Data Descriptor

Genomic Typing, Antimicrobial Resistance Gene and Virulence Factor Dataset for the Important Pathogenic Bacteria *Klebsiella pneumoniae*

Andrey Shelenkov *, Anna Slavokhotova, Yulia Mikhailova, Vasiliy Akimkin

Posted Date: 24 September 2024

doi: 10.20944/preprints202409.1714.v1

Keywords: *Klebsiella pneumoniae*; healthcare-associated infections; genomic epidemiology; whole genome sequencing; cgMLST; global clones; antibiotic resistance; virulence factors

 doi:10.20944/preprints202409.1714.v

Data Descriptor

Genomic Typing, Antimicrobial Resistance Gene and Virulence Factor Dataset for the Important Pathogenic Bacteria *Klebsiella pneumoniae*

Andrey Shelenkov *, Anna Slavokhotova, Yulia Mikhaylova and Vasiliy Akimkin

Central Research Institute of Epidemiology, Novogireevskaya str., 3a, 111123 Moscow, Russia
* Correspondence: fallandar@gmail.com

Abstract: The infections of bacterial origin represent a significant problem to public healthcare worldwide both in clinical and community settings. Recent decade was marked by limiting treatment options for bacterial infections due to growing antimicrobial resistance (AMR) acquired and transferred by various bacterial species, especially the ones causing healthcare-associated infections, which has become a dangerous issue noticed by the World Health Organization. Numerous reports shown that the spread of AMR is often driven by several species-specific lineages usually called the 'global clones of high risk'. Thus, it is essential to track the isolates belonging to such clones and investigate the mechanisms of their pathogenicity and AMR acquisition. Currently, the whole genome-based analysis is increasingly used for these purposes, including epidemiological surveillance and analysis of mobile elements involved in resistance transfer. However, in spite of the exponential growth of available bacterial genomes, their representation usually lack uniformity and availability of supporting metadata, which creates a bottleneck for such investigations. In this dataset, we provide the results of a comprehensive genomic epidemiology analysis of 61,857 genomes of a dangerous bacterial pathogen *Klebsiella pneumoniae*. Important typing information including multilocus sequence typing (MLST)-based sequence types (STs), assignment of the isolates to known global clones, capsular (KL) and lipooligosaccharide (O) types, presence of CRISPR-Cas systems, and cgMLST profiles are given, and the presence of AMR and virulence genes within the genomes is also reported. This dataset will be useful for researchers in the field of *K. pneumoniae* genomic epidemiology, resistance analysis and prevention measure development.

Keywords: *Klebsiella pneumoniae*; healthcare-associated infections; genomic epidemiology; whole genome sequencing; cgMLST; global clones; antibiotic resistance; virulence factors

Dataset: The dataset is available at https://doi.org/10.5281/zenodo.11069018

1. Summary

Infections of bacterial origin represent one of the most serious problems in global healthcare. The treatment of such infections is complicated due to spread of antimicrobial drug resistance in clinical pathogenic bacteria, which leads to a limited set of available treatment options [1]. However, antimicrobial resistance (AMR) can also be acquired by the non-hospital bacterial populations, thus making this problem global and becoming an issue in community settings [2].

The spread of AMR within particular bacterial species is often facilitated by several groups or lineages usually called 'clonal groups', 'global clones' or 'international clones of high risk'. An important role of such clones has been demonstrated for some of the most successful and widespread nosocomial pathogens like *Klebsiella pneumoniae* [3,4], *Acinetobacter baumannii* [5] and *Pseudomonas aeruginosa* [6]. Therefore, epidemiological surveillance of multidrug-resistant (MDR) bacteria and

 doi:10.20944/preprints202409.1714.v

developing effective prevention measures against them should involve checking whether particular isolates belong to such global clones.

Isolate typing and assignment to a particular clone can be based on several characteristics revealed by molecular biology techniques, but now the whole genome sequencing (WGS) is becoming a gold standard for this and many other investigations due to unprecedented amount of data it produces and its high cost-effectiveness [7,8]. Currently, several hundred thousand genomes of pathogenic bacteria are available in public databases, and increasing availability of WGS will facilitate further growth in amounts of such data.

The fraction of genomes available in public databases for particular bacterial species is determined by its significance to public healthcare and the incidence of infections caused by this species. *K. pneumoniae* is responsible for a significant share of nosocomial infections worldwide [9], and the World Health Organization (WHO) listed carbapenem-resistant *K. pneumoniae* as one of the critical pathogens with the highest priority in new antibiotic development [10].

At present, more than 70,000 draft *K. pneumoniae* genomes are available at NCBI (https://www.ncbi.nlm.nih.gov/datasets/taxonomy/573/, accessed on 8 September 2024). However, the data provided there usually lacks isolate typing information and other metadata, including AMR gene presence, which should be derived using various computational tools by users. Another commonly used database PasteurMLST [11] contains epidemiologically related and typing information for more than 40,000 genomes (https://bigsdb.pasteur.fr/cgi-bin/bigsdb/bigsdb.pl?db=pubmlst_klebsiella_isolates, accessed on 8 September 2024). However, the extraction of AMR and virulence gene information from this database is not straightforward, especially for large number of isolates simultaneously. In addition, no mechanism is provided for quick download of multiple genomes satisfying several selection criteria at once.

