Methods and Applications

Klebsiella pneumoniae Genome Database: A Global Resource for Genomic Surveillance of Dissemination, Pathogenicity, and Antimicrobial Resistance

Haijian Zhou[1]; Chongye Guo[2,3]; Guomei Fan[2,3]; Jinrui Hu[1]; Zhigang Cui[1]; Xiaoli Du[1]; Linhuan Wu[2,3,#]; Biao Kan[1,#]

ABSTRACT

Introduction: To address the escalating public health threat of hypervirulent and antibiotic-resistant *Klebsiella pneumoniae* (KP), we developed the Klebsiella pneumoniae Genome Database (KPGD; http://nmdc.cn/gcpathogen/kp) to strengthen global genomic surveillance of this pathogen.

Methods: KPGD integrates 75,987 genome assemblies from 122 countries with standardized annotations of serotypes, sequence types, antibiotic resistance genes (ARGs), virulence factors (VFs), and mobile genetic elements (MGEs). The platform offers interactive visualization modules and integrated analytical tools that enable real-time epidemiological monitoring and one-stop genomic analysis, thereby supporting global efforts to track the dissemination of resistant and hypervirulent KP (HvKp) and to inform infection control and antimicrobial stewardship strategies.

Results: Longitudinal analyses revealed that the emergence of HvKp is driven by the sustained expansion of carbapenem-resistant high-risk lineages under selection pressure from restricted, higher-tier antibiotics. Conjugative ARG-bearing plasmids carrying key resistance determinants largely mediate this expansion. In contrast, selection by first-line, narrower-spectrum antibiotics appears to favor the dissemination of virulence plasmids (predominantly IncFIB types) as a compensatory mechanism to offset resistance-associated fitness costs.

Conclusion: These findings collectively underscore the need for surveillance systems that simultaneously monitor high-risk lineages and the dissemination of ARGs and VFs — particularly via self-transmissible plasmids — to better understand and anticipate bacterial adaptation under diverse antibiotic pressures.

Klebsiella pneumoniae (KP) ranks among the leading causes of healthcare-associated infections worldwide, with high morbidity and mortality driven by carbapenem-resistant and hypervirulent strains The World Health Organization (WHO) listed KP as a critical pathogen on its Bacterial Priority Pathogens List in both 2017 and 2024 (*1*). Of particular concern are hypervirulent KP (hvKp) and carbapenem-resistant hvKp (CR-hvKp). High-risk clones such as NDM-positive ST147-KL64 (with mortality rates approaching 40%) (*2*–*3*) and ST23-KL1 (carrying rmpA/rmpA2, iroB, and/or iucA) are expanding rapidly — at an annual global growth rate of 59% and of 30% in China (*4*). These trends underscore the urgent need for systematic surveillance of carbapenem-resistant KP (CRKP), hvKp, and CR-hvKp.

Genomic surveillance provides high-resolution genetic insights that enable real-time reconstruction of transmission chains and prospective threat assessment. By elucidating pathogen evolution, dissemination patterns, and clinically relevant phenotypes, it strengthens public health responses (*5*–*6*). Clarifying the evolutionary trajectories of antimicrobial resistance (AMR) and virulence determinants further optimizes clinical decision-making and supports targeted infection control (*7*). This approach represents a paradigm shift from traditional phenotype-based monitoring, enabling outbreak tracing, epidemic forecasting, early warning of high-risk strains, and real-time assessment of emerging AMR and hypervirulence.

Discovering, surveying, and assessing the emergence and spread of hypervirulent KP clones and sequence types requires a comprehensive database of global KP genomes. Although public databases such as the National Center for Biotechnology Information (NCBI) are available, existing resources lack the integration of epidemiological metadata with standardized genomic annotations. To address these critical gaps in global KP surveillance, we established the Klebsiella pneumoniae Genome Database

(KPGD), which integrates sequence types, serotypes, antibiotic resistance genes (ARGs), virulence factors (VFs), mobile genetic elements (MGEs), and associated epidemiological metadata — including collection time and geographic origin. KPGD carries significant public health implications by enabling early warning of emerging high-risk clones, supporting outbreak investigations, and assisting public health agencies in monitoring the dissemination of hypervirulent and carbapenem-resistant KP.

METHODS

Data Source

KPGD integrates 75,987 KP genomes collected over more than 40 years from 122 countries. Of these, 71,132 publicly available genomes were retrieved from the National Center for Biotechnology Information (NCBI)(*8*), and 4,855 sequences were collected through national surveillance conducted by the Chinese Pathogen Identification Net (China PIN), which performs hospital-based sampling for febrile respiratory syndrome; KP isolates obtained from lower respiratory tract specimens testing positive by nucleic acid detection were cultured and sequenced. These 4,855 sequences originated from 170 cities across 29 provincial-level administrative divisions (PLADs) in China (Supplementary Figure S1, available at http://weekly.chinacdc.cn/), spanning a period of over 40 years. A major component of the database is the Chinese collection of 17,251 genomes, comprising 12,396 NCBI sequences combined with those from the China PIN. All 75,987 genome assemblies underwent uniform quality assessment using CheckM (Queensland University of Technology, Queensland, Australia, v1.2.2) (*9*). Genomes with completeness≥95%, contamination≤5%, and consistency≥95% were retained for downstream analysis.

