:
Guido Werner,
Robert Koch Institute, Germany

Reviewed by:
Majid Validi,
Shahrekord University of Medical Sciences, Iran
Scott Wesley Long,
Houston Methodist Hospital, United States
Guido Voets,
Check-Points B.V., Netherlands

***Correspondence:**
Sabine Kienesberger
sabine.kienesberger@uni-graz.at

† These authors have contributed equally to this work

‡Present address:
Christian Petternel,
Institute of Laboratory Diagnostics and Microbiology, Klinikum-Klagenfurt am Wörthersee, Klagenfurt, Austria

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *08 April 2021*
Accepted: *28 May 2021*
Published: *02 July 2021*

Citation:
Cosic A, Leitner E, Petternel C, Galler H, Reinthaler FF, Herzog-Obereder KA, Tatscher E, Raffl S, Feierl G, Högenauer C, Zechner EL and Kienesberger S (2021) Variation in Accessory Genes Within the Klebsiella oxytoca Species Complex Delineates Monophyletic Members and Simplifies Coherent Genotyping. Front. Microbiol. 12:692453. doi: 10.3389/fmicb.2021.692453

INTRODUCTION

The genus *Klebsiella* belongs to the family of Enterobacteriaceae and comprises multiple species. *Klebsiella pneumoniae* and species of the *Klebsiella oxytoca* complex are currently responsible for most human illnesses. Accurate identification of these pathogens is thus important for diagnosis, treatment, and epidemiological surveillance of infections. The *K. oxytoca* species complex (KoSC) can be resolved into nine distinct phylogroups (Ko1 to Ko9), which are believed to comprise six species (Merla et al., 2019). Phylogroups are based on specific beta-lactamase ($bla_{OXY-(1-9)}$) gene variants unique to this complex (Fevre et al., 2005; Merla et al., 2019) and each major phylogroup (except Ko5, Ko7 and Ko9) represents a unique species. Currently, *K. spallanzanii* (Ko3 or Ko9, bla_{OXY-3} or bla_{OXY-9}) and *K. huaxiensis* (Ko8, bla_{OXY-8}) can be differentiated biochemically from other complex members (Merla et al., 2019). In contrast, a bacterial isolate identified as *K. oxytoca* by classical polyphasic typing strategies can actually be one of several species: *K. oxytoca* (Ko2, bla_{OXY-2}), *K. michiganensis* (Ko1 or Ko5, bla_{OXY-1} or bla_{OXY-5}), *K. pasteurii* (Ko4, bla_{OXY-4}), or *K. grimontii* (Ko6, bla_{OXY-6}). Phylogroup Ko7 (bla_{OXY-7}) has been described based on a single strain and represents a sub-group of Ko6 (Izdebski et al., 2015).

Members of the KoSC are versatile pathogens and have been isolated from wound-, intra- abdominal-, urinary-, and lower respiratory tract infections (Herzog et al., 2014; Singh et al., 2016). Outbreaks have environmental sources (Hendrik et al., 2015) and subsequent dissemination of phylogroup members in healthcare settings is a major problem – especially in neonatal intensive care units (Rønning et al., 2019; Chapman et al., 2020; Chen et al., 2020). Outbreaks involve strains with extended-spectrum beta-lactamases (Lowe et al., 2012; Vasaikar et al., 2017), extended-spectrum activity of the chromosomal beta-lactamase (Fournier et al., 1995; Decré et al., 2004) and carbapenemases (Validi et al., 2016). Isolates can also carry resistance to fluoroquinolones and tetracyclines limiting therapeutic options (Brisse et al., 2000). KoSC members are also residents of the human gastrointestinal tract and under certain conditions act as pathobionts in adults and children. Intestinal enrichment with KoSC members is frequent in preterm neonates. *Klebsiella* overgrowth in infants due to antibiotic therapy is associated with late-onset sepsis, meningitis, and necrotizing enterocolitis, but the importance of cytotoxin production to disease is not yet clear (Carrie et al., 2019; Chen et al., 2020; Paveglio et al., 2020). Difficulties distinguishing these closely related complex members have not only led to inconsistent species classification in recent literature and databases but also make it impossible to draw conclusions regarding the clinical relevance of the individual species.

Here we report refined phylogenetic analyses of this closely related complex and delineate a uniform system of species classification that is equally valid for its monophyletic members.

METHODS

Collection and Characterization of Environmental Strains

Sample Collection

Samples were taken from activated sludge ($n = 29$), surface water ($n = 2$), minced meat ($n = 36$), fresh retail chicken ($n = 30$), plant roots and soil ($n = 31$). The food was obtained from supermarkets and butcher shops in Austria. The activated sludge samples were collected from two different Austrian sewage treatment plants with a maximum load of <10,000 population equivalent (PE) and >100,000 PE, respectively. Plant roots and associated soil samples were taken from *Fabaceae*, *Pinaceae*, *Solanaceae*, *Rutaceae*, *Poaceae*, and *Vitaceae*.

Identification of *K. oxytoca*

Food samples were prepared according to the International Organization for Standardization, 2003 ISO 6887-2:2003. For enrichment 25 g of the meat samples were mixed with 225 ml peptone broth 1% (Oxoid Ltd., Basingsoke, England), homogenized, then cultured 16 to 24 h at 37°C under shaking. 1 g of plant roots and soil samples were mixed with 25 ml peptone broth 1% and enriched overnight. Activated sludge and surface water samples were screened without enrichment. Serial dilutions were cultured on chromID^{TM} CPS® Agar (bioMérieux, Marcy-l'Etoile, France) for 24-48 h at 37°C. Colonies were scored based on the colour reaction of CPS media. Green to turquoise colonies were identified as *Klebsiella* spp., *Enterobacter* spp., *Serratia* spp., *Citrobacter* spp. and *Enterococcus* spp. pink to burgundy colonies as *E. coli* and light brown to brown colonies as *Proteus* spp. For pure cultures green to turquoise colonies were transferred to blood agar and Endo agar (24 h, 37°C). The isolate identification was carried out biochemically with VITEK® 2 and/or with MALDI-TOF MS analysis, using VITEK® MS (bioMérieux).

Genotypic and Phenotypic Toxin Assessment

All environmental isolates were screened for *npsB* presence and cytotoxicity. A 231 bp region of *npsB* was amplified by colony PCR using primer pair 1 + 2 (**Table 1**) in 30 cycles and primer annealing at 60°C. Cytotoxicity of supernatants of *K. oxytoca* spent medium was tested after 16 h culture. Viability of HeLa cells was used for the colorimetric 3-[4,5-dimethyl-2-thiazolyl] 2,5-diphenyl-2H-tetrazolium bromide (MTT) assay of mitochondrial reductase activity as previously described (Joainig et al., 2010).

Beta-Lactamase Gene Typing and Detection of Accessory Genes

Identification of beta-lactamase variants bla_{OXY-1} and bla_{OXY-2} utilized primer pairs 3 + 4 and 5 + 6, respectively, as previously described (Fevre et al., 2005) but with annealing at 55°C. The sequence of amplicons generated using oligonucleotides 7 and 8 determined all other bla_{OXY} variants. Amplification of *leupAB* utilized primers 11 + 12 (1,235 bp), of *orfABC* primers 13 + 14 (2,546 bp), and of *orfA′* primers 15 + 16 (591 bp). Absence of an insertion was confirmed by colony PCR over the conserved insertion site with primer pair 9 + 10

TABLE 1 | Oligonucleotides used in PCR-screening and sequencing.

#	Primer name	Sequence 5′-3′	Binding Site (nt position)*	Reference sequence	Fragment size	References
1	npsB_f	ctcgacgttttatctctgctg	20117..20137	AHC-6**	231 bp	This study
2	npsB_r	ttcctgaagtatctgccctgc	20327..20347			This study
3	OXY1-A	gtggcgtaaaaccgccctg	5001082..5001100	CAV1374	425 bp	Fevre et al., 2005
4	OXY1-B	gtccgccaaggtagctaatc	5001487..5001506			Fevre et al., 2005
5	OXY2-A	aaggctggagattaacgcag	4572112..4572131	NCTC13727	155 bp	Fevre et al., 2005
6	OXY2-B	gcccgccaaggtagccgatg	4572247..4572266			Fevre et al., 2005
7	OXY-E	ggttt**T**ggtaactgtgacggg	5304405..5304937	NCTC13727	1,098 bp	Fevre et al., 2005
8	OXY-G	cagagt**G**cagagtgttgcag	5304405..5304937			Fevre et al., 2005
9	core_f	gagatcccaagttctttagcaatgg	227821..227845	NCTC13727	175 bp	This study
10	core_r	cagcgcctggaaaacgtgct	227976..227995			This study
11	leupA_f	atgaagatagcgattcacaac	315403..315425	CAV1374	1,235 bp	Li et al., 2020
12	leupB_r	gcgtggtcttttagctgttc	316620..316639			Li et al., 2020
13	orfA_f	gcagtgatttaaatactcttgcgg	53092..53115	JKo3	2,525 bp	This study
14	orfC_r	ccgatacctccagtaatgcgc	55617..55637			This study
15	B2_v2_1	cggcttacgcacaaagaagcc	74467..74487	ARO112, cont20	591 bp	This study
16	B2_v2_2	atgtttcttgaagaacgtagg	75037..75075			This study
17	gapA_fwd	gttttcccagtcacgacgttgtatgaagtatgactccactcacgg	4012172..4012193	CAV1374	680 bp	Herzog et al., 2014
18	gapA_rev	ttgtgagcggataacaatttcaacgcctttcattgcgccttcggaa	4012827..4012851			Herzog et al., 2014
19	infB_fwd	gttttcccagtcacgacgttgtactctctgctggactacattcg	1034476..1034496	CAV1374	463 bp	Herzog et al., 2014
20	infB_rev	ttgtgagcggataacaatttccgctttcagctccagaacttc	1024918..1034938			Herzog et al., 2014
21	mdh_fwd	gttttcccagtcacgacgttgtacccaactgccttcaggttcag	964459..964479	CAV1374	704 bp	Herzog et al., 2014
22	mdh_rev	ttgtgagcggataacaatttc**C**cttc**C**acgtaggcgcattcc	965141..965162			Herzog et al., 2014
23	pgi_fwd	gttttcccagtcacgacgttgtagagaaaaacctgccggtgctgctg	21866..21889	CAV1374	701 bp	Herzog et al., 2014
24	pgi_rev	ttgtgagcggataacaatttccggttaatcag**G**ccgttagtggagc	21189..21213			Herzog et al., 2014
25	phoE_fwd	gttttcccagtcacgacgttgtaacct**GG**cgcaa**C**accga**T**ttcttc	5304914..5304937	CAV1374	533 bp	Herzog et al., 2014
26	phoE_rev	ttgtgagcggataacaatttcttcagctggttgattttgtaatccac	5304405..5304430			Herzog et al., 2014
27	rpoB_fwd	gttttcccagtcacgacgttgtaggcgaaatggcggaaaacca	111644..111663	CAV1374	1,076 bp	Herzog et al., 2014
28	rpoB_rev	ttgtgagcggataacaatttcgagtcttcgaagttgtaacc	110588..110607			Herzog et al., 2014
29	tonB_fwd	gttttcccagtcacgacgttgtactctatacttcggtacatcaggtt	2890556..2890579	CAV1374	589 bp	Herzog et al., 2014
30	tonB_rev_2	ttgtgagcggataacaatttcgtttacccggttcatagcgcc	2891124..2891144			This study
31	tonB_rev	ttgtgagcggataacaatttccctgtttggcggccagcacctggt	2686207..2686230	NCTC13727***	740 bp	Herzog et al., 2014
32	MLST_seq_fwd	gttttcccagtcacgacgttgta	n/a	n/a	n/a	Herzog et al., 2014
33	MLST_seq_rev	ttgtgagcggataacaatttc	n/a	n/a	n/a	Herzog et al., 2014

Nucleotide (nt) position of respective primer binding site in reference sequence. **Tilimycin/tilivalline biosynthesis gene cluster (accession number: HG425356.1). ***Only first 4 nucleotides on the 3′ end of the primer bind in CAV1374 in the first reading frame after tonB (in the fructosamine kinase family protein), therefore primer binding site is given for NCTC13727. underlined: binding site for MLST_seq primers; **bold: primer binding site ambiguity in given reference sequence.*

(174 bp fragment). Reactions were performed in 35 cycles with 60°C annealing.

MLST and Phylogenetic Analysis

Isolates were typed as described (Herzog et al., 2014) using oligonucleotides listed in **Table 1**. Isolates that failed to produce a PCR product for *tonB* with primer pair 29 + 31 were reanalyzed using primer pair 29 + 30. Oligonucleotide 30 (tonB_rev_2) binds at the 3′ end of *tonB* in contrast to the originally published tonB_rev primer, which binds distally to the *tonB* reading frame. The PCR product thus obtained still allows analysis of the full allelic region used for typing. Oligonucleotides 32 and 33 were used for sequencing the resulting PCR products. New alleles, newly identified STs and all typed isolates were submitted to pubmlst.org/organisms/klebsiella-oxytoca. The concatenated sequence alignment of all MLST loci was used to infer the phylogenetic relationship among the strains analyzed in this study. The Neighbor-joining tree was built in MEGAX (Kumar et al., 2018) with Tamurai-Nei model and gamma distribution, and utilized *K. pneumoniae* for tree rooting (Herzog et al., 2014).

Comparative Analysis of Whole Genome Sequences

Genome Extractions, Reference Set Compilation, and Validation

Workflow is illustrated in **Supplementary Figure 1**.

In June 2020, all available KoSC genomes independent of their assembly status were downloaded from NCBI GenBank and analyzed (n = 316). First, a subset of these genomes comprising clinical, animal, and environmental isolates were selected based on completeness and overall quality, with threshold set at 30 scaffolds maximum (**Supplementary Table 1**). Additional strains

from recent publications were included independent of their assembly status (see section "Result" for further information) (**Supplementary Table 1**). The average nucleotide identities (ANI) for this subset (n = 62) were calculated using ANIm based on MUMmer (NUCmer) v.3.23 (Kurtz et al., 2004) and PYANI (Pritchard et al., 2016) was used for score visualization (**Figure 3**). Comparative phylogenetic analysis, beta-lactamase-, and accessory gene typing were then applied to these 62 strains to establish a robust reference data set (**Figure 4**). To validate our findings we assessed phylogeny, core gene and accessory gene distribution of all 316 available KoSC genomes independent of assembly status and together with 24 other *Klebsiella* genomes (**Figure 5** and **Supplementary Table 2**). In a final step, we validated our findings by short read sequence typing of non-assembled KoSC genomes (n = 41) of clinical origin (NCBI BioProject PRJEB5065) (Moradigaravand et al., 2016, 2017).

Phylogenetic Analysis

Comparative phylogenetic analysis of whole genome sequences utilized three different approaches. (I) Phylogenetic analysis was done with PhyloPhlAn –version 3.0.51 (May 11, 2020) (Segata et al., 2013) using the amphora2 database (136 marker genes). The diversity was set to low and the default supermatrix configuration file was applied which utilizes five external tools: diamond, mafft, trimal, FastTreeMP, raxmlHPC.

(II) Core genome phylogeny of the biochemically identical KoSC members was created after extracting core genomes in the gbff format with the BPGA –version 1.3 (Chaudhari et al., 2016), with the sequence identity cut-off set at clustering 0.5 (50%). Core sequences were then aligned in MEGAX using MUSCLE with cluster method set as UPGMA. After initial alignment, the core tree was built in MEGAX. The construct was built using the Neighbor Joining method with Poisson model and gamma distribution set at one.

(III) Sequence types (STs) for GenBank entries were determined using pubmlst.org/organisms/klebsiella-oxytoca and a Neighbor-joining tree was built as described for the environmental strains except that the tree was midpoint rooted.

Blast Analyses and *in silico* PCR Analysis

Klebsiella genomes were analyzed with a standard online nucleotide Blast (query cover: 100%, *e*-value cut-off: 1e-30) using *leupABCD* as a query sequence. Screening of core genome insertion sites, genes and intergenic regions was done with blastn, where chromosomal and plasmid sequences for each strain were combined and built into a blast database. Assembly IDs of genome sequences used in blast analyses are listed in **Supplementary Table 2**.

Insertion site evaluation was performed with a blast query of 9,264 bp including *leupABCD*, the adjacent transporter and the two open reading frames (ORFs) flanking this region. The query sequence was retrieved from *K. oxytoca* CAV1374, nucleotide position 312527 to 321790. The five blastn outcomes obtained were defined as follows. Variant A: blast identities (>92.5%) for the ORFs flanking the leupeptin biosynthesis genes and no coding sequence in between; Variant B: blast identities (>94.64%) over the whole query sequence; Variant C: blast identity (>93.74%) for the flanking ORFs, separated approximately by 4,200 bp indicating an unknown insertion. Variant D: blast identity (>92.30%) for the flanking ORFs with an insertion of approximately 870 bp; Variant E: no significant identity to the query.

To compare PCR results from environmental samples with blast analysis of genomic data, we used the corresponding amplified gene fragment as the query sequence in blastn analysis, except for a full-length *npsB* query. Only results with homologies of >98.5% (*leupAB*, *orfABC*, *orfA*') or >93,5% (*npsB*) over the full query sequence were scored as positive. Strains were also analyzed for their bla_{OXY} variants using each full-length gene sequence as a query. We performed *in silico* PCR analysis for *leupAB*, *orfABC*, and *orfA*' on whole genome sequences using a perl script and 100% primer binding stringency or allowed one base mismatch per primer sequence (**Supplementary Table 2**).

Short Read Sequence Mapping

Raw sequencing reads of 41 KoSC strains and 30 non-KoSC strains (2 *Acinetobacter baumannii*, 2 *Escherichia coli*, 3 *K. pneumoniae*, 7 *K. quasipneumoniae*, 2 *Klebsiella spp*. and 14 *K. variicola*) were downloaded from NCBI BioProject PRJEB5065 (Moradigaravand et al., 2016, 2017). Short Read Sequence Typing for Bacterial Pathogens (SRST2) (Inouye et al., 2014) was utilized for read-mapping (fastq format) and allele calling. As query sequences we used the same gene fragments (*leupAB*, *orfABC*, and *orfA*') utilized for blast-based gene typing. To identify the conserved core genome site without variant insertion, we utilized the 175 bp fragment generated with primers 9 + 10. Full-length genes were used for *npsB* and variant $bla_{\mathrm{OXY1\text{-}9}}$ calling. Default parameters were used with coverage threshold set at 90. Results were manually evaluated and hits with >99% coverage and >0.90 identity score were determined positive and included in **Supplementary Table 3**. For alleles with multiple hits (i.e., blaOXY) the highest homology variant is shown.

Relevant Assembly IDs and Accession Numbers

Assembly IDs of genomes used in this study can be found in **Supplementary Table 2**. Accession numbers of bla_{OXY} variants: AJ871864 ($bla_{\mathrm{OXY-1}}$), AJ871866 ($bla_{\mathrm{OXY-2}}$), AF491278 ($bla_{\mathrm{OXY-3}}$), AY077481 ($bla_{\mathrm{OXY-4}}$), AJ871871 ($bla_{\mathrm{OXY-5}}$), AJ871875 ($bla_{\mathrm{OXY-6}}$); KT001254.1 ($bla_{\mathrm{OXY-7}}$); MN030566 ($bla_{\mathrm{OXY-8}}$); and MN030564.1 ($bla_{\mathrm{OXY-9}}$); Non-ribosomal peptide synthetase gene (*npsB*) of the *K. oxytoca* AHC-6 tilimycin/tilivalline biosynthesis gene cluster (HG425356.1) nucleotide position 17,602 to 21,972. Leupeptin biosynthesis gene cluster (nucleotide position 315,353 to 321,183) of strain CAV1374 (CP011636.1); locus tags for *leupA* (AB185_06775), *leupB* (AB185_06780), *leupC* (AB185_06785), and *leupD* (AB185_06790). *OrfABC* (nucleotide position 52,982 to 56,916) of strain JKO3 (AP014951.1); locus tags for *orfA* (KOJKO3_c0051), *orfB* (KOJKO3_c0052), *orfC* (KOJKO3_c0053), and the transporter (KOJKO3_c0054). *OrfA*′ and transporter (nucleotide position 74,455 to 75,057) of strain ARO112 (GCA_009757395.1, contig20), locus tags for *orfA*′ (GQ640_RS25055) and the transporter (GQ640_RS24900).

RESULTS

Core Phylogeny and Accessory Gene Content Are Highly Congruent for a Population of Environmental KoSC Isolates

To explore the presence of phylogroup-specific genes in a diverse population we first characterized a collection of 63 KoSC strains isolated from various environmental sources (**Figure 1**). Identification was performed by routine biochemical analyses and MALDI-TOF mass spectrometry. We then genotyped the strains' chromosomally encoded bla_{OXY} variants (**Figure 1**). PCR typing for bla_{OXY-1} and bla_{OXY-2} was sufficient, but reliable assessment of other bla_{OXY} variants required sequencing. Most strains were assigned to phylogroups Ko1 (bla_{OXY-1}), Ko2 (bla_{OXY-2}), and Ko6 (bla_{OXY-6}). A single Ko5 and one Ko4 (bla_{OXY-4}) isolate were also identified. Thus, we conclude that our collection comprised *K. oxytoca* (bla_{OXY-2}), *K. michiganensis* (bla_{OXY-2} & bla_{OXY-5}), *K. pasteurii* (bla_{OXY-4}), and *K. grimontii* (bla_{OXY-6}) strains.

A multi-locus sequence typing (MLST) scheme has been developed for *K. oxytoca* (Herzog et al., 2014) that also resolves phylogeny of *K. michiganensis* (bla_{OXY-1}), *K. pasteurii* (bla_{OXY-4}) and *K. grimontii* (bla_{OXY-6}) (Fevre et al., 2005; Izdebski et al., 2015; Moradigaravand et al., 2017). However, correspondence of blaOXY grouping and MLST typing has not been explored extensively. We applied MLST to all isolates to compare with blaOXY-based phylogrouping and to investigate the diversity of our collection. Allelic variation at seven housekeeping gene loci identified 57 distinct sequence types (ST). Several of the STs match STs previously identified for clinical isolates. We then inferred a phylogenetic tree for this population and in good agreement with the population structure of clinical isolates (Herzog et al., 2014; Moradigaravand et al., 2017), relationships of the environmental strains cluster into two major clades (MLST-cluster A and B). Cluster B further splits into the sub-groups B1 and B2 (**Figure 1**). In summary, we obtained correspondence of MLST based phylogeny and blaOXY typing in this diverse population.

We next surveyed the population for accessory gene content. Knowledge of whether human KoSC isolates carry genes for production of the enterotoxins tilimycin and tilivalline (Dornisch et al., 2017; Unterhauser et al., 2019) is clinically relevant. Application of a PCR-based screen for the presence of a non-ribosomal peptide synthetase gene (*npsB*) to this collection showed that *npsB* affiliated predominately with cluster A and sub-group B2 of this population but was also present in B1 (**Figure 1**). All isolates negative for PCR product were also negative phenotypically for toxin production, but from 30 *npsB* + isolates, 7 were not cytotoxic (**Figure 1**). The observed general distribution of *npsB* effectively disqualified the locus as a phylogroup-specific marker.

Previously, we observed that the biosynthesis genes (*leupAB*) of protease inhibitor leupeptin (Li et al., 2020) have a characteristic distribution along MLST-based KoSC phylogeny. We used PCR to screen the collection for *leupAB* and consistent with our previous findings (Li et al., 2020), leupeptin biosynthesis genes were detected solely in strains of sub-cluster B1 and thus only in *K. michiganensis* (bla_{OXY-1}). To interrogate the correlation between leupeptin biosynthesis genes and *K. michiganensis* in more detail, we performed an online blastn search using *leupABCD* as a query sequence and "*Klebsiella*" as search set. The locus was detected in whole genome sequences classified as *K. oxytoca* and *K. michiganensis* but, concordant with our findings for the environmental collection, these genomes all carry bla_{OXY-1} suggesting a misclassification of database entries. To obtain better discrimination we surveyed the insertion site of *leupABCD*. We noted that the position is conserved in every *leupABCD* positive genome; therefore we next asked whether *leupABCD* negative strains carry distinct gene loci at this site. We utilized a >9kb query sequence comprising the conserved flanking open reading frames (*orf1* and *orf2*; **Figure 2A**) and spanning the leupeptin biosynthesis genes (**Figure 2B**) for blast analysis. We sought to focus specifically on *K. oxytoca* and performed blast analysis and blaOXY typing only on database entries with the identifier "*K. oxytoca*" available in NCBI (n = 177, June 2020). Five different outcomes were obtained for the genome region of interest (**Figure 2**). Strains either carry (I) the conserved core genes without insertion (variant A), (II) the conserved core genes with the leupeptin biosynthesis genes inserted (variant B), (III) the conserved core genes and a gene cluster we designated here as *orfABC* (variant C), (IV) core genes and a truncated version of the *orfABC* locus, *orfA′* (variant D), or (V) no homology to the query. Importantly, despite our aim to analyze exclusively genomes belonging to the species *K. oxytoca*, we identified strains of all phylogroups. We next asked whether each insertion variant correlated with a specific blaOXY phylogroup. In 176 of the 177 genomes we observed perfect correlation between the two loci (**Supplementary Table 2**). Variant A strains (no insertion) carry bla_{OXY-2}, variant B strains (leupeptin biosynthesis genes) carry bla_{OXY-1} or bla_{OXY-5}, variant C strains (*orfABC*) carry bla_{OXY-6}, and the single observed variant D strain (*orfA′*) was positive for bla_{OXY-4}. Only *K. oxytoca* strain K3678 represents an exception (bla_{OXY-1} but no leupeptin genes).