In order to facilitate the rapid construction of *K. pneumoniae* isolate subsets for particular purpose, we developed the dataset containing typing information, including assignment to global clones, for the whole set of the *K. pneumoniae* isolates available at NCBI (accessed on 14 December 2023), which included 61,857 genomes. The information available contains multilocus sequence typing (MLST)-based types (STs), capsular polysaccharide (KL, or K) and lipooligosaccharide (O) types, core genome MLST (cgMLST) profiles, and the data regarding the presence of CRISPR-Cas systems in each of the isolates. The information regarding the presence of AMR genes providing resistance to various classes of antibiotics, as well as the set of virulence factors, is also available.

The combination of several typing schemes, e.g., ST, KL and O types, was previously shown to provide significantly higher resolution than any of these schemes used alone, while cgMLST possessed the highest discrimination ability [12].

This dataset was used for selection of appropriate reference genomes for comparison purposes in our previous research [3,13] and found to be very convenient for these purposes, so we decided making it publicly available for researchers working on *K. pneumoniae* genomic analysis.

We believe that the dataset will be useful for genomic and epidemiological studies of *K. pneumoniae*, including selection of the proper reference isolate sets for any types of genome-based investigations. The precomputed data will be especially useful for the researchers working in the promising field of genomic epidemiology of this important and highly dangerous pathogen.

 The updates are scheduled to be provided at least once a year.

2. Data Description

2.1. Data Structure

The dataset contains four tables provided in various formats: xlsx (all except cgMLST), tab-delimited txt and pdf (for summary table only). The files in xlsx format can be further processed by users in table processing software, e.g., the filters can be added to select the subsets of interest, sorting parameters can be changed, various graphs can be built, statistical processing can be performed etc.

 doi:10.20944/preprints202409.1714.v

Text format (txt) is intended for computational processing by various bioinformatics tools, while pdf format presents the summary file containing main isolate typing results in human-readable form.

The tables, which will be further described below, include the following:

- A summary table (table_summary) containing typing information for all isolates, such as MLST ST, KL- and O-types, assignments of the isolates to known clonal groups (global clones) and the possible presence and type of CRISPR-Cas systems in the genomes of the isolates
- An AMR gene table (table_amr) containing the information on the presence of genes known to provide, when properly expressed, the AMR to various classes of antimicrobial drugs for all isolates
- A virulence gene table (table_vfdb) containing the information on the presence of genes encoding virulence factors
- A table containing cgMLST profiles (table_cgmlst) for all isolates, which can be used for extended comparison and clustering purposes

2.1.1. Typing Summary Table

The format and exemplary data for the summary table are presented in Table 1.

Table 1. Exemplary data and column information for the typing summary table.

Assembly code	Clonal group	MLST ST	KL type	O type	CRISPR-Cas
GCA_000009885.1	CG23	23	KL1	O1/O2v2	I-E*
GCA_000255975.2	NA[1]	48	KL124	O1/O2v1	NF[2]
GCA_005508835.1	NA	ND[2]	KL15	O4	UNKN

[1] 'NA' (not available) indicates that the isolate did not belong to any known clonal group, [2] 'ND' (not determined) indicates that ST could not be determined; [3] 'NF'(not found) – CRISPR-Cas system was not found in the genome.

The first column contains the assembly code from Genbank, which uniquely identifies a particular K. pneumoniae genome assembly.

'CG' stands for 'clonal group' and indicates the assignment of a particular isolate to known clonal groups. Such an assignment is based on the ST according to the previously published data [4,14]. If an isolate was not assigned to any CG, this column contains 'NA' (not available) designation.

Third column contains an ST defined as a combination of seven loci (gapA, infB, mdh, pgi, phoE, rpoB and tonB genes) of a typing scheme [15]. Each variant of a particular locus is numbered sequentially, and the combination of seven locus variants constitutes a ST, which has its own number. For example, the combination of gapA_2, infB_1, mdh_1, pgi_1, phoE_9, rpoB_4 and tonB_12 was defined as ST23. Locus variants, as well as the definitions of corresponding STs, can be found in the Institut Pasteur database (https://bigsdb.pasteur.fr/cgi-bin/bigsdb/bigsdb.pl?db=pubmlst_klebsiella_seqdef, accessed on 12 August 2024). 'ND' in this column means that ST was not determined due to either low sequencing quality or the presence of a novel MLST allele, or novel allele combination, not uploaded to the databases yet.

KL- and O-type show the typing classes based on the corresponding sets of genes, respectively. Capsular polysaccharide is an essential factor determining bacterial virulence and their susceptibility to phages, which makes it a useful epidemiological marker [16]. The typing of capsular polysaccharide (cps) gene cluster in K. pneumoniae is based on the wzi or wzc gene sequences, while O-locus typing was defined by sequence identity in the conserved wzm and wzt genes. Each distinct gene cluster found between the flanking genes is assigned a unique number, or code, identifying the locus type, and these data can be found in public databases [17].