Analysis Methods

Genomes in KPGD were curated and analyzed using the Global Catalogue of Pathogens (gcPathogen) frameworks and tools (*10*–*11*) (Figure 1 and Supplementary Material, available at http://weekly.chinacdc.cn/). Hyper-VFs (iroB, iucA, rmpA, and rmpA2) and resistance genes were manually curated on

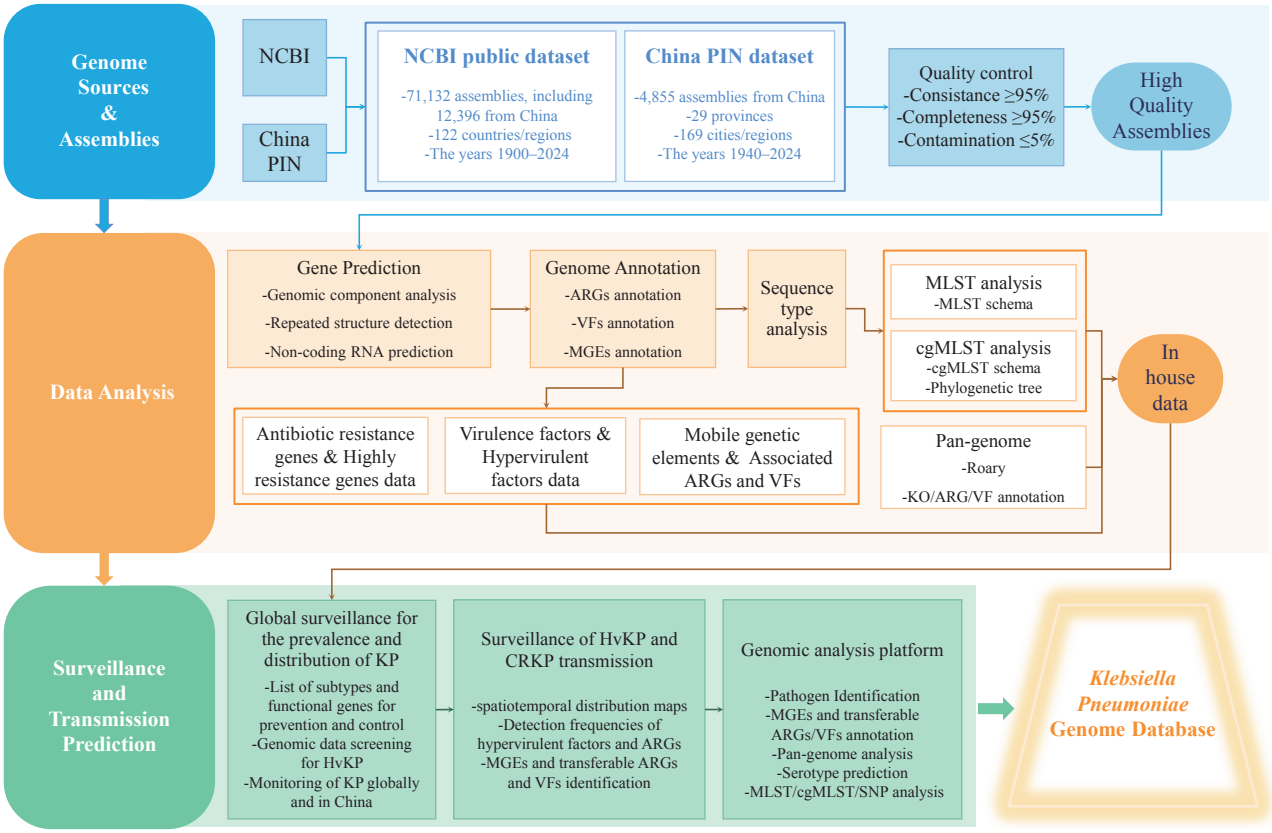


FIGURE 1. Data processing pipeline of the KPGD.
Abbreviation: KPGD=*Klebsiella pneumoniae* Genome Database; China PIN=Chinese Pathogen Identification Net.

the basis of a systematic literature review.

RESULTS

Overview of the KPGD Web Interface and Analytical Modules

The KPGD web platform (Figure 2A) provides comprehensive statistical summaries, including total genome counts and distributions of K/O antigens, hvKp strains, ARGs, sampling countries, and collection years. The "Genome Data" module supports advanced queries by WHO regions, serotypes, sequence types (STs), ARGs, and VFs. Search outputs appear in an interactive tabular format on a secondary page, integrating curated metadata with analytical results. The knowledge graph (Figure 2B) visualizes collaborative networks of research institutions and investigators, along with relevant publications and patents.

The AMR module identifies high-frequency, widely distributed ARGs across hosts and ecological niches. The MGE module maps mobile genetic elements harboring critical ARGs and hyper-VFs, enabling assessment of their horizontal transfer potential. The hvKp module captures the dynamics and epidemiological trends of hypervirulent strains (Figure 2C), while the pan-genome analysis module resolves core and accessory gene repertoires, including conserved resistance determinants across lineages.