Based on these findings we applied a discriminatory PCR screen using primer-binding sites shown in **Figure 2** to survey the alternate accessory gene sets carried by the environmental isolates (**Figure 1**). All B1 strains (including the bla_{OXY-5} isolate) contain *leupAB* (variant B) and all but one B2 isolate carry *orfABC* (variant C). This bla_{OXY-4} isolate (KS3/2) harbored the truncated *orfA′* (variant D). All cluster A strains were negative for *leupAB*, *orfABC*, and *orfA′* and amplicons from primers binding in the core genes confirmed that the region lacked insertion (variant A). Sequence similarity search with the newly identified *orf* A, *orfB*, and *orfC* and subsequent analyses of the conserved domains they contain imply a role in carbohydrate metabolism.

Results summarized in **Figure 1** show that core genome phylogeny, bla_{OXY} variations, and accessory gene distribution correspond in this population. Thus, we conclude that not only variations in the core genome differentiate the species of the KoSC but also species-specific gene loci. We hypothesize that the discrepancies observed in correlating bla_{OXY} and accessory

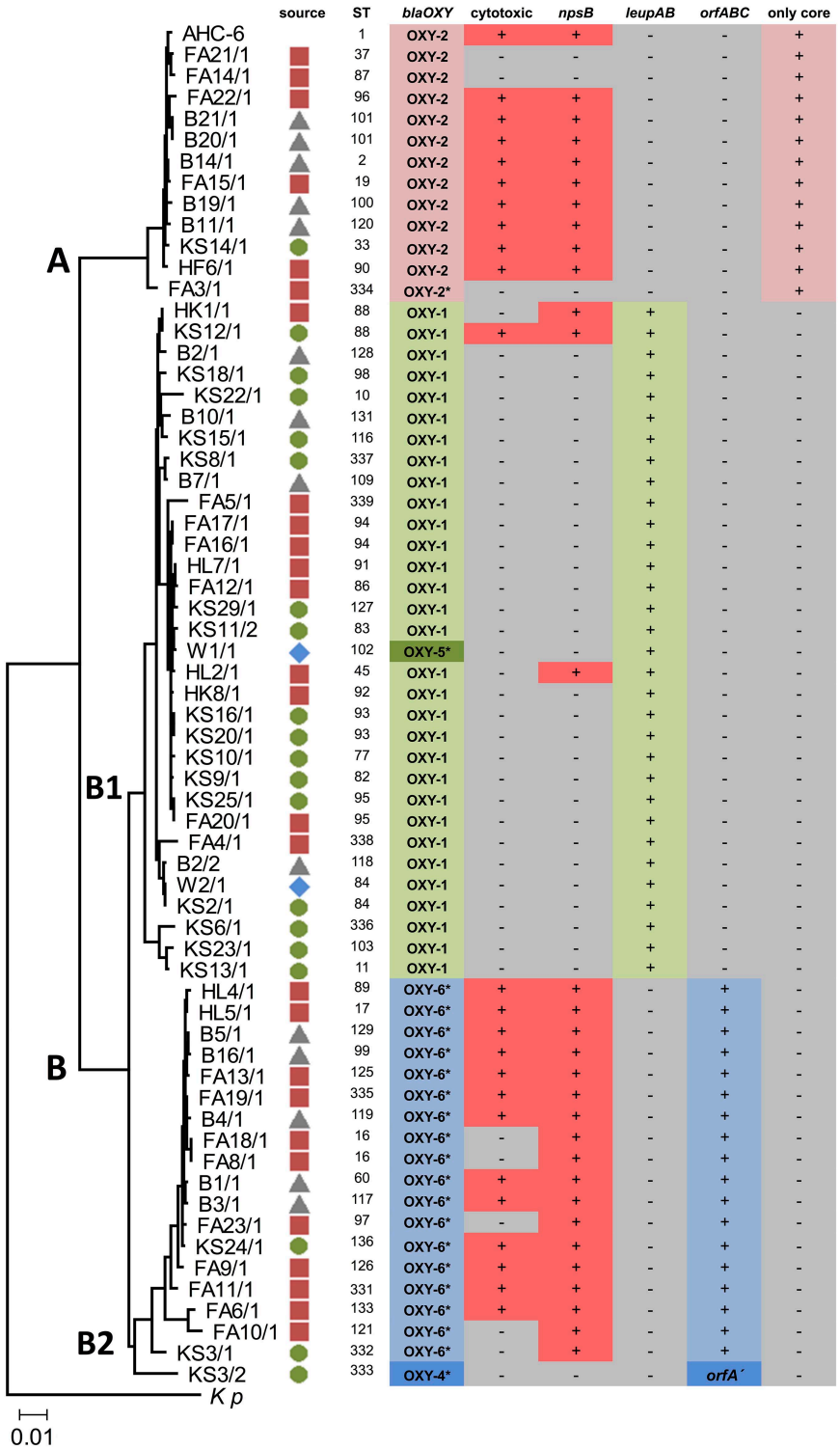


FIGURE 1 | Neighbor-joining tree showing the genetic relatedness of 63 environmental isolates. Source symbols next to strain designation indicate isolation site (■ food; ▲ soil; ● sewage sludge; ◆ surface water). The tree is based on the concatenated sequences of seven housekeeping gene loci used in MLST. *K. pneumoniae* (*Kp*) was used for tree rooting. Clusters (A,B) and sub-clusters (B1,B2) are indicated. The scale bar represents 0.01 substitutions per site. MLST reference strain *K. oxytoca* AHC-6 (sequence type, ST1) is included (top). bla_{OXY} variants (synonymous to phylogroups) were determined using PCR and where noted (*) amplicon sequencing. Cytotoxicity was verified by MTT assays. Positive (+) and negative (-) PCR results for *npsB*, *leupAB*, *orfABC/orfA′* or "only core" are indicated.

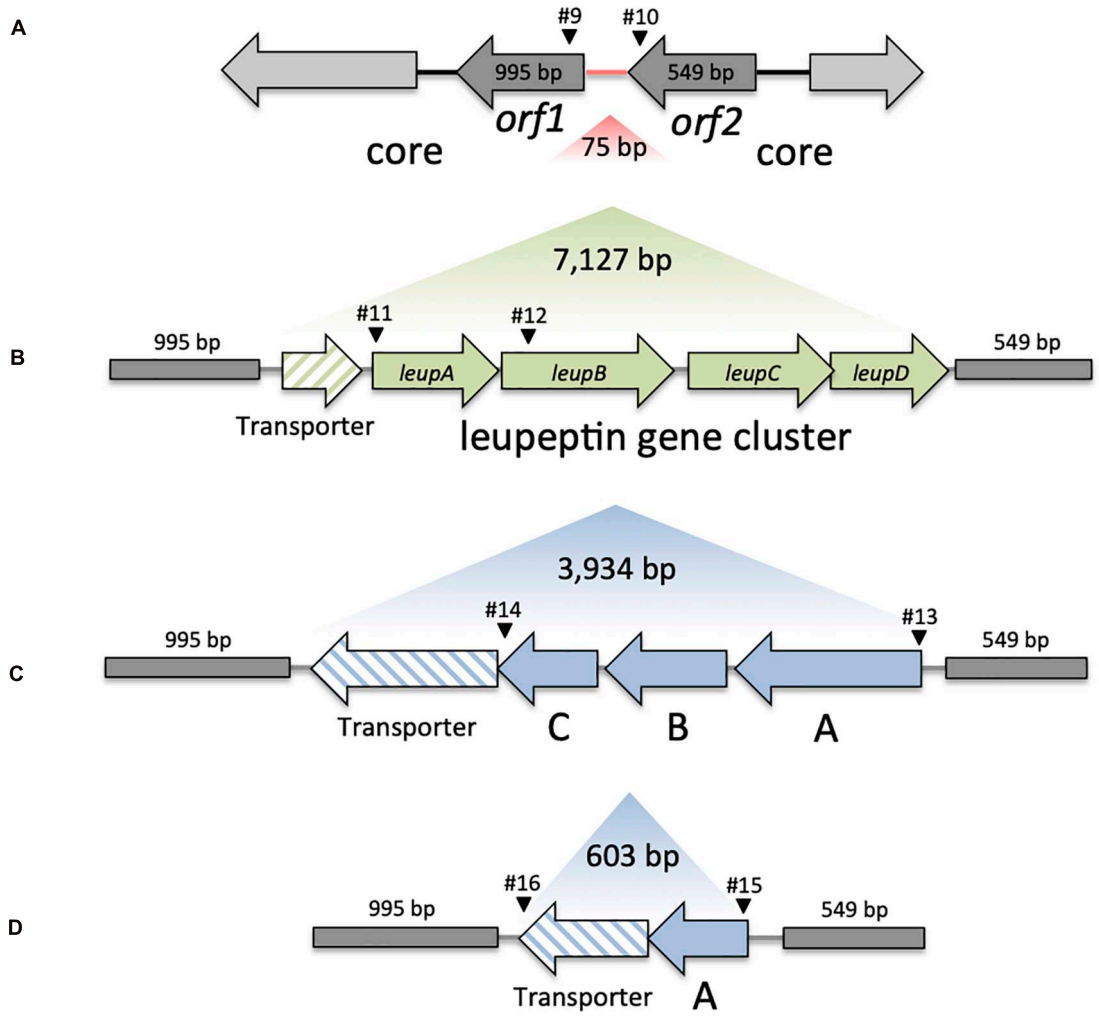


FIGURE 2 | Genomic variation in KoSC strains. **(A)** Overview of the core region lacking insertion (Variant A). Region between *orf1* and *orf2* indicated in red represents the insertion site for panel **(B)** Variant B, the transporter and *leupABCD*, **(C)** Variant C, transporter and *orfABC* and **(D)** Variant D, the truncated transporter and truncated *orfA'*. Sizes of the region lacking insertion, core genes, and respective gene clusters are indicated in bp. Black arrowheads indicate approximate binding sites and primer numbers applied in PCR screens of environmental test set (**Table 1**).

gene markers with species classification in database entries may result from misclassification. We further propose that the robustness of these correlations should facilitate typing and re-typing of isolates and will also consolidate species boundaries within the complex.

Average Nucleotide Identity Reveals Discrepancies in Standing *K. oxytoca*, *K. michiganensis,* and *K. grimontii* Classification

Inconsistencies in taxonomic assignment of deposited sequences of *K. oxytoca* has been noted by others previously (Chen et al., 2020) and the observed incongruence in classification is not unexpected as some strains were classified before the phylogroups were separated into species. However, also recent database submissions are misidentified. The problem is perpetuated by incorrectly classified references strains whose use confounds consistent taxonomy of KoSC members.

To move forward, we thought to provide a verified reference set of consistently classified KoSC members. We performed an in-depth analysis of all KoSC genomes retrieved from databases and applied a comparative approach to develop the discriminatory power of typing methods. A survey of database entries deposited for KoSC members retrieved none with the identifier "*K. spallanzanii*" or "*K. pasteurii*" and only 4 sequences of *K. huaxiensis*. We then downloaded all publicly available genome sequences for *K. oxytoca*, *K. michiganensis*, *K. grimontii* and *K. huaxiensis* and identified four *K. spallanzanii* genomes (bla_{OXY-3}) and 20 *K. pasteurii* (bla_{OXY-4}) among them. To generate a robust test set (**Supplementary Table 1**), we selected all genomes with complete status. To cover phylogroups more evenly we also enrolled genomes with more than 30 contigs where necessary. We further added the following strains from

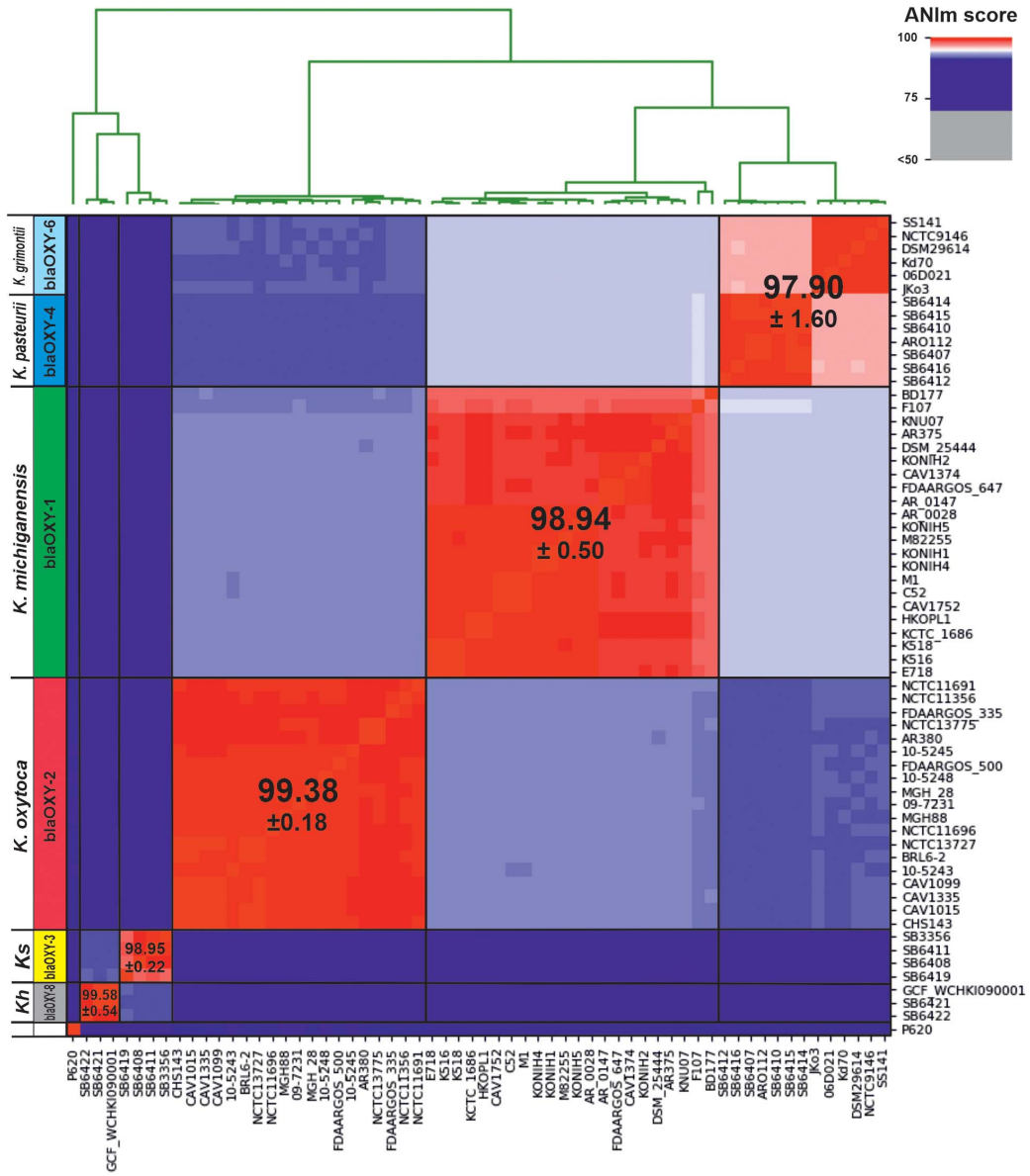


FIGURE 3 | Average nucleotide identity (ANI) detects 5 distinct phylogenetic groups for the KoSC. Here we show genetic distance between 62 genome sequences affiliated with one of the complex species. Scores were calculated with ANIm based on MUMmer alignments and visualized using PYANI (Pritchard et al., 2016). Observed clustering clearly correlates with blaOXY-based phylogroups as indicated horizontally. Affiliated species designations are shown and the average ANI score for each group is indicated. *Ks, K. spallanzanii; Kh, K. huaxiensis*.

recent publications regardless of their assembly status: *K. oxytoca* strain DSM29614 (Gallo et al., 2018), *K. grimontii* reference strain 06D021 (Passet and Brisse, 2018), the identified *K. spallanzanii* strains (Merla et al., 2019), *K. michiganensis* strain Kd70 (Dantur et al., 2018), *K. michiganensis* ARO112 (Oliveira et al., 2020), and reference strain DSM25444, which was the first *K. michiganensis* strain described (Saha et al., 2013).

Average nucleotide identities (ANI) of a large set of KoSC members has not been reported, thus we chose to first evaluate species boundaries in the KoSC test set assembled above using this unbiased whole-genome similarity metric regardless of the existing taxonomy. ANI resolved clusters in 5 major groups (**Figure 3**). blaOXY group and accessory gene typing of each test strain strongly correlated with the observed ANI affiliation (**Figure 3** and **Supplementary Table 1**). *K. pasteurii* (bla_{OXY-4}) and *K. grimontii* (bla_{OXY-6}) are highly related. These are not differentiated when the ANI-based > 5% sequence discontinuity threshold for species separation is applied. Instead, a sub-clustering correlating with the respective blaOXY group is apparent. In summary, the average identity for *K. oxytoca* (bla_{OXY-2}) was 99.38 (SD ± 0.21), for *K. michiganensis,* (bla_{OXY-1}) 98.94 (SD ± 0.56), and for *K. grimontii* (bla_{OXY-6})

and *K. pasteurii* (bla_{OXY-4}) combined 97.90 (SD ± 0.94). Scores for the limited number of analyzed *K. spallanzanii* (bla_{OXY-3}) were 98.95 (SD ± 0.56), and for *K. huaxiensis* (bla_{OXY-8}) 99.58 (SD ± 0.56). Inter-group comparisons in this analysis support monophyly of the four biochemically identical species as they are more closely related to each other than to *K. spallanzanii* and *K. huaxiensis* (**Figure 3**). The putative *K. oxytoca* strain P620 was not typeable for blaOXY and is phylogenetically distant to all investigated phylogroups.

Next, we asked whether species separation can also be achieved using discrete sets of core genes. Robust phylogenetic trees were generated for the same test set of KoSC genomes using a variety of comparisons. The first tree (**Figure 4A**) is based on coding information for 136 reference proteins (Amphora2, supermatrix) and included as an outer-group other closely related *Klebsiella*/*Raoultella* species. The resulting tree clusters showed excellent agreement with ANI results. However, in contrast to ANI, this analysis based on selected reference proteins suggests a close relationship of Ko2 (*K. oxytoca*) and Ko3 (*K. spallanzanii*) strains. Notably, the assessment also shows that strains *R. ornithinolytica* S12 and *K. aerogenes* NCTC6944 should be reassigned to *R. terrigena* and *K. pneumoniae*, respectively. This classification was validated by ANI.

Finally, we generated phylogenetic trees for the biochemically identical species based on a large portion of the core genome (2,785 proteins) *versus* just seven conserved MLST core genes (**Figures 4B,C**). The comparison resulted in identical clustering except for strains AR380, BD177, F107, and NCTC13775. Notably, not all strains can be typed using MLST particularly bla_{OXY-4} and bla_{OXY-6} strains.

The comparison confirmed converging outcomes using multiple typing methods based on a wide spectrum of markers. The relevant conclusion is that MLST cluster A represents *K. oxytoca*, sub-group B1, *K. michiganensis*, and sub-group B2 comprises *K. pasteurii*, and *K. grimontii*, which, nonetheless could be clearly separated into distinct groups in every analysis performed. Thus, the phylogenies determined from STs of this reference set of genomes (**Figure 4C** and **Supplementary Table 1**) as well as the STs of the environmental test set (**Figure 1**) can provide a standard for species identification of KoSC isolates where MLST results are available.

Unifying Classification Scheme for Distinction of the Biochemically Identical KoSC Members

We next performed a validation test by analyzing and visualizing the relationship of all 316 KoSC genomes including *K. spallanzanii* and *K. huaxiensis* and closely related *Klebsiella* and *Raoultella* species available in June 2020 (**Supplementary Table 2**). The inferred phylogenetic tree (Amphora2, supermatrix) showed a clustering in three major branches (**Figure 5**). Next, we interrogated the correlation of branching with presence of blaOXY variants and accessory gene markers. Blast analysis using bla_{OXY} variants, *leupAB*, *orfABC,* and *orfA′* as query sequences established correlation of main clusters and phylogroups as well as with the MLST-tree branches A, B1, and B2 (**Figure 5** and **Supplementary Table 2**). In good agreement with previous findings (Li et al., 2020), Ko5 (bla_{OXY-5}) strains cluster in branch B as a distinct group (intermediate cluster) but isolates share the *leupAB* locus present in sub-group B1 strains. Interestingly, we identified a selection of Ko1 (bla_{OXY-1}) strains within cluster A. These isolates were positive for *leupAB* and therefore share features with sub-group B1 strains. According to ANI, these intermediate isolates are *K. michiganensis* strains and not *K. oxytoca* strains. Similarly, some *K. michiganensis* strains were affiliated with sub-group B2, yet they still carry *leupAB* and not *orfABC* or *orfA′*. This finding highlights the monophyletic relationship of these complex members as phylogenetic trees inferred from concatenated core genes cannot always fully resolve phylogroup boundaries. However, species separation is achieved based on a combination of blaOXY and accessory gene typing.

In summary, we argue that the results of the numerous approaches taken here provide sufficient evidence to support reclassification of 94 of 316 KoSC strains in the databases (**Supplementary Table 2**). Notably, 18 isolates did not cluster in any of the major branches and represent other closely related *Klebsiella*/*Raoultella* species (**Supplementary Table 2**).

Validation of Rapid, Reliable Species Identification

To test the capacity for rapid reliable typing using the accessory gene content described above, we performed *in silico* PCRs on this validation set of all 316 KoSC genomes and the 24 other *Klebsiella* genomes (**Supplementary Table 2**) with the primer pairs used to detect *leupAB*, *orfABC*, and *orfA′* in the environmental test set. Results were in good agreement with the outcome of blast analyses for the respective gene loci (**Supplementary Table 2**). Application of *in silico* PCR without allowing any mismatch detected *leupAB* in 89% (103/116) and, when only one mismatch per primer was allowed, in 95% (109/116) of the *K. michiganensis* genomes previously positive in *leupAB* blast analysis. With both primer stringencies *orfA*' was detected in 75% (15/20) of the *K. pasteurii* genomes previously positive in *orfA*' blast analysis and in 100% (38/38) of *K. grimontii* genomes previously positive in *orfABC* blast analysis. These *in silico* results confirm that primer-binding sites are highly conserved. Since we were able to type all environmental isolates experimentally with this PCR, we predict that the actual detection rate in practice will be higher than observed in the very stringent *in silico* PCR. Moreover, specificity appears to be given as *in silico* PCRs and blast analyses for the gene loci were negative for *K. spallanzanii*, *K. huaxiensis*, and other members of the *Klebsiella* group tested in this analysis (**Supplementary Table 2**).

