FUTURE MICROBIOLOGY
2026, VOL. 21, NO. 8, 701–712
https://doi.org/10.1080/17460913.2026.2617118

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Pan-genome analysis and phylogenetic characterization of *Klebsiella pneumoniae* from global isolates

Hatoon A. Niyazi[a,b], Hanouf A. Niyazi[a], Hind AbdulMajed[a], Noha Juma[a], Noof Helmi[a], Mona Alqarni[a], Bandar Hasan Saleh[a,b], Manal Zubair[a], Abdelbagi Alfadil[a], Wafaa Alhazmi[c], Ohood S. Alharbi[d], Waiel S. Halabi[e], Rawan Altalhi[f], Ehssan Moglad[g], Mohanned Talal Alharbi[h], Malaz Gazzaz[i], Turki M. Alharthi[j], Hisham N. Altayb[k] and Karem Ibrahem[a,b]

[a]Department of Clinical Microbiology and Immunology, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia; [b]Department of Clinical Microbiology Laboratory, King Abdulaziz University Hospital, Jeddah, Saudi Arabia; [c]Department of Medical Laboratory Sciences, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia; [d]Department of Microbiology and Parasitology, Faculty of Medicine, Umm Al-Qura University, Makkah, Saudi Arabia; [e]Department of Optometry, Faculty of Applied Medical Sciences, University of Jeddah, Jeddah, Saudi Arabia; [f]Department of Biological Sciences, College of Science, University of Jeddah, Jeddah, Saudi Arabia; [g]Department of Pharmaceutics, College of Pharmacy, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia; [h]Department of Basic Medical Sciences, College of Medicine, University of Jeddah, Jeddah, Saudi Arabia; [i]Pharmaceutical Practices Department, College of Pharmacy, Umm Al-Qura University, Makkah, Saudi Arabia; [j]Department of Clinical Laboratory Sciences, Faculty of Applied Medical Sciences, Umm Al-Qura University, Makkah, Saudi Arabia; [k]Department of Biochemistry, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT

Aims: This study aimed to investigate the global genetic diversity, evolutionary relationships, and antimicrobial resistance (AMR) profiles of *Klebsiella pneumoniae* by performing a comprehensive pan-genome and phylogenetic analysis across worldwide isolates.

Materials and methods: A total of 72,057 *K. pneumoniae* genomes were retrieved from the NCBI database, from which 91 high-quality representative genomes each from a unique country were selected based on completeness, metadata availability, and sequence quality. Genomic assemblies were assessed using QUAST, annotated with PROKKA, and analyzed for pan-genomic composition and phylogenetic relatedness using standard bioinformatics pipelines.

Results: The pan-genome revealed a large accessory component, reflecting extensive genomic plasticity and adaptability. QUAST analysis indicated significant variability in genome size and contig number, while PROKKA annotation identified diverse coding sequences, tRNA, rRNA, and AMR genes. Phylogenetic clustering demonstrated both geographically localized and globally disseminated lineages, suggesting regional adaptation and intercontinental transmission.

Conclusions: This study provides a global perspective on the genomic diversity and evolutionary patterns of *K. pneumoniae*. The widespread presence of AMR determinants underscores the urgent need for continuous genomic surveillance and integration of metagenomic approaches to improve monitoring, infection control, and therapeutic strategies against multidrug-resistant strains.

ARTICLE HISTORY
Received 26 September 2025
Accepted 24 December 2025

KEYWORDS
K. pneumoniae; AMR; global genomic surveillance; genomic diversity; mobile genetic elements

1. Introduction

K. pneumoniae was primarily identified by Carl Friedlander in 1882, when isolating the encapsulated bacterium from the lungs of pneumonia patients [1]. Previously, it was known as Friedlander's bacillus. Later, it was renamed *K. pneumoniae* in 1886. This gram-negative, encapsulated, non-motile bacterium is commonly found in the environment and can cause pneumonia in individuals with alcohol abuse or diabetes [2]. As a bacillus from the Enterobacteriaceae family, *K. pneumoniae* is a common part of the human gut microbiome but has gained recognition as an invasive pathogen, especially in healthcare environment [3]. This bacterium causes numerous diseases including pneumonia, urinary tract infections, bloodstream infections, and liver abscesses, particularly affecting immunocompromised individuals [4]. Human beings are the main reservoir for *K. pneumoniae*, with 5% to 38% of the population carrying it in their feces and 1% to 6% in their nasopharynx. In healthcare facilities, the patient's gastrointestinal tract and the hands of medical personnel are primary reservoirs of infection, often leading to nosocomial outbreaks [2,5].

K. pneumoniae is known for its thick polysaccharide capsule, which not only contributes to its mucoid appearance but also plays a crucial role in evading host immune defenses [6]. *K. pneumoniae* exhibits significant capsular diversity, which contributes to its adaptability and pathogenicity in various host environments [7]. In addition to its increasing drug resistance.

The therapeutic importance of *K. pneumoniae* has significantly increased in past few years owing to its exceptional capacity to obtain antibiotic resistance mechanisms. The increase of carbapenem-resistant *K. pneumoniae* (CRKP) strains is particularly concerning due to their limited treatment options and high

CONTACT Karem Ibrahem kaibrahem@kau.edu.sa Department of Clinical Microbiology and Immunology, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia

Article highlights

- Analyzed 72,057 *Klebsiella pneumoniae* genomes, selecting 91 representative genomes for global phylogenetic comparison.
- Included only high-quality, complete assemblies to ensure analytical accuracy.
- Pan-genome analysis revealed extensive genomic plasticity and widespread antimicrobial resistance genes.
- Phylogenetic study showed both regional clustering and global spread, including post-COVID-19 hypervirulent and carbapenem-resistant strains.
- Highlights the importance of integrating genomic and clinical surveillance to track resistance and inform global infection control strategies.

mortality rates [8]. The study by Sanchez et al. [9] finds that the surveillance from 1998–2010 total of 3,132,354 susceptibility findings showed significant increases in resistance to most antimicrobial agents, except tetracycline, indicating a concerning trend in medicine efficacy against the virus. This suggested cross resistance among CRKP, which was found to be lower for tetracycline and amikacin [9]. Additionally, Li et al. [10] showed pathogenic *K. pneumoniae* isolates, including hypermucoviscous and specific serotypes, with potential drug resistance and emergence of a carbapenem-resistant hypervirulent strain [10]. Recent reports have highlighted a global rise in CRKP) following the COVID-19 pandemic, indicating increased hospital transmission and antimicrobial pressure across regions [11]. In parallel, the emergence of hypervirulent *K. pneumoniae* (hvKP) strains, characterized by enhanced virulence and poor clinical outcomes, has become a major global concern [12]. The convergence of antimicrobial resistance (AMR) and hypervirulence in *K. pneumoniae* poses a critical threat to infection control and therapeutic management, emphasizing the need for genomic surveillance and early detection strategies.

In recent years, pan-genome analysis has emerged as a crucial tool for comprehending the genomic structure of bacterial species like *K. pneumoniae* [13]. This approach allows researchers to distinguish between core genes, which are common to all strains, and accessory genes, which are specific to individual isolates and often confer adaptive advantages such as antibiotic resistance or enhanced virulence [14]. These studies suggested the need for genetic insight, which is where pan-genome analysis becomes valuable.

As multidrug-resistant *K. pneumoniae* clones continue to spread globally, there is growing concern about their impact. A comprehensive study is crucial to understand this threat. Investigating the genetic diversity and population structure of the pathogen worldwide reveals important details about antibiotic resistance, virulence, and its ability to adapt to different environments. Therefore, the current study aimed to conduct a comprehensive pan-genome analysis and construct a phylogenetic tree of *K. pneumoniae* strains from diverse geographical regions. By analyzing 91 genomes, each representing a distinct country, the research pursued to investigate the genomic heterogeneity and evolutionary relationships within this species on a global scale. The study employed a suite of bioinformatics tools, for genome assembly quality assessment and functional annotation, and Roary for pan-genome analysis.