By integrating these genetic insights with spatiotemporal, host, and clinical metadata, KPGD facilitates early detection of dissemination events, tracks the evolutionary trajectories of resistance and hypervirulence, and monitors the spread of high-risk MGEs — thereby providing a platform for proactive public health intervention. For example, the high-risk clones ST147_KL64_O2a, ST23_KL1_O1ab, and ST45_KL24_O2a were confined to a limited number of countries before 2010 but had spread globally by 2024, demonstrating clear international transmission. Specifically, ST147_KL64_O2a was detected in only 13 countries prior to 2010; its geographic range expanded to 40 countries by 2015 and further to 47 countries by 2024. Similarly, ST23_KL1_O1ab exhibited a pronounced dissemination pattern, with its country-level distribution increasing from 13 countries in 2010 to 32 countries by 2024. Meanwhile, ST45_KL24_O2a was initially confined to 6 countries before 2010, underwent rapid dissemination to reach 28 countries by 2015, and continued cross-border transmission, expanding further to 32 countries by 2024.

Additionally, KPGD offers five integrated one-stop online tools (Supplementary Material) for pathogen identification, Single Nucleotide Polymorphism (SNP) analysis, serotype prediction, detection of MGEs and transferable ARGs and VFs, and Core Gene Multilocus Sequence Typing (cgMLST) analysis (Figure 2D).

Temporal Dynamics and Evolutionary Trends of High-Risk KP Lineages

Genomic surveillance through KPGD reveals a dynamic and evolving risk landscape characterized by distinct evolutionary trajectories among high-risk KP lineages under different antibiotic selection pressures, as stratified by the World Health Organization (WHO) Access, Watch, Reserve (AWaRe) classification (*12*) (Figure 3A). These findings underscore the need for lineage-specific surveillance and intervention strategies.

Clonal group analyses identified several high-risk lineages (Figure 3A), notably ST11_KL64_O2a and ST258_KL107_O2afg, detected at elevated frequencies (peaking at around 0.2, particularly after 2016) under "Watch" antibiotic selection (carbapenems and other β-lactams). ST11_KL64_O2a persisted from 2010 to 2024 with pronounced enrichment among carbapenem-resistant isolates, indicating adaptation to last-resort therapies. ST307_KL102_O2afg peaked among carbapenem-resistant strains before 2015 (>0.07), potentially reflecting increased use of imipenem and meropenem during that period (*13*), before declining rapidly to below 0.1. ST147_KL64_O2a exhibited marked post-2020 expansion (around 0.04), especially among carbapenem- and ampicillin-resistant isolates. The hypervirulent lineage ST23_KL1_O1ab remained at low frequencies (0–0.05), although its recent increase among imipenem-resistant isolates suggests progressive acquisition of resistance to broad-spectrum β-lactams.

At the population level, resistance frequencies displayed distinct temporal patterns. Carbapenem-resistance ARGs increased steadily, with β-lactam ARGs among ertapenem- and meropenem-resistant isolates rising from near zero in 2010 to 0.15–0.3 by 2024. In contrast, cephalosporin resistance declined overall (below 0.25), while penicillin β-lactam resistance fluctuated between 0.1 and 0.7. These trends underscore a shifting AMR landscape and highlight lineage-specific evolutionary strategies under defined

FIGURE 2. Features of the KPGD Website. (A) KPGD homepage. (B) The Knowledge Graph page depicting global KP research networks; (C) The hvKp subpage summarizing datasets, spatiotemporal patterns, regional and country-level counts, and ARG/VF frequencies by ST, country, and habitat. (D) Phylogenetic clustering of high-risk KP ST23 clones using the KPGD cgMLST tool.

Note: For (B), it includes authors, institutions, publications, and patents. For (D), the 2019 isolate NMDC60197270 from China PIN clustered within the same phylogenetic branch as isolates from Bangladesh (2016 and 2017), India (2020), and Ghana (2021), suggesting potential transmission from these regions. Furthermore, the 2024 isolate NMDC60192954 clustered predominantly with isolates from Japan (2019–2020), indicating possible transmission from Japan.

Abbreviation: KPGD=*Klebsiella pneumoniae* Genome Database; China PIN=Chinese Pathogen Identification Net; ST=sequence type.

antibiotic pressures, illustrating the adaptive diversification of high-risk KP clones in response to therapeutic interventions.