In a final step, we verified the validity of our approach using non-assembled genome sequences. Fastq files of phylogenetically characterized KoSC strains of clinical origin (Moradigaravand et al., 2017) were extracted from BioProject PRJEB5065 (Moradigaravand et al., 2016) and subjected to short read sequence typing. Analysis of this population (Validation set II: clinical) revealed that the characteristic distribution of accessory

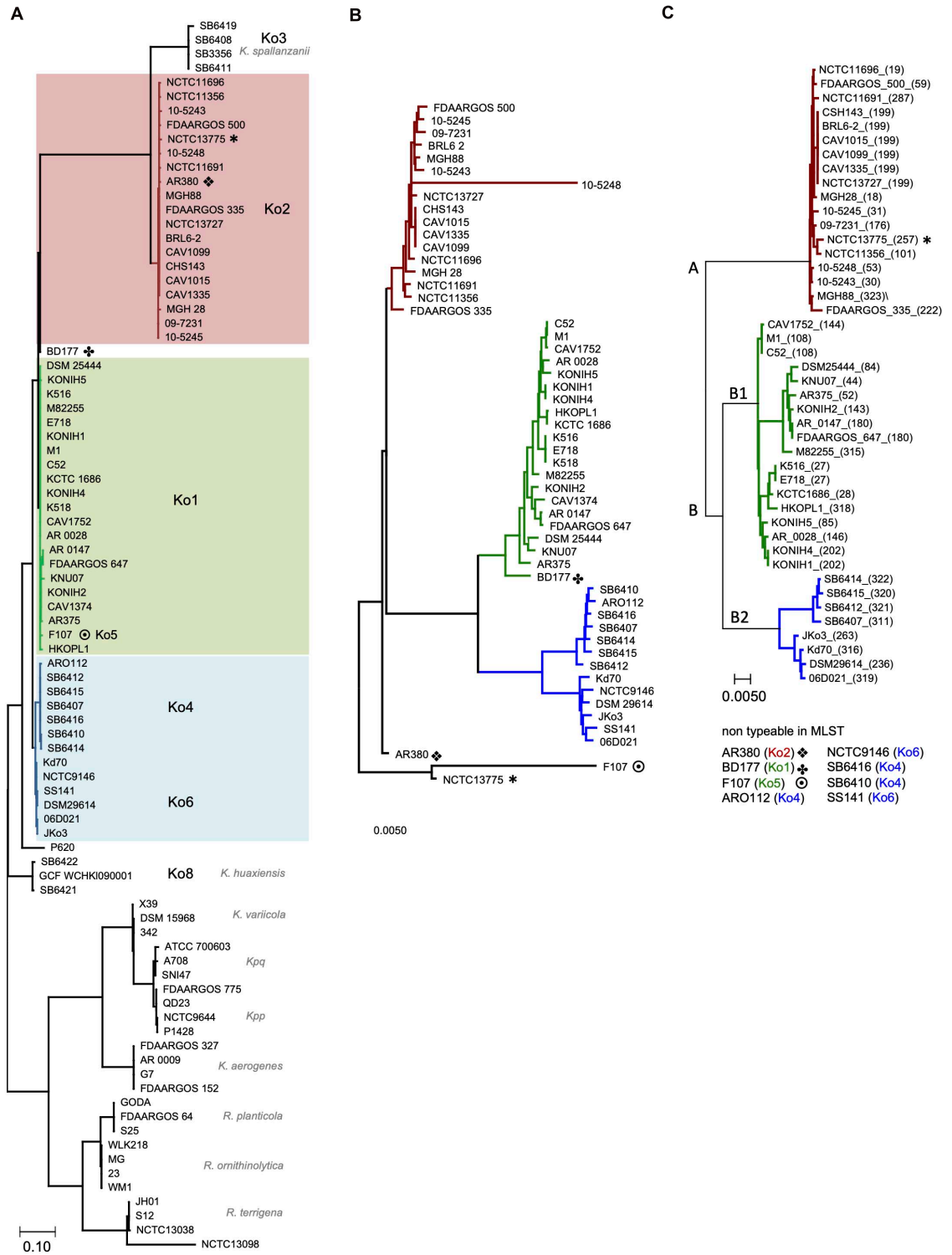


FIGURE 4 | Phylogenetic analysis shows congruency of typing methods. **(A)** Phylogenetic tree inferred from concatenated, partitioned alignment of 136 marker proteins (Amphora2, supermatrix). Mayor tree clusters, which correspond to phylogroups (Ko), are highlighted in color (red, Ko2; green, Ko1; blue, Ko4 and Ko6). *K. spallanzanii* (Ko3), *K. huaxiensis* (Ko8) are indicated and species affiliation for each sub-branch of the out-group is shown; The out-group was used to root the tree. **(B)** Neighbor-joining tree based on the shared core genome of analyzed strains. Phylogenetic relationship was resolved based on 2,785 proteins. **(C)** Neighbor-joining tree based on 7 concatenated housekeeping gene loci used for MLST typing. Major MLST clusters (**A** and **B**) and sub-clusters B1 and B2 are indicated. Sequence types (STs) are given in parenthesis following the strain designations. Non-typeable strains are listed below with their respective phylogroup. Trees in panels **(B,C)** are mid-point rooted. Scale bars are indicated for each tree. Strains that do not cluster congruently between the trees are indicated by symbols. * is one of the symbols used to indicate incongruent clustering.

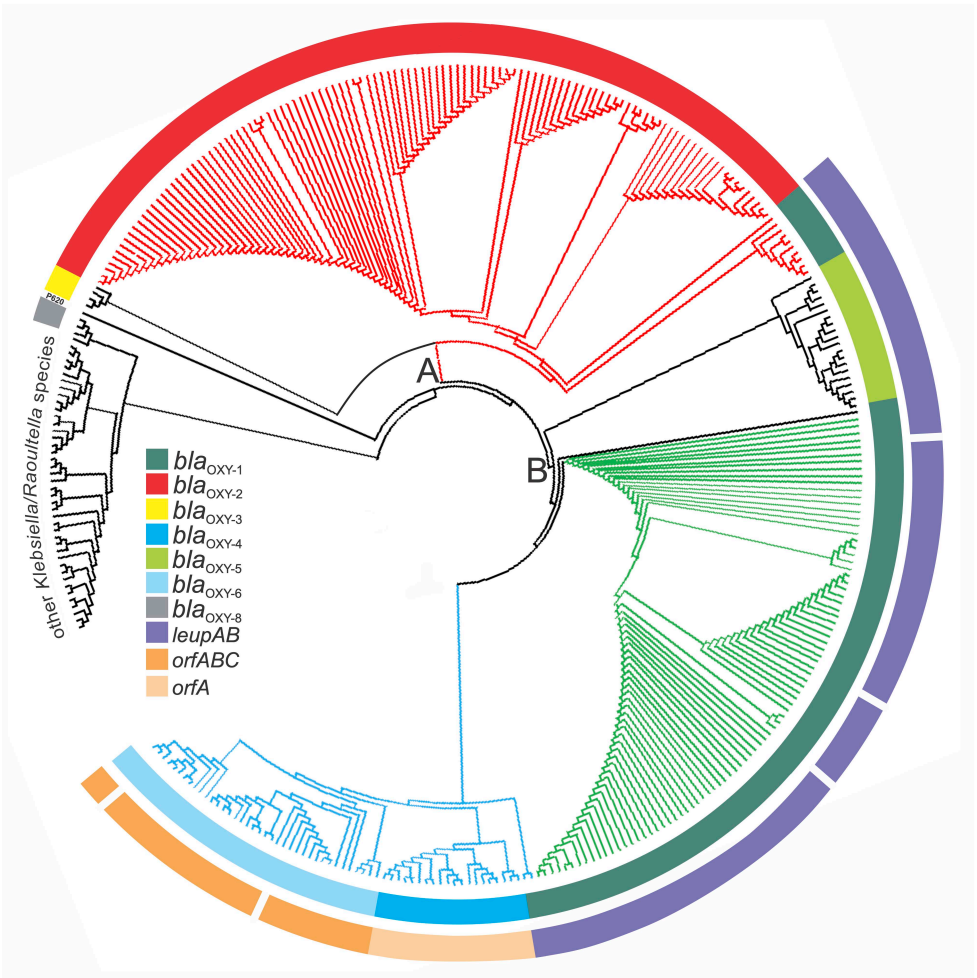


FIGURE 5 | Phylogenetic representation of 341 *Klebsiella* genomes retrieved from GenBank. Phylogenetic analysis was done with PhyloPhlAn. The three main branches are colored according to MLST clusters and phylogroups as previously described: Cluster A, red; sub-cluster B1, green, sub-cluster B2, blue. The inner ring is colored based the respective *bla*OXY variant present (representing phylogroups Ko1-Ko8) and the outer ring indicates the presence or absence of species-specific marker genes for *K. michiganensis* (*leupAB*) and *K. grimontii* (*orfABC*) or *K. pasteurii* (*orfA'*). Other *Klebsiella*/*Raoultella* species: *K. huaxiensis* and 18 of 313 database entries of KoSC members clustered outside the major tree branches A and B with 24 reference strains (**Supplementary Table 2**).

genes along the KoSC phylogeny is also conserved in strains of clinical origin. In agreement with previously published data (Moradigaravand et al., 2017) we identified the respective blaOXY alleles for each strain as well as the phylogroup-matching accessory genes (**Supplementary Table 3**) *leupAB*, *orfABC*, or the genomic core region lacking any insertion. No *K. pasteurii* (*orfA*') was identified in this strain set. All of the tested non-KoSC genomes lack *leupAB*, *orfABC*, *orfA*', and bla_{OXY} variants. One exceptional *K. quasipneumoniae* strain gave a positive result for the conserved "empty" core region but the isolate was not positive for any of the bla_{OXY} variants.

Based on these analyses, we conclude that simple assessment of four gene loci can be applied for reliable species affiliation of monophyletic members of the KoSC. In this classification scheme defining criteria for differentiation of the biochemically identical species are: *K. oxytoca* (bla_{OXY-2}, *leupAB*-, *orfABC*-, *orfA'*-), *K. michiganensis* ($bla_{OXY-1\&5}$, *leupAB* + , *orfABC*-, *orfA'*-), *K. pasteurii* (bla_{OXY-4}, *leupAB*-, *orfABC*-, *orfA'*+) and *K. grimontii* (bla_{OXY-6}, *leupAB*-, *orfABC*+, *orfA'*-).

DISCUSSION

Consistent, systematic classification of bacteria is challenging but highly important since nomenclature influences the way data gathered on organisms is interpreted. The KoSC comprises several closely related species of clinical and economic significance. To date *K. oxytoca, K. michiganensis, K. pasteurii* and *K. grimontii* cannot be distinguished morphologically, biochemically, or based on their 16s rRNA gene sequence. Moreover, they seem to inhabit the same ecological niches and share important virulence properties like the tilimycin/tilivalline biosynthetic gene cluster.

In this study we aimed to develop a uniform classification scheme with the capacity to define species of the KoSC with high resolution. We took a multi-pronged approach to determine the population structure of an uncharacterized collection of KoSC members of environmental origin. We applied genome-based species definition to the large volume of publicly available genomes. Congruency of phylogenies derived when indexing core and accessory genes helped validate genotypic relationships within the complex and also define clear genetic discontinuities. The data reveal a simple routinely applicable classification scheme with sufficient discriminatory power to resolve the biochemically identical *K. oxytoca*, *K. michiganensis*, *K. pasteurii* and *K. grimontii*. The data also expose discrepancies between the results of genome-based species definitions and standing nomenclature of the NCBI taxonomy database (**Supplementary Table 2**). Although stability in bacterial nomenclature is certainly desirable, reclassification of strains in this complex appears necessary. The community may choose to reclassify prior entries and possibly reinterpret clinical data gathered on organisms in the past. Such steps would help prevent continued confusion about organisms of the KoSC and clarify the clinical relevance of each species.

Further information gained from this study included insights into the distribution of KoSC members in the environment. We sought to recover strains from sewage sludge, soil, meat and surface water and ultimately cultured and isolated members of the complex from 68 of 128 environmental samples. Although the population size is too small to draw broad conclusions regarding environmental distribution of *K. oxytoca*, *K. michiganensis, K. pasteurii* and *K. grimontii* it is interesting to note that every second meat product analyzed was contaminated with at least one KoSC strain. Of these 32 isolates 19 were *npsB*+, implicating retail meat as a likely source of transmission of cytotoxin-producing *Klebsiella spp*. (**Figure 1**). Antibiotic-associated hemorrhagic colitis is caused by the overgrowth of toxigenic strains following use of antibiotics such as penicillin or amoxicillin (Högenauer et al., 2006). Transmission of toxin positive strains from food during antibiotic therapies might therefore increase patient risk for severe infections.

Klebsiella michiganensis strains can be isolated from the gastro-intestinal tract (Chen et al., 2020), blood and cerebrospinal fluid (Carrie et al., 2019) of infants, and they are affiliated with upper-respiratory diseases in adults (Herzog et al., 2014). In this study, we isolated *K. michiganensis* strains predominantly from sewage sludge. Previously, it was shown that sub-group B1 strains have the potential to synthesize protease inhibitor leupeptin, which might provide a colonization advantage in the lung (Li et al., 2020).

Klebsiella grimontii strains have been isolated from the gastro-intestinal tract of infants, adults, and animals as well as from diverse environmental sites (Passet and Brisse, 2018; Chen et al., 2020; Oliveira et al., 2020). Our analyses show that these *K. grimontii* strains include *K. pasteurii* (bla_{OXY-4}) isolates. Based on the ANI threshold 95, our analysis does not support a separation of phylogroups Ko4 and Ko6 strains into distinct species but strains of the two groups can clearly be differentiated based on core and accessory genes investigated in this study.

In summary, we show convergence of ANI grouping, blaOXY typing, and MLST phylogeny, thus previously obtained MLST data can be reinterpreted for species delineation. Routine laboratories will continue to identify strains following a conventional polyphasic strategy but resolution of that approach can now be improved with the rapid and simple PCR-typing strategy described here. The method allows robust species delineation without requiring whole genome sequencing or even bla_{OXY} sequencing. The available data allow us to propose that once an isolate is biochemically identified as *K. oxytoca*, PCR for bla_{OXY-1} *and* bla_{OXY-2} can reliably differentiate *K. michiganensis* (except bla_{OXY-5}) and *K. oxytoca* strains. Differentiation of *K. grimontii* (bla_{OXY-6}) and *K. pasteurii* (bla_{OXY-4}) or identification of bla_{OXY-5} strains requires bla_{OXY} gene sequencing. Yet, as an alternative in those cases, a simple PCR screen for *leupAB*, *orfABC*, and *orfA′* is sufficient to differentiate the species.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

EL, EZ, and SK conceived and designed the study. CP, HG, FR, GF, and CH conceived the environmental sampling. EL, CP, HG, KH-O, ET, SR, and SK isolated and characterized the environmental isolates. AC and SK collected and analyzed the data and performed the computational analysis. AC, EL, EZ, and SK wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported the doc.fund Molecular Metabolism (DOC 50) funded by the Austrian Science Fund (FWF), Land Steiermark, the City of Graz and the Hygiene Fund Graz.

ACKNOWLEDGMENTS

We are grateful to Thomas Rattei for critical reading of the manuscript and guidance in computational analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.692453/full#supplementary-material

Supplementary Figure 1 | Illustration of workflow. Five specific sets of samples/sequence data were sequentially analyzed **(A–E)**. Their respective

isolation and/or download source is indicated above. The main selection criteria for compiling the respective data sets can be found together with the number of strains/genomes analyzed in the blue boxes. Orange boxes identify which analysis/tool was performed/applied on each data set and lead over to the green boxes that indicate which figure/table presents the respective results. A more detailed description of data sets and methodology is provided in the "Methods" section.

Supplementary Table 1 | Summary of MLST and Blast results for the reference set of *Klebsiella oxytoca* species complex.

Supplementary Table 2 | Summary of Blast and in silico PCR results for 316 *K. oxytoca* species complex members and 24 other *Klebsiella/Raoultella* species.

Supplementary Table 3 | Summary of short read sequence typing of KoSC isolates.

REFERENCES

Brisse, S., Milatovic, D., Fluit, A. C., Verhoef, J., and Schmitz, F.-J. (2000). Epidemiology of Quinolone Resistance of *Klebsiella pneumoniae* and *Klebsiella oxytoca* in Europe. *Eur. J. Clin. Microbiol. Infect. Dis.* 19, 64–68.

Carrie, C., Walewski, V., Levy, C., Alexandre, C., Baleine, J., Charreton, C., et al. (2019). *Klebsiella pneumoniae* and *Klebsiella oxytoca* meningitis in infants. *Epidemiol. Clin. Features. Archiv. Pédiatrie* 26, 12–15. doi: 10.1016/j.arcped.2018.09.013

Chapman, P., Forde, B. M., Roberts, L. W., Bergh, H., Vesey, D., Jennison, A. V., et al. (2020). Genomic Investigation Reveals Contaminated Detergent as the Source of an Extended-Spectrum-β-Lactamase-Producing *Klebsiella michiganensis* Outbreak in a Neonatal Unit. *J. Clin. Microbiol.* 58, e1980–e1919.

Chaudhari, N. M., Gupta, V. K., and Dutta, C. (2016). BPGA- an ultra-fast pan-genome analysis pipeline. *Sci. Rep.* 6:24373.

Chen, Y., Brook, T. C., Soe, C. Z., O'Neill, I., Alcon-Giner, C., Leelastwattanagul, O., et al. (2020). Preterm infants harbour diverse *Klebsiella populations,* including atypical species that encode and produce an array of antimicrobial resistance- and virulence-associated factors. *Microbial. Genomics* 6:377.

Dantur, K. I., Chalfoun, N. R., Claps, M. P., Tórtora, M. L., Silva, C., Jure, Á, et al. (2018). The Endophytic Strain *Klebsiella michiganensis* Kd70 Lacks Pathogenic Island-Like Regions in Its Genome and Is Incapable of Infecting the Urinary Tract in Mice. *Front. Microbiol.* 9:1548. doi: 10.3389/fmicb.2018.01548

Decré, D., Burghoffer, B., Gautier, V., Petit, J.-C., and Arlet, G. (2004). Outbreak of multi-resistant *Klebsiella oxytoca* involving strains with extended-spectrum β-lactamases and strains with extended-spectrum activity of the chromosomal β-lactamase. *J. Antimicrob. Chemother.* 54, 881–888. doi: 10.1093/jac/dkh440

Dornisch, E., Pletz, J., Glabonjat, R. A., Martin, F., Lembacher-Fadum, C., Neger, M., et al. (2017). Biosynthesis of the Enterotoxic Pyrrolobenzodiazepine Natural Product Tilivalline. *Angew Chem. Int. Ed.* 56, 14753–14757. doi: 10.1002/anie.201707737

Fevre, C., Jbel, M., Passet, V., Weill, F.-X., Grimont, P. A. D., and Brisse, S. (2005). Six Groups of the OXY β-Lactamase Evolved over Millions of Years in *Klebsiella oxytoca*. *AAC* 49, 3453–3462. doi: 10.1128/aac.49.8.3453-3462.2005

Fournier, B., Lu, C. Y., Lagrange, P. H., Krishnamoorthy, R., and Philippon, A. (1995). Point mutation in the pribnow box, the molecular basis of beta-lactamase overproduction in *Klebsiella oxytoca*. *Antimicrob. Agents Chemother.* 39, 1365–1368. doi: 10.1128/aac.39.6.1365

Gallo, G., Presta, L., Perrin, E., Gallo, M., Marchetto, D., Puglia, A. M., et al. (2018). Genomic traits of *Klebsiella oxytoca* DSM 29614, an uncommon metal-nanoparticle producer strain isolated from acid mine drainages. *BMC Microbiol.* 18:198. doi: 10.1186/s12866-018-1330-5

Hendrik, T. C., Voor in 't holt, A. F., and Vos, M. C. (2015). Clinical and Molecular Epidemiology of Extended-Spectrum Beta-Lactamase-Producing *Klebsiella spp*.: A Systematic Review and Meta-Analyses. *PLoS One* 10:e0140754. doi: 10.1371/journal.pone.0140754

Herzog, K. A. T., Schneditz, G., Leitner, E., Feierl, G., Hoffmann, K. M., Zollner-Schwetz, I., et al. (2014). Genotypes of *Klebsiella oxytoca* Isolates from Patients with Nosocomial Pneumonia Are Distinct from Those of Isolates from Patients with Antibiotic-Associated Hemorrhagic Colitis. *J. Clin. Microbiol.* 52, 1607–1616. doi: 10.1128/jcm.03373-13

Högenauer, C., Langner, C., Beubler, E., Lippe, I. T., Schicho, R., Gorkiewicz, G., et al. (2006). *Klebsiella oxytoca* as a Causative Organism of Antibiotic-Associated Hemorrhagic Colitis. *N. Engl. J. Med.* 355, 2418–2426.

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

Izdebski, R., Fiett, J., Urbanowicz, P., Baraniak, A., Derde, L. P. G., Bonten, M. J. M., et al. (2015). Phylogenetic lineages, clones and β-lactamases in an international collection of *Klebsiella oxytoca* isolates non-susceptible to expanded-spectrum cephalosporins. *J. Antimicrob. Chemother.* 2015:dkv273. doi: 10.1093/jac/dkv273

Joainig, M. M., Gorkiewicz, G., Leitner, E., Weberhofer, P., Zollner-Schwetz, I., Lippe, I., et al. (2010). Cytotoxic Effects of *Klebsiella oxytoca* Strains Isolated from Patients with Antibiotic-Associated Hemorrhagic Colitis or Other Diseases Caused by Infections and from Healthy Subjects. *J. Clin. Microbiol.* 48, 817–824. doi: 10.1128/jcm.01741-09

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evolut.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12.

Li, J., Oh, J., Kienesberger, S., Kim, N. Y., Clarke, D. J., Zechner, E. L., et al. (2020). Making and Breaking Leupeptin Protease Inhibitors in Pathogenic Gammaproteobacteria. *Angew Chem. Int. Ed.* 59, 17872–17880. doi: 10.1002/anie.202005506

Lowe, C., Willey, B., O'Shaughnessy, A., Lee, W., Lum, M., Pike, K., et al. (2012). Outbreak of Extended-Spectrum β-Lactamase–producing *Klebsiella oxytoca* Infections Associated with Contaminated Handwashing Sinks1. *Emerg. Infect. Dis.* 18, 1242–1247. doi: 10.3201/eid1808.111268

Merla, C., Rodrigues, C., Passet, V., Corbella, M., Thorpe, H. A., Kallonen, T. V. S., et al. (2019). Description of *Klebsiella spallanzanii* sp. nov. and of *Klebsiella pasteurii* sp. nov. *Front. Microbiol.* 10:2360. doi: 10.3389/fmicb.2019.02360

Moradigaravand, D., Boinett, C. J., Martin, V., Peacock, S. J., and Parkhill, J. (2016). Recent independent emergence of multiple multidrug-resistant *Serratia marcescens* clones within the United Kingdom and Ireland. *Genome Res.* 26, 1101–1109. doi: 10.1101/gr.205245.116

Moradigaravand, D., Martin, V., Peacock, S. J., and Parkhill, J. (2017). Population Structure of Multidrug-Resistant *Klebsiella oxytoca* within Hospitals across the United Kingdom and Ireland Identifies Sharing of Virulence and Resistance Genes with *K. pneumoniae*. *Genome Biol. Evolut.* 9, 574–584. doi: 10.1093/gbe/evx019

Oliveira, R. A., Ng, K. M., Correia, M. B., Cabral, V., Shi, H., Sonnenburg, J. L., et al. (2020). *Klebsiella michiganensis* transmission enhances resistance to Enterobacteriaceae gut invasion by nutrition competition. *Nat. Microbiol.* 5, 630–641. doi: 10.1038/s41564-019-0658-4

Passet, V., and Brisse, S. (2018). Description of *Klebsiella grimontii* sp. nov. *Int. J. Systemat. Evolut. Microbiol.* 68, 377–381. doi: 10.1099/ijsem.0.002517

Paveglio, S., Ledala, N., Rezaul, K., Lin, Q., Zhou, Y., Provatas, A. A., et al. (2020). Cytotoxin-producing *Klebsiella oxytoca* in the preterm gut and its association with necrotizing enterocolitis. *Emerg. Microbes Infect.* 9, 1321–1329. doi: 10.1080/22221751.2020.1773743

Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., and Toth, I. K. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8, 12–24. doi: 10.1039/c5ay02550h

Rønning, T. G., Aas, C. G., Støen, R., Bergh, K., Afset, J. E., Holte, M. S., et al. (2019). Investigation of an outbreak caused by antibiotic−susceptible *Klebsiella oxytoca* in a neonatal intensive care unit in Norway. *Acta Paediatr.* 108, 76–82. doi: 10.1111/apa.14584

Saha, R., Farrance, C. E., Verghese, B., Hong, S., and Donofrio, R. S. (2013). *Klebsiella michiganensis* sp. nov., A New Bacterium Isolated from a Tooth Brush Holder. *Curr. Microbiol.* 66, 72–78. doi: 10.1007/s00284-012-0245-x

Segata, N., Börnigen, D., Morgan, X. C., and Huttenhower, C. (2013). PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* 4:2304.

Singh, L., Cariappa, M. P., and Kaur, M. (2016). *Klebsiella oxytoca*: An emerging pathogen? *Med. J. Armed Forces India* 72, S59–S61.

Unterhauser, K., Pöltl, L., Schneditz, G., Kienesberger, S., Glabonjat, R. A., Kitsera, M., et al. (2019). *Klebsiella oxytoca* enterotoxins tilimycin and tilivalline have distinct host DNA-damaging and microtubule-stabilizing activities. *Proc. Natl. Acad. Sci. U S A.* 116, 3774–3783. doi: 10.1073/pnas.1819154116

Validi, M., Soltan Dallal, M. M., Douraghi, M., Fallah Mehrabadi, J., and Rahimi Foroushani, A. (2016). Identification of *Klebsiella pneumoniae* Carbapenemase-producing *Klebsiella oxytoca* in Clinical Isolates in Tehran Hospitals, Iran by Chromogenic Medium and Molecular Methods. *Osong Public Health Res. Perspect.* 7, 301–306.