2. Materials and methods

2.1. Retrieval of K. pneumoniae genomes

The genomic data for *K. pneumoniae* strains were obtained from the National Centre for Biotechnology Information (NCBI) database [15]. The metadata of 72,057 *K. pneumoniae* genomes retrieved included accession number, assembly name, assembly type, collection date, geographic location, and total sequence length. These records were systematically filtered to ensure the inclusion of only high-quality and representative genomes. The filtering criteria comprised the following: (i) only assemblies labeled as "haploid" or "complete genome" were considered; (ii) genomes with abnormally short or excessively long total sequence lengths compared to the expected *K. pneumoniae* genome size (5.2–5.7 Mbp) were excluded; (iii) the most recent collection date from each country was selected to represent current genetic diversity; (iv) genomes lacking valid geographic metadata (geo_loc_name) or collection year were excluded; and (v) duplicates or assemblies with incomplete metadata were removed. After applying these filters, 91 high-quality genomes, each representing a unique country, were retained for subsequent comparative genomic and phylogenetic analyses.

2.2. Genome quality assessment using QUAST

The quality of the assembled genomes was evaluated using the Quality Assessment Tool for Genome Assemblies (QUAST) [16]. This approach ensured that the selected genomes that will be used for further investigation were high in quality, free from contamination, and exhibited acceptable assembly metrics. QUAST provided critical parameters such as N50 values, total genome length, contig count, and GC content, facilitating the identification of low-quality genomes to ensure the reliability of comparative genomic analyses.

2.3. Genome annotation with PROKKA

Genome annotation is essential for identifying coding sequences (CDS) and genomic structure of the studied genomes. The selected genomes were annotated using the PROKKA program [17]. The selected genomes were annotated using the PROKKA tool, which identified protein-coding genes, ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), and other genomic elements. The PROKKA results provided data on the number of CDS, ribosomal RNA genes, tRNA genes, and overall genomic characteristics for each *K. pneumoniae* strain."

The AMR gene profiles were derived directly from the annotation results generated by the PROKKA pipeline. PROKKA incorporates reference datasets such as the ResFinder and ARG-ANNOT databases to annotate known resistance determinants. The AMR-related annotations obtained through PROKKA were filtered to retain only those with high-confidence functional assignments. Each genome's annotation file was systematically parsed to extract AMR-

related coding sequences (CDS), and the total count of these genes was compiled for comparative analysis. Although the AMR gene prediction relied on computational annotation rather than re-analysis with recent resistance databases, it provided a consistent framework for assessing the relative abundance and diversity of AMR determinants across the 91 genomes.

2.4. Pangenome analysis with roary

A pangenomic analysis was performed to identify the core and accessory genes present in the 91 genomes. The roary pangenome program was utilized to classify genes into orthologous groups across the genomes [18]. Roary generated a gene presence/absence matrix that outlined the distribution of gene clusters among all strains, enabling the categorization of genes as either core (present in all genomes) or accessory (present in a subset of genomes). Phylogenetic analysis relied on the genomic alignment that Roary generated.

2.5. Phylogenetic analysis and tree construction

A phylogenetic analysis was conducted using the core genomic alignment produced by Roary to investigate the evolutionary relationships among 91 *K. pneumoniae* genomes. The phylogenetic tree was constructed using FastTree [19], which employs a maximum likelihood approach to infer evolutionary relationships from large genomic datasets. The whole phylogenetic tree was illustrated to clearly represent the evolutionary relationships among the *K. pneumoniae* genomes. The Interactive Tree of Life (iTOL) program [20] facilitated detailed customization and annotation for visualization purposes. The phylogenetic tree was arranged in a circular format to allow for passable viewing of all 91 strains. Each strain was labeled with its genomic accession number and country of origin.

3. Results

3.1. Retrieval of K. pneumoniae genomes

The present study retrieved a total of 72,057 genomes of *K. pneumoniae*, that were primarily obtained from the NCBI database. After collecting these data, the data were sorted as per the most current sample from each nation. The mentioned process resulted in the further selection of 91 genomes, each representing a different country. These 91 genomes were listed in Table 1 based on their collection date and the country of occurrence. The genomes exhibited extensive geographic distribution, including countries from Africa, Asia, Europe, North America, South America, and Oceania. Countries such as the United States, United Kingdom, China, and India were prominently represented, demonstrating the global distribution of *K. pneumoniae*. Figure 1 illustrates a world map used to visually convey the global distribution of the selected genomes, focusing on the countries that contributed to the genome dataset. Additionally, the map demonstrates the widespread occurrence of *K. pneumoniae* and provides a foundation for subsequent genetic analysis. The large set of genomes gathered from various continents ensure a thorough understanding of the pathogen's genetic variance, simplifying further examination of genomic diversity and geographic clustering patterns.

An earlier study showed that a total of 245 *K. pneumoniae* genomes found positive for pks (clb gene island) [21] were found among 2709 genomes [22]. Out of these, 157 were sourced from the China National GeneBank Database (CNGBdb) [23] and 88 were found in the public database NCBI. Similarly, another study retrieved data from NCBI and obtained 373 full genomes and chromosomes for core genome extraction [24]. A total of 39 common *K. pneumoniae* strains were included in the small group sample for phylogenetic analysis. This sample included strains exhibiting pathogenicity, nitrogen fixing skills, and potential for industrial use. Another previous work found 157 lineages using whole-genome sequencing of a broad collection of 289 Kp genomes [25]. The public cgMLST database has 155 established CGs. The detection rate of new lineages indicates that the true number of existing lineages is probably in the thousands, greatly exceeding the established count. Thus, in this current study the genomes of *K. pneumoniae* were collected from various nations to provide significant contemporary, significant insights on the worldwide and evolutionary spread of *K. pneumoniae*.

3.2. Assembly quality assessment

The genome assembly qualities were evaluated by employing QUAST to analyze vital metrics, including total genome length, GC content, N50 values, and the number of contigs exceeding 1000 bp. The outcome of the QUAST observed in Table 2 highlights the overall quality and consistency of the assemblies for the 91 *K. pneumoniae* genomes. The total genome lengths ranged from 5.4 Mbp to 5.5 Mbp, consistent with the expected size for *K. pneumoniae*. The GC content among all genomes was uniform, ranging from 55.5% to 57.5%, with the majority of values concentrated around 57%, a trait typical of this species. This uniformity in GC content is a hallmark of this species' genomic stability and confirms the reliability of the assemblies. The average GC content of *K. pneumoniae* was found to be around 58% in a prior study [24]. While the current study found a slightly lower GC content, it remains within the typical range observed for this species, accounting for possible strain variation.

Figure 2 shows a heatmap depicting the distribution of GC content across genomes, indicating slight variability among genomes from different nations, while the Belarus strain shows a marginally reduced GC content. Moreover, N50 values, reflecting assembly continuity, varied throughout the genomes, primarily between 200,000 bp and 300,000 bp. Certain assemblies, particularly from countries like as Switzerland and Sudan, had greater N50 values, signifying improved continuity and less fragmented contigs. Figure 2 illustrates the N50 distribution throughout the genomes, emphasizing the disparities in assembly quality among the different datasets. Lower N50 values, represented by purple, indicate incomplete assemblies. Whereas higher N50 values,

Table 1. Summary of *Klebsiella pneumoniae* genome retrieval by country and collection date from NCBI.

