Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Global genomic landscape of hypervirulent Klebsiella pneumoniae: hypervirulence biomarkers, mobile virulence elements, and antimicrobial resistance

Sergio Morgado

sergio.morgado@ioc.fiocruz.br

Instituto Oswaldo Cruz, FIOCRUZ

Erica Fonseca

Instituto Oswaldo Cruz, FIOCRUZ

Ana Carolina Vicente

Instituto Oswaldo Cruz, FIOCRUZ

Research Article

Keywords: hvKp, Epidemiology, Biomarkers, Convergence, Spatiotemporal, Virulence, ICEKp

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11424210/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Hypervirulent *Klebsiella pneumoniae* (hvKp) is an important cause of invasive infections and has raised increasing concern because of its potential convergence with antimicrobial resistance. However, estimates of hvKp prevalence vary substantially among studies because of heterogeneous genomic definitions. Here, we assessed the global genomic epidemiology of hvKp using a stringent five-biomarker definition and investigated its temporal distribution, antimicrobial resistance profile, and association with additional virulence-associated mobile elements, such as integrative and conjugative elements (ICEs).

Methods

A comprehensive dataset of 79,746 non-redundant *K. pneumoniae* genomes retrieved from GenBank was analyzed. The hvKp classification required the simultaneous presence of five plasmid-encoded biomarkers: *iroB*, *iucA*, *peg-344*, *rmpA*, and *rmpA2*. Bioinformatic tools were utilized to determine sequence types (STs), capsular polysaccharides (KL), genomic islands, and resistome profiles, alongside spatiotemporal statistical analyses.

Results

The strict definition identified 3,447 hvKp genomes (4.32%). The major hvKp lineages were ST23, ST65, ST11, ST86, and ST412, primarily exhibiting KL1 and KL2 capsules. Temporally, hvKp genomes spanned from 1911 to 2024; however, the relative proportion of hvKp has remained stable in the last decade. Approximately 65% of hvKp genomes lacked acquired antimicrobial resistance genes, whereas 25% exhibited multidrug resistance and 19% carried carbapenemase genes. ICEKp elements carrying yersiniabactin siderophore were identified in 75% of hvKp genomes, predominantly ICEKp10 and ICEKp3, and 67% of ICEKp-positive genomes also carried the colibactin genotoxin locus.

Conclusions

The prevalence of genomically defined hvKp is strongly influenced by the biomarkers used for classification, with a stringent five-biomarker criterion yielding substantially lower estimates than studies using fewer markers. Our findings further indicate that hvKp is a long-standing and globally distributed pathotype, while its relative prevalence and antimicrobial resistance burden have remained largely stable in the contemporary genomic era. It was noticed a convergence between hvKp and virulence-associated ICEKp elements, highlighting the complex and lineage-specific architecture of virulence in hvKp.

Background

Klebsiella pneumoniae has posed as one of the most critical antimicrobial-resistant bacterial pathogens worldwide and represents one of the fastest-growing threats associated with antibiotic resistance [1]. In addition, it ranks among the leading causes of mortality attributed to antimicrobial-resistant infections. In this context, carbapenem-resistant *K. pneumoniae* has been classified by the World Health Organization (WHO) as a critical priority pathogen, underscoring the urgent need for enhanced surveillance, infection control measures, and effective therapeutic strategies [2].

Hypervirulent *K. pneumoniae* (hvKp) emerged as a pathotype exhibiting increased virulence traits compared with classical *K. pneumoniae* (cKp), being characterized by a hypermucoviscous phenotype and initially identified in the 1980s [3]. Since then, hvKp has increasingly been reported worldwide, particularly in Asia, causing a variety of infections [4]. This hypervirulent has been attributed to pLVPK-like plasmids, which harbor critical virulence genes for this phenotype [5, 6]. Although virulent, hvKp is usually associated with low levels of antimicrobial resistance [7].

Conversely, different from hvKp strains, multidrug-resistant cKp strains typically lack or have very few virulence traits. However, this dichotomy appears to have shifted in recent years, as evidenced by reports of MDR-hvKp strains [8–10]. These high-risk clones presenting hypervirulence and multidrug resistance traits represent a significant public health concern [11]. This convergence can occur through multiple evolutionary pathways, including the acquisition of resistance plasmids by hvKp, the acquisition of virulence plasmids by MDR cKp, or the emergence of hybrid plasmids carrying both resistance and virulence determinants in either hvKp or MDR cKp strains [12].

In addition to plasmid-mediated evolution, other mobile genetic elements significantly shape the pathogenic landscape of these high-risk clones. Integrative and conjugative elements of *K. pneumoniae* (ICEKp) are characterized by the presence of the *ybt* locus, which is involved in the biosynthesis of yersiniabactin, an important siderophore that contributes to *K. pneumoniae* virulence [13]. Several ICEKp variants may harbor additional virulence traits, such as ICEKp1, which carries the salmochelin siderophore (*iro* locus) and a variant of the capsule regulator *rmpA*; and ICEKp10, which contains the genotoxin colibactin (*clb/pks* locus) [13, 14]. Experimental studies have shown that yersiniabactin enhances bacterial growth and persistence during pulmonary infection, while colibactin induces host-cell DNA damage and promotes intestinal colonization, systemic dissemination, and central nervous system invasion. Moreover, a recent study demonstrated that ICEKp10 may promote the killing of commensal bacterial taxa during interspecies interactions in the gut [15]. Although ICEKp variants are not the primary determinants of the classical hypervirulence phenotype, ICEKp acquisition by an hvKp lineage can substantially amplify its pathogenic potential and invasiveness [16].

Given this complex and overlapping landscape of virulence determinants, defining true hvKp remains challenging due to the absence of a standardized consensus. While murine models represent the gold standard, their application is impractical for large-scale screening. Currently, the most effective alternative relies on detecting the simultaneous presence of five plasmid-encoded biomarkers: *rmpA*, *rmpA2*, *iroB*, *iucA*, and *peg-344*, which strongly indicates the possession of an intact hvKp-associated

virulence plasmid [17]. However, several studies continue to report hvKp without applying rigorous definitions [13, 18–24], leaving the global epidemiological landscape of hvKp incomplete. Therefore, our aim was to address this gap by performing a large-scale genomic epidemiological analysis of all publicly available *K. pneumoniae* genomes to identify and characterize those harboring the five hypervirulence biomarkers.

Methods

A comprehensive dataset of *Klebsiella* pneumoniae genomes was retrieved from GenBank (n = 80,411; January 2025) and analyzed together with the available metadata. After removing redundant assemblies (same biosample), a total of 79,746 genomes were retained for analysis. The classification of genomes as hypervirulent was based on the simultaneous presence of five biomarkers (*iro*B, *iuc*A, *peg*-344, *rmp*A, and *rmp*A2), as proposed by Russo et al. (2025).

A search for virulence and resistance genes was conducted using abricate v1.0.1 (http://github.com/tseemann/abricate) with the VFDB and CARD databases, respectively, and Kleborate v3.1.3 [26], which was also utilized to identify the capsular polysaccharide (KL). The sequence type (ST) was obtained through mlst software (https://github.com/tseemann/mlst). Gene prediction and annotation were performed using Bakta v1.11.0 [27] and Prokka v1.14.6 [28]. The search for integrons was conducted using IntegronFinder v2.0.5 [29].