Vasaikar, S., Obi, L., Morobe, I., and Bisi-Johnson, M. (2017). Molecular Characteristics and Antibiotic Resistance Profiles of *Klebsiella* Isolates in Mthatha, Eastern Cape Province, South Africa. *Int. J. Microbiol.* 2017, 1–7. doi: 10.1155/2017/8486742

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

frontiers in Microbiology

ORIGINAL RESEARCH
published: 05 July 2021
doi: 10.3389/fmicb.2021.651520

Comparative Whole Genome Sequence Analysis and Biological Features of *Clostridioides difficile* Sequence Type 2‡

Xingxing Xu[1,2†], Qiao Bian[3†], Yun Luo[4], Xiaojun Song[5], Shan Lin[2], Huan Chen[6,7], Qian Liang[6,7], Meixia Wang[6,7], Guangyong Ye[1], Bo Zhu[1], Liang Chen[8,9], Yi-Wei Tang[10], Xianjun Wang[11] and Dazhi Jin[2,5]**

[1] Department of Clinical Laboratory, Women's Hospital, Zhejiang University School of Medicine, Hangzhou, China, [2] School of Laboratory Medicine, Hangzhou Medical College, Hangzhou, China, [3] Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China, [4] School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia, [5] Centre of Laboratory Medicine, Zhejiang Provincial People's Hospital, People's Hospital of Hangzhou Medical College, Hangzhou, China, [6] Key Laboratory of Microorganism Technology and Bioinformatics Research of Zhejiang Province, Hangzhou, China, [7] NMPA Key Laboratory for Testing and Risk Warning of Pharmaceutical Microbiology, Hangzhou, China, [8] Center for Discovery and Innovation, Hackensack Meridian Health, Nutley, NJ, Untied States, [9] Department of Medical Sciences, Hackensack Meridian School of Medicine, Nutley, NJ, Untied States, [10] Cepheid, Danaher Diagnostic Platform, Shanghai, China, [11] Department of Clinical Laboratory, Hangzhou First People's Hospital, Zhejiang University School of Medicine, Hangzhou, China

OPEN ACCESS

Edited by:
Guido Werner,
Robert Koch Institute (RKI), Germany

Reviewed by:
Jason Sahl,
Northern Arizona University,
United States
Patricia Severino,
Albert Einstein Israelite Hospital, Brazil

***Correspondence:**
Dazhi Jin
jind@hmc.edu.cn
Xianjun Wang
wangxj0525@126.com

[†]These authors have contributed equally to this work

[‡]This study was presented in part at the ASM Microbe 2019 (2019).

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: 20 January 2021
Accepted: 31 May 2021
Published: 05 July 2021

Citation:
Xu X, Bian Q, Luo Y, Song X, Lin S, Chen H, Liang Q, Wang M, Ye G, Zhu B, Chen L, Tang Y-W, Wang X and Jin D (2021) Comparative Whole Genome Sequence Analysis and Biological Features of Clostridioides difficile Sequence Type 2. Front. Microbiol. 12:651520. doi: 10.3389/fmicb.2021.651520

Clostridioides difficile sequence type 2 (ST2) has been increasingly recognized as one of the major genotypes in China, while the genomic characteristics and biological phenotypes of Chinese ST2 strains remain to be determined. We used whole-genome sequencing and phylogenetic analysis to investigate the genomic features of 182 ST2 strains, isolated between 2011 and 2017. PCR ribotyping (RT) was performed, and antibiotic resistance, toxin concentration, and sporulation capacity were measured. The core genome Maximum-likelihood phylogenetic analysis showed that ST2 strains were distinctly segregated into two genetically diverse lineages [L1 (67.0% from Northern America) and L2], while L2 further divided into two sub-lineages, SL2a and SL2b (73.5% from China). The 36 virulence-related genes were widely distributed in ST2 genomes, but in which only 11 antibiotic resistance-associated genes were dispersedly found. Among the 25 SL2b sequenced isolates, RT014 (40.0%, $n = 10$) and RT020 (28.0%, $n = 7$) were two main genotypes with no significant difference on antibiotic resistance ($\chi^2 = 0.024–2.667$, $P > 0.05$). A non-synonymous amino acid substitution was found in *tcdB* (Y1975D) which was specific to SL2b. Although there was no significant difference in sporulation capacity between the two lineages, the average toxin B concentration (5.11 ± 3.20 ng/μL) in SL2b was significantly lower in comparison to those in L1 (10.49 ± 15.82 ng/μL) and SL2a (13.92 ± 2.39 ng/μL) ($\chi^2 = 12.30$, $P < 0.05$). This study described the genomic characteristics of *C. difficile* ST2, with many virulence loci and few antibiotic resistance elements. The Chinese ST2 strains with the mutation in codon 1975 of the *tcdB* gene clustering in SL2b circulating in China express low toxin B, which may be associated with mild or moderate *C. difficile* infection.

Keywords: *Clostridioides difficile*, whole genome sequencing, ST2, *tcdB*, genomic characteristics

INTRODUCTION

Clostridioides difficile is an anaerobic, spore-forming Gram-positive bacillus that is able to colonize and proliferate in the human gut, especially following changes in the indigenous colonic microbiota after antibiotic use (Knight et al., 2015). The *C. difficile* genomes have been well documented with a high proportion of mobile genetic elements and an ultra-low level of genetic conservation (as low as 16%) (Knight et al., 2016), and its phylogenetics and evolutionary clades have been recognized as well (Knight et al., 2015). A whole-genome sequencing (WGS)-based phylogenetic tree showed that *C. difficile* has six main clades (Knight et al., 2015), in which clade 1 is the largest one with a variety of sequence types (STs), including ST2, which usually resulted in mild or moderate *C. difficile* infection (CDI) (Griffiths et al., 2010; Cheng et al., 2016). ST2 exhibits high genetic diversity with various PCR ribotypes (RT) (Griffiths et al., 2010; Jin et al., 2017), two toxin genes (*tcdA* and *tcdB*) are located on a 19.6-kb pathogenicity locus (PaLoc), and no binary toxins are found (Knight et al., 2015). ST2 also belongs to RT014, which is a successful lineage of *C. difficile* as significant reservoirs in both human and porcine populations in Australia, and there were obviously different genomic features and biological phenotypes in ST2 including antibiotic resistance elements and antibiotic resistance phenotype between different resources (Knight et al., 2016). However, the genomic characteristics and biological phenotypes remain to be determined in other regions.

Recent studies showed that the molecular epidemiology of *C. difficile* in China has its own characteristics with specific antibiotic resistance and genotype profiles as below (Tang et al., 2016; Jin et al., 2017; Luo et al., 2019). A meta-analysis displayed that the pooled incidence of toxigenic *C. difficile* in diarrhea patients in Mainland China was approximately 14%, and ST2 was one of the dominant genotypes in China (Tang et al., 2016). However, ST2 usually led to mild or moderate CDI with no other severe symptoms, making it be easily overlooked in clinical treatment (Jin et al., 2017; Luo et al., 2018). Even though antimicrobial resistance phenotypes of *C. difficile* ST2 in Australia have been described as reported previously, there is still a paucity of data on ST2 genomic characteristics and the biological characteristics, including toxin expression and sporulation in other regions.

WGS provides an ultra-fine scale resolution tool for analysis of genomic characteristics, evaluation of bacterial genetic diversity, identification of subtle genetic variability, identification of signatures of clonal transmission, and assessments of *C. difficile* epidemiology of strains implicated in infection recurrences and outbreaks (Didelot et al., 2012; Eyre and Walker, 2013; Martin et al., 2016). Here, a collection of 182 ST2 strains from different regions of human origin with isolation dates between 2011 and 2017 were studied. Genome characteristics associated with virulence and antibiotic resistance genotypes were studied *in silico*, and biological features, including antibiotic resistance phenotypes, toxin expression, and sporulation, were also measured *in vitro*.

MATERIALS AND METHODS

C. difficile Isolates and WGS Data

A total of 30 ST2 strains were isolated from CDI patients as part of different CDI surveillance programs conducted in different sites including Zhejiang (n = 18) between February 2013 and January 2017, Hebei (n = 10) between January 2011 and December 2014, Hunan (n = 2) in 2014. The remaining 10 ST2 strains were sourced from CDI patients as part of an international collaboration on molecular characteristics of *C. difficile*. Isolates originated from six sites related to hospitals, including Hong Kong (n = 2) in China, Pusan in South Korea (n = 1), Fukuoka in Japan (n = 1), Singapore (n = 3), and Sydney (n = 2) and Perth (n = 1) in Australia. All the 40 available ST2 isolates with no epidemiological relationships were included in this study by the time the study started. Clinical data were collected after the study was approved by the Ethics Committee of the Hangzhou Medical College. All available genomic data for ST2 strains were included in this study up to January 2017. The genome data of a total of 139 ST2 strains published between 2001 and 2016 were downloaded from the National Center for Biotechnology (NCBI[1]). Raw reads data of 3 ST2 strains (SRR1519431, SRR1519374, and SRR1519422) isolated in 1995 and 1997, respectively, were downloaded from the NCBI database (up to January 2017). The information of the 182 strain was provided in **Supplementary Material**.

PCR Ribotyping

Six reference *C. difficile* strains (ATCC 43255, ATCC 43598, BAA-1870, BAA-1803, BAA-1801, and ATCC 700057) were used as controls. PCR ribotyping was performed by using PCR followed by capillary gel electrophoresis described previously (Indra et al., 2008). The 16S rRNA gene primers were labeled at the 5′ end with carboxyfluorescein. After PCR amplification, PCR fragments were analyzed using an ABI 3100 genetic analyzer (Applied Biosystems, Foster City, CA, United States) with a 36-cm capillary loaded with a POP4 gel (Applied Biosystems). The size of each peak was determined using GeneMapper ID-X 1.3 (Applied Biosystems). RT assignment was performed after the data were deposited in the WEBRIBO database[2].

Whole-Genome Sequencing and Assembly

Genomic DNA was extracted as previously described (Stabler et al., 2009). WGS libraries were created using TruePrep DNA Library Prep^{TM} Kit V2 (Illumina, Santiago, CA, United States). WGS was performed using the Illumina Hiseq X-ten with 150-base paired-end reads. The sequence data were processed and quality controlled according to a standard pipeline as previously described (Preston et al., 2014). Briefly, FASTQ-formatted sequencing reads were quality controlled with a minimum quality Phred score of 30 (as a rolling average over 4 bases) using trimmomatic v0.36. Trimmomatic v0.36 was also used to remove

[1]https://www.ncbi.nlm.nih.gov

[2]https://webribo.ages.at/

adapters and low-quality sequences, and 63.051 Gb clean bases were finally generated (1.616 Gb/per isolate, Q20 ≥ 95%) (Bolger et al., 2014). Genome data for the 40 isolates from this study, and the downloaded raw reads for three genomes, were *de novo* assembled using Velvet (version 1.2.10). Optimal *k*-mers fell between 47 and 93 bp, according to the mean value for median contig size of genome assembly (n50). The genomic sequences of the remaining 139 strains were downloaded from the NCBI.

SNP Detection and Identification

All genome sequences were aligned to the *C. difficile* W0022a reference genome (GCF_002812625.1) and SNPs were identified using MUMmer (v3.23) with default parameters (Kurtz et al., 2004; Li et al., 2016). Following the removal of SNPs within 5 bp of the location interval (less than five bases existed between any two SNPs) by using the perl script (file name: filter_dist.pl.[3]), high-quality SNPs were annotated according to demographic information including clades and locations.

Recombination Detection and Phylogenetic Analysis

Gubbins was used to detect recombination in newly built whole-genome sequences (Croucher et al., 2015) as previously reported (Knight et al., 2015, 2016), and SNPs located in the recombination regions were identified and removed by using the perl script (file name: gubbinssns.vcf_to_genotype.pl., see text footnote 3). After that, an alignment of non-recombinant SNPs was obtained. A total of 182 *C. difficile* genomes were used to generate core genes by using BLAST with the thresholds of 80% nucleotide sequence identity and 80% of the query length. A total of 1,685 genes were defined as the core gene sets using BLASTp with an E-value of 1e-10, out of which the final SNPs were detected from all non-recombinant SNPs. The Maximum likelihood (ML) tree topology and branch length were inferred using IQ-TREE multicore v1.6.6. The recombination/mutation (*r/m*) ratio was calculated within the deep-branching phylogeny, which gives the relative probability that a nucleotide has changed as a result of recombination relative to a point mutation. Two phylogenetic trees before and after removal of SNPs located in the recombination regions by using Gubbins were compared, and the consistency indexes for two phylogenetic trees were analyzed by using the phangorn package in R (Schliep, 2011).

Antibiotic Resistance Genes and Virulence Factors

Antibiotic resistance-associated genes were determined using RGI (version 5.1.0) analysis and the CARD antibiotic resistance gene database[4] (Alcock et al., 2020). Virulence loci were determined by BLAST (Ye et al., 2006; Johnson et al., 2008) analysis of genome sequences in the virulence factors database (VFDB[5]), which aims to provide up-to-date knowledge of virulence factors from various bacteria and serves as a comprehensive warehouse of bacterial pathogenesis knowledge for the scientific community (Liu et al., 2019). The cutoff values were set as 90% nucleotide identity and 90% of the query length for gene coverage to screen virulence loci (Dingle et al., 2019).

Antibiotic Susceptibility Testing

Twelve antibiotics including vancomycin, metronidazole, moxifloxacin, erythromycin, clindamycin, rifampicin, levofloxacin, gatifloxacin, ciprofloxacin, fusidic acid, tetracycline, and piperacillin-tazobactam (PIP-TAZ) were used to test the minimum inhibition concentration (MIC) of the 40 isolates by the agar dilution method according to standard clinical and laboratory guidelines (CLSI) (Clinical and Laboratory Standards Institute, 2011). The reference strain (ATCC 700057) was included in each test as a control. MIC breakpoints were chosen according to a previous report (Jin et al., 2017).

Detection of *C. difficile* Toxin B Cytotoxicity and Sporulation Capacity

Clostridioides difficile toxin B concentration was quantitatively detected using the real-time cell analysis (RTCA) system according to a previously reported method (Ryder et al., 2010). The toxin B concentration was calculated by a formula which was derived based on the testing results of a panel of purified toxin B standards with known concentrations ranging from 0.02 to 200 ng/μL. The sporulation capacity was measured as reported previously (Qin et al., 2017). A count between 30 and 300 spores on each plate was considered significant.

Statistical Analysis

Data analysis was performed using Statistical Package for Social Sciences (SPSS, Chicago, IL, United States) version 22.0. The difference in toxin concentration and sporulation capacity was analyzed using a *t*-test or nonparametric test. A *P*-value of <0.05 was considered statistically significant.

RESULTS

Genomic Characteristics of *C. difficile* ST2

All 182 WGS data were analyzed after sequence quality control and mapping to the *C. difficile* strain W0022a (NCBI accession: GCF_002812625.1). A total of 14,795 high-quality SNPs were identified after the resulting alignments. Gubbins analysis revealed 10,324 spatially clustered SNPs within 1,352 homologous recombination events. After removing these SNPs, 4,471 high-quality biallelic SNPs were extracted. Of these, 1,249 non-rare (core) SNPs (27.9%, 1,249/4,471) were present in the core genome, which included 966 (77.3%, 966/1,249) non-synonymous SNPs. The *r/m* ratio was approximately 2.31, which was determined through 10,324 SNPs divided by 4,471 SNPs. The phylogenetic tree was explored by high-resolution core genome SNPs analysis as briefly shown in **Figure 1A**. Firstly, the *tcdB* gene was conserved across all ST2 strains, and 98.4% (179/182) of ST2 had an intact *tcdA* gene, and the remainder

[3]https://github.com/ChaseFor/Gubbin-snp

[4]https://card.mcmaster.ca/home

[5]http://www.mgc.ac.cn/VFs/main.htm

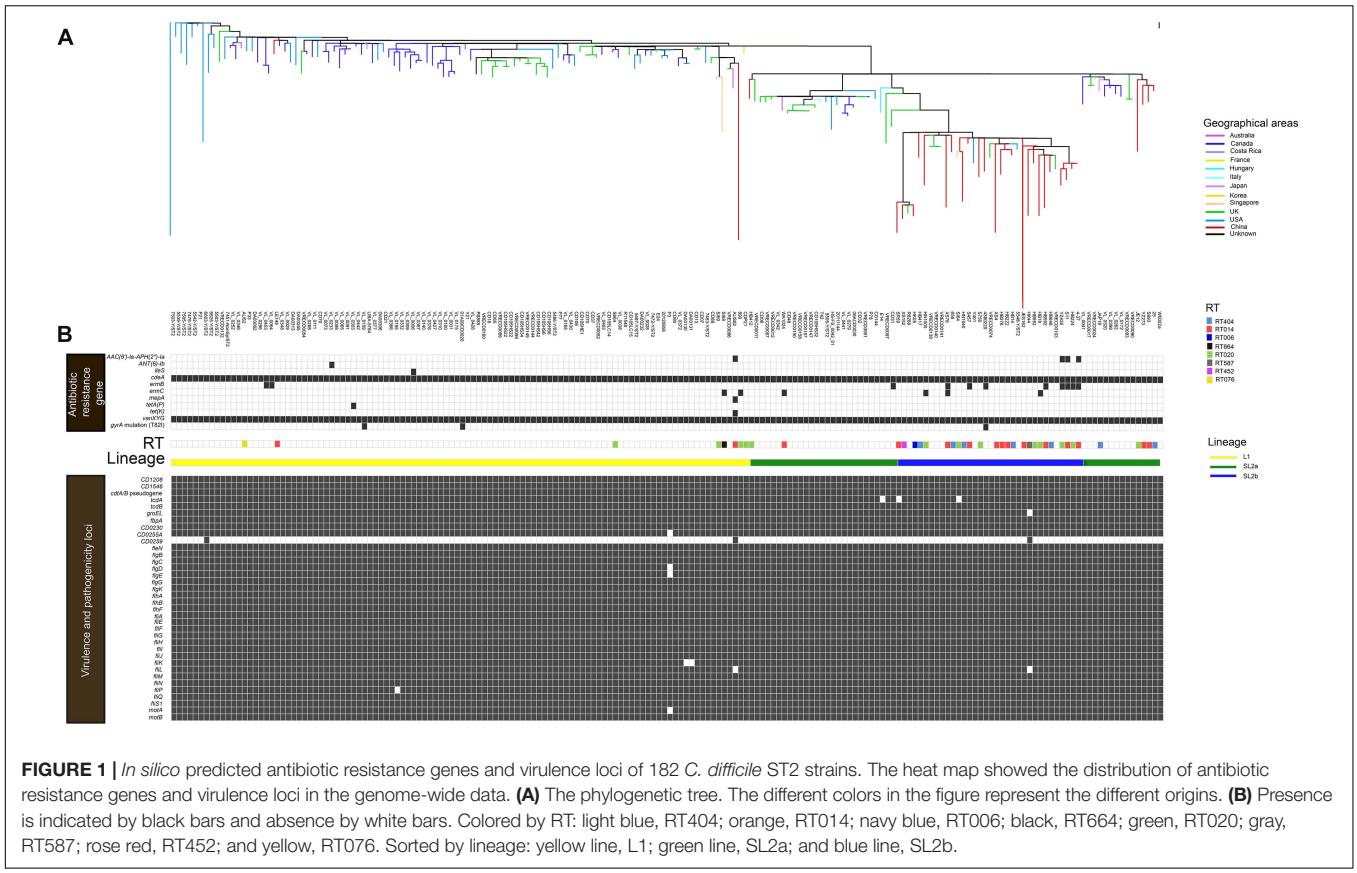


FIGURE 1 | *In silico* predicted antibiotic resistance genes and virulence loci of 182 *C. difficile* ST2 strains. The heat map showed the distribution of antibiotic resistance genes and virulence loci in the genome-wide data. **(A)** The phylogenetic tree. The different colors in the figure represent the different origins. **(B)** Presence is indicated by black bars and absence by white bars. Colored by RT: light blue, RT404; orange, RT014; navy blue, RT006; black, RT664; green, RT020; gray, RT587; rose red, RT452; and yellow, RT076. Sorted by lineage: yellow line, L1; green line, SL2a; and blue line, SL2b.

carried truncated *tcdA* genes with an in-frame deletion of 1,633, 1,732, and 2,371 bp, respectively. All strains harbored *cdtA/B* pseudogene located in a 4.2-kb pathogenicity locus (CdtLoc) as described (Gerding et al., 2014). The remaining 33 virulence-related genes were distributed in the 182 genomes. Of them, 25 virulence-related genes were found in all ST2 strains. The putative flagellar protein gene (*CD0259*), which was associated with *C. difficile* virulence (Aubry et al., 2012), exists in only three of the ST2 genomes. The remaining seven virulence loci including *groEL*, which plays a role in *C. difficile* cell adherence (Hennequin et al., 2001) and decreases *C. difficile* intestinal colonization (Péchiné et al., 2013), *CD0255A, flgE, fliK, fliL, fliP*, and *motA* were widely distributed in ST2 genomes (**Figure 1B**).

A total of 10 putative antibiotic resistance genes and one *gyrA* gene variant with amino acid substitution (T82I) were identified within the 182 genomes (**Figure 1B**). The *vanXY* variant gene, *vanXYG*, mediated aminoglycoside resistance, and the *cdeA* gene, associated with fluoroquinolone resistance, were found in all 182 isolates. The remaining nine genes were found in several genomes as described below. The tetracycline resistance elements, *tetA(P)* and *tet(K)*, were only found in two different *C. difficile* strains (one from Canada, another from Australia) (**Figure 1B**). For macrolide–lincosamide–streptogramin B (MLS_B) resistance, the *ermB* and *ermC* genes were scattered in 6.0% (11/182) and 3.3% (6/182) of strains, respectively. Only three strains carried the *gyrA* mutation (T82I). The *AAC(6′)-1e-APH(2″)-1a* cassette associated with aminoglycoside resistance was found in 2.2% (4/182) of strains. The above results showed that *C. difficile* ST2 genomes dispersedly harbored a small number of antibiotic resistance genes and associated mutations.

Phylogenetic Analysis of ST2 Isolates

A ML phylogeny indicated that the 182 ST2 genomes were obviously divided into two genetically diverse lineages, L1 and L2 (**Figure 2**), primarily differentiated by 18 unique SNPs (**Supplementary Table 1**). L2 could be further divided into two sub-lineages, SL2a and SL2b, based on eight specific SNPs (**Supplementary Table 2**). SL2b primarily comprised most of the *C. difficile* strains from China (25/34, 73.5%), while North America had a high proportion of strains in L1 (66.4%, 71/107) and there was a high diversity of strain sources in SL2a. Various RTs were dispersedly distributed in different sub-lineages as **Figure 1B** shown. The geographical distribution showed significant differences among L1, SL2a, and SL2b in this ML phylogeny. Interestingly, it has been found that SL2b has only one sublineage-specific non-synonymous mutation, which occurred in the *tcdB* gene (W0022a reference genome, base position: T671,314G; amino acid position: Tyr1975Asp). The *r/m* value of the Chinese SL2b was 2.60, which was higher than those of strains from SL2a (2.33) and SL1 (1.26). There were 28 homologous recombination regions found in all strains belonging to SL2, of which 14 were specific to SL2b; six homologous recombination regions existed in almost all of the strains (**Supplementary Figure 1** and **Supplementary Table 3**).