Accession	Assembly	Collection date	Country	Sequence Length
GCA_030845575.1	ASM3084557v1	2018	Afghanistan	5737974
GCF_006151695.1	ASM615169v1	2018	Algeria	5823864
GCF_037449135.1	ASM3744913v1	2022	Argentina	5540031
GCA_030294165.1	ASM3029416v1	2021	Australia	5594801
GCA_003583095.1	ASM358309v1	2016	Austria	5696674
GCF_036235415.1	ASM3623541v1	2023	Bangladesh	5361640
GCA_020005265.1	ASM2000526v1	2020	Belarus	5425462
GCF_036950405.1	ASM3695040v1	2019	Belgium	5456468
GCA_038433245.1	ASM3843324v1	2023	Brazil	5567630
GCF_963854595.1	KP31122023	2023	Bulgaria	5820489
GCA_022175105.1	PDT000750133.1	2017	Cambodia	5316747
GCA_026323435.1	PDT001495069.1	2020	Canada	5698674
GCF_032253935.1	ASM3225393v1	2022	Chile	5615879
GCA_041023815.1	ASM4102381v1	2024	China	5764993
GCF_036668115.1	ASM3666811v1	2023	Colombia	5735378
GCA_028614935.1	ASM2861493v1	2020	Croatia	4772599
GCA_016939815.1	ASM1693981v1	2019	Czech Republic	5698780
GCA_041020085.1	ASM4102008v1	2020	Czechia	5629108
GCA_028994575.1	ASM2899457v1	2023	Denmark	6051092
GCA_029955815.1	ASM2995581v1	2022	Ecuador	5351959
GCA_031174725.1	PDT001873519.1	2021	Egypt	5774152
GCA_024479345.1	PDT001370401.1	2020	Ethiopia	5681276
GCF_040137595.1	ASM4013759v1	2023	France	5509846
GCA_027908695.1	PDT001551536.1	2022	Germany	5892959
GCA_022968285.1	ASM2296828v1	2021	Ghana	5666331
GCA_041361325.1	ASM4136132v1	2024	Greece	5765284
GCA_024376135.1	PDT001366188.1	2018	Guadeloupe	5510566
GCA_022316305.1	PDT000111518.2	2012	Honduras	5811065
GCF_037167165.1	ASM3716716v1	2023	Hong Kong	5549664
GCF_032807725.1	ASM3280772v1	2023	Hungary	5743644
GCF_041679105.1	ASM4167910v1	2024	India	5530989
GCF_019433925.1	ASM1943392v1	2019	Iran	5680506
GCA_017916215.1	ASM1791621v1	2015	Iraq	5201616
GCA_022194705.1	PDT000715124.1	2011	Ireland	5765626
GCA_024499835.1	ASM2449983v1	2022	Israel	5636227
GCA_036421895.1	ASM3642189v1	2023	Italy	5707905
GCA_030878345.1	PDT001701438.1	2020	Japan	5519986
GCA_037053885.1	ASM3705388v1	2017	Jordan	5561739
GCA_019915325.1	ASM1991532v1	2021	Kazakhstan	5576938
GCA_037628085.1	ASM3762808v1	2022	Kenya	5431963
GCA_036881205.1	ASM3688120v1	2016	Kuwait	5552261
GCA_022207725.1	PDT000651023.1	2015	Laos	5272122
GCF_032676305.1	ASM3267630v1	2023	Lebanon	5304825
GCA_026329135.1	PDT001494582.1	2020	Lithuania	5630027
GCA_041363805.1	ASM4136380v1	2022	Madagascar	5368848
GCA_019877925.1	ASM1987792v1	2016	Malawi	5525629
GCF_036903135.1	ASM3690313v1	2021	Malaysia	5404900
GCF_037104535.1	ASM3710453v1	2023	Mexico	5410133
GCA_900514015.1	18174_5#79	2013	Montenegro	5358822
GCA_026332835.1	PDT001494626.1	2020	Mozambique	5610601
GCA_022283225.2	PDT000431898.2	2016	Myanmar	5601929
GCA_041061405.1	ASM4106140v1	2023	Nepal	5965341
GCF_041023835.1	ASM4102383v1	2023	Netherlands	5413684
GCA_031051475.1	PDT001716702.1	2015	New Zealand	5471831
GCA_026326255.1	PDT001494799.1	2020	Nigeria	5203566
GCA_041863025.1	ASM4186302v1	2018	Norway	5543151
GCA_021875435.1	PDT001197729.1	2020	Oman	6154406
GCF_032842255.1	9055	2023	Pakistan	5287542
GCA_026305215.1	PDT001493239.1	2020	Paraguay	5421801
GCF_040687995.1	ASM4068799v1	2021	Peru	5898946
GCA_002108355.1	ASM210835v1	2013	Philippines	5454384
GCA_037082035.1	ASM3708203v1	2021	Poland	5522263
GCF_032461365.1	ASM3246136v1	2022	Portugal	5550563
GCA_027894895.1	PDT001557824.1	2021	Qatar	5529043
GCA_009710645.1	ASM971064v1	2019	Romania	5943681
GCF_040539535.1	ASM4053953v1	2024	Russia	5730640
GCA_037206085.1	ASM3720608v1	2023	Saudi Arabia	5722556
GCF_030414195.1	ASM3041419v1	2018	Senegal	5551901
GCF_029848135.1	ASM2984813v1	2017	Singapore	5157567
GCA_964254225.1	24–2119	2024	Slovakia	5471848
GCF_037004625.1	ASM3700462v1	2022	South Africa	5740521
GCF_033898565.1	ASM3389856v1	2022	South Korea	5389891
GCF_040763125.1	ASM4076312v1	2022	Spain	5592654
GCF_017898205.1	ASM1789820v1	2015	Sri Lanka	5444822

(*Continued*)

Table 1. (Continued).

Accession	Assembly	Collection date	Country	Sequence Length
GCA_022530725.1	ASM2253072v1	2021	Sudan	5675175
GCF_029929895.1	ASM2992989v1	2021	Sweden	6051072
GCF_023612395.1	ASM2361239v1	2019	Switzerland	5762752
GCA_039724115.1	ASM3972411v1	2022	Taiwan	5515513
GCA_009925385.1	ASM992538v1	2018	Tanzania	5481532
GCF_024705635.1	ASM2470563v1	2021	Thailand	5761774
GCF_025819055.1	ASM2581905v1	2016	Trinidad and Tobago	5592304
GCF_002140135.1	ASM214013v1	2015	Tunisia	5607827
GCA_023026145.1	PDT001289577.1	2019	Turkey	5688171
GCA_033336495.1	ASM3333649v1	2022	Uganda	5181822
GCF_009887895.1	ASM988789v1	2014	Ukraine	5637323
GCA_014433645.1	ASM1443364v1	2013	United Arab Emirates	6719317
GCA_033253045.1	ASM3325304v1	2023	United Kingdom	5923719
GCA_024612745.1	PDT001382059.1	2017	Uruguay	5572642
GCA_041857965.1	PDT002366229.1	2024	USA	5637052
GCA_041400745.1	ASM4140074v1	2021	Viet Nam	5765774
GCA_026303475.1	PDT001492934.1	2020	West Bank	5390763