Phylogenetic reconstruction was performed by identifying the *K. pneumoniae* core genome with Roary v3.13.0 [30], extracting single-nucleotide polymorphisms from the concatenated core genes via snp-sites v2.5.1 [31], and generating a maximum-likelihood tree using IQ-TREE 2 v2.4.0 [32] with 1,000 ultrafast bootstrap replicates. Tree visualization was conducted on the iTOL web platform [33].

Statistical analyses were performed using R v4.5.1. Proportional differences were assessed using chi-square tests of independence or Fisher's exact test when expected counts were small. Temporal associations were evaluated using Spearman's rank correlation and linear regression. For comparison of proportions between studies, a two-proportion z-test was used for the global comparison. Significance was set at $p < 0.05$.

Results

3.1. Genomic features and epidemiology of hvKp

Screening of *K. pneumoniae* genomes from Genbank (n = 79,746) identified 3,447 genomes harboring the five hypervirulence biomarkers. The majority of hvKp genomes was observed in humans (n = 3,150), with smaller numbers detected in animals (n = 17) and the environment (n = 19) (Table S1). Phylogenomic analysis of these non-human genomes alongside clinical genomes revealed a clear intermingling of strains, highlighting the One Health dimension of hvKp transmission (Figure S1). The 3,447 hvKp genomes encompassed 98 STs, most of them (1,386/3,447, ~ 40%) was attributed to ST23, followed by

ST65 (n = 321, ~ 9%), ST11 (n = 242, ~ 7%), ST86 (n = 233, ~ 6%), and ST412 (n = 174, ~ 5%) (Table S1). The proportion of hvKp per ST was highly variable, with higher prevalence in clonal-groups 23, 65, 86, 218, 412, 592, 1764 (Fig. 1). Indeed, clonal-group 23 (CG23) is the most common associated with the hvKp pathotype [7]. *K. pneumoniae* ST11 is typically associated with multidrug resistance [34]. Interestingly, while it represents the second most common ST among hvKp in our dataset, only a minor fraction of the overall ST11 genomes exhibit the hvKp genotype (Fig. 1). With respect to the capsule locus of the hvKp, most cases exhibited KL1 (n = 1,492; ~43%), followed by KL2 (n = 794; ~23%) and KL57 (n = 383; ~11%) (Table S1). KL1 was present in 24 STs, but was predominantly found in ST23 (95%), while the KL2 was identified in 20 STs, primarily distributed in ST65 (42%) and ST86 (30%).

The 3,447 hvKp genomes encompassed all continents, covering 66 countries and regions (Africa: n = 32; Asia: n = 2,577; Europe: n = 337; North America: 89; Oceania: n = 44; Latin America: n = 9; and not recorded: n = 359). The prevalence of hvKp in Asia is notable, with China (n = 1,763), Japan (n = 222), Singapore (n = 189), and Taiwan (n = 117) being the top source countries of these genomes. These are followed by Ireland (n = 81) and the USA (n = 78) (Table S1). Indeed, the proportion of hvKp in Asia (2,577/29,365 ~ 8.7%) is higher than in all other continents combined (870/50,381; ~1.7%), which supports the interpretation of Asia as a hotspot in our dataset (χ^2=2,228; p < 2.2e-16; Z = 47.2).

The hvKp genomes with available temporal metadata (2,838/3,447) span the period from 1911 to 2024 (Table S1). A notable increase in the number of hvKp genomes was observed over the years, reaching a pronounced peak in 2021, predominantly driven by samples from Asia, followed by a gradual decline through 2024 (Fig. 2A).

Analysis of all *K. pneumoniae* genomes with available temporal metadata (n = 71,509/79,746) (1900–2025), revealed a significant positive correlation between year and the proportion of hvKp genomes (Spearman's ρ = 0.36, p = 0.009), indicating a long-term increase in hvKp frequency. However, when restricting the analysis to the last decade (≥ 2014), the correlation was not significant (ρ=-0.05, p = 0.86), suggesting that the apparent increase mainly reflects the expansion of genome numbers rather than a genuine epidemiological rise (Fig. 2B). Indeed, when analyzing global regions separately (Africa, Asia excluding China, China, Europe, Latin America, North America, and Oceania) from 2014 onwards, no significant temporal associations were detected, indicating overall stability in hvKp proportions. The only exception was a negative correlation observed for Asia (excluding China) (ρ=-0.79, p = 0.003), suggesting a decline in the relative frequency of hvKp genomes in this region during recent years. In our dataset, hvKp was identified in 4.32% (3,447/79,746) of genomes. Although China exhibited a regional peak of 10.62%, other global regions accounted for less than 2% (Table 1 and Fig. 3).

Table 1
Comparative of the proportion (%) of hvKp worldwide between the two studies

Regions	hvKp proportions (%)	
	This study (n = 3,447/79,746)	Heng et al., (2024) (n = 1,955/16,797)
Latin America	0.46	0.82
Africa	1.16	1.85
Asia-Pacific (China excluded)	5.74	14.58
China	10.62	38.39
North America	0.50	1.57
Europe and Central Asia	1.57	5.87
Overall	4.32	11.99

3.2. Resistome of hvKp

Of the 3,447 hvKp genomes, most (n = 2,271; ~65%) lacked acquired antibiotic resistance genes (ARGs), whereas ~ 25% (n = 881) exhibited a multidrug-resistant (MDR) genotype (carrying acquired ARGs against at least three antibiotic classes; Table S1 and Table S2); furthermore, over half of these MDR isolates (n = 493; ~14% of the total hvKp) carried carbapenemase genes (bla_{IMP}, bla_{KPC}, bla_{NDM}, bla_{VIM-1}, and bla_{OXA}; Table S2). The majority of MDR-hvKp belonged to ST23 (n = 269/1,386) and ST11 (n = 242/242), mainly presenting the KL1 (n = 294), KL64 (n = 239), or KL2 (n = 140) capsule (Table S1). In fact, ST11 was the main top STs exhibiting high proportions of MDR genotype (Fig. 4). Moreover, distinct ST11 lineages were observed, with genomes from China (n = 221) clustering separately from those from Taiwan (n = 16) and Italy (n = 1) (Figure S2).

Curiously, the majority of MDR-hvKp ST11 genomes carried carbapenemase genes (predominantly bla_{KPC-2}), whereas most MDR-hvKp ST23 genomes did not harbor carbapenemases (Table S2). The geographical distribution of the MDR-hvKp included China (n = 441), Japan (n = 75), Czechia (n = 66), and Singapore (n = 65), with higher occurrence between 2016–2019 (Table S1). With respect to current clinically relevant ARGs, the first occurrence of carbapenemases was in 2008 (human/China), and 667/3,447 (~ 19%) hvKp genomes carried one or two carbapenemase genes, 17/3,447 (< 1%) carried the mobilized colistin resistance gene (*mcr*), and six carried both carbapenemase and *mcr* genes (Table S2). The details of these ARGs are in supplementary information.