Temporal dynamics of key carbapenem-resistance

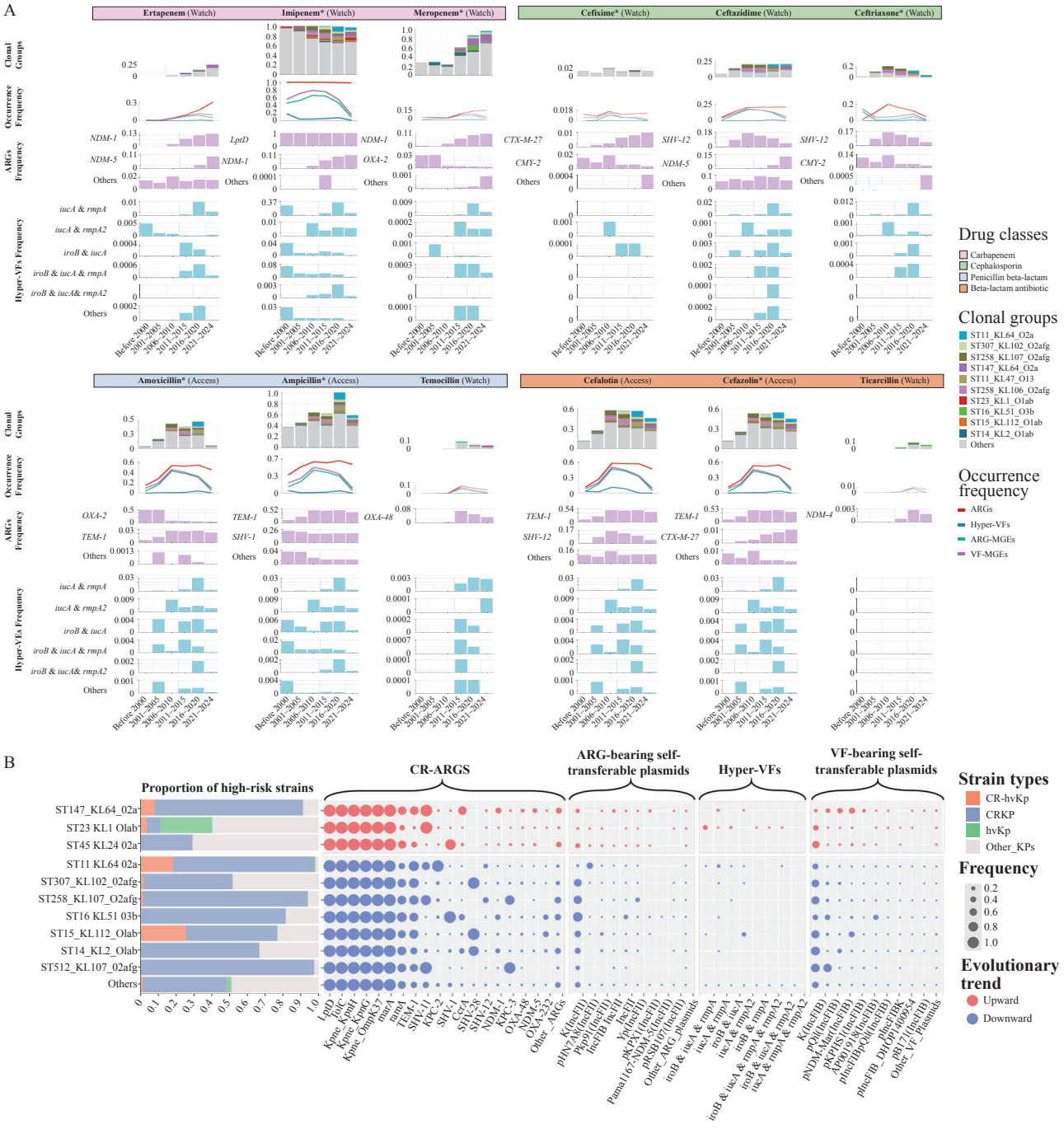


FIGURE 3. Risk Profiles and Evolutionary Trajectories of KP. (A) Longitudinal trends in ARGs, hyper-VFs, and major clonal lineages among KP isolates (before 2000–2024). (B) Risk assessment of the top 10 clonal lineages based on resistance and virulence determinants.

Note: For (A), Antibiotics are classified according to the WHO AWaRe framework ("Access" and "Watch"). Stacked bars represent the frequencies of the top 10 clonal lineages (with "Others" denoting remaining types). Line plots depict temporal changes in overall resistance and virulence frequencies under antibiotic pressure, while area charts summarize the occurrence of key ARGs and hyper-VFs. For (B), Bars represent different clonal groups, with colors indicating high-risk clones, including CR-hvKP, CRKP, and hvKP. Bubble plots illustrate the frequencies of carbapenem-resistance and hypervirulence genes, as well as the frequencies of self-transferable plasmids associated with these determinants. Red bubbles indicate an increasing trend for the corresponding clonal group, whereas blue bubbles indicate a decreasing trend.

Abbreviation: WHO=World Health Organization.

ARGs ("Watch" antibiotics) further emphasize this expansion (Figure 3A). NDM-1 emerged at 0.038 (2010–2015) and peaked at 0.13 (2022–2024). NDM-5 rose from undetectable levels to 0.11 by 2022–2024. In contrast, ARGs associated with "Access" antibiotics (e.g., TEM-1 in amoxicillin/ampicillin-resistant isolates) remained stable (around 0.1–0.5), consistent with their role in first-line therapy. Other ARGs (OXA-2, CMY-2, and CTX-M-27) persisted at low prevalence. Notably, the near-universal presence of LptD (frequency≈1.0) among imipenem-resistant isolates suggests a potential role in KP survival under carbapenem pressure. These ARGs were predominantly disseminated by conjugative IncFII-type plasmids, including K(IncFII), pHN7A8, and pKP91, whose frequencies fluctuated widely (0 to 0.6), underscoring their central role in the rapid horizontal transfer of carbapenem resistance.

Hyper-VF diversity was closely associated with resistance profiles across KP populations (Figure 3A). Among imipenem-resistant isolates ("Watch" antibiotic), hyper-VF richness was pronounced. Before 2000, the dominant combination (iroB, iucA, and rmpA) peaked at 0.08 and then declined steadily. Between 2011 and 2020, a two-factor combination (iucA and rmpA) predominated (around 0.04), surpassing all others. By 2024, all hyper-VF combinations had decreased to around 0.005. These hyper-VFs were largely disseminated by self-transmissible IncFIB-type virulence plasmids, including K(IncFIB) and pQil, peaking at around 0.5.

As shown in Figure 3B, ST23_KL1_O1ab harbored a substantially higher proportion of hvKP strains than other lineages. CR-hvKP strains were most frequently detected in ST15_KL2_O1ab, ST11_KL64_O2a, and ST147_KL64_O2a. Notably, the three clonal lineages displaying an increasing trend carried relatively low frequencies of ARG-bearing self-transferable plasmids. In contrast, all other clonal lineages (with decreasing trends) carried the ARG-bearing self-transferable plasmid K(IncFII), with the highest frequency exceeding 0.8. However, all lineages carried the VF-bearing self-transferable plasmid K(IncFIB), with a maximum frequency of approximately 0.6.

Time-series analyses revealed that fluctuations in virulence plasmid prevalence closely aligned with antibiotic selection regimes. Higher plasmid frequencies were associated with "Access" antibiotics, particularly first-generation cephalosporins (e.g., cefalotin and cefazolin). This suggests that virulence plasmid dissemination may be favored under first-line antibiotic pressure, potentially compensating for resistance-associated fitness costs and enhancing KP persistence.

CONCLUSIONS

This study establishes the Klebsiella pneumoniae Genome Database (KPGD) as a global resource for genomic surveillance, addressing the public health threats posed by the dissemination, pathogenicity, and antimicrobial resistance of hypervirulent and antibiotic-resistant lineages. By integrating 75,987 genomes from 122 countries with epidemiological metadata and standardized bioinformatic pipelines, KPGD enables high-resolution tracking of KP transmission, evolution, and risk patterns across decades and continents. These datasets can also be integrated into national monitoring systems, such as the China PIN, to support epidemic analysis, risk assessment, and early warning of high-risk KP clones.

Genomic surveillance revealed distinct evolutionary trajectories among high-risk lineages under differential antibiotic selection pressures. Clones including ST11_KL64_O2a and ST258_KL107_O2afg persisted under carbapenem-dominated "Watch" antibiotics, driven by horizontal acquisition of carbapenemase genes (e.g., NDM-1 and NDM-5) via conjugative IncFII-type plasmids. Temporal fluctuations in plasmid detection correlated with changes in antibiotic usage, underscoring their role as vectors for resistance dissemination.

The expansion of NDM-positive ST147_KL64_O2a represents a specific example of a high-risk clone in which carbapenem resistance and hypervirulence markers co-occur. This does not imply that NDM universally drives hypervirulence. Future studies should investigate region-specific associations between carbapenemase genes and hypervirulence markers.

Concurrently, a dynamic interplay between resistance and pathogenicity was observed. Under first-line "Access" antibiotic pressure, KP strains gained a selective advantage through acquisition of IncFIB-type virulence plasmids, facilitating the emergence of clones exhibiting both high pathogenicity and antimicrobial resistance (*14*).

Collectively, these findings provide a genomic framework for guiding public health strategies. KPGD serves as a foundational global resource for identifying emerging high-risk clones, elucidating the molecular drivers of their success, and tracking the horizontal

transfer of ARGs and VFs worldwide. Ensuring equitable access to such genomic platforms will be essential for coordinated international responses to the growing threat of hypervirulent and antimicrobial-resistant KP.

To this end, KPGD is an open and freely accessible web platform available at http://nmdc.cn/gcpathogen/kp. Unlike static genomic collections or one-time analytical studies, KPGD provides interactive real-time querying and integrated analytical modules for the global research community without registration or payment. Users worldwide can filter genomes by WHO region, serotype, ST, ARG, VF, collection time, and geographic origin.

Conflicts of interest: No conflicts of interest.

Funding: Supported by the National Key Research and Development Program of China (2023YFC2308800), the Self-Supporting Program of Guangzhou National Laboratory (SRPG22007), the Major Project of Guangzhou National Laboratory (GZNL2024A01025), and the Capital's Funds for Health Improvement and Research (CFH2024-1G-4361).

doi: 10.46234/ccdcw2026.092

[#] Corresponding authors: Linhuan Wu, wulh@ac.cn; Biao Kan, kanbiao@icdc.cn.

[1] National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention & Chinese Academy of Preventive Medicine, Beijing, China; [2] Chinese National Microbiology Data Center (NMDC), Beijing, China; [3] Beijing Research Center for Respiratory Infectious Diseases; Beijing Key Laboratory of Surveillance, Early Warning and Pathogen Research on Emerging Infectious Diseases, Beijing, China.

Submitted: April 18, 2026
Accepted: April 25, 2026
Issued: May 01, 2026

REFERENCES

1. WHO Bacterial Priority Pathogens List. Bacterial pathogens of public health importance to guide research, development and strategies to prevent and control antimicrobial resistance. Geneva: World Health Organization; 2024. Licence: CC BY-NC-SA 3.0 IGO. https://iris.who.int/server/api/core/bitstreams/1a41ef7e-dd24-4ce6-a9a6-1573562e7f37/content.
2. Di Pilato V, De Angelis LH, Aiezza N, Baccani I, Niccolai C, Parisio EM, et al. Resistome and virulome accretion in an NDM-1-producing ST147 sublineage of *Klebsiella pneumoniae* associated with an outbreak in Tuscany, Italy: a genotypic and phenotypic characterisation. Lancet Microbe 2022;3(3):e224 – 34. https://doi.org/10.1016/S2666-5247(21)00268-8.
3. Martin MJ, Corey BW, Sannio F, Hall LR, MacDonald U, Jones BT, et al. Anatomy of an extensively drug-resistant *Klebsiella pneumoniae* outbreak in Tuscany, Italy. Proc Natl Acad Sci USA 2021;118(48):e2110227118. https://doi.org/10.1073/pnas.2110227118.
4. Zhou HJ, Guo CY, Cui ZG, Hu JR, Du XL, Sun Y, et al. The epidemiology and hypervirulence of *Klebsiella pneumoniae* ST23 unveil epidemic risks in China and worldwide. Lancet Reg Health West Pac 2025;62:101661. https://doi.org/10.1016/j.lanwpc.2025.101661.
5. Lytras S, Lamb KD, Ito J, Grove J, Yuan K, Sato K, et al. Pathogen genomic surveillance and the AI revolution. J Virol 2025;99(2):e01601 – 24. https://doi.org/10.1128/jvi.01601-24.
6. David S, Reuter S, Harris SR, Glasner C, Feltwell T, Argimon S, et al. Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial spread. Nat Microbiol 2019;4(11):1919 – 29. https://doi.org/10.1038/s41564-019-0492-8.
7. Loconsole D, Sallustio A, Sacco D, Santantonio M, Casulli D, Gatti D, et al. Genomic surveillance of carbapenem-resistant *Klebsiella pneumoniae* reveals a prolonged outbreak of extensively drug-resistant ST147 NDM-1 during the COVID-19 pandemic in the Apulia region (Southern Italy). J Glob Antimicrob Resist 2024;36:260 – 6. https://doi.org/10.1016/j.jgar.2024.01.015.
8. Kitts PA, Church DM, Thibaud-Nissen F, Choi J, Hem V, Sapojnikov V, et al. Assembly: a resource for assembled genomes at NCBI. Nucl Acids Res 2016;44(D1):D73 – 80. https://doi.org/10.1093/nar/gkv1226.
9. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 2015;25(7):1043 – 55. https://doi.org/10.1101/gr.186072.114.
10. Guo CY, Chen Q, Fan GM, Sun Y, Nie JY, Shen ZH, et al. gcPathogen: a comprehensive genomic resource of human pathogens for public health. Nucl Acids Res 2024;52(D1):D714 – 23. https://doi.org/10.1093/nar/gkad875.
11. Fan GM, Guo CY, Zhang Q, Liu DM, Sun QL, Cui ZG, et al. A secure visualization platform for pathogenic genome analysis with an accurate reference database. Biosafety Health 2024;6(4):235 – 43. https://doi.org/10.1016/j.bsheal.2024.07.003.
12. Sharland M, Zanichelli V, Ombajo LA, Bazira J, Cappello B, Chitatanga R, et al. The WHO essential medicines list AWaRe book: from a list to a quality improvement system. Clin Microbiol Infect 2022;28(12):1533 – 5. https://doi.org/10.1016/j.cmi.2022.08.009.
13. Wang LG, Chen H, Zhang YY, Tian Y, Hu XY, Wu J, et al. Global antibiotic consumption and regional antimicrobial resistance, 2010-21: an analysis of pharmaceutical sales and antimicrobial resistance surveillance data. Lancet Glob Health 2025;13(11):e1880 – 91. https://doi.org/10.1016/S2214-109X(25)00308-0.
14. Carvalho AG, Belém MGL, Rodrigues RS, Da Silva MEP, dos Santos Dorneles NW, Da Silva Lima NC, et al. Serotype distribution, virulence factors, and antimicrobial resistance profiles of *Streptococcus agalactiae* (Group B Streptococcus) isolated from pregnant women in the Brazilian Amazon. BMC Microbiol 2025;25(1):361. http://10.1186/s12866-025-04077-2.

SUPPLEMENTARY MATERIALS

Bioinformatic analysis of the genomes in the *Klebsiella pneumoniae* Genome Database (KPGD) was performed using the analysis tools within the gcPathogen one-stop analysis system (*1*), including pathogen identification, multilocus sequence typing (MLST), core genome multilocus sequence typing (cgMLST), pan-genome analysis, and genomic annotation for ARGs, VFs, and mobile genetic elements (MGEs) (Figure 1). A suite of one-stop online tools has been integrated into KPGD. The following section describes their primary workflows and the underlying software.

The Pathogen Identification Tool primarily employs bioinformatic software such as Kraken2 (*2*) and RNAmmer (*3*) for sequence similarity comparisons against reference libraries, generating optimal alignments for species identification. The results page displays the identified species, basic biological information, the corresponding reference genome, and detailed 16S rDNA data. The 16S rDNA BLAST search uses an e-value threshold of ≤1e-5, and the top 10 hits are retained based on bit-score. For ANI-based analysis, a consistency threshold of >95% is required for species assignment. If multiple species are detected, Kraken2 k-mer analysis is triggered; a species is considered confirmed if its k-mer abundance exceeds 80% and the second-ranked species abundance falls below 3%.