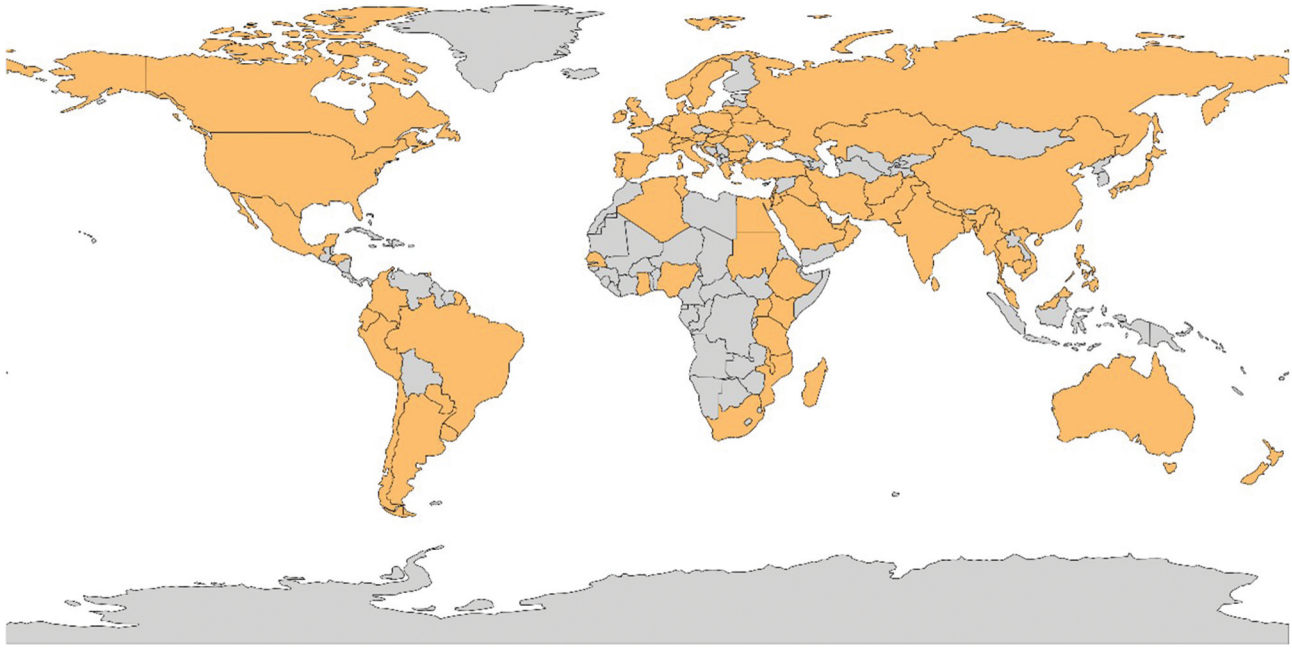

Figure 1. The most recent genomic data for each place was obtained from the 91 nations listed in orange. The global impact of the virus is underscored by this visualization, which displays the extensive dispersion of *Klebsiella pneumoniae* across varied geographic locations, spanning multiple continents.

represented by yellow, indicate more continuous and complete assemblies. The N50 heatmap indicates that, despite a few significant outliers, most genomes have been assembled to a high grade. Genomes with lower N50 values may require additional sequencing or refinement to improve continuity.

The genome assemblies exhibit high quality, marked by minimal fragmentation and stable GC content, thereby enabling their use in forthcoming genomic studies. The heatmaps for GC content and N50 values (Figure 2) provide a visual comparison of these metrics across the different countries in the dataset, demonstrating the robustness and consistency of the assemblies. The amount of GC impacts gene expression, DNA replication, and genomic stability. Genome values in this study are consistent, indicating good quality of assembly. It suggested that there is no contamination or sequencing mistake, which can be caused by variations in GC content.

3.3. Functional annotation

The 91 genomes of *K. pneumoniae* were used for annotation following the quality assessment of the genomes. PROKKA annotation tool was used to perform the functional annotation for listing the coding sequences (CDS), tRNA, rRNA, and other genomic features. This study was essential in providing detailed insights into the functional abilities and genetic makeup of the engineered genomes. A comprehensive set of annotations was generated for each genome, including the identification of various genomic features including hypothetical proteins, antibiotic resistance genes, and mobile genetic elements.

Table 2. Genome Assembly Statistics from QUAST Analysis for 91 *Klebsiella pneumoniae* Genomes.

Accession	Assembly	Country	Total length	contigs	GC (%)	N50	L50
GCA_030845575.1	ASM3084557v1	Afghanistan	5737974	4	56.94	5340491	1
GCF_006151695.1	ASM615169v1	Algeria	5823864	170	56.86	122891	15
GCF_037449135.1	ASM3744913v1	Argentina	5530195	5	57.23	154589	5
GCA_030294165.1	ASM3029416v1	Australia	5594801	3	57.14	5420112	1
GCA_003583095.1	ASM358309v1	Austria	5617365	223	57.2	357940	5
GCF_036235415.1	ASM3623541v1	Bangladesh	5356714	80	57.18	371239	6
GCA_020005265.1	ASM2000526v1	Belarus	4594873	3074	54.77	1909	651
GCF_036950405.1	ASM3695040v1	Belgium	5456468	91	57.1	191227	12
GCA_038433245.1	ASM3843324v1	Brazil	5566354	106	57.28	141399	14
GCF_963854595.1	KP31122023	Bulgaria	5352450	6	57.15	5589618	1
GCA_022175105.1	PDT000750133.1	Cambodia	5316747	84	57.46	198055	8
GCA_026323435.1	PDT001495069.1	Canada	5693556	250	57.08	49481	34
GCF_032253935.1	ASM3225393v1	Chile	5615879	3	56.85	5334759	1
GCA_041023815.1	ASM4102381v1	China	5755120	116	57.01	160332	13
GCF_036668115.1	ASM3666811v1	Colombia	5726944	103	57.11	129808	15
GCA_028614935.1	ASM2861493v1	Croatia	4513170	1292	56.47	4726	281
GCA_016939815.1	ASM1693981v1	Czech Republic	5698780	5	57.17	5312419	1
GCA_041020085.1	ASM4102008v1	Czechia	5620465	66	57.09	374649	6
GCA_028994575.1	ASM2899457v1	Denmark	6051092	4	56.51	5410656	1
GCA_029955815.1	ASM2995581v1	Ecuador	5344542	36	57.45	322222	6
GCA_031174725.1	PDT001873519.1	Egypt	5768371	259	56.98	51822	34
GCA_024479345.1	PDT001370401.1	Ethiopia	5668320	147	57.22	94039	18
GCF_040137595.1	ASM4013759v1	France	5509590	133	57.13	5235716	1
GCA_027908695.1	PDT001551536.1	Germany	5889652	112	56.78	186206	11
GCA_022968285.1	ASM2296828v1	Ghana	5661211	86	56.74	205953	9
GCA_041361325.1	ASM4136132v1	Greece	5757733	91	57.02	180988	11
GCA_024376135.1	PDT001366188.1	Guadeloupe	5510566	68	57.2	213862	9
GCA_022316305.1	PDT000111518.2	Honduras	5811065	97	57.18	220631	10
GCF_037167165.1	ASM3716716v1	Hong Kong	5546161	120	56.78	355645	16
GCF_032807725.1	ASM3280772v1	Hungary	5743644	2	57.19	5515480	1
GCF_041679105.1	ASM4167910v1	India	5525870	76	57.28	227165	6
GCF_019433925.1	ASM1943392v1	Iran	5680506	100	56.85	157973	12
GCA_017916215.1	ASM1791621v1	Iraq	5201616	1	57.52	5201616	1
GCA_022194705.1	PDT000715124.1	Ireland	5763402	100	57.15	224849	8
GCA_024499835.1	ASM2449983v1	Israel	5633714	80	56.84	252604	8
GCA_036421895.1	ASM3642189v1	Italy	5700989	80	56.76	175537	11
GCA_030878345.1	PDT001701438.1	Japan	5517065	63	57.23	372708	6
GCA_037053885.1	ASM3705388v1	Jordan	5554088	99	57.11	122560	14
GCA_019915325.1	ASM1991532v1	Kazakhstan	5576938	3	57.2	5319600	1
GCA_037628085.1	ASM3762808v1	Kenya	5426547	118	57.35	100849	18
GCA_036881205.1	ASM3688120v1	Kuwait	5540385	69	57.1	355747	5
GCA_022207725.1	PDT000651023.1	Laos	5272122	115	57.44	115752	14
GCF_032676305.1	ASM3267630v1	Lebanon	5290730	27	57.02	502078	4
GCA_026329135.1	PDT001494582.1	Lithuania	5629674	77	57.21	214696	9
GCA_041363805.1	ASM4136380v1	Madagascar	5367643	395	57.28	21983	77
GCA_019877925.1	ASM1987792v1	Malawi	5507460	157	57.06	100621	16
GCF_036903135.1	ASM3690313v1	Malaysia	5400979	63	57.24	226052	9
GCF_037104535.1	ASM3710453v1	Mexico	5405867	38	57.39	347819	12
GCA_900514015.1	18174	Montenegro	5820489	6	57.28	345343	5
GCA_026332835.1	PDT001494626.1	Mozambique	5591568	407	57.3	29478	59
GCA_022283225.2	PDT000431898.2	Myanmar	5600218	68	57.3	414034	6
GCA_041061405.1	ASM4106140v1	Nepal	5965341	8	56.86	5467994	1
GCF_041023835.1	ASM4102383v1	Netherlands	5410997	97	56.92	191072	10
GCA_031051475.1	PDT001716702.1	New Zealand	5470173	78	57.47	260669	8
GCA_026326255.1	PDT001494799.1	Nigeria	5200319	196	57.63	48171	36
GCA_041863025.1	ASM4186302v1	Norway	5543151	5	57.01	5314429	1
GCA_021875435.1	PDT001197729.1	Oman	6154406	146	56.57	146525	15
GCF_032842255.1	9055	Pakistan	5280415	67	57.66	174292	11
GCA_026305215.1	PDT001493239.1	Paraguay	5417522	181	57.34	55416	32
GCF_040687995.1	ASM4068799v1	Peru	5898946	3	57.3	5456495	1
GCA_002108355.1	ASM210835v1	Philippines	5454384	79	57.07	225222	9
GCA_037082035.1	ASM3708203v1	Poland	5522263	4	57.33	5242798	1
GCF_032461365.1	ASM3246136v1	Portugal	5539426	98	57.13	167970	11
GCA_027894895.1	PDT001557824.1	Qatar	5521710	123	57.25	104964	16
GCA_009710645.1	ASM971064v1	Romania	5898537	150	56.56	216943	8
GCF_040539535.1	ASM4053953v1	Russia	5730640	8	57.26	119523	16
GCA_037206085.1	ASM3720608v1	Saudi Arabia	5716691	87	56.8	287088	8
GCF_030414195.1	ASM3041419v1	Senegal	5551901	2	57.16	5270056	1
GCF_029848135.1	ASM2984813v1	Singapore	5153501	25	56.86	487783	4
GCA_964254225.1	24–2119	Slovakia	5468213	105	57.16	137705	15
GCF_037004625.1	ASM3700462v1	South Africa	5717595	60	57.24	203544	9
GCF_033898565.1	ASM3389856v1	South Korea	5389891	32	56.52	722111	4
GCF_040763125.1	ASM4076312v1	Spain	5592654	90	56.67	5263212	1
GCF_017898205.1	ASM1789820v1	Sri Lanka	5439636	56	57.16	362475	6

(*Continued*)

Table 2. (Continued).

Accession	Assembly	Country	Total length	contigs	GC (%)	N50	L50
GCA_022530725.1	ASM2253072v1	Sudan	5675175	1	57.2	5675175	1
GCF_029929895.1	ASM2992989v1	Sweden	6046043	32	57.09	5106042	1
GCF_023612395.1	ASM2361239v1	Switzerland	5762752	7	57.09	5372651	1
GCA_039724115.1	ASM3972411v1	Taiwan	5515513	2	57.07	5267549	1
GCA_009925385.1	ASM992538v1	Tanzania	5477594	66	57.32	405535	5
GCF_024705635.1	ASM2470563v1	Thailand	5761774	6	57.16	5422063	1
GCF_025819055.1	ASM2581905v1	Trinidad and Tobago	5579203	70	56.94	270013	9
GCF_002140135.1	ASM214013v1	Tunisia	5580021	745	56.94	13248	112
GCA_023026145.1	PDT001289577.1	Turkey	5683861	80	56.99	284057	8
GCA_033336495.1	ASM3333649v1	Uganda	5173683	29	57.59	362309	6
GCF_009887895.1	ASM988789v1	Ukraine	5633430	60	57.09	272065	8
GCA_014433645.1	ASM1443364v1	United Arab Emirates	6701960	269	55.96	225143	8
GCA_033253045.1	ASM3325304v1	United Kingdom	5923719	5	56.57	5422745	1
GCA_024612745.1	PDT001382059.1	Uruguay	5572198	103	57.52	119763	14
GCA_041857965.1	PDT002366229.1	USA	5636130	67	57.16	264472	8
GCA_041400745.1	ASM4140074v1	Viet Nam	5762568	18	57.02	5267471	1
GCA_026303475.1	PDT001492934.1	West Bank	5390763	148	57.26	60652	28

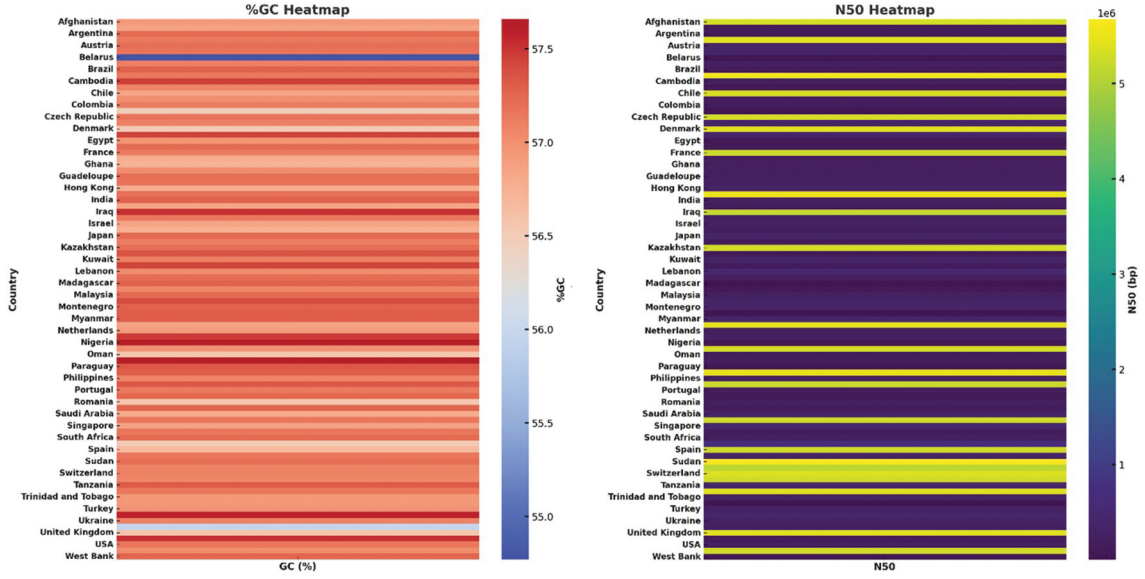

Figure 2. (a) Heatmaps of %GC Content and (b) N50 Across 91 *Klebsiella pneumoniae* Genomes.

Table 3. Summary of Prokka annotations across 91 *Klebsiella pneumoniae* Genomes.

Feature	Minimum	Maximum	Average
Total Coding Sequences (CDS)	4800	5500	5000
Transfer RNA (tRNA)	60	85	73
Ribosomal RNA (rRNA)	3	5	4
Antimicrobial Resistance Genes (AMR)	10	60	35
Mobile Genetic Elements (Plasmids, IS)	2	8	4

Various *K. pneumoniae* strains exhibit a considerable level of heterogeneity in gene content, as indicated by the range of CDS between 4800 and 5500 shown in Table 3. A genome of *K. pneumoniae* typically contains around 5000 CDS, suggesting that most strains share a similar gene content with some diversity caused by accessory genes like virulence or AMR genes. A typical bacterial genome generally contains 60 to 85 transfer RNA (tRNA) genes, which are crucial for protein translation. These genomes likely contain an extensive array of tRNA genes, with an average count of 73 tRNA genes. This diversity is important for enabling translation in different environments. Similar to other bacterial species, *K. pneumoniae* generally has between 3 and 5 rRNA genes spread out over its genome.

The quantity of AMR genes exhibits significant variability, ranging from 10 to 60, with a mean of 35 AMR genes per genome. The significant variation indicates the different resistance profiles of *K. pneumoniae* strains, which are recognized for obtaining and accumulating resistance genes via horizontal gene transfer. Mobile genetic elements, including plasmids and insertion sequences (IS), serve as crucial mechanisms for horizontal gene transfer, promoting the dissemination of resistance genes and virulence factors. The average of 4 mobile genetic components indicates a moderate degree of genomic plasticity, with strains possibly able to obtain new characteristics through horizontal gene transfer. The findings demonstrate variation in gene composition across the genomes, underlining the diversity of both the core and accessory genomes in *K. pneumoniae*. Each genome had around 5000

coding sequences, indicating variations in genome size and the number of gene annotations.

A previous study indicated that the conserved genome across these clinical strains contained 4558 coding sequences (CDS), exceeding the core content identified in larger collections of *K. pneumoniae* strains, which is roughly 1700 CDS [26]. An independent investigation of the assembly of *K. pneumoniae* UABC-Str0120 yielded 6,057,252 base pairs across 524 contigs, with a GC content of 56.6%. The genome consists of 6290 coding sequences, including 6014 coding sequences (CDS), 3 rRNAs, 48 tRNAs, 9 non-coding RNAs, and 216 pseudogenes [27].

3.4. Phylogenetic tree and evolutionary relationships

Figure 3 illustrates the phylogenetic tree, depicting the evolutionary relationships among strains based on genetic similarities. The clustering of closely related strains suggests a recent common ancestor, while distantly related strains on separate branches exhibit higher evolutionary divergence. Comprehending the global evolution of *K. pneumoniae* can be improved by analyzing the evolutionary distances among the genomes illustrated by the phylogenetic tree.

A diverse range of countries from Europe, Asia, the Middle East, Africa, and North America are represented in this clade of *K. pneumoniae*. Examples of these nations are Sweden, Norway, the Netherlands, and Poland. Additional countries encompass Saudi Arabia, Afghanistan, Sri Lanka, Kuwait, and Laos. The distribution of closely related strains across continents indicates international transmission or the global spread of particular lineages of *K. pneumoniae*. This could be due to international travel, trade, or medical tourism.

The existence of strains from diverse locations within the same clade over multiple branches indicates a recent common ancestor and perhaps transnational transmission pathways. The coexistence of strains from the United States, Malawi, and Norway within the same lineage indicates intercontinental transmission. Another clade, comprising strains from Sri Lanka, Kuwait, and Laos, suggests close genetic linkages potentially resulting from regional transmission in Asia. Similarly, the clustering of strains from Australia and New Zealand may indicate a common evolutionary lineage

Figure 3. Circular phylogenetic tree showing the evolutionary relationships of 91 Klebsiella pneumoniae genomes from different countries. The strains are labeled by genome accession numbers and their corresponding country of origin.

or regional transmission between these adjacent nations. Considering the geographical closeness of Argentina, Brazil, and Uruguay, together with the existence of analogous strains within the same lineage, it is conceivable that regional transmission has transpired throughout South America. Closely related strains may spread more easily between these nations because of shared healthcare practices or commercial routes.

The familiarity of strains from Qatar, Switzerland, Singapore, and Portugal indicates possible global transmission, particularly if these strains include similar traits like antibiotic resistance genes or plasmids that have permitted their dissemination. Qatar, Singapore, and Switzerland serve as epicenters of global travel and healthcare, potentially simplifying the transmission of these strains between nations.

The inclusion of Iran, Oman, Trinidad & Tobago, and Kenya in an adjacent part of the tree implies potential international transmission, maybe through air travel or other global networks. The close genetic relationships among strains from Oman, Brazil, and China are examples of potential clonal proliferation of a geographically spread lineage. A multidrug-resistant clone may have spread across continents if these strains of bacteria share resistance genes. Another possible explanation for the clustering of strains from Argentina, Colombia, and Peru is regional clonal growth, facilitated by shared healthcare systems or regional trade.

Strains from Poland, France, and the West Bank are genetically close to each other, which could indicate strong evolutionary relationship and regional or even transcontinental transmission, particularly in areas with extensive trade or healthcare infrastructure. It is possible that this bacterium has spread internationally due to the genetic similarity between strains from Europe (Germany, Slovakia, and Denmark) and Africa (Uganda, Senegal, and Kenya). A particular lineage may have spread throughout Europe, as indicated by the short branch lengths noted among strains from Germany, Russia, and Denmark. This might indicate a common infectious agent or a recent evolutionary split that has reached different parts of the area, maybe affecting medical facilities. The similarity between strains from Mexico, Lebanon, and South Korea suggests the possibility of a new lineage that has traversed continents. *K. pneumoniae* is a common cause of hospital-acquired infections (HAIs), hence its spread could be associated with healthcare institutions or foreign travel.

There may be evidence of shared evolutionary history or cross-border transmission among strains from France, Poland, and the West Bank, as they cluster closely together. Similar healthcare procedures or regional trade linkages may be driving the circulation of this common lineage within European and Middle Eastern hospitals. The clustering of Sri Lanka, Afghanistan, and Saudi Arabia may indicate regional transmission throughout Asia and the Middle East. The genetic closeness of these strains could be explained by common treatment practices or trading routes. A clonal lineage may have emerged and disseminated throughout these far regions, as there are significant genetic connections across strains from Mexico, South Korea, and Lebanon. This suggested global spread of a multidrug-resistant lineage if these bacteria exhibit common resistance characteristics. Strains found in the United States, Malawi, and Bulgaria raise the likelihood of global transmission. It showed roots in healthcare facilities that treat patients with antibiotic-resistant diseases.

Previous studies have shown that *K. pneumoniae* strains that have resistance genes to carbapenem and blaNDM pose an endangered public health risk [28]. The treatment of individuals infected by these strains is difficult since they are hypervirulent and highly resistant. One study documented the global evolutionary spread of these strains, highlighting major epidemics in South America, Asia, and Europe [29], with strains closely resembling those observed in the current study. Another study found that some *K. pneumoniae* strains have evolved to be both resistant and hypervirulent [30]. Strains from different parts of the world, including China, Saudi Arabia, and Brazil, displayed strong genetic links, suggesting clonal proliferation, in this study. This aligns with the trends shown in our phylogenetic tree, especially for regionally distributed strains.

Genome Medicine conducted a comprehensive investigation of over 13,000 *K. pneumoniae* isolates from 103 distinct countries. The study primarily concentrated on the global distribution of particular sequence types (STs), including ST11, ST258, ST15, and ST307. The dissemination of CRKP is ascribed to certain sequence types (STs), recognized as high-risk lineages. Analogous to the phylogenetic analysis, our study discerned that most strains belong to a limited number of dominant drug-resistant clones. A supplementary analysis, analyzing 16,537 publicly accessible *K. pneumoniae* genomes, highlighted the extensive prevalence of ST258 and ST231 clones in countries such as the Philippines, Nigeria, India, and Colombia [31].

4. Discussion

The rapid growth of publicly available genome sequences has been driven by the widespread adoption of next-generation and, more recently, third-generation sequencing technologies. Consequently, elucidating the evolutionary relationships among *K. pneumoniae* strains using such extensive genomic datasets remains challenging. In this study, an extensive pan-genome analysis and phylogenetic characterization were conducted on *K. pneumoniae* isolates representing diverse geographical regions. The analysis of 72,057 genomes retrieved from the NCBI database, followed by the selection of 91 representative genomes from distinct countries, provided an unbiased global overview of the genetic diversity, population structure, and evolutionary dynamics of *K. pneumoniae*. The pan-genomic analysis revealed a large accessory genome, consistent with previous reports indicating a high level of genomic plasticity and horizontal gene transfer in this species. Prior studies have estimated the core genome of *K. pneumoniae* to range between 1700 and 4900 genes depending on the dataset size and strain diversity, which aligns with the variability observed in the present study [32–35].

The phylogenetic clustering of geographically related isolates observed in this study suggests potential regional transmission within healthcare environments and localized outbreaks. Strains from Europe, Asia, and the Middle East often shared phylogenetic proximity, indicating possible common ancestry or transcontinental clonal expansion. The global distribution of these lineages aligns with earlier investigations that reported high-resolution genomic similarity among geographically distinct *K. pneumoniae* isolates [28]. The inclusion of one representative genome per country ensured balanced global representation, allowing meaningful cross-continental comparisons. Although this approach limited intra-regional resolution, it provided a foundational framework for understanding global phylogenetic and evolutionary relationships, which can be further refined through future studies involving a higher number of isolates per region.

The present analysis also emphasizes that the accessory genome plays a crucial role in the species' adaptability and persistence. The variation in the number of AMR genes and mobile genetic elements across genomes reflects the genomic versatility of *K. pneumoniae* to acquire new traits under antibiotic pressure. The findings of the present study are consistent with the recent global trends of increasing carbapenem-resistant and hypervirulent *K. pneumoniae* lineages. Studies have reported a rapid post-COVID-19 escalation in CRKP infections, particularly in healthcare settings with high antibiotic utilization [11]. Moreover, the worldwide prevalence of hypervirulent *K. pneumoniae* strains has been associated with severe infections and poor prognosis [12]. These observations reinforce the importance of integrating genomic monitoring with clinical surveillance to mitigate the dual challenges of AMR and hypervirulence. Several studies have demonstrated that horizontal gene transfer via plasmids, transposons, and insertion sequences drives the dissemination of AMR determinants across clinical and environmental isolates [36,37]. Such genomic plasticity allows the pathogen to survive under diverse ecological and therapeutic conditions, underscoring the need for continuous global genomic surveillance.

Although the present study primarily focused on pan-genomic and phylogenetic analyses, integrating multilocus sequence typing (MLST), clade, and clonal complex identification in future research would provide finer resolution for tracing phylogenetic relatedness and transmission dynamics. These methods can identify high-risk sequence types, such as ST11, ST258, and ST307, that are known to facilitate global dissemination of carbapenem resistance [38,39]. Incorporating MLST and clonal lineage analyses with pan-genomic frameworks will further enhance the understanding of *K. pneumoniae* evolution and epidemiological spread across different regions.

Earlier studies have proposed a comprehensive genetic framework for *K. pneumoniae* encompassing more than 150 phylogenetically distinct lineages, many of which exhibit multidrug resistance and high virulence potential [32]. The species possesses an extensive accessory genome with nearly 30,000 protein-coding genes, several of which are associated with iron acquisition, capsule biosynthesis, and hypermucoviscosity, contributing to enhanced pathogenicity and adaptability [7]. The present results are consistent with these findings and provide additional evidence for widespread genomic diversity and lineage-specific adaptation.

This study provides important insights into the global distribution, genetic heterogeneity, and evolutionary dynamics of *K. pneumoniae* through the integration of comparative genomics and phylogenetic analyses [35]. These findings are crucial for understanding the mechanisms of antibiotic resistance dissemination and for guiding effective infection control and therapeutic strategies. The AMR gene information in this study was obtained from the PROKKA annotation output, which offers a general overview of resistance determinants based on integrated reference datasets. Curated databases such as the Comprehensive Antibiotic Resistance Database (CARD) or ResFinder provide greater specificity and coverage for AMR gene identification and should be incorporated in future studies for enhanced accuracy. Moreover, combining genomic data with transcriptomic and proteomic analyses could further elucidate the expression and regulation of resistance and virulence genes, thereby strengthening the molecular understanding of *K. pneumoniae* pathogenesis.

5. Conclusion

This study comprehensively analyzed 72,057 *K. pneumoniae* genomes retrieved from the NCBI database to explore their genetic diversity, evolutionary relationships, and global distribution. The findings revealed substantial variability in genome size, GC content, and contig number, as indicated by the QUAST assembly statistics, reflecting both genomic plasticity and differences in assembly quality. Functional annotation using PROKKA identified key genomic features, including coding sequences, tRNAs, rRNAs, and AMR genes, highlighting the widespread occurrence of resistance determinants across globally distributed strains. The phylogenetic analysis delineated both closely related regional clusters and highly divergent ancestral lineages, suggesting the coexistence of local transmission and long-range dissemination driven by clonal expansion and horizontal gene transfer.

Collectively, these results emphasize the extensive genetic heterogeneity and global dissemination potential of *K. pneumoniae*, underscoring its capacity to adapt and persist in diverse environments. The detection of AMR-associated genes reinforces the growing clinical threat posed by multidrug-resistant and hypervirulent lineages. This study demonstrates the importance of continuous global genomic surveillance to monitor the evolution, resistance dynamics, and transmission pathways of *K. pneumoniae*. Future work should integrate metagenomic and transcriptomic analyses of clinical samples to investigate microbial community interactions, virulence expression, and resistance regulation, thereby providing a more comprehensive understanding of this pathogen's role in human health and disease management.

Acknowledgments

Generative AI tools were not used for literature review, study design, data collection, data analysis, interpretation of results, manuscript drafting, editing, figure creation, or reference management.

Author contributions

Conceptualization was carried out by Abdelbagi Alfadil, Karem Ibrahem, Hisham N. Altayb. Methodology was developed by Hisham N. Altayb, Karem Ibrahem. Data collection and investigation were performed by Turki M. Alharthi, Mona Abdulrahman Alqarni, Bandar Hasan Saleh, Ohood S. Alharbi, Hisham N. Altayb and Waiel S. Halabi. Data curation and formal analysis were completed by Ehssan Moglad, Malaz Gazzaz and Mohanned Talal Alharbi. Resources and supervision were provided by Abdelbagi Alfadil and Wafaa Alhazmi. Writing of the original draft was prepared by Hatoon A. Niyazi, Hanouf A. Niyazi, Karem Ibrahem and Hisham N. Altayb. Review and editing were conducted by Hatoon A. Niyazi, Ohood S. Alharbi, Wafaa Alhazmi, Hind AbdulMajed, Wafaa Alhazmi, Manal A. Zubair, Noha A. Juma, Noof R. Helmi, Rawan Altalhi. Project administration was undertaken by Karem Ibrahem and Hisham N. Altayb. All authors have read and approved the final manuscript.

Disclosure statement

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Funding

The authors declare that this study was conducted without any external financial support. All work was carried out independently by the authors.

Writing assistance

No writing assistance was utilized in the production of this manuscript.

Reviewer disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

Ethics approval and consent to participate

This study was conducted using genome sequence data retrieved exclusively from publicly available databases (National Center for Biotechnology Information, NCBI). No human participants, animals, or identifiable personal data were involved, and no new biological samples were collected. All data were fully anonymized and accessed in accordance with the respective database usage policies. Consequently, ethical approval and informed consent were not required for this study.

Data availability statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request. We are committed to transparency and collaboration and are willing to share data with researchers for the purposes of verification, further analysis, or additional studies related to this work. Please contact the corresponding author to discuss the details of data access.

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Chang D, Sharma L, Dela Cruz CS, et al. Clinical epidemiology, risk factors, and control strategies of Klebsiella pneumoniae infection. Front Microbiol. 2021;12. doi: 10.3389/fmicb.2021.750662
2. Ashurst JV, Dawson A. Klebsiella pneumonia. StatPearls [internet]. Treasure Island (FL): StatPearls Publishing; 2024 [cited 2024 Sep 19]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK519004/
3. Calfee DP. Recent advances in the understanding and management of Klebsiella pneumoniae. F1000Res. 2017;6:1760. doi: 10.12688/f1000research.11532.1
4. Abbas R, Chakkour M, Zein El Dine H, et al. General overview of Klebsiella pneumonia: epidemiology and the role of Siderophores in its pathogenicity. Biol (Basel). 2024;13(2):78. doi: 10.3390/biology13020078
5. Hasan T, Shlash S, Jasim S, et al. The epidemiology of Klebsiella Pneumoniae: a review. Ann Rom Soc Cell Biol. 2021;25:8848–8860.
6. Paczosa MK, Mecsas J. Klebsiella pneumoniae: going on the offense with a strong defense. Microbiol Mol Biol Rev. 2016;80(3):629–661. doi: 10.1128/MMBR.00078-15
7. Gomez-Simmonds A, Uhlemann A-C. Clinical implications of genomic adaptation and evolution of carbapenem-resistant Klebsiella pneumoniae. J Infect Dis. 2017;215(suppl_1):S18–S27. doi: 10.1093/infdis/jiw378

•• **This review provides foundational insight into the genomic evolution of carbapenem-resistant K. pneumoniae and its clinical consequences, making it highly relevant for interpreting global phylogenomic patterns.**

8. Nagaraj G, Shamanna V, Govindan V, et al. High-resolution genomic profiling of carbapenem-resistant Klebsiella pneumoniae isolates: a multicentric retrospective Indian study. Clin Infect Dis. 2021;73 (Suppl 4):S300–S307.
9. Sanchez GV, Master RN, Clark RB, et al. Klebsiella pneumoniae antimicrobial drug resistance, United States, 1998–2010. Emerg Infect Dis. 2013;19(1):133–136. doi: 10.3201/eid1901.120310
10. Li H-F, Zhang L-X, Zhang W-L, et al. Study on virulence genes, drug resistance and molecular epidemiology of Klebsiella pneumoniae with high virulence in Inner Mongolia, China. IDR. 2023;16:1133–1144. doi: 10.2147/IDR.S391468
11. Liu C, Guo J, Fan S, et al. An increased prevalence of carbapenem-resistant hypervirulent Klebsiella pneumoniae associated with the COVID-19 pandemic. Drug Resist Updat. 2024;77:101124. doi: 10.1016/j.drup.2024.101124

•• **This study offers critical evidence linking recent global healthcare pressures to the rise of carbapenem-resistant hypervirulent lineages, directly supporting the contemporary relevance of our findings.**

12. Tang Y, Du P, Du C, et al. Genomically defined hypervirulent Klebsiella pneumoniae contributed to early-onset increased mortality. Nat Commun. 2025;16(1):2096. doi: 10.1038/s41467-025-57379-4
13. Barh D, Soares SC, Tiwari S, et al. Pan-genomics: applications, challenges, and future prospects. Academic Press; 2020.
14. Sundaresan AK, Gangwar J, Murugavel A, et al. Complete genome sequence, phenotypic correlation and pangenome analysis of uropathogenic Klebsiella spp. AMB Expr. 2024;14(1):78. doi: 10.1186/s13568-024-01737-w
15. Sayers EW, Beck J, Bolton EE, et al. Database resources of the National center for Biotechnology Information. Nucleic Acids Res. 2021;49(D1):D10–D17. doi: 10.1093/nar/gkaa892
16. Gurevich A, Saveliev V, Vyahhi N, et al. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072–1075. doi: 10.1093/bioinformatics/btt086
17. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–2069. doi: 10.1093/bioinformatics/btu153
18. Page AJ, Cummins CA, Hunt M, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31 (22):3691–3693. doi: 10.1093/bioinformatics/btv421
19. Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PLOS ONE. 2010;5(3):e9490. doi: 10.1371/journal.pone.0009490

20. Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293–W296. doi: 10.1093/nar/gkab301
21. Nougayrède J-P, Homburg S, Taieb F, et al. *Escherichia coli* induces DNA Double-strand breaks in eukaryotic cells. Science. 2006;313 (5788):848–851. doi: 10.1126/science.1127059
22. Wang Z, Liu Y, Liu P, et al. Genomic and clinical characterization of Klebsiella pneumoniae carrying the pks island. Front Microbiol. 2023;14:1189120. doi: 10.3389/fmicb.2023.1189120
23. Chen FZ, You LJ, Yang F, et al. Cngbdb: China National GeneBank DataBase. Yi Chuan. 2020;42(8):799–809. doi: 10.16288/j.yczz.20-080
24. Zhou X, Chu Q, Li S, et al. A new and effective genes-based method for phylogenetic analysis of *Klebsiella pneumoniae*. Infect Genet Evol. 2022;100:105275. doi: 10.1016/j.meegid.2022.105275
25. Wyres KL, Holt KE. *Klebsiella pneumoniae* population genomics and Antimicrobial-Resistant Clones. Trends Microbiol. 2016;24 (12):944–956. doi: 10.1016/j.tim.2016.09.007
26. Cheung BH, Alisoltani A, Kochan TJ, et al. Genome-wide screens reveal shared and strain-specific genes that facilitate enteric colonization by Klebsiella pneumoniae. bioRxiv. 2023;14(6):e02128–23.
27. Arauz-Cabrera J, Marquez-Salazar D, Delgadillo-Valles R, et al. Genomic Profile of a Multidrug-Resistant Klebsiella pneumoniae Strain Isolated from a Urine Specimen. Curr Microbiol. 2024;81 (9):276. doi: 10.1007/s00284-024-03802-w
28. Wang Q, Liu Y, Chen R, et al. Genomic insights into the evolution and mechanisms of carbapenem-resistant hypervirulent Klebsiella pneumoniae co-harboring blaKPC and blaNDM: implications for public health threat mitigation. Ann Clin Microbiol Antimicrob. 2024;23(1):27. doi: 10.1186/s12941-024-00686-3
29. Arcari G, Carattoli A. Global spread and evolutionary convergence of multidrug-resistant and hypervirulent Klebsiella pneumoniae high-risk clones. Pathog Glob Health. 2023;117(4):328–341. doi: 10.1080/20477724.2022.2121362
30. Hala S, Malaikah M, Huang J, et al. The emergence of highly resistant and hypervirulent Klebsiella pneumoniae CC14 clone in a tertiary hospital over 8 years. Genome Med. 2024;16(1):58. doi: 10.1186/s13073-024-01332-5
31. Argimón S, David S, Underwood A, et al. Rapid genomic characterization and global surveillance of *Klebsiella* using pathogenwatch. Clin Infect Dis. 2021;73(Supplement_4):S325–S335. doi: 10.1093/cid/ciab784
32. Holt KE, Wertheim H, Zadoks RN, et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. Proc Natl Acad Sci USA. 2015;112(27). doi: 10.1073/pnas.1501049112
33. Ejaz H, Wang N, Wilksch JJ, et al. Phylogenetic analysis of Klebsiella pneumoniae from hospitalized children, Pakistan. Emerg Infect Dis. 2017;23(11):1872–1875. doi: 10.3201/eid2311.170833
34. Chung the H, Karkey A, Pham Thanh D, et al. A high-resolution genomic analysis of multidrug-resistant hospital outbreaks of Klebsiella pneumoniae. EMBO Mol Med. 2015;7(3):227–239. doi: 10.15252/emmm.201404767
35. Caputo A, Merhej V, Georgiades K, et al. Pan-genomic analysis to redefine species and subspecies based on quantum discontinuous variation: the Klebsiella paradigm. Biol Direct. 2015;10(1):55. doi: 10.1186/s13062-015-0085-2
    - **This study provides a theoretical and methodological foundation for pan-genome analysis in Klebsiella, directly supporting the analytical framework used in our work.**
36. Castaneda-Barba S, Top EM, Stalder T. Plasmids, a molecular cornerstone of antimicrobial resistance in the one health era. Nat Rev Microbiol. 2024;22(1):18–32. doi: 10.1038/s41579-023-00926-x
37. Plasmids carrying antimicrobial resistance genes in enterobacteriaceae. Journal of antimicrobial chemotherapy. Oxford Academic; [cited 2025 Nov 12]. Available from: https://academic.oup.com/jac/article/73/5/1121/4822282
38. Kilian M, Scholz CFP, Lomholt HB. Multilocus sequence typing and phylogenetic analysis of propionibacterium acnes. J Clin Microbiol. 2012;50(4):1158–1165. doi: 10.1128/JCM.r06129-11
39. Hong W, Yang Z, Wu G, et al. Integrating serotyping, MLST, and phenotypic data: decoding the evolutionary drivers of salmonella pathogenicity and drug resistance. Appl Environ Microbiol. 2025;91 (10):e01511–25. doi: 10.1128/aem.01511-25