The Final column shows the presence of a CRISPR-Cas system in the isolate. Clustered regularly interspaced short palindromic repeat (CRISPR) arrays and CRISPR-associated genes (cas) function as a variable genetic element and form bacterial adaptive immune systems. CRISPR-Cas systems are divided into six major types (I-VI) and several subtypes (A-I, K, U) based on a combination of phylogenetic, comparative genomic, and structural analysis [18,19]. 'NF' in this column indicates that

 doi:10.20944/preprints202409.1714.v

CRISPR-Cas system was not revealed in a particular isolate genome, 'UNKN' shows the detected presence an incomplete or untypeable system, and in other cases a system type is shown.

2.1.2. AMR and Virulence Gene Tables

Another part of the dataset includes information regarding the presence of various genes known to confer AMR in the investigated K. pneumoniae genomes. The first column shows the assembly code, which is identical to the one from the typing summary table; the second column gives the number of AMR genes revealed in a particular isolate, while other columns describe the presence of a particular gene in the first row and its sequence similarity level with the corresponding allele from the Resfinder database. The absence of a gene is marked with a dot for better readability.

An example is provided in Table 2. Only a few genes are presented in this example.

Table 2. Exemplary data and column information for antimicrobial resistance gene table.

Assembly code	NUM_FOUND	*oqxA*	*oqxB*	…	*blaTEM-1B*	*catA1*
GCA_000016305.1	19	100.00	100.00	…	99.8	100.00
GCA_000219965.2	13	100.00	100.00	…	.	.

The presence of a particular gene known to provide a resistance to some antimicrobial drug by itself does not confirm the phenotypic resistance to this drug since this gene, for example, might not be expressed [12,20]. However, this information is essential for the estimation of the AMR potential and spread within bacterial population of interest.

The format of the virulence gene table is the same as the one for AMR gene table except for the set of genes included.

The information presented in AMR and virulence gene tables can be useful for the comparative analysis of pathogenicity within particular groups of the isolates. Genetically and epidemiologically close isolates can possess different AMR and virulence gene repertoire, and thus this information is essential for the selection of proper reference sets.

2.1.3. cgMLST Profiles

The fourth part of the dataset includes cgMLST profiles for all isolates. The cgMLST typing scheme is similar to MLST in the sense that it enumerates gene variants and uses their combination to form a profile, but the striking difference is that cgMLST relies on the set of all conservative genes within particular species (usually, the genes appeared in more than 90% of known isolates). cgMLST scheme for K. pneumoniae includes 2358 loci, and the allele variants for these loci are available in a regularly updated database at cgmlst.org (https://www.cgmlst.org/ncs/schema/schema/2187931/, accessed on 12 August 2024).

cgMLST profiles can be used in the cluster analysis of a selected set of isolates for estimation of their genomic similarity and closeness and, possibly, obtain some valuable epidemiological or evolutionary insights. The threshold of 18 differences in cgMLST loci was proposed to check whether two K. pneumoniae isolates belonged to a single strain [21], but less or more strict criteria can be used depending on the specific investigation goal.

In this dataset, cgMLST profiles are provided in a table format. The first column contains the same assembly identifier as other tables, while the other columns show the numbers representing the variants of genes appeared in a header row. Several special designations can appear besides numbers, namely, N - indicates a novel allele variant not present in the database; 0? indicates a locus is missing in the assembly (probably, due to the low quality of initial sample or sequencing problems); "-" points an allele is partially covered; "+" represents multiple possible alleles, in which case the most probable is left.

2.2. General Descriptive Statistics

 doi:10.20944/preprints202409.1714.v

Some general statistics based on the dataset is provided below. Previously it was shown that a Genbank set of genomes cannot be considered representative for the whole worldwide K. pneumoniae population since it was strongly biased towards multidrug-resistant or other clinically relevant isolates from particular regions [22]. At the same time, the descriptive statistics on the distribution of particular STs, CGs, AMR, virulence genes etc. can provide useful information for the purposes of reference set selection and making comparisons.

Below we will refer to any Genbank assembly record containing a complete or partial genome as an ''isolate'' for simplicity, although some different assemblies can in fact represent the same isolate, or some genomic records may contain only a part of the isolate genome.

The summary of top three dominating characteristics in each feature category is given in Table 3.

Table 3. Top three dominating representatives of the features investigated in K. pneumoniae dataset.