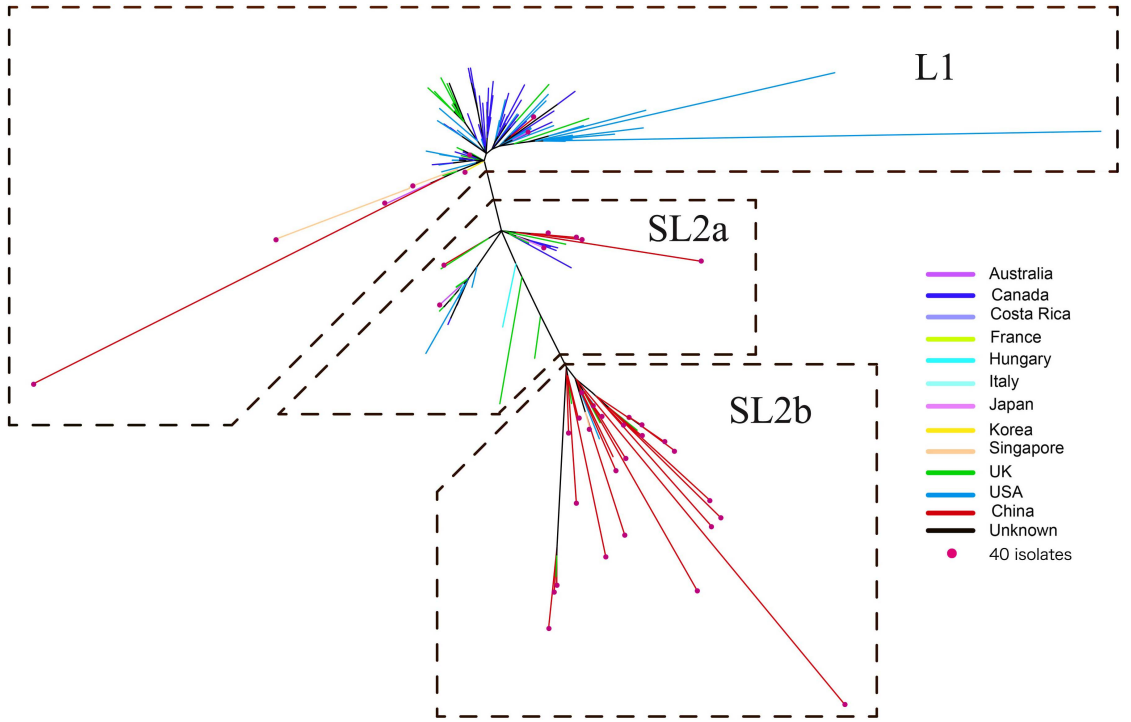


FIGURE 2 | Maximum likelihood phylogeny tree of *C. difficile* ST2. The different colors represented the different origins; the red dot marker was used for 40 *C. difficile* isolates by whole genome sequencing in this study.

In vitro Antibiotic Susceptibility Testing

Antibiotic susceptibility testing was performed in the 40 isolates sequenced in this study (**Table 1**). Summary MIC distributions for 12 antibiotic agents, by RT lineage, were presented in **Figure 3**. The rate of resistance to MLS_B was 90.0% (36/40). The rates of resistance to the fourth-generation fluoroquinolones, gatifloxacin (15.0%, 6/40) and moxifloxacin (10.0%, 4/40) were lower than those to the third-generation fluoroquinolones, levofloxacin (82.5%, 33/40) (χ^2 = 36.47 and 42.29, $P < 0.001$) and ciprofloxacin (92.5%, 37/40) (χ^2 = 48.32 and 54.48, $P < 0.001$). No isolates were resistant to vancomycin or metronidazole. The rates of resistance to MLS_B, fluoroquinolone, and fusidic acid (62.5%, 25/40) were distinctly higher than those to rifampicin (5.0%, 2/40) (χ^2 = 29.57, $P < 0.001$), tetracycline (7.5%, 3/40) (χ^2 = 26.59, $P < 0.001$), and piperacillin (2.5%, 1/40) (χ^2 = 32.82, $P < 0.001$). All the RT020 isolates were susceptible to PIP-TAZ, and resistant to MLS_B and fluoroquinolone. Of them, 61.5% (8/13), 7.7% (1/13), and 7.7% (1/13) were resistant to fusidic acid, and rifampicin, tetracycline, respectively. All of the RT014 isolates were susceptible to rifampicin; 86.7% (13/15) and 93.3% (14/15) were resistant to fluoroquinolone and MLS_B, respectively; 60.0% (9/15) were resistant to fusidic acid; 6.7% (1/15) were resistant to tetracycline and piperacillin. The multidrug-resistant (MDR: resistant to ≥3 of these agents) rate was 65.0% (26/40), noted to be high in this study. Even though no significant differences in the antibiotic patterns were found among the different RTs and SLbs, L1 and SL2a, with no strains resistant to rifampin, tetracycline, PIP-TAZ, and moxifloxacin, they obviously presented different resistance patterns from SL2b (**Supplementary Table 4**).

Toxin Expression and Sporulation

The clinical information for the 40 patients is shown in **Table 1**. All the 40 ST2 isolates had no epidemiological relationships in spite that some of them belonged to the same RT. Some of the isolates were from the same province but not from the same location. Thus, no same clones existed in these ST2 genomes. The RTCA data showed that the average concentrations of toxin B in groups L1, SL2a, and SL2b were 10.49 ± 15.82, 13.92 ± 2.39, and 5.11 ± 3.20 ng/μL, respectively (**Figure 4**). *C. difficile* isolates in SL2a had significantly higher toxin concentration abilities than those in SL2b (t = 6.709, $P < 0.001$). Spore resuscitation abilities were represented in the mode of InterQuartile Range [M (P25, P75)]. The numbers for *C. difficile* sporulation in L1, SL2a and SL2b were 1.1×10^6 (5.2×10^5, 2.3×10^6)/mL, 1.2×10^6 (7.8×10^4, 1.3×10^6)/mL, and 6.6×10^5 (2.5×10^5, 1.2×10^6)/mL, respectively (**Figure 4**). There were no statistical differences in sporulation in the three groups (χ^2 = 2.13, P = 0.346).

DISCUSSION

Clostridioides difficile ST2 is one of the most dominant genotypes associated with mild CDI or asymptomatic carriage in China (Tian et al., 2016; Zhang et al., 2016; Jin et al., 2017), whereas

TABLE 1 | Clinical information of 40 *C. difficile* ST2 isolate.

Patient characteristics	*N* = 40
Demographics	
Age [mean, median (range)] (year)[a]	51.0 (3–95)
Gender[a] Male [*n* (%)]	17 (42.5%)
Isolation place [*n* (%)]	
China	
Zhejiang	18 (45.0%)
Hubei	10 (25.0%)
Hunan	2 (5.0%)
Hong Kong	2 (5.0%)
Japan	1 (2.5%)
South Korea	1 (2.5%)
Singapore	3 (7.5%)
Australia	3 (7.5%)
Isolation place [*n* (%)]	
2011–2013	12 (30.0%)
2014	18 (45.0%)
2015–2017	10 (25.0%)
Ward type [*n* (%)]	
Gastroenterology	13 (32.5%)
Infectious disease	4 (10.0%)
Oncology	3 (7.5%)
Respiratory	2 (5.0%)
Neurology	2 (5.0%)
Hematology	2 (5.0%)
Geriatrics	1 (2.5%)
Urology	1 (2.5%)
Outpatient	2 (5.0%)
Health Checkup	5 (12.5%)

[a]*Information from five strains in Hong Kong and Singapore was not available.*

the biological phenotypes and genomic characteristics of Chinese ST2 strains remain unexplored thus far. The ST2 lineage might be overshadowed by increasing prevalence of ST37/RT017 worldwide and outbreaks of ST1/RT027 lineage. As the ML phylogenetic trees on ST1/RT027 and ST37/RT017, our data revealed two distinct lineages of *C. difficile* ST2 with multiple independent clonal expansions after removal of SNPs located in the recombination regions. We also compared two phylogeny trees before and after the removal of SNPs located in the recombination regions by using Gubbins. The ML phylogeny tree before removal of SNPs located in the recombination regions showed similarity with SL1, SL2a, and SL2b (data not shown). The results showed that the consistency indexes of the ML phylogenetic trees with and without removal of SNPs located in the recombination regions were 0.9796078 and 0.937613, which was analyzed by using the phangorn package in R, indicating that removal of recombinant SNPs made the remaining SNPs more consistent with the resulting phylogeny. Notably, further analysis indicated that there was a sub-lineage of SL2b divided under L2, most of which (73.5%, 25/34) originated from China, and the average toxin B concentration in SL2b was significantly lower in comparison to those in L1 and SL2a.

ST2 and ST37 are the predominant genotypes in China (Tang et al., 2016; Jin et al., 2017). The T82I gene mutation in *gyrA* found in our study has been reported in the global outbreak of RT027 and RT017 (He et al., 2013; Cairns et al., 2017). The T82I substitution in RT027 and RT017 was globally distributed in both of their two sublineages but was found *in silico* in three ST2 genomes. In addition, a substitution in the *gyrB* gene, also associated with fluoroquinolone resistance, and the substitutions in the rifampin resistance-determining region of the *rpoB* gene (He et al., 2013; Cairns et al., 2017) were not found in ST2 genomes as the RT014 lineage shown (Knight et al., 2016), indicating that ST2 genomes had their own specific molecular characteristics with low diversity of antibiotic resistance-related gene mutations.

The data showed the average toxin production in SL2b was significantly lower than those in L1 and SL2a, and then a sublineage specific non-synonymous mutation in the *tcdB* gene (Y1975D) was found to be located in the C-terminus receptor-binding domain (RBD) of *tcdB*, which is a critical region for interaction with host epithelial cell membranes (Carter et al., 2012; Orth et al., 2014). The demographic information of all 40 isolates was checked, and no epidemiological relationships were found, even in the same RT and the same location. Thus, the ST2 genomes with the mutation of Y1975D were not from the same clone. The *tcdB* gene showed high genetic diversity with classification into eight subtypes (*tcdB*1-8). The subtype *tcdB*1 has the maximum within-subtype variation and consists of three clusters (*tcdB*1a, *tcdB*1b, and *tcdB*1c) as shown in our previous study, and all ST2 strains carried *tcdB*1a (Shen et al., 2020). Most of the *C. difficile* strains from North America, East Asia, and European countries express *tcdB*1, which was closely related to human and animal diseases with more cytotoxicity than other subtypes (Shen et al., 2020). The sequences of RT027-*tcdB*-RBD are genetically divergent from other genotypes, and probably associated with rapid cell entry (Dingle et al., 2011). Although the biological function of the gene mutation (Y1975D) in *tcdB* was unknown as this mutation was not located in the 712bp amplicon previously analyzed (Dingle et al., 2011), the structure of the *tcdB* protein might be impacted, leading to the prevention of toxin B binding to cell surface receptors and decreasing *C. difficile* ST2 cytotoxicity. Thus, we speculated that it was reasonable that CDI led by *C. difficile* ST2 in SL2b had mild or moderate clinical symptoms, and the level of pathological lesion was low, especially for strains in China. However, the mutation in the *tcdB* gene-specific to SL2b from China was not found to be significantly associated with toxin B concentration, probably due to the sampling bias in this study. Thus, more studies should be performed to clarify the relationship between this SNP and low toxin B cytotoxicity and examine its contribution to the decreased ST2 strain virulence.

Clostridioides difficile sporulation capacity was associated with bacterial viability, successfully determining the transmission ability of epidemic clones (Qin et al., 2017; Zhu et al., 2018). In this study, no significant difference was found on sporulation capacity of *C. difficile* ST2 between SL2b and other lineages. A previous report provided the first epidemiological evidence on an outbreak of epidemic

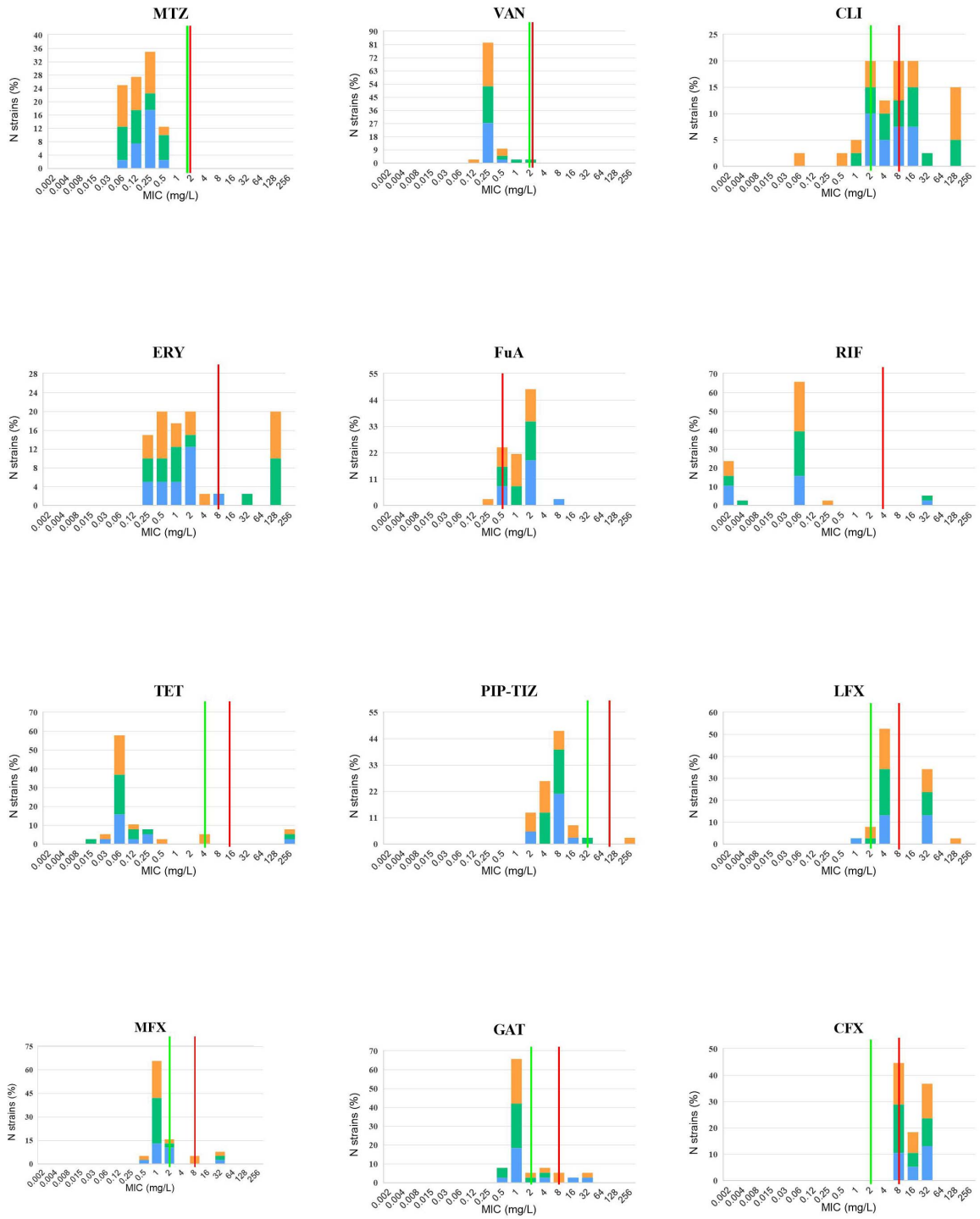


FIGURE 3 | *In vitro* 12 antibiotic susceptibility with susceptible and resistant breakpoints. MIC distributions for 12 antibiotic agents against 40 *C. difficile* isolates, orange: RT014, green: RT020, and blue: RT664, RT006, RT076, RT404, RT452, and RT587. VAN, vancomycin; MTZ, metronidazole; ERY, erythromycin; CLI, Clindamycin; RIF, rifaximin; TET, tetracycline; PIP-TAZ, piperacillin-tazobactam; LFX, levofloxacin; MFX, moxifloxacin; GAT, gatifloxacin; CFX, ciprofloxacin. Where available, established susceptible and resistant breakpoints were indicated by vertical green and red lines, respectively.

C. difficile genotype ST81 in a general teaching hospital in China (Qin et al., 2017). The data showed that ST81 had similar sporulation capacity in comparison to ST2, but significantly less toxin production than ST2 in China (Qin et al., 2017). Obviously, ST2 strains had a potentially high risk of transmission, leading to a nosocomial outbreak, and the

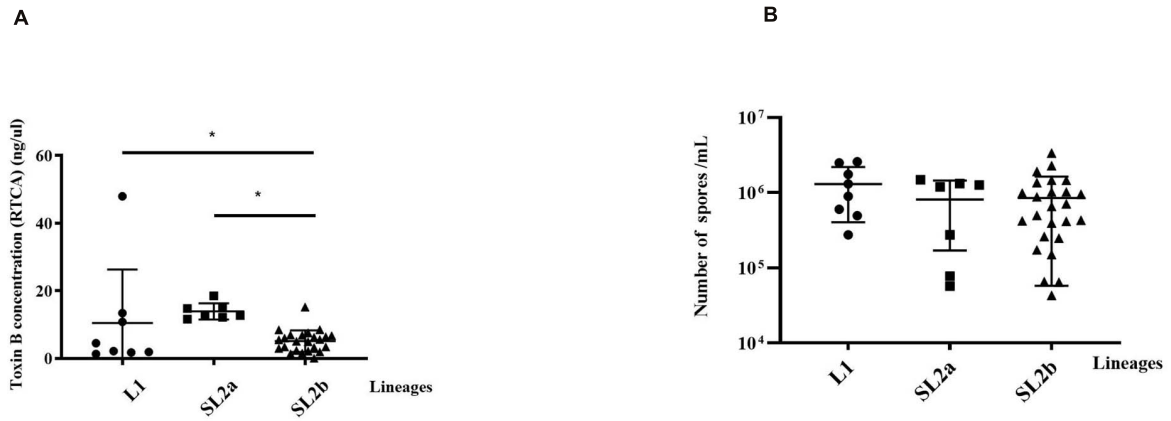


FIGURE 4 | Comparison of toxin production and Sporulation determination in the 40 *C. difficile* ST2 strains with different sub-lineages. **(A)** Toxin B concentrations by the RTCA system; **(B)** sporulation capacity by the heat-induced experiment. The toxin concentrations and sporulation capacity were shown as the mean ± standard deviation. Significant differences were marked with $^*P < 0.05$.

mild clinical symptoms were easily overlooked by clinicians. Moreover, *C. difficile* ST2 with high sporulation capacity might contribute to cross-species transmission from animal to human. Therefore, *C. difficile* ST2 should be closely monitored in future surveillance.

Antibiotic resistance elements are the key factors driving genetic diversity and epidemiological changes in *C. difficile* (Knight et al., 2015). In our study, only two antibiotic resistance loci (*cdeA* and *vanXYG*) existed in all ST2 strains and other antibiotic resistance loci were dispersedly distributed in ST genomes. Similar data were also presented in the previous study on RT014 lineage, mainly including ST2, ST13, and ST49 (Knight et al., 2016). In comparison to other STs in RT014, the ST2 genome has fewer antibiotic resistance genes and associated SNPs. The previous study also showed that only one ST2 strain simultaneously presented phenotypic resistance to tetracycline and MLS_B with positivity for *tetM* and *ermB*, and ST2, except two strains that were resistant to aminoglycoside-streptothricin, but without any resistance cassettes (Knight et al., 2016). Similarly, our data showed that only two strains carried *tetA(P)* and *tet(K)*, and two aminoglycoside-streptothricin resistance cassettes [*AAC(6′)-1e-APH(2″)-1a* and *ANT(6)-1b*] existed in each of five strains. However, *ermB* was present in the 11 strains, of which 83.3% (10/12) were from SL2b. The rates of resistance to the third-generation fluoroquinolones were significantly higher than those to the fourth-generation fluoroquinolones. According to *in silico* antibiotic resistance analysis, the T82I SNP in *gyrA* was only found in three strains, but *cdeA* existed in all ST2 strains. Therefore, we predicted that the *cdeA* gene was the main reason for third-generation fluoroquinolones resistance in *C. difficile* ST2, and but did not mediate resistance to fourth-generation fluoroquinolones. Based on the MIC data reported here, only 30.6% (11/36) of MLS_B resistant isolates contained the *ermB* or *ermC* gene, suggesting that these antibiotic resistance genes alone did not always lead to phenotypic resistance, or other alternative antibiotic resistance mechanisms might exist (Knight et al., 2019). We did not find vancomycin-resistant strains in this study despite finding the *vanXYG* gene in all 182 isolates, demonstrating that *in silico* vancomycin genotyping was a poor predictor of vancomycin phenotype in *C. difficile*.

There is still a paucity of antibiotic resistance data for *C. difficile* ST2 strains in China. In this study, we found all ST2 isolates susceptible to the first-line human CDI therapies vancomycin and metronidazole. The antibiotic resistance pattern of *C. difficile* ST2 observed in this study differed dramatically from that in our previous studies (Jin et al., 2017). The resistance rate for clindamycin (17/18, 94.4%) was markedly higher than that reported previously, but that for fusidic acid, erythromycin, rifampicin, and tetracycline were lower than in previous studies (Jin et al., 2017). The antibiotic resistance pattern in ST strains observed in this study also differed dramatically from that in a systematic review and meta-analysis in China in 2016 (Tang et al., 2016), in which data on erythromycin, clindamycin, and rifampicin showed higher resistance rates than those in ST2, but our data reported here shows high rates of resistance to fusidic acid and levofloxacin in comparison to the systematic review data (Tang et al., 2016). However, the above review data did not provide detailed information about RTs, thus more data on antibiotic resistance specific to *C. difficile* ST2 should be investigated in future studies.

Our study had some limitations as described below. Firstly, the number of isolates investigated ($n = 40$) was low relative to the phylogenetic analysis of ST2 genomes, and the number of strains from other regions or countries is limited. Thus, greater numbers of strains from various regions in China and other countries would enhance the understanding of the phylogenetic analysis of ST2 genomes. Secondly, clinical information and biological features from the other 142 strains were unavailable, including RT types and phenotypic antibiotic resistance. The antibiotic resistance phenotypic-genotypic concordance was still unclear in *C. difficile* ST2. We are going to collecting more ST2 strains from different provinces in China and other countries in order

to confirm the biological features of ST2 strains in SL2b in the near future.

CONCLUSION

Our study revealed two distinct lineages in *C. difficile* ST2 genomes with many virulence loci and few antibiotic resistance elements. SL2b was exclusively identified with a sub-lineage-dependent genome mutation (Y1975D) in *tcdB*, mainly in *C. difficile* strains from China expressing low toxin B, which might be associated with mild or moderate CDI.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. The genomic data of the 39 isolates sequenced in this study were deposited in the NCBI database under study accession number PRJNA591265. The accession number of *C. difficile* W0022a is GCF_002812625.1.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Hangzhou Medical College. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

DJ, Y-WT, and XW conceived the study, designed the experiments, and revised the manuscript. XX, QB, and XS collected the samples and performed the experiments. XX, YL, QB, QL, and MW analyzed the data. QB, SL, BZ, and GY performed the statistical analysis. XX, YL, and QB drafted the manuscript. DJ, LC, and Y-WT supervised the study. All authors edited and approved the final version of the manuscript. All corresponding authors had full access to all the data in this study and had final responsibility for the decision to submit for publication.

FUNDING

This work was supported in part by the Program for Major Science and Technology Medicine and Healthcare in Zhejiang (WKJ-ZJ-2107) and by a Key Research and Development Program of Shandong (2019JZZY011018).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.651520/full#supplementary-material

REFERENCES

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

Aubry, A., Hussack, G., Chen, W., KuoLee, R., Twine, S. M., Fulton, K. M., et al. (2012). Modulation of toxin production by the flagellar regulon in Clostridium difficile. *Infect. Immun.* 80, 3521–3532. doi: 10.1128/iai.00224-12

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

Cairns, M. D., Preston, M. D., Hall, C. L., Gerding, D. N., Hawkey, P. M., Kato, H., et al. (2017). Comparative genome analysis and global phylogeny of the toxin variant clostridium difficile PCR Ribotype 017 reveals the evolution of two independent sublineages. *J. Clin. Microbiol.* 55, 865–867. doi: 10.1128/jcm.01296-16

Carter, G. P., Rood, J. I., and Lyras, D. (2012). The role of toxin A and toxin B in the virulence of Clostridium difficile. *Trends Microbiol.* 20, 21–29. doi: 10.1016/j.tim.2011.11.003

Cheng, J. W., Xiao, M., Kudinha, T., Kong, F., Xu, Z. P., Sun, L. Y., et al. (2016). Molecular epidemiology and antimicrobial susceptibility of *Clostridium difficile* Isolates from a university teaching hospital in China. *Front. Microbiol.* 7:1621. doi: 10.3389/fmicb.2016.01621

Clinical and Laboratory Standards Institute (2011). *Methods for Antimicrobial Susceptibility Testing of Anaerobic Bacteria. Approved Standard*, Seventh Edn. Wayne, PA: Clinical and Laboratory Standards Institute.

Croucher, N. J., Page, A. J., Connor, T. R., Delaney, A. J., Keane, J. A., Bentley, S. D., et al. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* 43:e15. doi: 10.1093/nar/gku1196

Didelot, X., Eyre, D. W., Cule, M., Ip, C. L., Ansari, M. A., Griffiths, D., et al. (2012). Microevolutionary analysis of Clostridium difficile genomes to investigate transmission. *Genome Biol.* 13:R118.

Dingle, K. E., Didelot, X., Quan, T. P., Eyre, D. W., Stoesser, N., Marwick, C. A., et al. (2019). A role for tetracycline selection in recent evolution of agriculture-associated *Clostridium difficile* PCR Ribotype 078. *mBio* 10:e2790-18.