While most hvKp genomes do not harbor acquired ARGs, since 2002 some genomes from distinct STs have acquired resistomes. From 2013 onward, there was a marked increase in the number of MDR genomes (Table 2 and Fig. 5). To investigate temporal trends in the resistome, we analyzed whether the

average and median number of resistance genes per hvKp genome, as well as the number of hvKp genomes exhibiting MDR genotype, changed over time (Table 2 and Fig. 5). Spearman's rank correlation analysis showed no significant association between the average ($\rho = 0.29$, $p = 0.17$) or median ($\rho = 0.08$, $p = 0.69$) number of genes and year, indicating that resistome size per genome has not increased significantly over time. Although the number of hvKp genomes carrying ARGs showed a strong positive correlation with year ($\rho = 0.83$, $p = 8.46e\text{-}07$), the proportion of hvKp genomes with ARGs did not exhibit a significant temporal association ($\rho = 0.29$, $p = 0.17$). A similar pattern was observed for MDR-hvKp genomes, where the number ($\rho = 0.81$, $p = 2.69e\text{-}06$) but not the proportion ($\rho = 0.20$, $p = 0.35$) increased over time.

Table 2
hvKp genomes and their resistome over the years

Years	#Median ARGs	#Average ARGs	#Genomes (n = 2,838/3,447)	#Genomes with ARG (n = 1,023/1,176)	#Genomes MDR (n = 765/881)	Resistomes
1911–2001	-	-	13	0	0	
2002	0	4.4	5	2	2	*aac aadA ant aph armA*
2003	0	0	3	0	0	*cat ctx-m dfr erm floR*
2004	0	1.28	7	1	1	*fosA mcr mphA mphE oxa*
2005	6.5	5.5	8	6	6	*qac*Edelta qnr *sul tem tet*
2006	4	4.25	4	3	3	
2007	0	0.7	30	2	1	
2008	0	1.69	13	4	3	*aac aadA ant aph armA*
2009	0	1.5	30	6	4	*arr cat cmlA cmy ctx-m*
2010	0	5.17	23	7	4	*dfr dha erm floR fosA*
2011	0	2.18	33	5	4	*imp lap mphA mphE oxa*
2012	0	0.86	60	6	4	*qac*Edelta *qnrA sul tem tet veb*
2013	0	2.7	96	39	29	*aac aadA ant aph armA arr cat*
2014	0	2.32	152	47	34	*cmlA cmy ctx-m dfrA dha ereA*
2015	0	3.01	164	53	40	*erm floR fosA imp kpc lap mcr*
2016	1	4	218	112	82	*mphA mphE ndm oxa per*
2017	0	4.24	322	126	105	*qac*Edelta *qnrA sul tem tet veb vim*
2018	2	5.82	207	118	96	*aac aadA ant aph armA arr*

Years	#Median ARGs	#Average ARGs	#Genomes (n = 2,838/3,447)	#Genomes with ARG (n = 1,023/1,176)	#Genomes MDR (n = 765/881)	Resistomes
2019	1	6.12	296	156	125	*cat cml*A *cmy ctx-m dfr*A *dha ere*A
2020	0	2.08	299	85	55	*erm flo*R *fos*A *imp kpc lap mcr*
2021	0	1.51	469	85	63	*mph*A *mph*E *msr*E *ndm oxa*
2022	0	2.37	217	78	52	*qac*Edelta *qnr sul tem tet vim*
2023	0	3.14	154	73	46	*aac aad*A *ant aph arr cat cmy ctx-m dfr*A *flo*R *fos*A *kpc lap mph*A *ndm oxa qac*Edelta1 *qnr sul tem tet*
2024	2	5.6	15	9	6	

The temporal analysis included only genomes with available temporal metadata, with the number of genomes indicated in the column headers. The resistome corresponding to each temporal period is delineated by the colored background.

3.3. Integrative and conjugative elements in hvKp

In addition to the five hvKp biomarkers, several other genes are associated with *K. pneumoniae* virulence (Lam et al., 2018), some of which are mobilized by genomic platforms such as integrative and conjugative elements (ICEs). Overall, 75% (2,618/3,447) of the hvKp genomes harbored ICEKp elements, all of them containing the *ybt* locus. These ICEs predominantly belonged to the ICEKp10 (1,727/2,618; 65%) and ICEKp3 (598/2,618; 22%) variants. Furthermore, 67% (1,769/2,618) of these ICE-carrying genomes co-harbored the *ybt* locus and the colibactin genotoxin locus (Table S2).

These loci displayed a marked lineage-specific distribution (Fig. 6). ST23, the predominant hvKp lineage, exhibited a highly conserved profile characterized by *ybt1*-ICEKp10 and *clb2*, with 97% (1,379/1,421) and 94% (1,328/1,421) of the ST23 genomes carrying these elements, respectively. ST65 was predominantly associated with *ybt17*-ICEKp10 (174/211) and *clb3* (193/211). In contrast, ST11 exhibited a high prevalence of *ybt9*-ICEKp3 (215/245) but completely lacked the *clb* locus. Notably, ST86 demonstrated the highest diversity of ICEKp types (types 3, 4, 5, 6, 7, and 12) but also lacked the colibactin locus (Table S2).

Discussion

Improved accuracy in identifying hvKp is essential for understanding its epidemiology and for assessing the clinical and public health implications of this pathotype. In this study, we performed a large-scale genomic analysis of 79,746 nonredundant *K. pneumoniae* genomes and identified hvKp based on the simultaneous presence of five biomarkers that have been proposed as the most accurate genomic predictors of the hvKp phenotype [17]. This definition identified 3,447 hvKp genomes, corresponding to 4.32% of the total dataset. Although the absolute number of hvKp genomes increased markedly over the study period, analyses restricted to the contemporary genomic era showed no significant increase in the proportion of hvKp genomes over time. Thus, the apparent expansion of hvKp is largely attributable to the increasing availability of *K. pneumoniae* genomes rather than to a recent global increase in the relative frequency of this pathotype. Nevertheless, this global pattern does not exclude regional variation, as some geographic areas may experience localized outbreaks or endemic circulation of hvKp, resulting in prevalence estimates higher than those observed at the global level [24, 35, 36].

The substantially lower prevalence estimated here than in previous large-scale genomic studies highlights the impact of the criteria used to define hvKp [20, 22, 23]. Heng et al. (2024) performed a comprehensive analysis of the *K. pneumoniae* population structure using a dataset of 16,797 genomes, classifying 1,955 genomes (~ 11%) as hvKp based on the presence of at least one of the five biomarkers. In another recent study, Gan et al. (2025) analyzed 55,684 genomes, identifying hvKp in 15.2% of the isolates using only two biomarkers. Consequently, these approaches may have introduced bias, potentially overestimating the prevalence of hvKp. In contrast, our analysis, applying a more stringent criterion requiring all five biomarkers, revealed a significantly lower proportion of hvKp globally (4.32%), including in Asia and China (Table 1 and Fig. 3).

Despite the lower prevalence estimated in our study, the overall population structure of hvKp was consistent with previous reports [20, 22, 37], which observed ST23, ST65, ST11, ST86, and ST412 as the major hvKp lineages, and KL1 and KL2 as the predominant capsular types, except for one study that identified KL64 as the most prevalent [23]. In fact, one of these reports considered the five biomarkers and exhibited similar results, except by the absence of hvKp ST11 in its results [37]. Our dataset comprised 98 STs, compared with 139 reported by Jiang et al. (2025), which may reflect differences in sampling as well as the more stringent definition used here. The strong association between particular clonal groups and the five-marker genotype further indicates that hypervirulence is not uniformly distributed across the *K. pneumoniae* population. For example, although ST11 represented one of the most frequent hvKp STs in our dataset, only a relatively small fraction of the overall ST11 population exhibited the five-marker hvKp genotype. This observation is consistent with reports primarily associating ST11 with antimicrobial resistance rather than hypervirulence, although a subset of ST11 has acquired the hypervirulent genotype [7, 38].