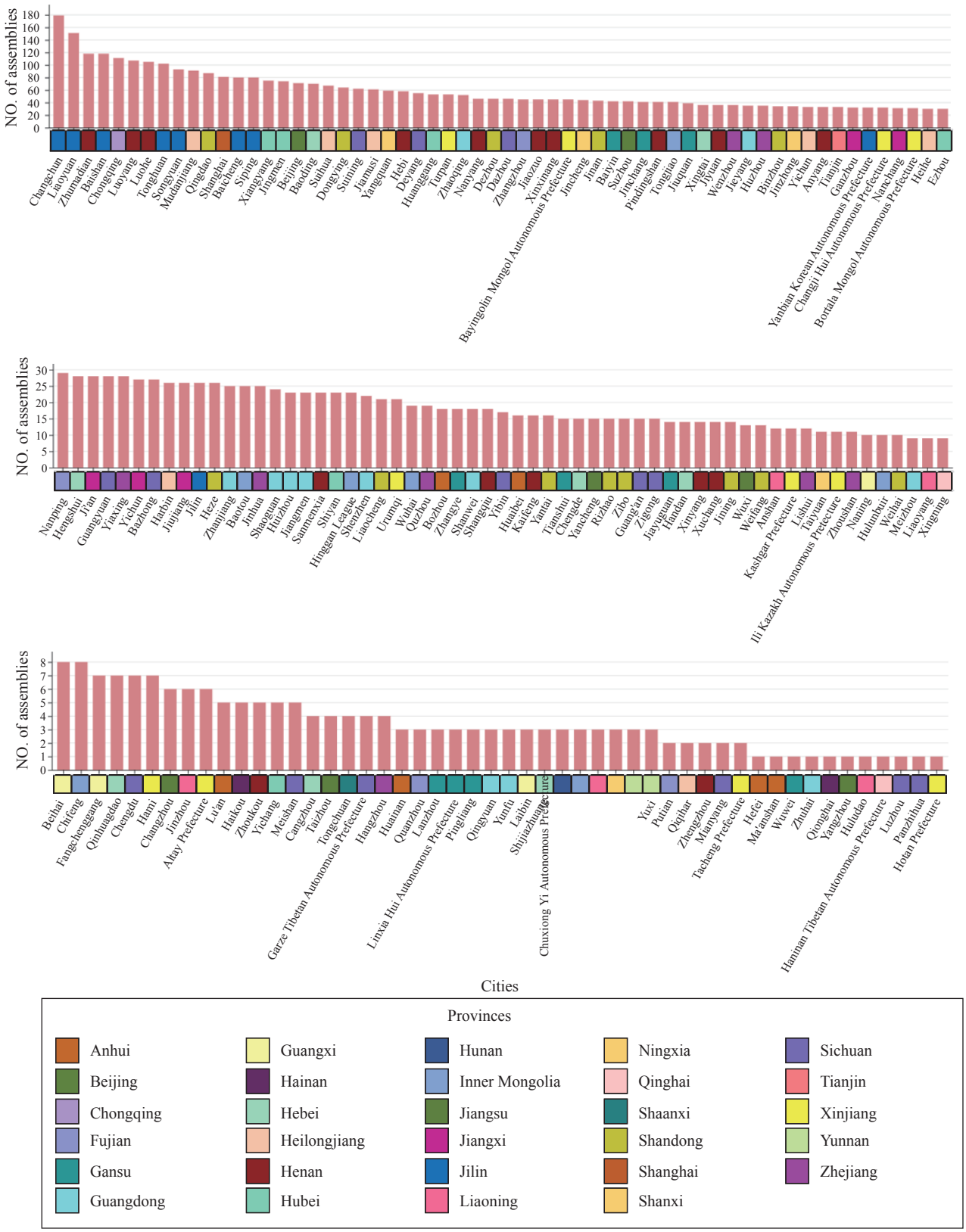
Genomic annotation is performed using the integrated annotation tool of the gcPathogen platform (*1*). It conducts genomic component analysis using PILER-CR v1.06 (*4*) for CRISPR array recognition, followed by repeat structure detection using TRF v4.07b (*5*). Non-coding RNA prediction is then carried out with tRNAscan-SE v1.4 (*6*) and RNAmmer v1.2. Finally, gene prediction is performed using Prodigal v2.6.3 (*7*). The predicted genes can be annotated against 11 commonly used functional databases, including KEGG (*8*), COG (*9*), NCBI-nr (*10*), CARD (*11*), CAZy (*12*), PHI (*13*), SwissProt (*14*), VFDB (*15*), Pfam (*16*), MetaCyc (*17*), and AntiSMASH (*18*). The alignment parameters are set as follows: query coverage >80%, subject coverage >80%, and sequence identity >90%.

The SNP Analysis Tool employs iVar (*19*) for SNP calling, Gubbins (*20*) to construct SNP matrices (excluding recombinant regions), and IQ-TREE 2 (*21*) for phylogenetic tree construction, with mutation site details displayed on the results page.

The Serotype Prediction Tool applies Kleborate (*22*), with the result file containing the species match degree, data quality, ST (sequence type), serotype, and other relevant details.

The MGE and Transferable ARGs and VFs Detection Tool leverages annotation software such as ISEScan (*23*) and MobileElementFinder (*24*) to annotate insertion sequences (IS), integrative conjugative elements (ICE), integrons (IN), plasmids, transposons (Tn), and their associated ARGs and VFs, presenting their types, names, genomic locations, and lengths. An ARG or VF is considered potentially transferable if it meets either of the following criteria: (i) the same IS element is present within 10 kb both upstream and downstream of the gene; or (ii) the gene is located within the sequence range of an ICE, integron, plasmid, phage, or transposon. These criteria are applied to assess horizontal transfer potential.

Built on chewBBACA (*25*), the cgMLST Analysis Tool enables schema creation and allele calling for both complete and draft genomes. With ≥3 genus- or species-level genomes uploaded, it generates visualized phylogenetic trees and result files. The cgMLST analysis tool enables phylogenetic clustering of closely related isolates. When these clusters are combined with spatiotemporal metadata (collection time and geographic origin), they can be used to infer potential transmission chains and support outbreak investigations.