Dingle, K. E., Griffiths, D., Didelot, X., Evans, J., Vaughan, A., Kachrimanidou, M., et al. (2011). Clinical Clostridium difficile: clonality and pathogenicity locus diversity. *PLoS One* 6:e19993. doi: 10.1371/journal.pone.0019993

Eyre, D. W., and Walker, A. S. (2013). Clostridium difficile surveillance: harnessing new technologies to control transmission. *Expert Rev. Anti Infect. Ther.* 11, 1193–1205. doi: 10.1586/14787210.2013.845987

Gerding, D. N., Johnson, S., Rupnik, M., and Aktories, K. (2014). Clostridium difficile binary toxin CDT: mechanism, epidemiology, and potential clinical importance. *Gut Microb.* 5, 15–27. doi: 10.4161/gmic.26854

Griffiths, D., Fawley, W., Kachrimanidou, M., Bowden, R., Crook, D. W., Fung, R., et al. (2010). Multilocus sequence typing of Clostridium difficile. *J. Clin. Microbiol.* 48, 770–778.

He, M., Miyajima, F., Roberts, P., Ellison, L., Pickard, D. J., Martin, M. J., et al. (2013). Emergence and global spread of epidemic healthcare-associated Clostridium difficile. *Nat. Genet.* 45, 109–113.

Hennequin, C., Porcheray, F., Waligora-Dupriet, A., Collignon, A., Barc, M., Bourlioux, P., et al. (2001). GroEL (Hsp60) of Clostridium difficile is involved in cell adherence. *Microbiology (Reading)* 147, 87–96. doi: 10.1099/00221287-147-1-87

Indra, A., Huhulescu, S., Schneeweis, M., Hasenberger, P., Kernbichler, S., Fiedler, A., et al. (2008). Characterization of *Clostridium difficile* isolates using capillary gel electrophoresis-based PCR ribotyping. *J. Med. Microbiol.* 57, 1377–1382. doi: 10.1099/jmm.0.47714-0

Jin, D., Luo, Y., Huang, C., Cai, J., Ye, J., Zheng, Y., et al. (2017). Molecular epidemiology of *clostridium difficile* infection in hospitalized patients in Eastern China. *J. Clin. Microbiol.* 55, 801–810. doi: 10.1128/jcm.01898-16

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36, W5–W9.

Knight, D. R., Elliott, B., Chang, B. J., Perkins, T. T., and Riley, T. V. (2015). Diversity and evolution in the genome of *Clostridium difficile*. *Clin. Microbiol. Rev.* 28, 721–741. doi: 10.1128/cmr.00127-14

Knight, D. R., Kullin, B., Androga, G. O., Barbut, F., Eckert, C., Johnson, S., et al. (2019). Evolutionary and genomic insights into *Clostridioides difficile* Sequence Type 11: a diverse zoonotic and antimicrobial-resistant lineage of global one health importance. *mBio* 10:e00446-19.

Knight, D. R., Squire, M. M., Collins, D. A., and Riley, T. V. (2016). Genome analysis of *Clostridium difficile* PCR Ribotype 014 Lineage in Australian pigs and humans reveals a diverse genetic repertoire and signatures of long-range interspecies transmission. *Front. Microbiol.* 7:2138. doi: 10.3389/fmicb.2016.02138

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12.

Li, P., Li, X., Gu, Q., Lou, X. Y., Zhang, X. M., Song, D. F., et al. (2016). Comparative genomic analysis of Lactobacillus plantarum ZJ316 reveals its genetic adaptation and potential probiotic profiles. *J. Zhejiang Univ. Sci. B* 17, 569–579. doi: 10.1631/jzus.b1600176

Liu, B., Zheng, D., Jin, Q., Chen, L., and Yang, J. (2019). VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res.* 47, D687–D692.

Luo, Y., Cheong, E., Bian, Q., Collins, D. A., Ye, J., Shin, J. H., et al. (2019). Different molecular characteristics and antimicrobial resistance profiles of *Clostridium difficile* in the Asia-Pacific region. *Emerg. Microbes Infect.* 8, 1553–1562. doi: 10.1080/22221751.2019.1682472

Luo, Y., Zhang, W., Cheng, J. W., Xiao, M., Sun, G. R., Guo, C. J., et al. (2018). Molecular epidemiology of *Clostridium difficile* in two tertiary care hospitals in Shandong Province. China. *Infect. Drug Resist.* 11, 489–500. doi: 10.2147/idr.s152724

Martin, J. S., Monaghan, T. M., and Wilcox, M. H. (2016). *Clostridium difficile* infection: epidemiology, diagnosis and understanding transmission. *Nat. Rev. Gastroenterol. Hepatol.* 13, 206–216. doi: 10.1038/nrgastro.2016.25

Orth, P., Xiao, L., Hernandez, L. D., Reichert, P., Sheth, P. R., Beaumont, M., et al. (2014). Mechanism of action and epitopes of *Clostridium difficile* toxin B-neutralizing antibody bezlotoxumab revealed by X-ray crystallography. *J. Biol. Chem.* 289, 18008–18021. doi: 10.1074/jbc.m114.560748

Péchiné, S., Hennequin, C., Boursier, C., Hoys, S., and Collignon, A. (2013). Immunization using GroEL decreases *Clostridium difficile* intestinal colonization. *PLoS One* 8:e81112. doi: 10.1371/journal.pone.0081112

Preston, M. D., Assefa, S. A., Ocholla, H., Sutherland, C. J., Borrmann, S., Nzila, A., et al. (2014). *PlasmoView*: a web-based resource to visualise global *Plasmodium falciparum* genomic variation. *J. Infect. Dis.* 209, 1808–1815. doi: 10.1093/infdis/jit812

Qin, J., Dai, Y., Ma, X., Wang, Y., Gao, Q., Lu, H., et al. (2017). Nosocomial transmission of *Clostridium difficile* Genotype ST81 in a General Teaching Hospital in China traced by whole genome sequencing. *Sci. Rep.* 7:9627.

Ryder, A. B., Huang, Y., Li, H., Zheng, M., Wang, X., Stratton, C. W., et al. (2010). Assessment of *Clostridium difficile* Infections by quantitative detection of tcdB Toxin by use of a real-time cell analysis system. *J. Clin. Microbiol.* 48, 4129–4134. doi: 10.1128/jcm.01104-10

Schliep, K. P. (2011). phangorn: phylogenetic analysis in R. *Bioinformatics* 27, 592–593. doi: 10.1093/bioinformatics/btq706

Shen, E., Zhu, K., Li, D., Pan, Z., Luo, Y., Bian, Q., et al. (2020). Subtyping analysis reveals new variants and accelerated evolution of *Clostridioides difficile* toxin B. *Commun. Biol.* 3:347.

Stabler, R. A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., et al. (2009). Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biol.* 10:R102.

Tang, C., Cui, L., Xu, Y., Xie, L., Sun, P., Liu, C., et al. (2016). The incidence and drug resistance of *Clostridium difficile* infection in Mainland China: a systematic review and meta-analysis. *Sci. Rep.* 6:37865.

Tian, T. T., Zhao, J. H., Yang, J., Qiang, C. X., Li, Z. R., Chen, J., et al. (2016). Molecular characterization of *Clostridium difficile* isolates from human subjects and the environment. *PLoS One* 11:e0151964. doi: 10.1371/journal.pone.0151964

Ye, J., McGinnis, S., and Madden, T. L. (2006). BLAST: improvements for better sequence analysis. *Nucleic Acids Res.* 34, W6–W9.

Zhang, X., Wang, X., Yang, J., Liu, X., Cai, L., and Zong, Z. (2016). Colonization of toxigenic *Clostridium difficile* among ICU patients: a prospective study. *BMC Infect. Dis.* 16:397. doi: 10.1186/s12879-016-1729-2

Zhu, D., Sorg, J. A., and Sun, X. (2018). *Clostridioides difficile* biology: sporulation, germination, and corresponding therapies for *C. difficile* infection. *Front. Cell Infect. Microbiol.* 8:29. doi: 10.3389/fcimb.2018.00029

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

frontiers in Microbiology

ORIGINAL RESEARCH
published: 27 July 2021
doi: 10.3389/fmicb.2021.647565

Antimicrobial Resistance Profiling and Phylogenetic Analysis of *Neisseria gonorrhoeae* Clinical Isolates From Kenya in a Resource-Limited Setting

Meshack Juma[1†], Arun Sankaradoss[2†], Redcliff Ndombi[1], Patrick Mwaura[1], Tina Damodar[2], Junaid Nazir[2], Awadhesh Pandit[2], Rupsy Khurana[2], Moses Masika[1], Ruth Chirchir[1], John Gachie[1], Sudhir Krishna[2,3], Ramanathan Sowdhamini[2], Omu Anzala[1] and Iyer S. Meenakshi[2]**

[1] KAVI Institute of Clinical Research, University of Nairobi, Nairobi, Kenya, [2] National Centre for Biological Sciences, Tata Institute of Fundamental Research (TIFR), Bengaluru, India, [3] School of Interdisciplinary Life Sciences, Indian Institute of Technology Goa, Ponda, India

Background: Africa has one of the highest incidences of gonorrhea. *Neisseria gonorrhoeae* is gaining resistance to most of the available antibiotics, compromising treatment across the world. Whole-genome sequencing (WGS) is an efficient way of predicting AMR determinants and their spread in the population. Recent advances in next-generation sequencing technologies like Oxford Nanopore Technology (ONT) have helped in the generation of longer reads of DNA in a shorter duration with lower cost. Increasing accuracy of base-calling algorithms, high throughput, error-correction strategies, and ease of using the mobile sequencer MinION in remote areas lead to its adoption for routine microbial genome sequencing. To investigate whether MinION-only sequencing is sufficient for WGS and downstream analysis in resource-limited settings, we sequenced the genomes of 14 suspected *N. gonorrhoeae* isolates from Nairobi, Kenya.

Methods: Using WGS, the isolates were confirmed to be cases of *N. gonorrhoeae* (n = 9), and there were three co-occurrences of *N. gonorrhoeae* with *Moraxella osloensis* and *N. meningitidis* (n = 2). *N. meningitidis* has been implicated in sexually transmitted infections in recent years. The near-complete *N. gonorrhoeae* genomes (n = 10) were analyzed further for mutations/factors causing AMR using an in-house database of mutations curated from the literature.

Results: We observe that ciprofloxacin resistance is associated with multiple mutations in both gyrA and parC. Mutations conferring tetracycline (*rpsJ*) and sulfonamide (*folP*) resistance and plasmids encoding beta-lactamase were seen in all the strains, and *tet(M)*-containing plasmids were identified in nine strains. Phylogenetic analysis clustered the 10 isolates into clades containing previously sequenced genomes from Kenya and countries across the world. Based on homology modeling of AMR targets, we see that the mutations in GyrA and ParC disrupt the hydrogen bonding with quinolone drugs and mutations in FolP may affect interaction with the antibiotic.

OPEN ACCESS

Edited by:
Guido Werner, Robert Koch Institute (RKI), Germany

Reviewed by:
Leonor Sánchez-Busó, Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana (FISABIO), Spain
Sebastian Banhart, Robert Koch Institute (RKI), Germany

***Correspondence:**
Iyer S. Meenakshi
meenakshis@ncbs.res.in
Omu Anzala
oanzala@kaviuon.org

†These authors share first authorship

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

Received: *12 January 2021*
Accepted: *31 May 2021*
Published: *27 July 2021*

Citation:
Juma M, Sankaradoss A, Ndombi R, Mwaura P, Damodar T, Nazir J, Pandit A, Khurana R, Masika M, Chirchir R, Gachie J, Krishna S, Sowdhamini R, Anzala O and Meenakshi IS (2021) Antimicrobial Resistance Profiling and Phylogenetic Analysis of Neisseria gonorrhoeae Clinical Isolates From Kenya in a Resource-Limited Setting. Front. Microbiol. 12:647565. doi: 10.3389/fmicb.2021.647565

Conclusion: Here, we demonstrate the utility of mobile DNA sequencing technology in producing a consensus genome for sequence typing and detection of genetic determinants of AMR. The workflow followed in the study, including AMR mutation dataset creation and the genome identification, assembly, and analysis, can be used for any clinical isolate. Further studies are required to determine the utility of real-time sequencing in outbreak investigations, diagnosis, and management of infections, especially in resource-limited settings.

Keywords: nanopore sequencing, whole-genome sequencing, antimicrobial resistance, AMR prediction, sequence-typing

INTRODUCTION

Gonorrhea is one of the most common sexually transmitted infections (STIs), which is caused by *Neisseria gonorrhoeae* (gonococci). It is estimated to have affected 87 million people (between the ages 15 and 49) worldwide, 11.4 million in the African region, in 2016 [World Health Organisation (WHO), 2018]. Gonorrhea can cause epididymitis in men and pelvic inflammatory disease in women, which can result in infertility and ectopic pregnancies. It is also implicated in neonatal ophthalmia, which can result in blindness, and increased acquisition and transmission of other STIs (Unemo and Dillon, 2011). WHO included *N. gonorrhoeae* in its list of AMR "priority pathogens" in 2017 [World Health Organisation (WHO), 2017] and has projected a 90% reduction in global gonorrhea incidence by 2030 [World Health Organisation (WHO), 2018]. Although *N. gonorrhoeae* infections are treated with antibiotics, gonococci have acquired resistance to most of the antibiotics used to treat gonorrhea, and despite two clinical trials, no vaccines are available (Wi et al., 2017). Increasing antimicrobial resistance threatens effective treatment and control.

Many strains are developing resistance to ceftriaxone and azithromycin, which are used as the last options of the first-line therapies of gonorrhea (Unemo and Nicholas, 2012). However, with isolates showing increased MICs to both these antibiotics, this is not viable as a long-term treatment (Papp et al., 2017). Many countries in the African region have reported decreased susceptibility to drugs like ceftriaxone, azithromycin, or ciprofloxacin (Wi et al., 2017). Gentamycin has been used as an alternative therapy in some places in Africa, although a few cases of resistance have been reported in other parts of the world (Chisholm et al., 2011). Whole-genome sequencing (WGS) allows for timely detection and elucidation of AMR determinants in bacteria (Collineau et al., 2019). High-throughput sequencing has many advantages, like long read lengths, short time, and reduction of bias introduced through amplification by PCR steps (Goldstein et al., 2019). WGS has been used to investigate quinolone-resistant gonorrhea outbreaks (Duncan et al., 2011; Kivata et al., 2019) in Kenya. It has also been used to analyze isolates resistant to ciprofloxacin, azithromycin, and ESCs in other countries (Chaudhry et al., 2002; Eyre et al., 2018; Golparian et al., 2018).

MinION, a mobile sequencing device from ONT (Oxford, United Kingdom), sequences DNA by monitoring the transfer of individual DNA molecules through various types of pores, resulting in very long and unbiased sequence reads, as there is no amplification or chemical reactions used during sequencing (Golparian et al., 2018). The mobile nature of the MinION offers many advantages and the sequencer has been evaluated for use in resource-limited settings and on the International Space Station (ISS) (Castro-Wallace et al., 2017; Elliott et al., 2020). MinION has also been shown to be effective for diagnostics, high-quality assemblies, and AMR surveillance in bacteria. However, the high error rate for the sequencer limits its use. Golparian and co-workers sequenced 14 clinical isolates of *N. gonorrhoeae* and evaluated multiple methods for genome assemblies using MinION with Illumina reads (Golparian et al., 2018). Street and co-workers used MinION to obtain a *de novo* assembly of *N. gonorrhoeae* isolated from patient urine samples (Street et al., 2020). Zhang and co-workers showed that AMR-profiling results for *N. gonorrhoeae* were comparable when using only MinION-based assemblies or MinION-Illumina hybrid assemblies (Zhang et al., 2020). Naidenov and co-workers used Nanopore-only sequencing to obtain complete genomes and carried out AMR profiling of two new strains from the *Elizabethkingia* genus (Naidenov et al., 2019). Sanderson and co-workers used nanopore sequencing data and showed that it is possible to generate accurate consensus bacterial genomes from metagenomic sequencing data without using Illumina short reads (Sanderson et al., 2020).

In a previous study from Kenya, 22 multi-drug-resistant gonococcal isolates from heterosexual patients (2 females and 20 males) were sequenced using Illumina MiSeq for understanding *gyrA* and *parC* mutations in ciprofloxacin resistance (Kivata et al., 2019). Cehovin et al. sequenced 103 genomes from 73 patients from coastal Kenya, mainly men who had sex with men, using Illumina HiSeq and analyzed the plasmids conferring antibiotic resistance (Cehovin et al., 2018). These studies in Kenya have used the Illumina platform and mainly focused on isolates from male patients. We sequenced 10 isolates derived from women who visited STI clinics in Nairobi between 2012 and 2017, using only MinION sequencing in Kenya and evaluated methods for assembly and AMR profiling of genomes. We have also carried out phylogenies with previously deposited sequences in PubMLST, from Kenya, and other countries. Here, we assess the possibility of using MinION data alone for WGS, assembly, and comparative analysis. This study demonstrates the utility of the

portable MinION sequencer for genome-based analysis in regular clinical setups.

MATERIALS AND METHODS

Bacterial Isolates and Culture Conditions

Neisseria gonorrhoeae isolates were derived from high vaginal swabs collected between January 2012 and December 2017. The stocked organisms were stored at −70°C and retrieved and cultured on Modified Thayer–Martin agar. The plates were incubated at 37°C and 5% carbon dioxide (under raised CO_2) as per the standard protocol.

Isolates were identified by Gram stain, Catalase test, Oxidase test, and carbohydrate-utilization studies of glucose, maltose, fructose, and sucrose, and confirmed by analytical profile index testing kit (API NH, bioMérieux). All gonococcus isolates were re-stocked in 20% glycerol-tryptic soy broth for long-term storage at −70°C for external quality control, sequencing, and molecular characterization. Standard *N. gonorrhoeae* isolates, i.e., WHO K, WHO P, WHO O, WHO R, and WHO M (Unemo et al., 2016), were used to test the media for viability and colonial characteristics. Phenotypic characterization was performed by characterizing lactamase production determined using nitrocefin solution. In the present study, clinical isolates from 10 patients were sequenced (**Supplementary Table 1**).

Antibiotic Susceptibility Testing

Ciprofloxacin, spectinomycin, cefixime, ceftriaxone, gentamycin, and cefoxitin strips were used to set *E*-test according to the Clinical and Laboratory Standards Institute (CLSI). An antimicrobial gradient diffusion (*E*-test) method was used to determine the minimum inhibitory concentration (MIC) of *N. gonorrhoeae*, and interpretation of susceptibilities for selected antimicrobial agents was carried out as recommended by CLSI (**Supplementary Table 2**).

Isolation of Genomic DNA

All the colonies from Thayer–Martin agar (37°C in a humidified 5% CO_2 environment for 36 h) with the phenotypic characteristics of *N. gonorrhoeae* were picked up to create a culture suspension in 1X PBS. The genomic DNA was isolated using the Qiagen DNA isolation kit using the manufacturer's specifications. Extracted DNA was purified using AMPure XP beads (Beckman Coulter, United Kingdom), eluted in 50 μl of nuclease-free water, and quantified using a QiaXpert (Qiagen). Pipetting was minimized to reduce shearing of the DNA prior to sequencing.

ONT Library Preparation and MinION Sequencing

DNA library preparation for Nanopore sequencing was carried out using Ligation Kit SQK-LSK109 (ONT). Fragmented DNA was repaired and dA-tailed using the NEBNext FFPE DNA Repair Mix and NEBNext Ultra II End Repair/dA-Tailing Module (New England BioLabs). An individual barcode was added to dA-tailed DNA by using the barcoding extension kit EXP-NPB104 in accordance with the ONT protocols with NEB Blunt/TA Ligase Master Mix (New England BioLabs). Each barcoded DNA was pooled in equimolar amounts, and an adapter was attached using the NEBNext Quick Ligation Module (New England BioLabs). The MinION flow cell and reagents were shipped from Bangalore, India, to Nairobi at 4°C. The number of active pores was checked before loading. The samples were pooled using equimolar pooling, the library was loaded into the SpotON flow cell R9.5 (FLO-MIN106), and sequencing was carried out on MinKNOW using the 48-h script.

Base-Calling, Read Trimming, and Processing

Guppy (v 3.2.2) (ONT) was used for base-calling of fast5 files. Reads with a mean q score (quality) greater than 7 and a read length greater than 500 bp were used and trimmed for adaptor sequences and barcodes using qcat (v1.1.0)[1] from ONT. The files corresponding to each barcode (sample) were analyzed as follows.

Assembly and Assessment

The selected raw reads were corrected, trimmed, and assembled using Canu (v1.8)[2] using the parameters genome Size = 2.1 m and minReadLength = 500 (Koren et al., 2017). Minimap (v 2.17)[3] and Miniasm (v0.3)[4] were used for the self-mapping of the raw reads and the concatenation of the alignments to get the *de novo* assembly, respectively (Li, 2016). Two rounds of error correction of this draft assembly were carried out with Racon (v1.4.3)[5] using raw reads (Vaser et al., 2017). The above assemblies were polished using the raw nanopore reads with Nanopolish (v0.11.1)[6] (Loman et al., 2015).

Species identification was carried out using the Ribosomal Multilocus Sequence Typing rMLST (Jolley et al., 2012) approach from BIGSdb and PubMLST to identify the causative organism and co-infections (Jolley et al., 2018). We used the *de novo* assemblies for the above step. The database uses 53 genes encoding the bacterial ribosome protein subunits (*rps* genes) for species assignment. We also used blastn with 16S rRNA and 23S rRNA as the queries for species confirmation. The plasmid sequences were identified from the *de novo* assemblies. The plasmid sequences were used for detecting the AMR determinants.

We also carried out a guided genome mapping using reference genome *N. gonorrhoeae* FA1090 (NCBI:txid485). The trimmed raw reads were aligned to the reference genome using the bwa-mem option from bwa (v0.7.12)[7] (Li and Durbin, 2010). Samtools (v1.9)[8] was used for deriving the bam alignment file and the consensus genome (bcftools), after normalizing the indels and variant calling (Li et al., 2009). The commands samtools mpileup

[1] https://github.com/nanoporetech/qcat
[2] https://github.com/marbl/canu
[3] https://github.com/lh3/minimap2
[4] https://github.com/lh3/miniasm
[5] https://github.com/isovic/racon
[6] https://github.com/jts/nanopolish
[7] http://bio-bwa.sourceforge.net/
[8] http://www.htslib.org/

with the options *-Ou -f* and bcftools calls with the options *-mv -Oz –ploidy 1* were used for variant calling. The options bcftools *filter –IndelGap 5* and bcftools *consensus -H "A"* were used for deriving the consensus genome. Mummer (v3.9.4)[9] (Kurtz et al., 2004) and bedtools (v2.25.0)[10] (Quinlan and Hall, 2010) were used for calculating the correspondence of the assemblies with the reference genome, genome coverage, and sequencing depth. The base positions with zero coverage were extracted using awk commands, and bedtools maskfasta option was used to derive the draft genomes; missing nucleotides were marked in the genome as "N."

Annotation

The gene prediction and annotation were carried out using the RAST server[11] using the reference genome FA1090 for guiding the gene prediction (Overbeek et al., 2013).

Genome-Based MLST Analysis

Gene profiles for sequence typing of *N. gonorrhoeae* were downloaded from different databases like NG-STAR (*N. gonorrhoeae* Sequence Typing for Antimicrobial Resistance)[12] (Demczuk et al., 2017), PubMLST (Public databases for multi-locus sequence-typing)[13] (Bennett et al., 2007), and NG-MAST (*N. gonorrhoeae* multi-antigen sequence typing)[14] (Martin et al., 2004). NG-MAST uses two highly polymorphic loci, the outer-membrane porin (*por*) and transferrin binding protein β-subunit *tbpB*, for sequence typing. NG-STAR uses variants in seven antimicrobial resistance determinants (*penA*, *mtrR*, *porB*, *ponA*, *gyrA*, *parC*, and 23S rRNA) to three classes of antibiotics (cephalosporins, macrolides, and fluoroquinolones), and MLST (database PubMLST) uses seven housekeeping genes (*abcZ*, *adk*, *aroE*, *fumC*, *gdh*, *pdhC*, and *pgm*) for sequence typing. Profiles were created and blastn (*E*-value 10^{-7}) from standalone BLAST+ [15](Altschul, 2005) was used to assign the genes to the different alleles for MLST genes.

Identification of Mutations Causing AMR

Around 85 mutations in 14 genes reported to be involved in AMR were checked for in the sequenced strains, including promoter mutations resulting in overexpression of efflux proteins (Unemo and Shafer, 2011). Additionally, we checked for the presence of four drug efflux pumps, two genes from conjugative plasmids, and plasmid-mediated AMR determinants (**Supplementary File**). Mutations were manually screened for, in the protein sequences identified using the RAST server in all the genomes. Sequencing depth at the corresponding gene loci was calculated from the results of the samtools depth command. We compared our results with that of three publicly available databases for AMR, CARD (Comprehensive Antibiotic Resistance Database), ARIBA (Antimicrobial Resistance Identification By Assembly) (Hunt et al., 2017), and Pathogenwatch (Balloux et al., 2018; Alcock et al., 2019). Resistance genes from plasmids were identified using blastn (*E*-value 10^{-7}) against NCBI Bacterial Antimicrobial Resistance Reference Gene Database[16] (NCBI Resource Coordinators, 2017).