Feature	Top 3 representatives
Clonal Group	NA, CG258, CG147
Sequence type	ST11, ST258, ST147
KL-type	KL64, KL107, KL2
O-type	O1/O2v1, O1/O2v2, O3b
AMR genes	*oqxAB, fosA6, sul1*
Virulence genes	*ompA, fur, acrAB*
CRISPR-Cas system	NF, I-E, I-E*

Totally, 37.3% of the isolates belonged to known CGs, with CG258 accounting for 25.3% of all K. pneumoniae genomes from Genbank and 68% of K. pneumoniae isolates belonging to CGs, respectively. CG147 was the second largest CG, which included 6.2%/16.6% of all isolates/CG isolates. CG258 (including ST11 and ST258, among others), is known to be the dominant carbapenemase-producing K. pneumoniae clone worldwide, which raises global concerns in healthcare systems of many countries [23]. Thus, it is not surprising that it has the largest number of the sequenced isolates in Genbank, and two STs belonging to CG258 are also the most represented there with 15.5% for ST11 and 8% for ST258, respectively. ST147, a most prominent member of CG147, which is also recognized as a globally distributed highly resistant clone [24], accounted for 5.3% of the isolates.

In total, 2402 distinct STs were revealed in Genbank for K. pneumoniae, while ST could not be determined for 6.4% of the isolates due to low-quality sequencing and/or lack of coverage for the corresponding genomic regions. However, 1069 of these STs were represented by a single isolate only, and 389 top STs included ten or more isolates, totally enclosing 86% of the Genbank K. pneumoniae genomes.

Top three KL types included KL64 (12.9% of the isolates), KL107 (8.5%) and KL2 (5.5%). KL2 is considered to be predominantly associated with hypervirulent K. pneumoniae isolates [25]. The abundance of KL64 and KL107 is also not surprising since these types were often associated with CG258 and CG147, and ST11-KL64 and ST258-KL107 were even highlighted as a separate important multidrug-resistant clones [26,27]. The diversity of KL types was high with 151 distinct types revealed.

The range of lipopolysaccharide types was less diverse with 12 distinct types, three of which were O1 variants. O1/O2v1 (37.3%) and O1/O2v2 (34.9%) together accounted for about 72% of the isolates, which corresponds to previous reports [28,29], and O3b was revealed in about 8% of the isolates.

The median number of AMR genes possessed by the isolates was 13, with the numbers ranging from 1 to 43. This number was never equal to zero since intrinsic blaSHV gene variants were also included.

The most abundant AMR genes were oqxAB encoding efflux pump conferring fluoroquinolone resistance. These genes were revealed in about 90% of the isolates, which corresponds to the recent

 doi:10.20944/preprints202409.1714.v

clinical data [30]. High frequency of occurrence was also observed for fosA6 (fosfomycin resistance, 61.6%) and sul1 (sulfamethoxazole resistance, 44.4%). The most widespread beta-lactamase-encoding genes, except for intrinsic blaSHV, were $bla_{CTX\text{-}M\text{-}15}$ and $bla_{TEM\text{-}1B}$ accounting for about 38% each, while carbapenemase-encoding $bla_{NDM\text{-}1}$ (about 10%) was the most abundant in the corresponding class. However, it should be noted again that isolate distribution in Genbank is skewed towards MDR strains and thus cannot be considered representative for the whole population. On the other hand, these data provides useful insights into the procedure of relevant reference set selection for various K. pneumoniae investigations.

The number of virulence genes in the dataset varied from 27 to 131 with a median equal to 73. Among these, ompA (outer membrane protein, porin), fur (ferric uptake regulator) and acrAB (efflux pump) were revealed in virtually all isolates. Worth noting, 13.4% of the isolates carried rmpA, one of the major hypervirulence markers [31]. Totally, 428 distinct virulence genes were found, but only 137 of them occurred in a significant number (>=10) of the isolates.

CRISPR-Cas systems were revealed in about 29% (18199) of the isolates, with I-E occurring slightly more often than I-E* (14.2% vs. 10.9% for all isolates and 48.2% vs. 37% for the isolates with CRISPR-Cas, respectively). These findings correspond to other recent reports, in which CRISPR-Cas prevalence in K. pneumoniae was estimated to be about 25% [32,33]. Besides this, 2644 isolates included CRISPR-Cas systems which were incomplete or had undetectable type with alternative cas gene sequence. Interestingly, 67 isolates included two types of systems I-E and I-E* simultaneously, 15 of which belonged to ST1758-KL27-O4.

Another interesting fact was that about 92% of CG147 possessed CRISPR-Cas system, while CG258 virtually lacked it. The members of CG147 were previously shown to be abundant with CRISPR-Cas systems and even to possess both this system and antibiotic resistance-encoding plasmids simultaneously [34]. The absence of CRISPR-Cas in CG258 was also reported previously and linked to dissemination of IncF type plasmids in this group [35].

The full data describing the distribution of all the characteristics discussed above is presented in Table S1.

3. Methods

We retrieved 61,857 genomic sequences of *K. pneumoniae* from Genbank (https://www.ncbi.nlm.nih.gov/genbank/, accessed on 14 December 2023), for which the assembly level was defined as 'Complete Genome', 'Chromosome', or 'Scaffold', and for which MLST confirmed the correct species identification.