SUPPLEMENTARY FIGURE S1. Distribution of 4,855 China PIN-derived sequences across 29 PLADs and 169 cities in China.
Abbreviation: PLAD=provincial-level administrative division.

REFERENCES

1. Fan GM, Guo CY, Zhang Q, Liu DM, Sun QL, Cui ZG, et al. A secure visualization platform for pathogenic genome analysis with an accurate reference database. Biosafety Health 2024;6(4):235 – 43. https://doi.org/10.1016/j.bsheal.2024.07.003.
2. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol 2019;20(1):257. https://doi.org/10.1186/s13059-019-1891-0.
3. Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucl Acids Res 2007;35(9):3100 – 8. https://doi.org/10.1093/nar/gkm160.
4. Edgar RC. PILER-CR: fast and accurate identification of CRISPR repeats. BMC Bioinformatics 2007;8(1):18. https://doi.org/10.1186/1471-2105-8-18.
5. Gant TW, Sauer UG, Zhang SD, Chorley BN, Hackermüller J, Perdichizzi S, et al. A generic Transcriptomics Reporting Framework (TRF) for 'omics data processing and analysis. Regul Toxicol Pharmacol 2017;91(S1):S36 – 45. https://doi.org/10.1016/j.yrtph.2017.11.001.
6. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2. 0: improved detection and functional classification of transfer RNA genes. Nucl Acids Res 2021;49(16):9077 – 96. https://doi.org/10.1093/nar/gkab688.
7. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11(1):119. https://doi.org/10.1186/1471-2105-11-119.
8. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucl Acids Res 2000;28(1):27 – 30. https://doi.org/10.1093/nar/28.1.27.
9. Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, Koonin EV. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucl Acids Res 2021;49(D1):D274 – 81. https://doi.org/10.1093/nar/gkaa1018.
10. Sayers EW, Cavanaugh M, Clark K, Pruitt KD, Schoch CL, Sherry ST, et al. GenBank. Nucl Acids Res 2022;50(D1):D161 – 4. https://doi.org/10.1093/nar/gkab1135.
11. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucl Acids Res 2020;48(D1):D517 – 25. https://doi.org/10.1093/nar/gkz935.
12. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucl Acids Res 2014;42(D1):D490 – 5. https://doi.org/10.1093/nar/gkt1178.
13. Urban M, Cuzick A, Seager J, Wood V, Rutherford K, Venkatesh SY, et al. PHI-base: the pathogen-host interactions database. Nucl Acids Res 2020;48(D1):D613-20. http://dx.doi.org/10.1093/nar/gkz904.
14. McMillan LEM, Martin ACR. Automatically extracting functionally equivalent proteins from SwissProt. BMC Bioinformatics 2008;9(1):418. https://doi.org/10.1186/1471-2105-9-418.
15. Liu B, Zheng DD, Zhou SY, Chen LH, Yang J. VFDB 2022: a general classification scheme for bacterial virulence factors. Nucl Acids Res 2022;50(D1):D912 – 7. https://doi.org/10.1093/nar/gkab1107.
16. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: the protein families database in 2021. Nucl Acids Res 2021;49(D1):D412 – 9. https://doi.org/10.1093/nar/gkaa913.
17. Karp PD, Riley M, Paley SM, Pellegrini-Toole A. The MetaCyc database. Nucl Acids Res 2002;30(1):59 – 61. https://doi.org/10.1093/nar/30.1.59.
18. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, Van Wezel GP, Medema MH, et al. antiSMASH 6. 0: improving cluster detection and comparison capabilities. Nucl Acids Res 2021;49(W1):W29 – 35. https://doi.org/10.1093/nar/gkab335.
19. Castellano S, Cestari F, Faglioni G, Tenedini E, Marino M, Artuso L, et al. iVar, an interpretation-oriented tool to manage the update and revision of variant annotation and classification. Genes 2021;12(3):384. https://doi.org/10.3390/genes12030384.
20. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucl Acids Res 2015;43(3):e15. https://doi.org/10.1093/nar/gku1196.
21. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 2020;37(5):1530 – 4. https://doi.org/10.1093/molbev/msaa015.
22. Lam MMC, Wick RR, Watts SC, Cerdeira LT, Wyres KL, Holt KE. A genomic surveillance framework and genotyping tool for *Klebsiella pneumoniae* and its related species complex. Nat Commun 2021;12(1):4188. https://doi.org/10.1038/s41467-021-24448-3.
23. Xie ZQ, Tang HX. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. Bioinformatics 2017;33(21):3340 – 7. https://doi.org/10.1093/bioinformatics/btx433.
24. Johansson MHK, Bortolaia V, Tansirichaiya S, Aarestrup FM, Roberts AP, Petersen TN. Detection of mobile genetic elements associated with antibiotic resistance in *Salmonella enterica* using a newly developed web tool: MobileElementFinder. J Antimicrob Chemother 2021;76(1):101 – 9. https://doi.org/10.1093/jac/dkaa390.
25. Silva M, Machado MP, Silva DN, Rossi M, Moran-Gilad J, Santos S, et al. chewBBACA: a complete suite for gene-by-gene schema creation and strain identification. Microb Genom 2018;4(3):e000166. https://doi.org/10.1099/mgen.0.000166.