Homology Modeling

To understand the basis of antibiotic resistance, wild-type and mutant proteins implicated in AMR were modeled using templates from other bacteria. Coordinates of heteroatoms like antibiotics and DNA are present in the templates used. Hence, the heteroatom modeling module from Modeler (v9.23) (Sánchez and Sali, 2000) was used to build a multi-chain model with symmetry restraints. Clustal Omega[17] was used for the alignment of query protein sequences with the template protein sequence, with manual correction of the alignment (Sievers and Higgins, 2014). The residues in different chains were separated using "/" and the symbol "." was used to indicate the ligand molecules in the alignment.

Multiple-Sequence Alignment and Phylogenetics

Phylogenetic tree analysis was carried out using 123 previously sequenced *N. gonorrhoeae* strains from Kenya and an additional 130 strains from different countries across the world, standard WHO strains, and the reference genome. The Genome Comparator module from Bacterial Isolate Genome Sequence Database (BIGSdb)[18] was used to obtain the sequences and generate core genome alignments (cgMLST1649-) using the MAFFT module within BIGSdb (Jolley et al., 2018). We used a dataset of strains from different WHO geographical regions including strains from Kenya and WHO reference strains. The parameters used were minimum percentage identity 70%, minimum percentage alignment 50%, and blastn search to annotate loci in genomes with ≥50% loci (word size of 20). Incomplete loci were not used for pairwise comparison and pairwise missing loci and paralogous loci were excluded from the analysis. The *N. meningitidis* strain (PubMLST ID: 12672|053442, cc4821) was included as an outgroup. The alignments were checked for redundancy in sequences of core genes; the sequences with 100% redundancy and the sequences missing all the gene loci analyzed were removed. Phylogenetic trees were derived based on Maximum likelihood approach using RAxML software (v8.2.12) using GTR + GAMMA substitution model with 1000 bootstraps (Kozlov et al., 2019). The best ML search trees were identified for both the datasets (only Kenyan strains and combined dataset of all 255 strains) annotated with details like WHO geographical region and antibiotic sensitivity from BIGSdb and Cehovin et al. (2018) using iTOL (Letunic and Bork, 2016).

The workflow for consensus genome used in the study has been illustrated in **Supplementary Figure 1A**, and the

[9]http://mummer.sourceforge.net/

[10]https://bedtools.readthedocs.io/en/latest/

[11]https://rast.nmpdr.org/

[12]https://ngstar.canada.ca/

[13]https://pubmlst.org/

[14]http://www.ng-mast.net/

[15]https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/

[16]https://www.ncbi.nlm.nih.gov/bioproject/PRJNA313047/

[17]https://www.ebi.ac.uk/Tools/msa/clustalo/

[18]https://pubmlst.org/software/database/bigsdb/

subsequent analysis workflow has been shown in **Supplementary Figure 1B**. The workflow for the *de novo* assemblies has been shown in **Supplementary Figure 1C**.

RESULTS

Overview of Sequenced Data

We obtained between 534 and 2,387 Mb of reads from the MinION sequencing runs for each sample. The longest reads ranged from 156 to 406 kb with an average length of 1.5–2.7 kb (**Table 1**). We sequenced clinical isolates from 14 patients and obtained 12 *N. gonorrhoeae* genome sequences. Out of these, 10 genomes were near-complete with a good sequencing depth and were used for subsequent analysis.

Genome Assembly Statistics and Species Identification

Overview of the mappability of the reads from different strains to the reference genome has been depicted in **Supplementary Figure 2**. Assemblies were compared and selected based on the following criteria: the lowest number of mismatches, misassemblies, contigs, and the highest fraction of genome coverage using QUAST (Gurevich et al., 2013) (**Supplementary Tables 3A–C**). The length of the assemblies and the number of CDS identified were fairly the same across different methods. No plasmid or transposon gene insertions were observed in the genomes assembled *de novo* and the consensus genome. Similar to previous reports, we observed frameshifts with the *de novo* assemblies (Golparian et al., 2018). The consensus genome (FA1090 was used for read mapping) was used for further analysis as the genome coverage, number of complete genomic features, number of indels, and number of mismatches were better.

TABLE 1 | Read and assembly statistics.

Strain IDs	Genome coverage (%)	Sequencing depth	# Mapped reads
3	99.03	523.604	308,387
12	99	377.339	216,691
18	98.78	272.294	175,050
57	98.81	1255.388	2,272,143
61	98.97	63.726	43,553
81	99.59	86.3	182,347
100A	98.68	848.073	1,460,162
196	98.2	22.6	143,725
240	99.59	86.3	70,185
274	98.88	240.667	180,375
285	98.31	129.085	116,342
298	98.67	219.616	212,198

*The number of mapped reads for each genome was obtained using samtools flagstat and the genome assembly statistics were obtained using QUAST. The genome coverage was not affected by the number of reads or the sequencing depth. The genomes assembled from strains 81 and 196 had low sequencing depth and lower coverage in the genes surveyed in the study and were not used for further analysis. The strains 4 and 11 were identified to be N. meningitidis genomes from the ribosomal MLST analysis and were not analyzed further in the study (**Supplementary Table 1**).*

Among the 14 samples, 2 samples showed the presence of *N. meningitidis*, 9 samples showed the presence of *N. gonorrhoeae*, and 3 samples showed the presence of both *N. gonorrhoeae* and *M. osloensis* (**Supplementary Table 1**). *N. meningitidis* has been implicated in STI in recent years, while *M. osloensis* is also a commensal of the urogenital tract (Slack, 2010; Johnson et al., 2016).

Although high error rates have been reported with the use of the MinION sequencer, we have used long sequencing time (48 h) and obtained high depths of sequencing (63-1255X; **Table 1**). Previous studies have shown that sequencing using MiNION for longer time duration increases the sequencing output and the error rate can be minimized by using depth and consensus (Tyler et al., 2018; Plesivkova et al., 2019).

Sequence Typing of the Strains

The alleles used in sequence typing (MLST, NG-MAST, and NG-STAR) were assigned to the genes from different strains, and we identified few novel alleles. The novel allele sequences have been deposited in the PubMLST database and the assigned allele numbers have been mentioned (**Supplementary Tables 4A–C**). The sequencing depth at each of these loci was found to be very high (>50–600) (**Supplementary Table 4D**). This is higher than the recommended depth for SNP detection using Nanopore sequencing by Sanderson and co-workers (20X) (Sanderson et al., 2020).

Antibiotic Resistance and Mutations in AMR Genes

The MIC results with six different antibiotics have been provided in **Table 2**.

A dataset of AMR targets was identified using literature (**Supplementary Table 5A**). These included mutations in the promoter region of repressors resulting in overexpression of efflux proteins, mutations in drug targets that result in higher MICs for the drugs either individually or in combination with other mutations, plasmids for drug resistance, and mosaic alleles in genes and promoter regions reported from drug-resistant strains of *N. gonorrhoeae* from across the world. Out of these, 28 mutations (in 11 genes) were observed in the strains (**Table 3**). The sequencing depth of *gyrA*, *parC*, *mtrR*, *ponA1*, *penA*, *porB*, *folP*, *rplD, pilQ, rpoB*, and *rpsJ* genes was high (**Supplementary Tables 4D, 5B**) and there were no missing bases in these genes. Other databases used for AMR profiling (Pathogenwatch, ARIBA, and CARD) missed a few of the mutations we identified (**Supplementary Table 5C**). The ARIBA database has not been updated; hence, it may be missing some recently identified AMR-associated mutations (Hunt et al., 2017). CARD collects AMR determinants from several bacteria and is not specific to *N. gonorrhoeae*; it may include mutations causing AMR in other bacteria but not gonococci (Alcock et al., 2019), whereas compared to these databases, Pathogenwatch is frequently updated; we identified the mutation *rplD* G70D (implicated in macrolide resistance) in three strains, but this mutation has been missed by Pathogenwatch (Ma et al., 2020b).

TABLE 2 | MIC values for different antibiotics.

Strain ID	Ceftriaxone (μg/ml)	Ciprofloxacin (μg/ml)	Spectinomycin (μg/ml)	Cefixime (μg/ml)	Gentamycin (μg/ml)	Cefoxitin (μg/ml)
3	0.016 (S)	0.032 (S)	0.75 (S)	0.016 (S)	0.064 (S)	0.016 (S)
12	0.016 (S)	3 (R)	8 (S)	<0.016 (S)	8 (I)	0.016 (S)
18	<0.016 (S)	0.032 (S)	0.75 (S)	0.016 (S)	0.064 (S)	0.016 (S)
57	<0.032 (S)	3 (R)	4 (S)	0.016 (S)	4 (S)	0.016 (S)
61	0.016 (S)	3 (R)	8 (S)	<0.016 (S)	3 (S)	0.016 (S)
100A	<0.016 (S)	1 (I)	8 (S)	<0.016 (S)	4 (S)	<0.016 (S)
240	<0.016 (S)	0.75 (I)	2 (S)	<0.016 (S)	1 (S)	0.016 (S)
274	<0.016 (S)	<0.016 (S)	4 (S)	<0.016 (S)	3 (S)	0.016 (S)
285	0.047 (S)	1 (I)	4 (S)	<0.016 (S)	1.5 (S)	0.125 (S)
298	0.016 (S)	0.75 (I)	4 (S)	<0.016 (S)	3 (S)	0.016 (S)

*The MIC values (μg/ml) and the interpretations for different antibiotics for the 10 strains. S – Susceptible, R – Resistant (highlighted in red), I – Intermediate (highlighted in yellow). The interpretation of MIC was inferred from CSLI breakpoints (**Supplementary Table 2**).*

We found six isolates carrying AMR mutations in penA C-terminal region (penicillin and cephalosporin resistance determinant) and four isolates with mutations in *ponA1* (penicillin and cephalosporin resistance determinant), respectively. We detected the D345 insertion and F504L mutation in *penA* (Zapun et al., 2016), which could be involved in penicillin and cephalosporin resistance (**Table 3**). Five isolates harbored *gyrA* resistance mutations, two of which also had a *parC* mutation. One strain had a mutation in only the *parC* gene; however, this is a silent mutation that was among the set of mutations observed in a cephalosporin-resistant isolate (Chaudhry et al., 2002). One of the ciprofloxacin-resistant strain contained a *gyrA* mutation, consistent with previous observations of these mutations in susceptible (Tanaka et al., 2000; Chaudhry et al., 2002) isolates. These mutations, in the Quinolone Resistance Determining Region (QRDR) of both the proteins, have been shown to contribute to the high MICs for ciprofloxacin (Belland et al., 1994). Interestingly, one of the *gyrA* mutations we observed, D95A/G, and the mutation E91G in ParC have been reported to be specific to Kenyan isolates (Kivata et al., 2019). Specific mutations in *porB1b* (penB AMR determinant), implicated in decreased influx of antimicrobials through the porin PorB (Lee et al., 2010; Chen and Seifert, 2013), were identified in 6 of the 10 isolates. All 10 strains harbored the V57M mutation in RpsJ, which has been shown to contribute to tetracycline resistance and R228S mutation in FolP, which results in sulfonamide resistance (Yun et al., 2012). Three strains harbored mutations in *rplD* gene, which is implicated in azithromycin resistance (Ma et al., 2020b). Eight strains had mutations in *pilQ* gene (PenC), implicated in penicillin, ESC, and tetracycline resistance, while four strains had the H522N mutation in *rpoB*, which is implicated in rifampicin resistance. Plasmid-mediated AMR determinants [high-level resistance to benzylpenicillin and tetracycline -p*bla*TEM and p*tet(M)*] were detected in all strains except one where we detected only the *tet(M)*-containing plasmid. We did not identify *mef* or *erm* containing conjugative plasmids/transposons, which are reported to cause resistance to macrolides (Roberts et al., 1999) or resistance-conferring mutations for spectinomycin, in either the *de novo* assemblies or consensus genome.

Understanding the Basis of Antibiotic Resistance

We investigated the basis of drug resistance for the mutations in the proteins GyrA, ParC, FolP, porB1b, PonA1, and RpoB (**Supplementary Figures 3–8**, template details in **Supplementary Table 6**) through homology modeling using templates with co-crystallized antibiotic structures. The results have been described in **Supplementary File** (Section 8).

Phylogenetic Relationship With Other *N. gonorrhoeae* Genomes From Kenya, Other WHO Geographical Regions, and WHO Strains

To understand the genetic relatedness of isolates from Kenya with isolates from other geographical regions and WHO reference isolates, we constructed a phylogenetic tree using the alignments derived from the core genome MLST (cgMLST). After filtering for sequence redundancy and loci completeness, we retained 238 sequences from different geographical regions across the world (Americas, Europe, Western Pacific, Africa and WHO reference strains) for the phylogenetic analysis. The metadata for the strains like geographical region and antibiotic resistance details have been indicated in the phylogenetic trees (details in **Supplementary Table 7**).

We observed eight clusters of sequences from across the world (clusters I–VIII). Clades I and II consisted of only Kenyan sequences, while sequences from different geographical regions and Kenya co-clustered in the rest of the clades (**Figure 1**). Cehovin et al. had observed in an earlier study that strains from Kenya were distinct from strains sequenced from USA and the UK (Cehovin et al., 2018). Consistent with this observation, we see Kenyan sequence-specific clades, but we also observe clades where sequences from Kenya are mixed from sequences from other geographical regions. The clades were more or less consistent in the phylogenetic tree constructed using sequences from all over the world and using sequences from only Kenya (**Figure 2**). Out of the strains sequenced in the study, strain 3 clustered in clade I; strain 298 in clade V; clade 274 in clade VI; strains 12, 18, 61, 57, and 100A in clade VII; and strains 240 and

285 in clade VIII. Each cluster consisted of strains belonging to different multi-locus sequence types (MLSTs). Strains belonging to certain STs were specific to a cluster (ST1588 – cluster VII, ST1893 – cluster II) while strains from some other STs were found in multiple clusters (ST1599 in cluster III, IV, V, and VIII and ST1902 in clusters III and VIII) (metadata in **Supplementary Table 7**). Broader sampling and analysis of clinical isolates from Kenya can help us determine the extent of genetic diversity within the *N. gonorrhoeae* strains.

DISCUSSION

In this study, we assessed the feasibility of nanopore sequencing and AMR gene annotation in a clinical microbiology laboratory setting. We found that DNA extraction followed by MinION sequencing facilitates efficient detection of chromosomal and plasmid-mediated AMR genes in clinical isolates. From the 14 isolates sequenced in this work, we obtained 12 *N. gonorrhoeae* genomes, out of which 10 had more than 98% coverage with high depth in the AMR and MLST gene regions analyzed in the study. We were able to detect mutations in multiple genes involved in AMR and we also focused on investigating the structure–function aspects of drug resistance mutations in some proteins.

The consensus genome mapping obtained using bwa-mem was chosen for analysis, due to higher genome coverage and lower number of indels and misassemblies compared to the reference genome. Although earlier studies have shown that bacterial MinION sequences show high mapping rates to the reference genome, with fewer indel rates, bwa-mem has found limited application with long error-prone reads (Finn et al., 2014; Laver et al., 2015; Gerstmans et al., 2020; Nguyen et al., 2020). We show that it is possible to obtain assemblies with good genome coverage for AMR determinant detection by using nanopore-only reads with a consensus genome mapping strategy. Previous studies have obtained hybrid genome assembly for gonococcal isolates by combining short-read (Illumina) and long-read (ONT) sequence data (Eyre et al., 2018; Golparian et al., 2018). This is another approach for deriving a consensus genome using MinION sequencing and, accordingly, an alternative to *de novo* assembly.

Although the error rate for single nanopore reads is higher, consensus approaches can result in highly reliable assemblies. Assemblies derived from MinION-1D sequencing have shown a good agreement between AMR testing and AMR prediction

TABLE 3 | AMR gene profile of the 10 strains.

		Strains										
		Reference	**3**	**12**	**18**	**57**	**61**	**100A**	**240**	**274**	**285**	**298**
Gene	**Mutation**											
gyrA	S91F/Y	S	S	**F**	**F**	**F**	S	**F**	**F**	S	**F**	S
	D95N/G/A	D	D	**A**	**A**	**A**	D	**A**	D	D	**A**	D
parC	E91K/G	E	E	**G**	E	E	E	E	**G**	E	**G**	**G**
***ponA1* (pbp1)**	L421P	L	**P**	L	**P**	**P**	L	**P**	**P**	L	**P**	L
rpsJ	V57M	V	**M**	**M**	**M**	**M**	**M**	**M**	**M**	**M**	**M**	**M**
***penA*/pbp2**	D345 insertion	–	**D**	**D**	**D**	**D**	**D**	**D**	**D**	**D**	**D**	**D**
	F504L	F	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**
	A510V	A	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**
	A516G	A	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**
	P551S/L	P	P	P	**S**	**S**	P	**S**	**L**	**S**	**S**	**S**
***porB1b* (penb)**	G120D/K	G	G	G	**D**	**D**	G	**D**	**D**	G	G	G
	A121D	A	A	A	**D**	**G**	A	**G**	A	**S**	**S**	**G**
	N122	N	N	N	N	**D**	N	**S**	N	**K**	**K**	N
***pilQ* (penC)**	QAATPAKQ insertion at 180	–	–	–	–	–	–	–	–	**QAATPAK**	–	–
	D526N	D	D	D	D	D	D	D	D	D	**G**	D
	N640S	N	N	N	N	N	N	N	N	N	**S**	N
folP	T66M	T	**M**	T	T	T	T	T	**M**	T	**M**	T
	R228S	R	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**
mtrR	A39T	A	**T**	**T**	**T**	**T**	**T**	**T**	A	A	A	A
	H105	H	H	H	H	H	H	H	**Y**	**Y**	**Y**	H
rplD	G70D/S/A/R	G	**D**	Y	G	Y	G	G	G	G	G	G
rpoB	H552N	H	**N**	H	**N**	**N**	H	**N**	H	H	H	H
Ciprofloxacin resistance		–	–	R	R	R	I	I	–	–	R	R
Gentamycin resistance		–	–	I	–	–	–	–	–	–	–	–

Mutations observed in the studied strains have been listed, mutated residues are indicated in bold, and mutations identified in this study have been underlined. The resistance profile for the strains as inferred through MIC has been indicated for ciprofloxacin and gentamycin. The isolates were sensitive to the other antibiotics that were tested as inferred from ***Table 2****. The newly identified mutations need to be experimentally investigated for their role in antibiotic resistance.*

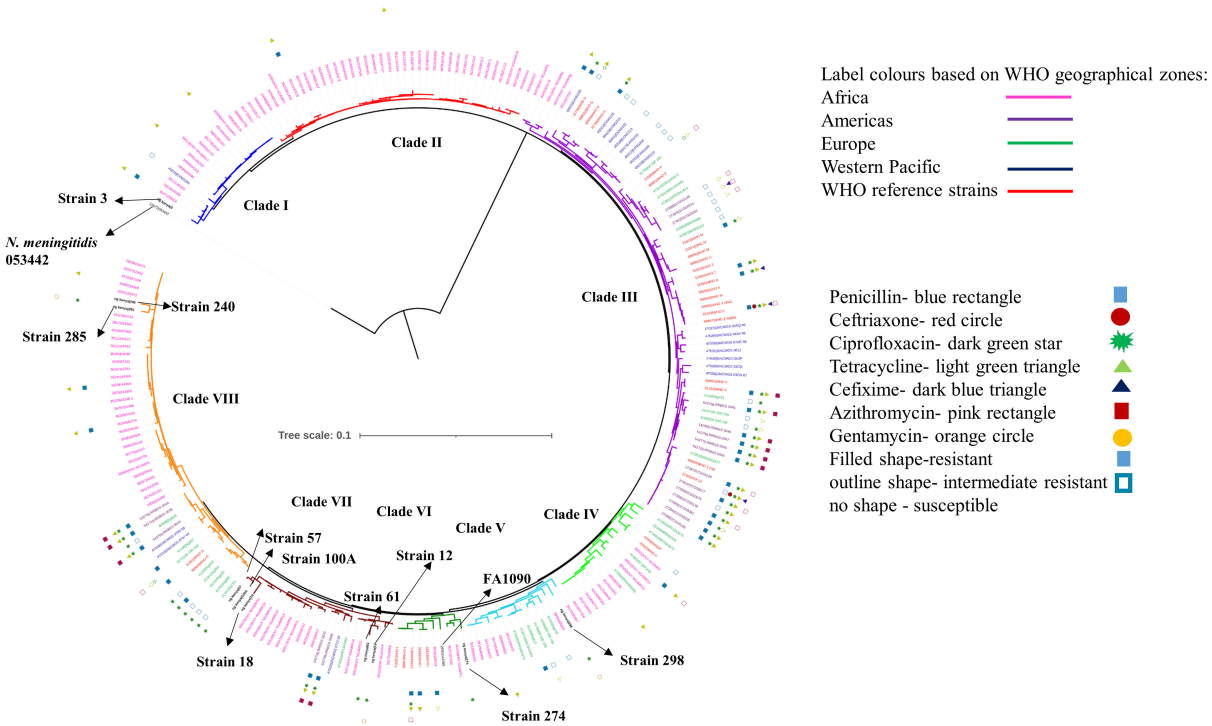


FIGURE 1 | Phylogenetic tree of the sequenced strains with strains from WHO geographical regions and WHO reference strains. The strains sequenced in this study have been indicated in black, bold text in the tree. The antibiotic resistance profile, wherever available, has been indicated as metadata (Penicillin – blue rectangle, Ceftriaxone – red circle, Ciprofloxacin – dark green star, Tetracycline – right-pointing light green triangle, Cefixime – left-pointing dark blue triangle, Azithromycin – pink rectangle, Gentamycin – orange circle, Filled shape – resistant, outline shape – intermediate resistant, no shape – susceptible). The clades inferred from the phylogeny analysis have been indicated in different colors and numbered I–VIII. The strains from different WHO geographical regions have been indicated in different color labels (Africa – Pink, Americas – Magenta, Europe – Green, Western Pacific – Blue, WHO reference strains – Red) The strains sequenced in this study, the reference strain for *N. gonorrhoeae* (FA1090), and an *N. meningitidis* strain (PubMLST id:053442) used as an outgroup for the analysis have been labeled and indicated with arrows.

results (Lemon et al., 2017; Sanderson et al., 2020). Studies on comparison of assemblies derived from Nanopore-only, Illumina, and hybrid assemblies have shown that the AMR prediction is comparable across all three methods (Chen et al., 2020). Recent advances in Nanopore sequencing and analysis workflows have demonstrated the utility of ONT sequencers for metagenomic analysis (Sanderson et al., 2020; Street et al., 2020). Additionally, there are workflows available for real-time sequence analysis that make the process faster.

Databases like PubMLST, CARD, ARIBA, and Pathogenwatch include profiles for resistance mediators like *penA*, *mtrR*, *penB*, *ponA*, 23S rDNA, *gyrA*, and *tetM* and do not include other determinants in their gonococcal AMR characterization module. We identified a set of AMR determinants from literature, including chromosomal gene mutations, promoter mutations, and the presence of plasmids with AMR genes. Plasmids containing genes encoding beta-lactamase and tetracycline resistance mediators [*tet*(M) and *bla*TEM] were identified in all but one strain (strain 12 contained only a TetM containing plasmid). This is consistent with a previous study that the plasmids are almost ubiquitous in Kenyan strains, thought to be because of the overuse of doxycycline for treating STIs (Cehovin et al., 2018). Among the reported mutations, we observed S91F and D95Y in the QRDR (residues 55–110) of GyrA and E91G in the ParC QRDR (residues 66–119). We also observed the silent mutations Y104 (codon change TAC − > TAT seen in all isolates except 298) and L131 (codon change of CTG − > CTA seen in isolates 3 and 240) in ParC reported to be associated with ciprofloxacin-resistant strains (Chaudhry et al., 2002). Novel mutations, V68A in the QRDR region of ParC (strain 285), E79 insertion in FolP (all isolates), and G70Y in RplD (strains 12 and 57), were also observed. However, whether these mutations have any effect on antibiotic resistance has to be experimentally validated. We were able to explain the basis of ciprofloxacin resistance in all strains except one, strain 298, which showed no mutations in gyrA, but had the E91G mutation in parC. We did not identify plasmids containing aminoglycoside resistance determinants like aminoglycoside N-acetyltransferases (*aat*) for gentamycin resistance (Garneau-Tsodikova and Labby, 2016).

Based on homology modeling using templates crystalized with antibiotics, we see that the mutations in GyrA disrupt the hydrogen bonding with ciprofloxacin, and mutations in ParC occur very close to Ser 88, which is involved in hydrogen bonding with moxifloxacin. We also observe that mutations in FolP at the substrate-binding site could affect the interaction with the antibiotic, as the residue at R228 interacts with sulfonamide.