The temporal distribution of hvKp further supports the long-standing presence of this pathotype. Genomes with available temporal metadata spanned 1911–2024, extending the period previously reported for genomic hvKp collections (1932–2021) [13, 22]. Interestingly, the three oldest hvKp genomes in our dataset (1911, 1920, and 1932) indicate that hvKp is not a recently emerged

phenomenon, being present well before the introduction of modern antimicrobial therapy and the global expansion of antimicrobial resistance. Indeed, Lam et al. (2018) estimated through phylogenomic inference that ST23 hvKp has been circulating since the late 1800s. Together, these observations support the view that hypervirulence is an established evolutionary trait of *K. pneumoniae* rather than a recent consequence of antimicrobial selective pressure. Despite this ancient evolutionary origin, the clinical manifestation of the hvKp phenomenon was only formally recognized in the 1980s and has been continuously reported ever since [39]. Interestingly, our analyses also revealed the convergence of hvKp with the ICEKp from 1997 onwards. This convergence generated lineages with an expanded virulence repertoire (e.g., ST23), as ICEKp, which typically encodes the yersiniabactin siderophore, is the most common mobile genetic element associated with invasive *K. pneumoniae* infections [13].

The resistome analysis provides other perspective concerning a convergence. Most hvKp genomes identified here (~ 65%) lacked acquired antimicrobial resistance genes, whereas ~ 25% exhibited an MDR genotype and ~ 19% carried carbapenemase genes. These estimates are lower than those reported by Jiang et al. (2025), who classified 41.58% of their hvKp collection as MDR and 33.29% as carbapenem-resistant, and differ from the 6.89% carbapenem-resistant hvKp reported by Heng et al. (2024). These differences may partly reflect variation in the definitions and datasets used to identify hvKp, emphasizing again the importance of standardized criteria for comparing across studies.

Importantly, our temporal analysis suggests that this convergence has not progressively intensified at the global population level. Although the number of ARG- and MDR-positive hvKp genomes has increased over time, their proportion and resistome size per genome have remained stable, suggesting that ARG acquisition in hvKp is driven by sporadic events (e.g., plasmid acquisition). Nevertheless, the identification of carbapenemase-carrying and MDR-hvKp lineages demonstrates that convergence can occur and remains an important clinical concern, particularly when hypervirulence becomes established in successful antimicrobial-resistant lineages. Indeed, the distribution of resistance was strongly lineage dependent. ST11 was particularly notable, as all ST11 genomes classified as hvKp in our dataset also exhibited an MDR genotype, and most carried carbapenemase genes. This association is consistent with previous studies identifying ST11 as a major lineage among carbapenem-resistant and carbapenem-resistant hypervirulent *K. pneumoniae* [20, 22].

Conclusions and limitations

This is the first comprehensive global genomic study of hvKp using all five hypervirulence biomarkers, allowing the most accurate assessment of the prevalence of these pathotypes without overestimation. Our data reveal that hvKp lineages have been circulating for more than a century, while their proportion relative to the overall *K. pneumoniae* population has remained largely stable. Notably, the hvKp pathotype has persisted across time and geographic regions since at least 1911. Overall, our results support the persistence of a virulence-resistance dichotomy within *K. pneumoniae*. Nevertheless, the long-standing presence of hvKp does not preclude the subsequent acquisition of additional virulence-associated

elements. In particular, the convergence of hvKp with ICEKp represents an additional layer of virulence complexity, potentially resulting in an expanded virulence repertoire.

Some limitations should be considered when interpreting the findings. First, sampling bias across time and geography is substantial, with genomes from Asia, particularly China, being overrepresented. As a result, the apparent concentration of hvKp in this region may partly reflect unequal sequencing efforts rather than true epidemiological prevalence. Second, metadata incompleteness and absence of genomes from early decades limited our ability to contextualize temporal and spatial patterns. Third, the quality and completeness of assemblies may lead to underdetection of the biomarkers and other genes.

Abbreviations

ARGs: antibiotic resistance genes

CG23: clonal-group 23

ICEKp: integrative and conjugative elements of *K. pneumoniae*

KL: capsular polysaccharide

MDR: multidrug-resistant

ST: sequence type

WHO: World Health Organization

cKp: classical *Klebsiella pneumoniae*

hvKp: hypervirulent *Klebsiella pneumoniae*

mcr: mobilized colistin resistance gene

Declarations

Funding

This study was financed by FAPERJ - Fundação Carlos Chagas Filho de Amparo à

Pesquisa do Estado do Rio de Janeiro, Processo SEI-260003/010175/2026.

CRediT authorship contribution statement

Sergio M. Morgado: Methodology, Formal analysis, Writing - Original Draft, Writing - Review & Editing. Érica L. Fonseca: Investigation, Writing - Original Draft, Writing - Review & Editing. Ana Carolina Vicente:

Conceptualization, Writing - Original Draft, Writing - Review & Editing. All authors have read and approved the manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication:

Not applicable.

Competing interests

The authors declare none.

Data availability

The authors confirm that the data supporting the findings of this study are available within the article or its supplementary materials.

References

1. Cravo Oliveira Hashiguchi T, Ait Ouakrim D, Padget M et al (2019) Resistance proportions for eight priority antibiotic-bacterium combinations in OECD, EU/EEA and G20 countries 2000 to 2030: a modelling study. Euro Surveill 24(20):1800445. 10.2807/1560-7917.ES.2019.24.20.1800445
2. Chen TA, Chuang YT, Lin CHA, Decade-Long (2024) Review of the Virulence, Resistance, and Epidemiological Risks of Klebsiella pneumoniae in ICUs. Microorganisms 12(12):2548. 10.3390/microorganisms12122548
3. Liu YC, Cheng DL, Lin CL (1986) Klebsiella pneumoniae liver abscess associated with septic endophthalmitis. Arch Intern Med 146:1913–1916
4. Al Ismail D, Campos-Madueno EI, Donà V et al (2025) Hypervirulent Klebsiella pneumoniae (hvKp): Overview, Epidemiology, and Laboratory Detection. Pathog Immun. ; 24;10(1):80–119. https://doi.org/10.20411/pai.v10i1.777
5. Tang HL, Chiang MK, Liou WJ et al (2010) Correlation between Klebsiella pneumoniae carrying pLVPK-derived loci and abscess formation. Eur J Clin Microbiol Infect Dis 29:689–698. https://doi.org/10.1007/s10096-010-0915-1
6. Chen YT, Chang HY, Lai YC et al (2004) Sequencing and analysis of the large virulence plasmid pLVPK of Klebsiella pneumoniae CG43. Gene 337:189–198. https://doi.org/10.1016/j.gene.2004.05.008
7. García-Cobos S, Oteo-Iglesias J, Pérez-Vázquez M (2025) Hypervirulent Klebsiella pneumoniae: Epidemiology outside Asian countries, antibiotic resistance association, methods of detection and

clinical management. Enferm Infecc Microbiol Clin 43:102–109. https://doi.org/10.1016/j.eimce.2024.12.008

8. Russo TA, Marr CM (2019) Hypervirulent Klebsiella pneumoniae. Clin Microbiol Rev 32:e00001. https://doi.org/10.1128/CMR.00001-19. 19
9. Dai P, Hu D (2022) The making of hypervirulent Klebsiella pneumoniae. Clin Lab Anal 36:e24743. https://doi.org/10.1002/jcla.24743
10. Chen J, Zhang H, Liao X et al (2023) Hypervirulent Klebsiella pneumoniae. Infect Drug Resist 16:5243–5249. https://doi.org/10.2147/IDR.S418523
11. Marr CM, Russo TA (2019) Hypervirulent Klebsiella pneumoniae: a new public health threat. Expert Rev Anti Infect Ther 17:71–73. https://doi.org/10.1080/14787210.2019.1555470
12. Tian D, Liu X, Chen W et al (2022) Prevalence of hypervirulent and carbapenem-resistant Klebsiella pneumoniae under divergent evolutionary patterns. Emerg Microbes Infect 11:1936–1949. https://doi.org/10.1080/22221751.2022.2103454
13. Lam MMC, Wyres KL, Duchêne S et al (2018) Population genomics of hypervirulent Klebsiella pneumoniae clonal-group 23 reveals early emergence and rapid global dissemination. Nat Commun 9:2703. https://doi.org/10.1038/s41467-018-05114-7
14. Lin TL, Lee CZ, Hsieh PF, Tsai SF, Wang JT (2008) Characterization of integrative and conjugative element ICEKp1-associated genomic heterogeneity in a Klebsiella pneumoniae strain isolated from a primary liver abscess. J Bacteriol 190(2):515–526. 10.1128/JB.01219-07
15. Tan YH, Arros P, Berríos-Pastén C, Wijaya I, Chu WHW, Chen Y, Cheam G, Mohamed Naim AN, Marcoleta AE, Ravikrishnan A, Nagarajan N, Lagos R, Gan YH (2024) Hypervirulent Klebsiella pneumoniae employs genomic island encoded toxins against bacterial competitors in the gut. ISME J. ;18(1):wrae054. 10.1093/ismejo/wrae054. Erratum in: ISME J. 2024;18(1):wrae143. doi: 10.1093/ismejo/wrae143
16. Wang S, Ding Q, Zhang Y, Zhang A, Wang Q, Wang R, Wang X, Jin L, Ma S, Wang H (2022) Evolution of Virulence, Fitness, and Carbapenem Resistance Transmission in ST23 Hypervirulent Klebsiella pneumoniae with the Capsular Polysaccharide Synthesis Gene wcaJ Inserted via Insertion Sequence Elements. Microbiol Spectr 10(6):e0240022. 10.1128/spectrum.02400-22
17. Russo TA, Alvarado CL, Davies CJ et al (2024) Differentiation of hypervirulent and classical Klebsiella pneumoniae with acquired drug resistance. MBio 15:e0286723. https://doi.org/10.1128/mbio.02867-23
18. Banerjee T, Wangkheimayum J, Sharma S et al (2023) Extensively Drug-Resistant Hypervirulent Klebsiella pneumoniae From a Series of Neonatal Sepsis in a Tertiary Care Hospital, India. Front Med (Lausanne) 8:645955. https://doi.org/10.3389/fmed.2021.645955
19. Spadar A, Perdigão J, Campino S et al (2023) Large-scale genomic analysis of global Klebsiella pneumoniae plasmids reveals multiple simultaneous clusters of carbapenem-resistant hypervirulent strains. Genome Med 15(1):3. https://doi.org/10.1186/s13073-023-01153-y

20. Heng H, Yang X, Ye L et al (2024) Global genomic profiling of Klebsiella pneumoniae: A spatio-temporal population structure analysis. Int J Antimicrob Agents 63:107055. https://doi.org/10.1016/j.ijantimicag.2023.107055
21. Wu Y, Pu F, Yan Z et al (2025) Geographic containment and virulence-resistance trade-offs drive the evolution of hypervirulent Klebsiella pneumoniae. Imeta 4(5):e70077. https://doi.org/10.1002/imt2.70077
22. Jiang X, Li S, Li C, Yin Z, Chen F, Hu L, Lu T, Liu X, Wang Y, Ma G, Wang X, Chen F, Zhou D (2025) The global genomic landscape of hypervirulent Klebsiella pneumoniae from 1932 to 2021. mLife 4(4):378–396. 10.1002/mlf2.70029
23. Gan L, Mao P, Tian Z et al (2025) Higher prevalence of hypervirulent Klebsiella pneumoniae isolates with high-risk multidrug resistance in Asia. J Infect Public Health 18(9):102834. 10.1016/j.jiph.2025.102834
24. Guo Y, Wang S, Zhan L et al (2017) Microbiological and Clinical Characteristics of Hypermucoviscous Klebsiella pneumoniae Isolates Associated with Invasive Infections in China. Front Cell Infect Microbiol. ;7:24. Published 2017 Feb 1. 10.3389/fcimb.2017.00024
25. Russo TA, Lebreton F, McGann PT (2025) A Step Forward in Hypervirulent Klebsiella pneumoniae Diagnostics. Emerg Infect Dis 31:1–3. https://doi.org/10.3201/eid3101.241516
26. Lam MMC, Wick RR, Watts SC et al (2021) A genomic surveillance framework and genotyping tool for Klebsiella pneumoniae and its related species complex. Nat Commun 12:4188. https://doi.org/10.1038/s41467-021-24448-3
27. Schwengers O, Jelonek L, Dieckmann MA et al (2021) Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. Microb Genom 7(11):000685. https://doi.org/10.1099/mgen.0.000685
28. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30(14):2068–2069. 10.1093/bioinformatics/btu153
29. Néron B, Littner E, Haudiquet M et al (2022) IntegronFinder 2.0: Identification and Analysis of Integrons across Bacteria, with a Focus on Antibiotic Resistance in Klebsiella. Microorganisms 10:700. https://doi.org/10.3390/microorganisms10040700
30. Page AJ, Cummins CA, Hunt M et al (2015) Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691–3693. https://doi.org/10.1093/bioinformatics/btv421
31. Page AJ, Taylor B, Delaney AJ et al (2016) SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genom 2:e000056. https://doi.org/10.1099/mgen.0.000056
32. Minh BQ, Schmidt HA, Chernomor O et al (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol 37:1530–1534. https://doi.org/10.1093/molbev/msaa015
33. Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296. https://doi.org/10.1093/nar/gkab301

34. Huang X, Yao X, Hou Y, Zhang D, Xie R, Shi C, Shang Y, Bi H, Song W, Hua L, Li C, Chen H, Wu B, Peng Z (2025) Global trends of antimicrobial resistance and virulence of Klebsiella pneumoniae from different host sources. Commun Med (Lond) 5(1):383. 10.1038/s43856-025-01112-1
35. Ye M, Tu J, Jiang J et al (2016) Clinical and Genomic Analysis of Liver Abscess-Causing Klebsiella pneumoniae Identifies New Liver Abscess-Associated Virulence Genes. Front Cell Infect Microbiol 6:165 Published 2016 Nov 29. 10.3389/fcimb.2016.00165
36. Nagendra D, Suresh V, Udupa AA, Srinivas T, Eshwara VK, Varma M, Prakash P, Rao S, Chaudhuri S (2026) Definitional ambiguity and the dual threat of Hypervirulent Klebsiella pneumoniae infections: a systematic review and meta-analysis. Infection 54(2):645–681. 10.1007/s15010-025-02708-4
37. Tang Y, Du P, Du C, Yang P, Shen N, Russo TA, Liu C (2025) Genomically defined hypervirulent Klebsiella pneumoniae contributed to early-onset increased mortality. Nat Commun 16(1):2096. 10.1038/s41467-025-57379-4
38. Chen T, Ying L, Xiong L, Wang X, Lu P, Wang Y, Shen P, Xiao Y (2024) Understanding carbapenem-resistant hypervirulent Klebsiella pneumoniae: Key virulence factors and evolutionary convergence. hLife 2:611–624. 10.1016/j.hlife.2024.06.005
39. Hetta HF, Alanazi FE, Ali MAS et al (2025) Hypervirulent Klebsiella pneumoniae: Insights into Virulence, Antibiotic Resistance, and Fight Strategies Against a Superbug. Pharmaceuticals (Basel) 18(5):724 Published 2025 May 15. 10.3390/ph18050724

Figures

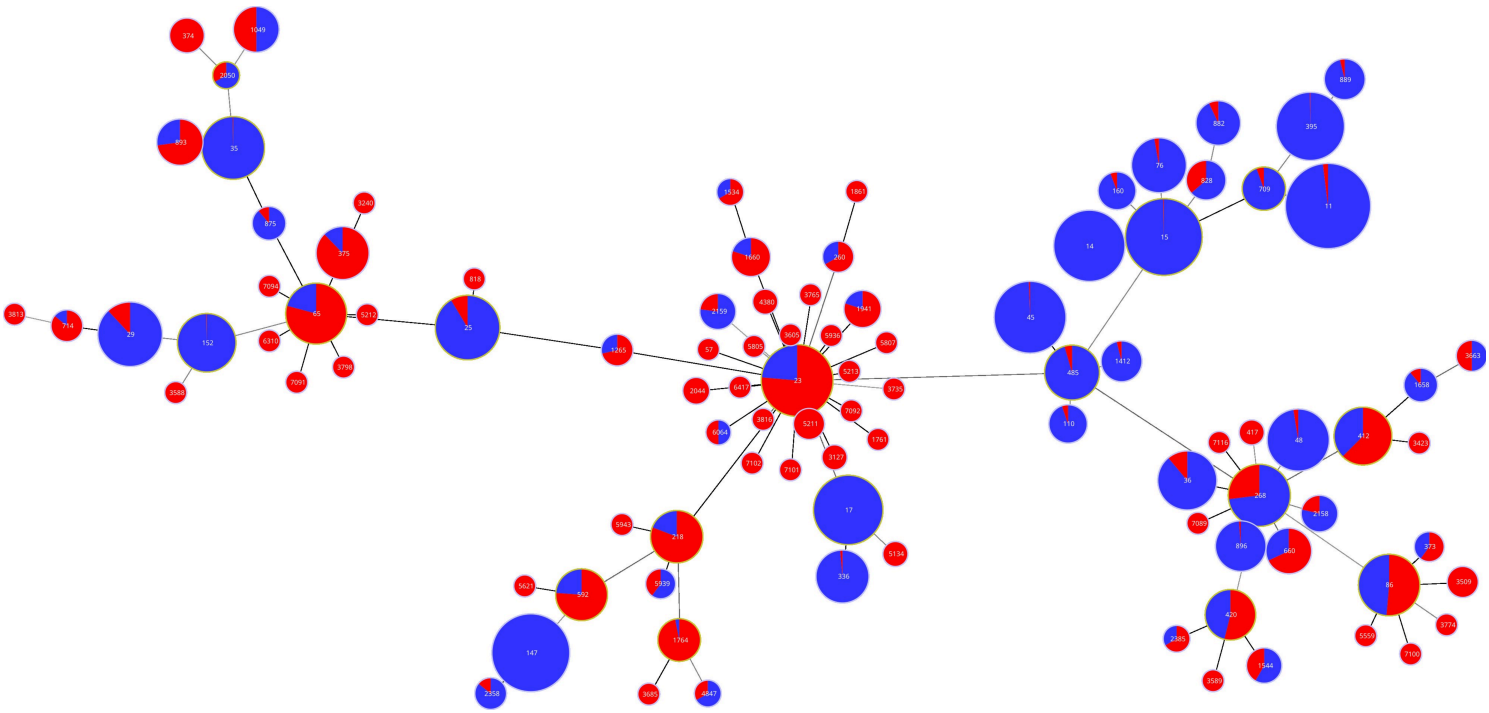


Figure 1

Distribution of hvKp across *K. pneumoniae* STs. A minimum spanning tree based on the allelic profiles of the associated STs shows the proportions of hvKp (red fans) and cKp (blue fans) genomes within each ST. Each circle represents an ST, and the lines indicate clonal groups connecting related STs.

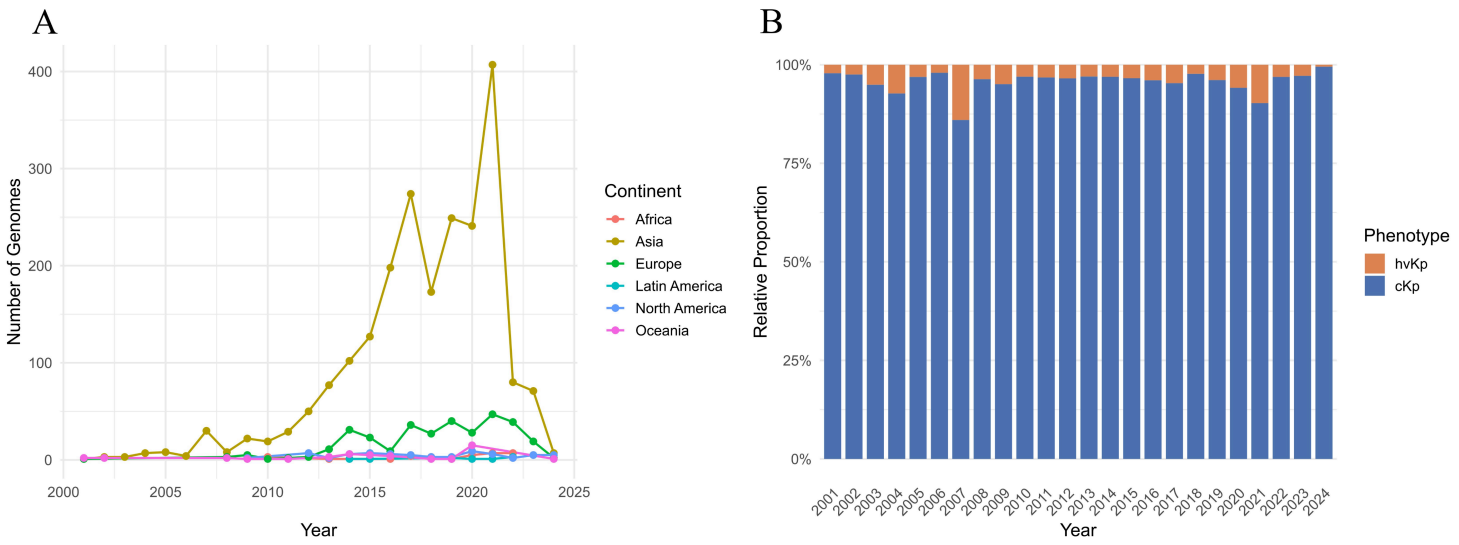


Figure 2

Temporal and geographical distribution of *K. pneumoniae* genomes. (A) Absolute number of hvKp genomes deposited per year, stratified by continent. **(B)** Stacked bar chart illustrating the annual relative proportion of hvKp compared to cKp genomes from 2001 to 2024.

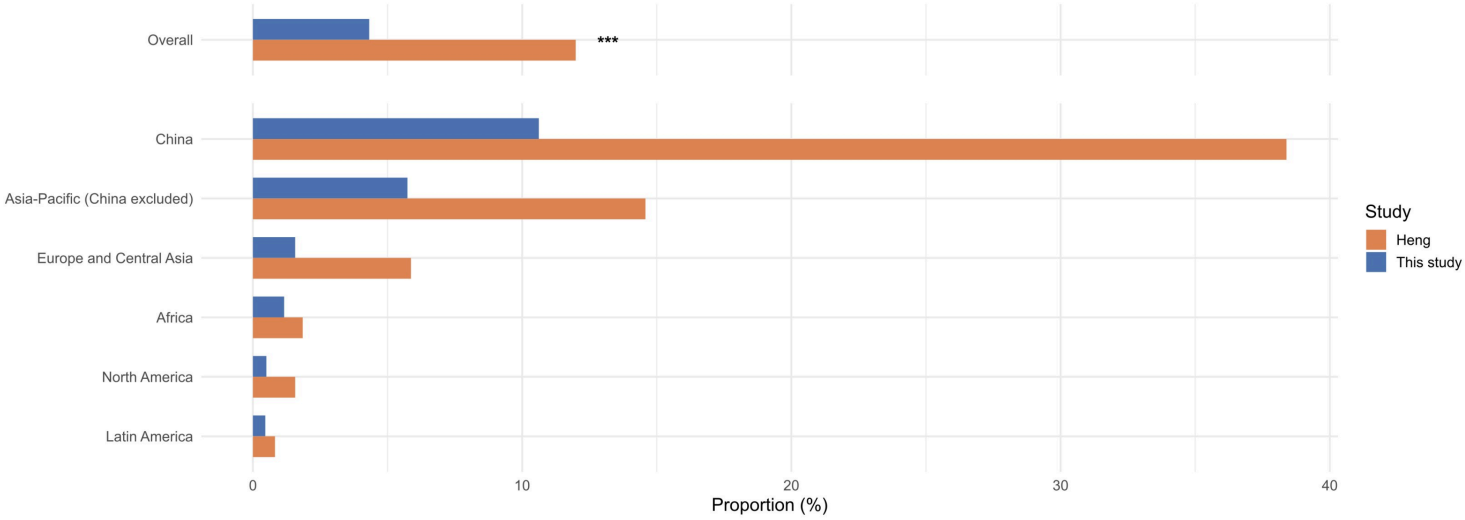


Figure 3

Regional prevalence of hvKp. Bar chart comparing the worldwide proportion of hvKp isolates by geographical region between the dataset from Heng et al. (2024) (orange bars) and the current study (blue bars). The Overall category represents the global proportion across all combined regions, with asterisks denoting statistical significance (*** $p < 0.05$) between the two studies.

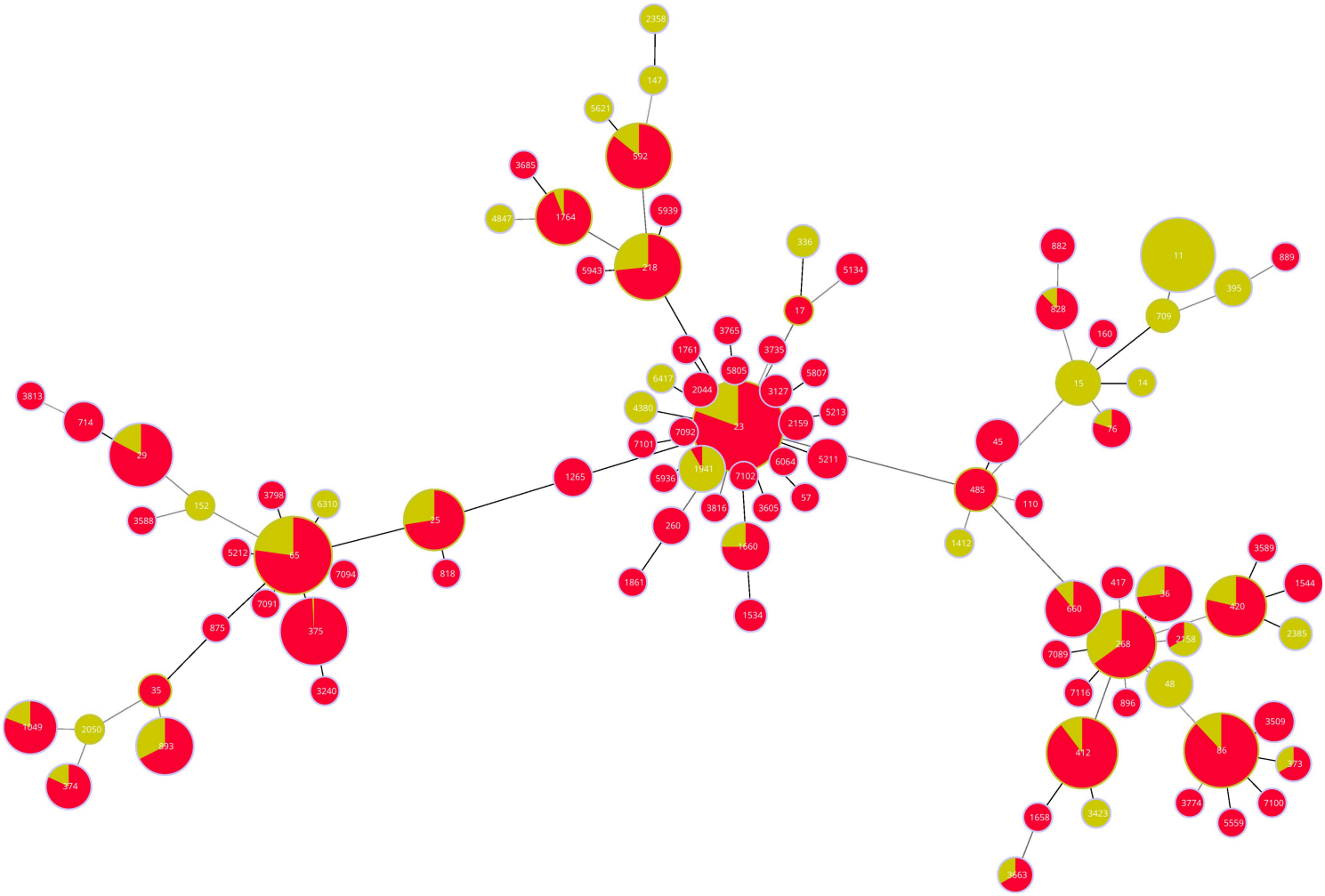


Figure 4

Distribution of MDR among hvKp STs. A minimum spanning tree based on the allelic profiles of the associated STs shows the proportions of MDR hvKp (yellow fans) and non-MDR hvKp (red fans) genomes within each ST. Each circle represents an ST, and the lines indicate clonal groups connecting related STs.

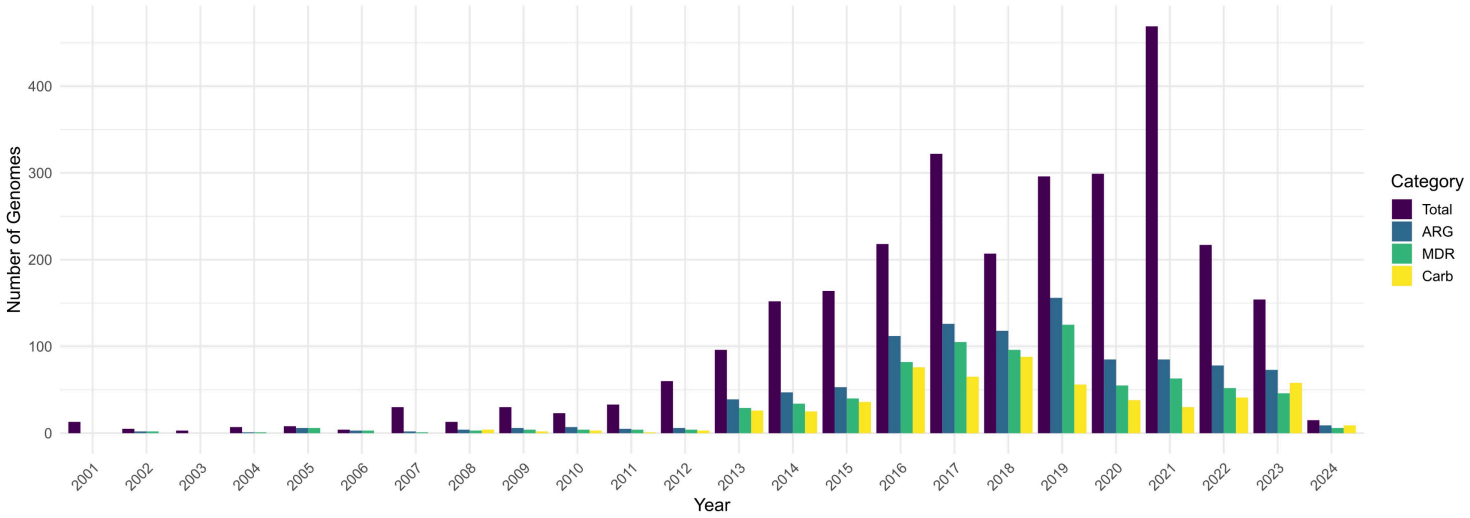


Figure 5

Temporal evolution of resistome profiles in hvKp. Bar chart depicting the absolute number of sequenced hvKp genomes per year from 2001 to 2024. The bars represent the total number of hvKp genomes (Total, purple), hvKp genomes carrying at least one acquired antibiotic resistance gene (ARG, blue), hvKp presenting MDR genotype (MDR, green), and hvKp harboring carbapenemase genes (Carb, yellow).

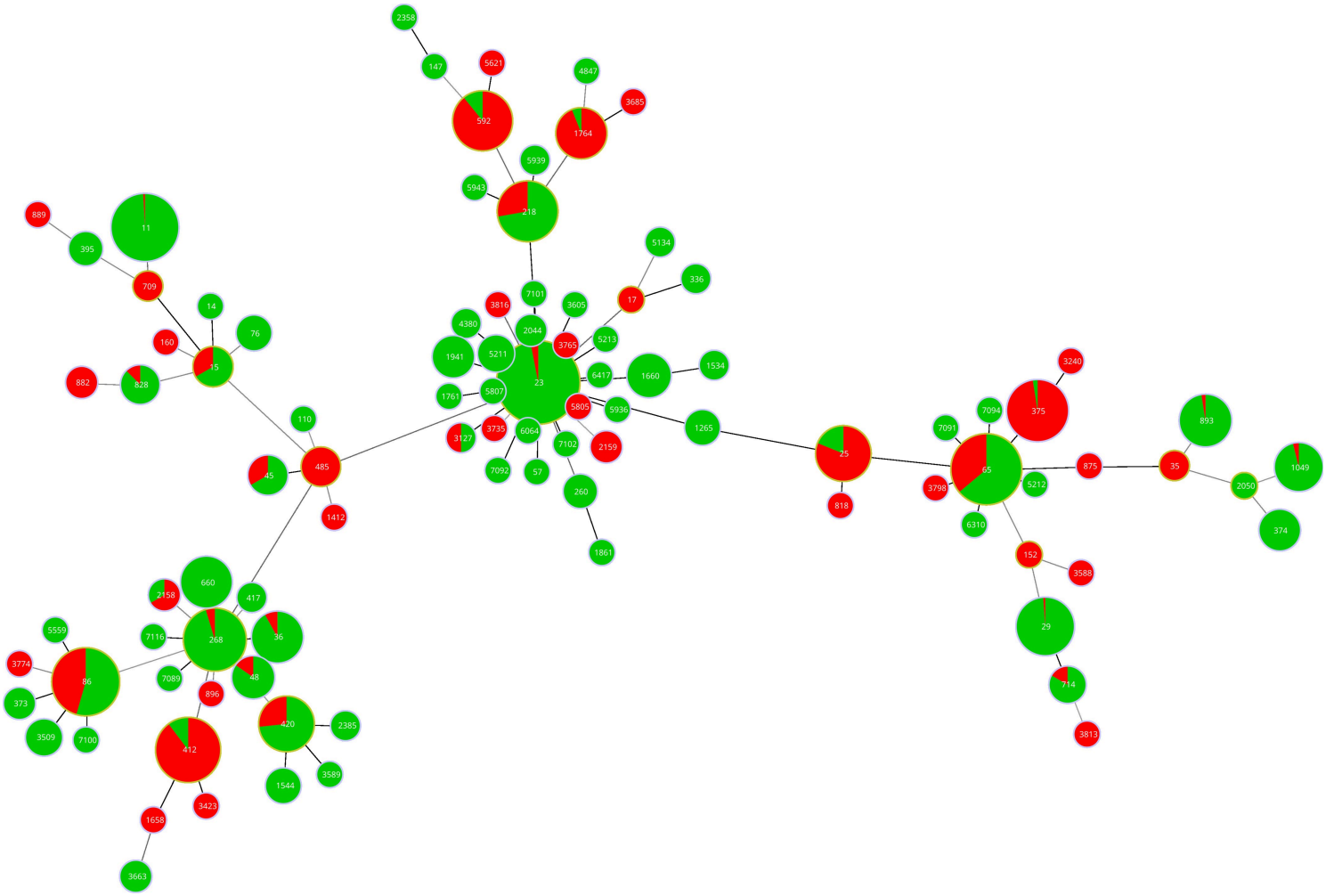


Figure 6

Distribution of ICEs among hvKp STs. A minimum spanning tree based on the allelic profiles of the associated STs shows the proportions of hvKp genomes carrying identified ICEKp (green fans) or lacking identified ICEKp (red fans) within each ST. Each circle represents an ST, and the lines indicate clonal groups connecting related STs.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.xls
- TableS3.xls
- TableS2.xls
- Supplementaryinformation.docx
- FigureS1.pdf
- FigureS2.pdf