MLST was performed using the Institut Pasteur database (https://bigsdb.pasteur.fr/cgi-bin/bigsdb/bigsdb.pl?db=pubmlst_klebsiella_seqdef, accessed on 12 August 2024) using the typing scheme described in [15].

The evolution of the multidrug resistance in *K. pneumoniae* is largely driven by the plasmid-mediated acquisition of resistance genes, and several hospital outbreak clones, also known as 'global clones' or 'clonal groups', were shown to play an important role in this process [14,36]. Thus, an assignment of the isolates to known global clones represents an important step in epidemiological surveillance and antimicrobial resistance spread investigation. In this dataset, such an assignment was based on MLST ST according to previously published data [4,14]. The list of STs belonging to different clonal groups is shown in Table 4 for reference purposes.

Table 4. Composition of *K. pneumoniae* clonal groups based on MLST ST.

Clonal group	STs
CG15	14, 15, 709
CG20	16, 17, 20, 22, 336, 1123
CG23	23, 26, 57, 163, 218, 260, 887, 2044
CG25	25
CG29	29, 711, 1109, 1271
CG34	34, 228, 592, 1444, 1454, 1521, 2141

 doi:10.20944/preprints202409.1714.v

CG35	35, 466, 1948
CG36	36, 268, 298, 433, 564, 1272, 2130
CG37	37, 177, 256, 309, 896, 1198, 1507, 1779
CG43	43, 101, 2017
CG45	45, 485
CG48	48
CG65	25, 65, 375
CG66	66, 2039, 2058
CG86	86, 373, 3994
CG101	101, 1685, 2016, 2017, 2502
CG111	111, 692
CG133	133, 420, 2127
CG147	147, 273, 392, 1709
CG152	105, 152
CG230	224, 230, 1307, 1583
CG231	231, 1190
CG253	253, 1138, 2116
CG258	11, 258, 340, 379, 395, 418, 437, 512, 833, 855, 1084, 1199
CG322	322, 491, 1377
CG380	380, 679
CG490	76, 199, 490, 495, 1263
CG540	4, 6, 504, 540, 1013
CG661	237, 661, 2113

The AMR genes were detected using Resfinder 4.6.0 software [37] (https://http://genepi.food.dtu.dk/resfinder, achttps://doi.org/10.1093/nar/gkab1107cessed on 10 August 2024, using default parameters).

The detection of capsule polysaccharide loci (KL) and lipooligosaccharide loci (O) was made using Kaptive v. 2.0.9 [17] with the default parameters (last update of the databases was on 12 March 2024).

Searching for virulence factors in the *K. pneumoniae* genomes was performed using VFDB [38] (http://www.mgc.ac.cn/VFs/main.htm, accessed on 20 August 2024, using the default parameters)

The presence of CRISPR-Cas systems in the genomes under study was analyzed using CRISPRCasFinder [39] version 4.2.20 with the following parameters: '-fast -rcfowce -ccvRep -vicinity 1200 –cas -useProkka'. The variants of CRISPR-Cas systems (I-E vs. I-E*) were determined according to the literature data [33].

The cgMLST profiles were built using MentaList software [40] (https://github.com/WGS-TB/MentaLiST, version 0.2.4, default parameters, accessed on 21 August 2024) using the scheme with 2358 loci obtained from cgmlst.org (https://www.cgmlst.org/ncs/schema/schema/2187931/, accessed on 12 August 2024).

Additional data processing, output formatting and preparation of the tables for the dataset were performed using the computational pipeline developed by us earlier [41,42].

4. User Notes

The description of possible dataset applications to genomic epidemiology investigations, including the commands for UNIX-based systems, will be provided on the webpage of the dataset in a 'how_to_use.txt' file.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Table S1: General characteristics of *Klebsiella pneumoniae* genomes from Genbank.

Author contributions: Conceptualization, A.Sh., A.S and Y.M.; data curation, A. Sh.; formal analysis, A.Sh., Y.M., V.A.; funding acquisition, V.A.; investigation, A. Sh., A.S., Y.M., and V.A.; methodology, A.Sh. and Y.M.;

project administration, V.A.; resources, A.Sh, A.S..; software, A.Sh.; supervision, V.A.; validation, A.Sh. and A.S.; visualization, A.Sh.; writing—original draft, A.Sh.; writing—review and editing, A.Sh., A.S., Y.M., V.A. All authors have read and agreed to the published version of the manuscript

Funding: This work was supported by the Ministry of Science and Higher Education of the Russian Federation within the framework of a grant in the form of a subsidy for the creation and development of the «World-class Genomic Research Center for Ensuring Biological Safety and Technological Independence under the Federal Scientific and Technical Program for the Development of Genetic Technologies», agreement No. 075-15-2019-1666

Data Availability Statement: The dataset is available at https://doi.org/10.5281/zenodo.11069018.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Salam, M.A.; Al-Amin, M.Y.; Salam, M.T.; Pawar, J.S.; Akhter, N.; Rabaan, A.A.; Alqumber, M.A.A. Antimicrobial Resistance: A Growing Serious Threat for Global Public Health. *Healthcare* **2023**, *11*, doi:10.3390/healthcare11131946.
2. Denissen, J.; Reyneke, B.; Waso-Reyneke, M.; Havenga, B.; Barnard, T.; Khan, S.; Khan, W. Prevalence of ESKAPE pathogens in the environment: Antibiotic resistance status, community-acquired infection and risk to human health. *International Journal of Hygiene and Environmental Health* **2022**, *244*, doi:10.1016/j.ijheh.2022.114006.
3. Shaidullina, E.R.; Schwabe, M.; Rohde, T.; Shapovalova, V.V.; Dyachkova, M.S.; Matsvay, A.D.; Savochkina, Y.A.; Shelenkov, A.A.; Mikhaylova, Y.V.; Sydow, K., et al. Genomic analysis of the international high-risk clonal lineage Klebsiella pneumoniae sequence type 395. *Genome Medicine* **2023**, *15*, doi:10.1186/s13073-023-01159-6.
4. Arcari, G.; Carattoli, A. Global spread and evolutionary convergence of multidrug-resistant and hypervirulent Klebsiella pneumoniae high-risk clones. *Pathog Glob Health* **2023**, *117*, 328-341, doi:10.1080/20477724.2022.2121362.
5. Zhang, X.; Li, F.; Awan, F.; Jiang, H.; Zeng, Z.; Lv, W. Molecular Epidemiology and Clone Transmission of Carbapenem-Resistant Acinetobacter baumannii in ICU Rooms. *Front Cell Infect Microbiol* **2021**, *11*, 633817, doi:10.3389/fcimb.2021.633817.
6. Zurita, J.; Sevillano, G.; Solis, M.B.; Paz, Y.M.A.; Alves, B.R.; Changuan, J.; Gonzalez, P. Pseudomonas aeruginosa epidemic high-risk clones and their association with multidrug-resistant. *J Glob Antimicrob Resist* **2024**, *38*, 332-338, doi:10.1016/j.jgar.2024.07.003.
7. Atxaerandio-Landa, A.; Arrieta-Gisasola, A.; Laorden, L.; Bikandi, J.; Garaizar, J.; Martinez-Malaxetxebarria, I.; Martinez-Ballesteros, I. A Practical Bioinformatics Workflow for Routine Analysis of Bacterial WGS Data. *Microorganisms* **2022**, *10*, doi:10.3390/microorganisms10122364.
8. Price, V.; Ngwira, L.G.; Lewis, J.M.; Baker, K.S.; Peacock, S.J.; Jauneikaite, E.; Feasey, N. A systematic review of economic evaluations of whole-genome sequencing for the surveillance of bacterial pathogens. *Microb Genom* **2023**, *9*, doi:10.1099/mgen.0.000947.
9. Lin, X.C.; Li, C.L.; Zhang, S.Y.; Yang, X.F.; Jiang, M. The Global and Regional Prevalence of Hospital-Acquired Carbapenem-Resistant Klebsiella pneumoniae Infection: A Systematic Review and Meta-analysis. *Open Forum Infect Dis* **2024**, *11*, ofad649, doi:10.1093/ofid/ofad649.
10. WHO. *WHO Bacterial Priority Pathogens List, 2024: bacterial pathogens of public health importance to guide research, development and strategies to prevent and control antimicrobial resistance.*; World Health Organization: Geneva, 2024.
11. Jolley, K.A.; Bray, J.E.; Maiden, M.C.J. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* **2018**, *3*, 124, doi:10.12688/wellcomeopenres.14826.1.
12. Shelenkov, A.; Mikhaylova, Y.; Yanushevich, Y.; Samoilov, A.; Petrova, L.; Fomina, V.; Gusarov, V.; Zamyatin, M.; Shagin, D.; Akimkin, V. Molecular Typing, Characterization of Antimicrobial Resistance, Virulence Profiling and Analysis of Whole-Genome Sequence of Clinical Klebsiella pneumoniae Isolates. *Antibiotics (Basel)* **2020**, *9*, doi:10.3390/antibiotics9050261.
13. Shelenkov, A.; Mikhaylova, Y.; Voskanyan, S.; Egorova, A.; Akimkin, V. Whole-Genome Sequencing Revealed the Fusion Plasmids Capable of Transmission and Acquisition of Both Antimicrobial Resistance and Hypervirulence Determinants in Multidrug-Resistant Klebsiella pneumoniae Isolates. *Microorganisms* **2023**, *11*, doi:10.3390/microorganisms11051314.

14. Wyres, K.L.; Wick, R.R.; Judd, L.M.; Froumine, R.; Tokolyi, A.; Gorrie, C.L.; Lam, M.M.C.; Duchene, S.; Jenney, A.; Holt, K.E. Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of Klebsiella pneumoniae. *PLoS Genet* **2019**, *15*, e1008114, doi:10.1371/journal.pgen.1008114.
15. Diancourt, L.; Passet, V.; Verhoef, J.; Grimont, P.A.; Brisse, S. Multilocus sequence typing of Klebsiella pneumoniae nosocomial isolates. *J Clin Microbiol* **2005**, *43*, 4178-4182, doi:10.1128/JCM.43.8.4178-4182.2005.
16. Arbatsky, N.P.; Shneider, M.M.; Dmitrenok, A.S.; Popova, A.V.; Shagin, D.A.; Shelenkov, A.A.; Mikhailova, Y.V.; Edelstein, M.V.; Knirel, Y.A. Structure and gene cluster of the K125 capsular polysaccharide from Acinetobacter baumannii MAR13-1452. *Int J Biol Macromol* **2018**, *117*, 1195-1199, doi:10.1016/j.ijbiomac.2018.06.029.
17. Lam, M.M.C.; Wick, R.R.; Judd, L.M.; Holt, K.E.; Wyres, K.L. Kaptive 2.0: updated capsule and lipopolysaccharide locus typing for the Klebsiella pneumoniae species complex. *Microb Genom* **2022**, *8*, doi:10.1099/mgen.0.000800.
18. Makarova, K.S.; Koonin, E.V. Annotation and Classification of CRISPR-Cas Systems. *Methods Mol Biol* **2015**, *1311*, 47-75, doi:10.1007/978-1-4939-2687-9_4.
19. Hryhorowicz, M.; Lipinski, D.; Zeyland, J. Evolution of CRISPR/Cas Systems for Precise Genome Editing. *Int J Mol Sci* **2023**, *24*, doi:10.3390/ijms241814233.
20. Stasiak, M.; Mackiw, E.; Kowalska, J.; Kucharek, K.; Postupolski, J. Silent Genes: Antimicrobial Resistance and Antibiotic Production. *Pol J Microbiol* **2021**, *70*, 421-429, doi:10.33073/pjm-2021-040.
21. Schurch, A.C.; Arredondo-Alonso, S.; Willems, R.J.L.; Goering, R.V. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. *Clin Microbiol Infect* **2018**, *24*, 350-354, doi:10.1016/j.cmi.2017.12.016.
22. Lam, M.M.C.; Wick, R.R.; Watts, S.C.; Cerdeira, L.T.; Wyres, K.L.; Holt, K.E. A genomic surveillance framework and genotyping tool for Klebsiella pneumoniae and its related species complex. *Nat Commun* **2021**, *12*, 4188, doi:10.1038/s41467-021-24448-3.
23. Liu, C.; Wu, Y.; Fang, Y.; Sang, Z.; Huang, L.; Dong, N.; Zeng, Y.; Lu, J.; Zhang, R.; Chen, G. Emergence of an ST1326 (CG258) Multi-Drug Resistant Klebsiella pneumoniae Co-harboring mcr-8.2, ESBL Genes, and the Resistance-Nodulation-Division Efflux Pump Gene Cluster tmexCD1-toprJ1 in China. *Front Microbiol* **2022**, *13*, 800993, doi:10.3389/fmicb.2022.800993.
24. Rodrigues, C.; Desai, S.; Passet, V.; Gajjar, D.; Brisse, S. Genomic evolution of the globally disseminated multidrug-resistant Klebsiella pneumoniae clonal group 147. *Microb Genom* **2022**, *8*, doi:10.1099/mgen.0.000737.
25. Kocsis, B. Hypervirulent Klebsiella pneumoniae: An update on epidemiology, detection and antibiotic resistance. *Acta Microbiol Immunol Hung* **2023**, *70*, 278-287, doi:10.1556/030.2023.02186.
26. Wang, J.; Feng, Y.; Zong, Z. The Origins of ST11 KL64 Klebsiella pneumoniae: a Genome-Based Study. *Microbiol Spectr* **2023**, *11*, e0416522, doi:10.1128/spectrum.04165-22.
27. Wu, J.W.; Quyen, T.L.T.; Hsieh, Y.C.; Chen, Y.Y.; Wu, L.T.; Pan, Y.J. Investigation of carbapenem-resistant Klebsiella pneumoniae in Taiwan revealed strains co-harbouring bla(NDM) and bla(OXA-48-like) and a novel plasmid co-carrying bla(NDM-1) and bla(OXA-181). *Int J Antimicrob Agents* **2023**, *62*, 106964, doi:10.1016/j.ijantimicag.2023.106964.
28. Choi, M.; Hegerle, N.; Nkeze, J.; Sen, S.; Jamindar, S.; Nasrin, S.; Sen, S.; Permala-Booth, J.; Sinclair, J.; Tapia, M.D., et al. The Diversity of Lipopolysaccharide (O) and Capsular Polysaccharide (K) Antigens of Invasive Klebsiella pneumoniae in a Multi-Country Collection. *Front Microbiol* **2020**, *11*, 1249, doi:10.3389/fmicb.2020.01249.
29. Mills, R.O.; Dadzie, I.; Le-Viet, T.; Baker, D.J.; Addy, H.P.K.; Akwetey, S.A.; Donkoh, I.E.; Quansah, E.; Semanshia, P.S.; Morgan, J., et al. Genomic diversity and antimicrobial resistance in clinical Klebsiella pneumoniae isolates from tertiary hospitals in Southern Ghana. *J Antimicrob Chemother* **2024**, *79*, 1529-1539, doi:10.1093/jac/dkae123.
30. Mihailovskaya, V.S.; Selivanova, P.A.; Kuznetsova, M.V. Prevalence of qacEΔ1, qacE, oqxA, oqxB, acrA, cepA and zitB genes among multidrug-resistant Klebsiella pneumoniae isolated in a cardiac hospital. *Journal of microbiology, epidemiology and immunobiology* **2024**, *101*, 502-511, doi:10.36233/0372-9311-548.
31. Russo, T.A.; Carlino-MacDonald, U.; Drayer, Z.J.; Davies, C.J.; Alvarado, C.L.; Hutson, A.; Luo, T.L.; Martin, M.J.; McGann, P.T.; Lebreton, F. Deciphering the relative importance of genetic elements in hypervirulent Klebsiella pneumoniae to guide countermeasure development. *EBioMedicine* **2024**, *107*, 105302, doi:10.1016/j.ebiom.2024.105302.
32. Alkompoz, A.K.; Hamed, S.M.; Zaid, A.S.A.; Almangour, T.A.; Al-Agamy, M.H.; Aboshanab, K.M. Correlation of CRISPR/Cas and Antimicrobial Resistance in Klebsiella pneumoniae Clinical Isolates Recovered from Patients in Egypt Compared to Global Strains. *Microorganisms* **2023**, *11*, doi:10.3390/microorganisms11081948.
33. Kannadasan, A.B.; Sumantran, V.N.; Vaidyanathan, R. A Global Comprehensive Study of the Distribution of Type I-E and Type I-E* CRISPR-Cas Systems in Klebsiella pneumoniae. *Indian J Community Med* **2023**, *48*, 567-572, doi:10.4103/ijcm.ijcm_486_22.

34. Botelho, J.; Cazares, A.; Schulenburg, H. The ESKAPE mobilome contributes to the spread of antimicrobial resistance and CRISPR-mediated conflict between mobile genetic elements. *Nucleic Acids Res* **2023**, *51*, 236-252, doi:10.1093/nar/gkac1220.
35. Tang, Y.; Fu, P.; Zhou, Y.; Xie, Y.; Jin, J.; Wang, B.; Yu, L.; Huang, Y.; Li, G.; Li, M., et al. Absence of the type I-E CRISPR-Cas system in Klebsiella pneumoniae clonal complex 258 is associated with dissemination of IncF epidemic resistance plasmids in this clonal complex. *J Antimicrob Chemother* **2020**, *75*, 890-895, doi:10.1093/jac/dkz538.
36. Peirano, G.; Chen, L.; Kreiswirth, B.N.; Pitout, J.D.D. Emerging Antimicrobial-Resistant High-Risk Klebsiella pneumoniae Clones ST307 and ST147. *Antimicrob Agents Chemother* **2020**, *64*, doi:10.1128/AAC.01148-20.
37. Bortolaia, V.; Kaas, R.S.; Ruppe, E.; Roberts, M.C.; Schwarz, S.; Cattoir, V.; Philippon, A.; Allesoe, R.L.; Rebelo, A.R.; Florensa, A.F., et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* **2020**, *75*, 3491-3500, doi:10.1093/jac/dkaa345.
38. Liu, B.; Zheng, D.; Zhou, S.; Chen, L.; Yang, J. VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res* **2022**, *50*, D912-D917, doi:10.1093/nar/gkab1107.
39. Couvin, D.; Bernheim, A.; Toffano-Nioche, C.; Touchon, M.; Michalik, J.; Neron, B.; Rocha, E.P.C.; Vergnaud, G.; Gautheret, D.; Pourcel, C. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res* **2018**, *46*, W246-W251, doi:10.1093/nar/gky425.
40. Feijao, P.; Yao, H.T.; Fornika, D.; Gardy, J.; Hsiao, W.; Chauve, C.; Chindelevitch, L. MentaLiST - A fast MLST caller for large MLST schemes. *Microb Genom* **2018**, *4*, doi:10.1099/mgen.0.000146.
41. Egorova, A.; Mikhaylova, Y.; Saenko, S.; Tyumentseva, M.; Tyumentsev, A.; Karbyshev, K.; Chernyshkov, A.; Manzeniuk, I.; Akimkin, V.; Shelenkov, A. Comparative Whole-Genome Analysis of Russian Foodborne Multidrug-Resistant Salmonella Infantis Isolates. *Microorganisms* **2021**, *10*, doi:10.3390/microorganisms10010089.
42. Shelenkov, A.; Petrova, L.; Fomina, V.; Zamyatin, M.; Mikhaylova, Y.; Akimkin, V. Multidrug-Resistant Proteus mirabilis Strain with Cointegrate Plasmid. *Microorganisms* **2020**, *8*, doi:10.3390/microorganisms8111775.