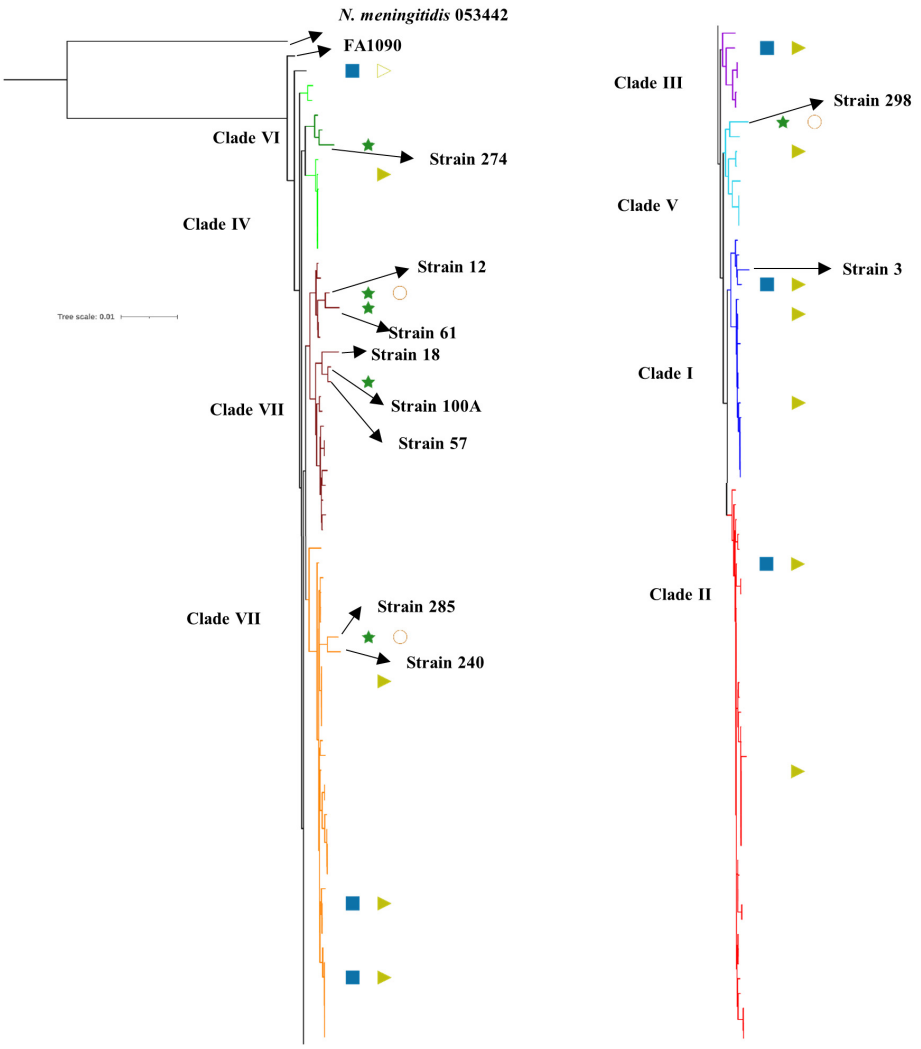


FIGURE 2 | Phylogenetic tree of the sequenced strains with previously reported strains from Kenya. The antibiotic resistance profile, wherever available, has been indicated as metadata (Penicillin – blue rectangle, Ciprofloxacin – dark green star, Tetracycline – right-pointing light green triangle, Gentamycin – orange circle, Filled shape – resistant, outline shape – intermediate resistant, no shape – susceptible). The clades inferred from the phylogeny analysis have been indicated in different colors and numbered I–VIII as per **Figure 1**. The strains sequenced in this study, the reference strain for *N. gonorrhoeae* (FA1090), and an *N. meningitidis* strain (PubMLST id:053442) used as an outgroup for the analysis have been labeled and indicated with arrows.

A phylogenetic study with previously sequenced Kenyan strains clustered our strains in the previously described cluster. From phylogenetic analysis using sequences from different countries, we observed five major clusters, with most of the Kenyan sequences (including the sequences from this study) clustered together with strains from different geographical regions.

We anticipate that the nanopore sequencing technology will expand the scope for rapid genome sequencing, which can help us understand the mechanistic basis of drug resistance and clinical management. AMR in *N. gonorrhoeae* is a result of a single to multiple mutations in same or different genes acting synergistally and additional unknown mutations/genes may be involved (Harrison et al., 2016; Ma et al., 2020a). Hence, there is a need for updated resistance gene databases and more studies on AMR genotype–phenotype correlations (Lemon et al., 2017). Recent advances have also shown promise for identifying clinical isolates and detecting AMR elements in real time (Schmidt et al., 2017; Sanderson et al., 2018). However, the read error rates mandate a consensus approach for read assembly and AMR allele detection. With future improvements in read accuracy and computational tools for base-calling, we believe that AMR detection can happen in clinical settings within hours of sequencing (Lemon et al., 2017).

CONCLUSION

Using a literature mining approach, we have predicted ciprofloxacin resistance using mutations in *gyrA*/*parC* in strains showing decreased susceptibility to ciprofloxacin.

However, few of the *gyrA* mutations can also occur in susceptible strains, and molecular tests using gene-based PCR or WGS can be used to complement culture-based antibiotic resistance testing. Culture-based testing can reflect mutations in unknown antibiotic targets, but cannot predict if a patient can develop resistance later on, based on pre-existing mutations.

In conclusion, in the first consensus genome for gonococci using only MinION, we show that using this approach, we can obtain near-complete genomes that were effectively used for AMR and phylogenetic analysis. We also show that currently available tools for AMR analysis of gonococci are not able to capture many mutations listed in the literature. We have also provided a dataset of over 100 existing mutations in different genes implicated in AMR, plasmids, and efflux pumps, which can be used by researchers across the world. This list will be continuously updated to keep up with the identification of new AMR targets/mutations.

Here, we demonstrate the potential of using MinION in resource-limited settings where NGS facility is unavailable. This can be used in settings with concerns about the export of samples/DNA for WGS to other countries. This approach can also be used across a range of bacterial genomes, even in case of metagenomic samples, where reads from a bacterial genome can be separated from one another, for subsequent analysis. Testing in clinical datasets can potentially provide a new approach to complement traditional methods for diagnosis, AMR surveillance, and public health management of *N. gonorrhoeae*. Further work is required in evaluating the MinION sequencer for outbreak prediction and clinical diagnosis of bacterial infection.

DATA AVAILABILITY STATEMENT

The raw reads generated for this study have been deposited in BioProject Accession: PRJNA660404 (https://www.ncbi.nlm.nih.gov). The biosample details are available under the IDs SAMN15960547, SAMN15960548, SAMN15960549, SAMN15960550, SAMN15960551, SAMN15960552, SAMN15960553, SAMN15960554, and SAMN15960555. The corresponding genomes and annotation files are available under the IDs CP061491, CP061490, CP061498, CP061488, CP061487, CP061486, CP061492, CP061485, CP061484, and CP061483.

ETHICS STATEMENT

The ethics approval for the study was obtained from the Kenyatta National Hospital – University of Nairobi Ethics and Research Committee, Nairobi, Kenya, for the use of anonymized samples collected from female patients attending local Sexually Transmitted Disease (STD) clinics.

AUTHOR CONTRIBUTIONS

MJ collected the isolates and carried out characterization, culturing, and sequencing. AS, TD, AP, and RK planned and carried out the sequencing. IM conceived and performed the bioinformatics analysis, and wrote the manuscript with inputs from SK, AS, MM, and MJ. RN, PM, MM, RC, JN, and JG performed the experiments. SK, RS, and OA provided intellectual support and valuable discussions. All authors listed have made a substantial contribution to the work and approved it for publication.

FUNDING

This work was supported by the capacity component of the "Indo Africa dengue sequencing to vaccine" grant from Narayana Murthy (Infosys) and the NCBS core funds to SK. We would like to acknowledge KAVI-ICR for the initial funding for the project. We thank NCBS (TIFR) and KAVI-ICR for infrastructural and financial support. RS acknowledges funding and support provided by JC Bose Fellowship (SB/S2/JC-071/2015) from Science and Engineering Research Board, India and Bioinformatics Centre Grant funded by Department of Biotechnology, India (BT/PR40187/BTIS/137/9/2021).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.647565/full#supplementary-material

REFERENCES

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

Altschul, S. F. (2005). "BLAST Algorithm," in *Encyclopedia of Life Sciences*, eds R. Bridgewater. (New York: John Wiley and Sons)

Balloux, F., Brønstad Brynildsrud, O., van Dorp, L., Shaw, L. P., Chen, H., Harris, K. A., et al. (2018). From Theory to Practice: translating Whole-Genome Sequencing (WGS) into the Clinic. *Trends Microbiol.* 26, 1035–1048. doi: 10.1016/j.tim.2018.08.004

Belland, R. J., Morrison, S. G., Ison, C., and Huang, W. M. (1994). *Neisseria gonorrhoeae* acquires mutations in analogous regions of gyrA and parC in fluoroquinolone-resistant isolates. *Mol. Microbiol.* 14, 371–380. doi: 10.1111/j.1365-2958.1994.tb01297.x

Bennett, J. S., Jolley, K. A., Sparling, P. F., Saunders, N. J., Hart, C. A., Feavers, I. M., et al. (2007). Species status of *Neisseria gonorrhoeae*: evolutionary and epidemiological inferences from multilocus sequence typing. *BMC Biol.* 5:35. doi: 10.1186/1741-7007-5-35

Castro-Wallace, S. L., Chiu, C. Y., John, K. K., Stahl, S. E., Rubins, K. H., McIntyre, A. B. R., et al. (2017). Nanopore DNA Sequencing and Genome Assembly on the International Space Station. *Sci. Rep.* 7:18022. doi: 10.1038/s41598-017-18364-0

Cehovin, A., Harrison, O. B., Lewis, S. B., Ward, P. N., Ngetsa, C., Graham, S. M., et al. (2018). Identification of Novel *Neisseria gonorrhoeae* Lineages Harboring Resistance Plasmids in Coastal Kenya. *J. Infect. Dis.* 218, 801–808. doi: 10.1093/infdis/jiy240

Chaudhry, U., Ray, K., Bala, M., and Saluja, D. (2002). Mutation patterns in gyrA and parC genes of ciprofloxacin resistant isolates of *Neisseria gonorrhoeae* from India. *Sex. Transm. Infect.* 78, 440–444. doi: 10.1136/sti.78.6.440

Chen, A., and Seifert, H. S. (2013). Structure-function studies of the *Neisseria gonorrhoeae* major outer membrane porin. *Infect. Immun.* 81, 4383–4391. doi: 10.1128/iai.00367-13

Chen, Z., Kuang, D., Xu, X., González-Escalona, N., Erickson, D. L., Brown, E., et al. (2020). Genomic analyses of multidrug-resistant *Salmonella* Indiana, Typhimurium, and Enteritidis isolates using MinION and MiSeq sequencing technologies. *PLoS One* 15:e0235641. doi: 10.1371/journal.pone.0235641

Chisholm, S. A., Quaye, N., Cole, M. J., Fredlund, H., Hoffmann, S., Jensen, J. S., et al. (2011). An evaluation of gentamicin susceptibility of *Neisseria gonorrhoeae* isolates in Europe. *J. Antimicrob. Chemother.* 66, 592–595. doi: 10.1093/jac/dkq476

Collineau, L., Boerlin, P., Carson, C. A., Chapman, B., Fazil, A., Hetman, B., et al. (2019). Integrating Whole-Genome Sequencing Data Into Quantitative Risk Assessment of Foodborne Antimicrobial Resistance: a Review of Opportunities and Challenges. *Front. Microbiol.* 10:1107. doi: 10.3389/fmicb.2019.01107

Demczuk, W., Sidhu, S., Unemo, M., Whiley, D. M., Allen, V. G., Dillon, J. R., et al. (2017). *Neisseria gonorrhoeae* Sequence Typing for Antimicrobial Resistance, a Novel Antimicrobial Resistance Multilocus Typing Scheme for Tracking Global Dissemination of N. gonorrhoeae Strains. *J. Clin. Microbiol.* 55, 1454–1468. doi: 10.1128/jcm.00100-17

Duncan, S., Thiong'o, A. N., Macharia, M., Wamuyu, L., Mwarumba, S., Mvera, B., et al. (2011). High prevalence of quinolone resistance in *Neisseria gonorrhoeae* in coastal Kenya. *Sex,. Transm. Infect.* 87:231. doi: 10.1136/sti.2010.048777

Elliott, I., Batty, E. M., Ming, D., Robinson, M. T., Nawtaisong, P., de Cesare, M., et al. (2020). Oxford Nanopore MinION Sequencing Enables Rapid Whole Genome Assembly of Rickettsia typhi in a Resource-Limited Setting. *Am. J. Trop. Med. Hyg.* 102, 408–414. doi: 10.4269/ajtmh.19-0383

Eyre, D. W., Sanderson, N. D., Lord, E., Regisford-Reimmer, N., Chau, K., Barker, L., et al. (2018). Gonorrhoea treatment failure caused by a *Neisseria gonorrhoeae* strain with combined ceftriaxone and high-level azithromycin resistance, England, February 2018. *Euro Surveill.* 23:1800323.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230.

Garneau-Tsodikova, S., and Labby, K. J. (2016). Mechanisms of Resistance to Aminoglycoside Antibiotics: overview and Perspectives. *Medchemcomm* 7, 11–27. doi: 10.1039/c5md00344j

Gerstmans, H., Grimon, D., Gutiérrez, D., Lood, C., Rodríguez, A., van Noort, V., et al. (2020). A VersaTile-driven platform for rapid hit-to-lead development of engineered lysins. *Sci. Adv.* 6:eaaz1136. doi: 10.1126/sciadv.aaz1136

Goldstein, S., Beka, L., Graf, J., and Klassen, J. L. (2019). Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing. *BMC Genomics.* 20:23. doi: 10.1186/s12864-018-5381-7

Golparian, D., Donà, V., Sánchez-Busó, L., Foerster, S., Harris, S., Endimiani, A., et al. (2018). Antimicrobial resistance prediction and phylogenetic analysis of *Neisseria gonorrhoeae* isolates using the Oxford Nanopore MinION sequencer. *Sci. Rep.* 8:17596. doi: 10.1038/s41598-018-35750-4

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

Harrison, O. B., Clemence, M., Dillard, J. P., Tang, C. M., Trees, D., Grad, Y. H., et al. (2016). Genomic analyses of *Neisseria gonorrhoeae* reveal an association of the gonococcal genetic island with antimicrobial resistance. *J. Infect.* 73, 578–587. doi: 10.1016/j.jinf.2016.08.010

Hunt, M., Mather, A. E., Sánchez-Busó, L., Page, A. J., Parkhill, J., Keane, J. A., et al. (2017). ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb. Genom.* 3:e000131.

Johnson, L., Weberman, B., Parker, N., and Hackert, P. (2016). *Neisseria meningitidis*: an Emerging Sexually Transmitted Infection. *Open Forum Infect. Dis.* 3:1301. doi: 10.1093/ofid/ofw172.1004

Jolley, K. A., Bliss, C. M., Bennett, J. S., Bratcher, H. B., Brehony, C., Colles, F. M., et al. (2012). Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* 158, 1005–1015. doi: 10.1099/mic.0.055459-0

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

Kivata, M. W., Mbuchi, M., Eyase, F. L., Bulimo, W. D., Kyanya, C. K., Oundo, V., et al. (2019). gyrA and parC mutations in fluoroquinolone-resistant *Neisseria gonorrhoeae* isolates from Kenya. *BMC Microbiol.* 19:76. doi: 10.1186/s12866-019-1439-1

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

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

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12.

Laver, T., Harrison, J., O'Neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., et al. (2015). Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol. Detect. Quantif.* 3, 1–8. doi: 10.1016/j.bdq.2015.02.001

Lee, S.-G., Lee, H., Jeong, S. H., Yong, D., Chung, G. T., Lee, Y. S., et al. (2010). Various penA mutations together with mtrR, porB and ponA mutations in Neisseria gonorrhoeae isolates with reduced susceptibility to cefixime or ceftriaxone. *J. Antimicrob. Chemother.* 65, 669–675. doi: 10.1093/jac/dkp505

Lemon, J. K., Khil, P. P., Frank, K. M., and Dekker, J. P. (2017). Rapid Nanopore Sequencing of Plasmids and Resistance Gene Detection in Clinical Isolates. *J. Clin. Microbiol.* 55, 3530–3543. doi: 10.1128/jcm.01069-17

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, 242–245.

Li, H. (2016). Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* 32, 2103–2110. doi: 10.1093/bioinformatics/btw152

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698

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

Loman, N. J., Quick, J., and Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods* 12, 733–735. doi: 10.1038/nmeth.3444

Ma, K. C., Mortimer, T. D., Duckett, M. A., Hicks, A. L., Wheeler, N. E., Sánchez-Busó, L., et al. (2020b). Increased power from conditional bacterial genome-wide association identifies macrolide resistance mutations in *Neisseria gonorrhoeae*. *Nat. Commun.* 11:5374. doi: 10.1038/s41467-020-19250-6

Ma, K. C., Mortimer, T. D., Duckett, M. A., Hicks, A. L., Wheeler, N. E., Sánchez-Busó, L., et al. (2020a). Increased power from bacterial genome-wide association conditional on known effects identifies *Neisseria gonorrhoeae* macrolide resistance mutations in the 50S ribosomal protein L4. *bioRxiv* [preprint].

Martin, I. M. C., Ison, C. A., Aanensen, D. M., Fenton, K. A., and Spratt, B. G. (2004). Rapid sequence-based identification of gonococcal transmission clusters in a large metropolitan area. *J. Infect. Dis.* 189, 1497–1505. doi: 10.1086/383047

Naidenov, B., Lim, A., Willyerd, K., Torres, N. J., Johnson, W. L., Hwang, H. J., et al. (2019). Pan-Genomic and Polymorphic Driven Prediction of Antibiotic Resistance in *Elizabethkingia*. *Front. Microbiol.* 10:1446. doi: 10.3389/fmicb.2019.01446

NCBI Resource Coordinators (2017). Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 45, D12–D17.

Nguyen, S. H., Cao, M. D., and Coin, L. (2020). Real-time resolution of short-read assembly graph using ONT long reads. *bioRxiv* [preprint].

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2013). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42, D206–D214. doi: 10.1093/nar/gkt1226

Papp, J. R., Abrams, A. J., Nash, E., Katz, A. R., Kirkcaldy, R. D., O'Connor, N. P., et al. (2017). Azithromycin Resistance and Decreased Ceftriaxone Susceptibility in *Neisseria gonorrhoeae*, Hawaii, USA. *Emerg. Infect. Dis.* 23, 830–832. doi: 10.3201/eid2305.170088

Plesivkova, D., Richards, R., and Harbison, S. (2019). A review of the potential of the MinIONTM single-molecule sequencing system for forensic applications. *Wiley Interdiscip. Rev. Forensic. Sci.* 1:e1323. doi: 10.1002/wfs2.1323

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Roberts, M. C., Chung, W. O., Roe, D., Xia, M., Marquez, C., Borthagaray, G., et al. (1999). Erythromycin-resistant *Neisseria gonorrhoeae* and oral commensal Neisseria spp. carry known rRNA methylase genes. *Antimicrob. Agents Chemother.* 43, 1367–1372. doi: 10.1128/aac.43.6.1367

Sánchez, R., and Sali, A. (2000). Comparative protein structure modeling. Introduction and practical examples with modeller. *Methods Mol. Biol.* 143, 97–129. doi: 10.1385/1-59259-368-2:97

Sanderson, N. D., Street, T. L., Foster, D., Swann, J., Atkins, B. L., Brent, A. J., et al. (2018). Real-time analysis of nanopore-based metagenomic sequencing from infected orthopaedic devices. *BMC Genom.* 19:714. doi: 10.1186/s12864-018-5094-y

Sanderson, N. D., Swann, J., Barker, L., Kavanagh, J., Hoosdally, S., Crook, D., et al. (2020). High precision *Neisseria gonorrhoeae* variant and antimicrobial resistance calling from metagenomic Nanopore sequencing. *Genome Res.* 30, 1354–1363. doi: 10.1101/gr.262865.120

Schmidt, K., Mwaigwisya, S., Crossman, L. C., Doumith, M., Munroe, D., Pires, C., et al. (2017). Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J. Antimicrob. Chemother.* 72, 104–114. doi: 10.1093/jac/dkw397

Sievers, F., and Higgins, D. G. (2014). Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol. Biol.* 1079, 105–116. doi: 10.1007/978-1-62703-646-7_6

Slack, M. P. E. (2010). "Chapter 172 - Gram-negative coccobacilli," in *Powderly WGBT-ID (Third E)*, eds J. Cohen and S. M. Opal (London: Content Repository Only), 1738–1756. doi: 10.1016/b978-0-323-04579-7.00172-6

Street, T. L., Barker, L., Sanderson, N. D., Kavanagh, J., Hoosdally, S., Cole, K., et al. (2020). Optimizing DNA Extraction Methods for Nanopore Sequencing of *Neisseria gonorrhoeae* Directly from Urine Samples. *J. Clin. Microbiol.* 58, e01822–19.

Tanaka, M., Nakayama, H., Haraoka, M., and Saika, T. (2000). . Antimicrobial resistance of *Neisseria gonorrhoeae* and high prevalence of ciprofloxacin-resistant isolates in Japan, 1993 to 1998. *J. Clin. Microbiol.* 38, 521–525. doi: 10.1128/jcm.38.2.521-525.2000

Tyler, A. D., Mataseje, L., Urfano, C. J., Schmidt, L., Antonation, K. S., Mulvey, M. R., et al. (2018). Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. *Sci. Rep.* 8:10931. doi: 10.1038/s41598-018-29334-5

Unemo, M., and Dillon, J.-A. R. (2011). Review and international recommendation of methods for typing neisseria gonorrhoeae isolates and their implications for improved knowledge of gonococcal epidemiology, treatment, and biology. *Clin. Microbiol. Rev.* 24, 447–458. doi: 10.1128/cmr.00040-10

Unemo, M., Golparian, D., Sánchez-Busó, L., Grad, Y., Jacobsson, S., Ohnishi, M., et al. (2016). The novel 2016 WHO *Neisseria gonorrhoeae* reference strains for global quality assurance of laboratory investigations: phenotypic, genetic and reference genome characterization. *J. Antimicrob. Chemother.* 71, 3096–3108. doi: 10.1093/jac/dkw288

Unemo, M., and Nicholas, R. A. (2012). Emergence of multidrug-resistant, extensively drug-resistant and untreatable gonorrhea. *Future Microbiol.* 7, 1401–1422. doi: 10.2217/fmb.12.117

Unemo, M., and Shafer, W. M. (2011). Antibiotic resistance in *Neisseria gonorrhoeae*: origin, evolution, and lessons learned for the future. *Ann. N. Y. Acad. Sci.* 1230, E19–E28.

Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746. doi: 10.1101/gr.214270.116

Wi, T., Lahra, M. M., Ndowa, F., Bala, M., Dillon, J.-A. R., Ramon-Pardo, P., et al. (2017). Antimicrobial resistance in *Neisseria gonorrhoeae*: global surveillance and a call for international collaborative action. *PLoS Med.* 14:e1002344. doi: 10.1371/journal.pmed.1002344

World Health Organisation (WHO) (2017). *Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics.* 74. Available at: https://www.who.int/medicines/publications/WHO-PPL-Short_Summary_25Feb-ET_NM_WHO.pdf (accessed January 12, 2021).

World Health Organisation (WHO) (2018). *Report On Global Sexually Transmitted Infection Surveillance.* 74. Available Online at: https://www.who.int/reproductivehealth/publications/stis-surveillance-2018/en/ (accessed January 12, 2021).

Yun, M.-K., Wu, Y., Li, Z., Zhao, Y., Waddell, M. B., Ferreira, A. M., et al. (2012). Catalysis and Sulfa Drug Resistance in Dihydropteroate Synthase. *Science* 335, 1110–1114. doi: 10.1126/science.1214641

Zapun, A., Morlot, C., and Taha, M.-K. (2016). Resistance to β-Lactams in Neisseria ssp Due to Chromosomally Encoded Penicillin-Binding Proteins. *Antibiotics* 5:35. doi: 10.3390/antibiotics5040035

Zhang, C., Wang, F., Zhu, C., Xiu, L., Li, Y., Li, L., et al. (2020). Determining antimicrobial resistance profiles and identifying novel mutations of *Neisseria gonorrhoeae* genomes obtained by multiplexed MinION sequencing. *Sci. China Life Sci.* 63, 1063–1070. doi: 10.1007/s11427-019-1558-8

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

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

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership