

Genomic insights into the rapid rise of *Pseudomonas aeruginosa* ST463: A high-risk lineage's adaptive strategy in China

Xu Dong , Yanghui Xiang^{a*}, Lanjuan Li^a, Ying Zhang^{a,b}, and Tiantian Wu 

^aState Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China; ^bJinan Microecological Biomedicine Shandong Laboratory, Jinan, China

ABSTRACT

High-risk lineages of *Pseudomonas aeruginosa* pose a serious threat to public health, causing severe infections with high mortality rates and limited treatment options. The emergence and rapid spread of the high-risk lineage ST463 in China have further exacerbated this issue. However, the basis of its success in China remains unidentified. In this study, we analyzed a comprehensive dataset of ST463 strains from 2000 to 2023 using whole genome sequencing to unravel the epidemiological characteristics, evolutionary trajectory, and antibiotic resistance profiles. Our findings suggest that ST463 likely originated from a single introduction from North America in 2007, followed by widespread domestic dissemination. Since its introduction, the lineage has undergone significant genomic changes, including the acquisition of three unique regions that enhanced its metabolism and adaptability. Frequent recombination events, along with the burden of bacteriophages, antibiotic resistance genes, and the spread of c1-type (*bla*_{KPC-2}) plasmid-carrying strains, have played crucial roles in its expansion in China. Mutation analysis reveals adaptive responses to antibiotics and selective pressures on key virulence factors, indicating that ST463 is evolving toward a more pathogenic lifestyle.

ARTICLE HISTORY

Received 1 October 2024
Revised 19 December 2024
Accepted 17 April 2025

KEYWORDS

Pseudomonas aeruginosa;
ST463; high-risk lineage;
bacterial evolution;
adaptability

Introduction

Pseudomonas aeruginosa is a ubiquitous Gram-negative bacterium belonging to the family *Pseudomonadaceae* [1]. As an opportunistic pathogen, it is a leading cause of morbidity and mortality in individuals with cystic fibrosis, chronic obstructive pulmonary disease (COPD) and those who are immunocompromised [2]. Due to its extensive intrinsic and acquired resistance mechanisms, *P. aeruginosa*, along with other notorious “ESKAPE” pathogens, was listed by the World Health Organization (WHO) as one of the critical pathogens constituting a major threat to global public health [3,4]. The emergence and spread of carbapenem-resistant *Pseudomonas aeruginosa* (CRPA) strains in recent years have further exacerbated this threat, as carbapenems are often the last resort for treating multidrug-resistant infections [5]. The limited treatment options against CRPA not only impose a heavy financial burden on healthcare systems but also raise widespread concerns about refractory hospital-acquired infections [6].


A recent epidemiological study on global CRPA showed that the most prevalent lineage in China was

CG463 (ST463) [7]. Since its initial report by Ge et al. in 2009 [8], an increasing number of studies have documented the emergence and spread of ST463 carrying the *bla*_{KPC-2} gene within China [9–11]. This Sequence Type (ST) has been found to simultaneously possess the *exoU* and *exoS* genes, which encode potent cytotoxins that contribute to its cytotoxicity and invasiveness. Our previous research has demonstrated the outstanding performance of ST463 in virulence and pathogenesis [12]. With its widespread resistance and significant pathogenicity, understanding the spread and impact of ST463 is now a public health imperative, highlighting the need for vigilant surveillance and immediate action.

Although several previous studies have investigated certain features of ST463, such as its plasmid profiles and drug resistance determinants [13,14], a comprehensive analysis of ST463 strains from China remains absent. Addressing this knowledge gap, our extensive genomic study on ST463 isolates aims to uncover the evolutionary history, transmission dynamics, and the pivotal adaptive mechanisms that

CONTACT Tiantian Wu  tianw@zju.edu.cn; Ying Zhang  yzhang207@zju.edu.cn

*These authors contributed equally to this work.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/21505594.2025.2497901>

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

have facilitated the rise and proliferation of ST463 in Chinese healthcare settings.

Results

The phylogeny of *P. aeruginosa* ST463

We utilized a collection of 275 *P. aeruginosa* ST463 genomes derived from public databases, previously published data, and strains collected by us in 2021–2023 (Supplementary Data 1). Hierarchical Bayesian clustering based on the genetic variation of 275 ST463 genomes delineated them into two major clades (clade 1 and clade 2) and further divided clade 2 into 4 clusters, which agreed with the phylogenetic groupings on the recombination-free maximum

likelihood (ML) phylogeny (Figure 1). The tree topology indicated that the clade 2 was independent of clade 1 and emerged from a single ancestor. Notably, the two clades exhibit marked differences in both size and geographical distribution. Specifically, clade 2 represents nearly the entire ST463 population (265/275, 96.36%), with samples exclusively from China. In contrast, clade 1, though smaller, predominantly consists of samples from North America. Therefore, we herein named the two clades as non_cn (non-China) and cn (China), and the four clusters as cn:1–4.

Consistent with previous reports [15,16], our findings indicate a significant concentration of ST463 in China's Eastern regions, predominantly in Zhejiang. An additional isolate was identified in Yunnan, expanding the known geographical distribution (Figure 1). To

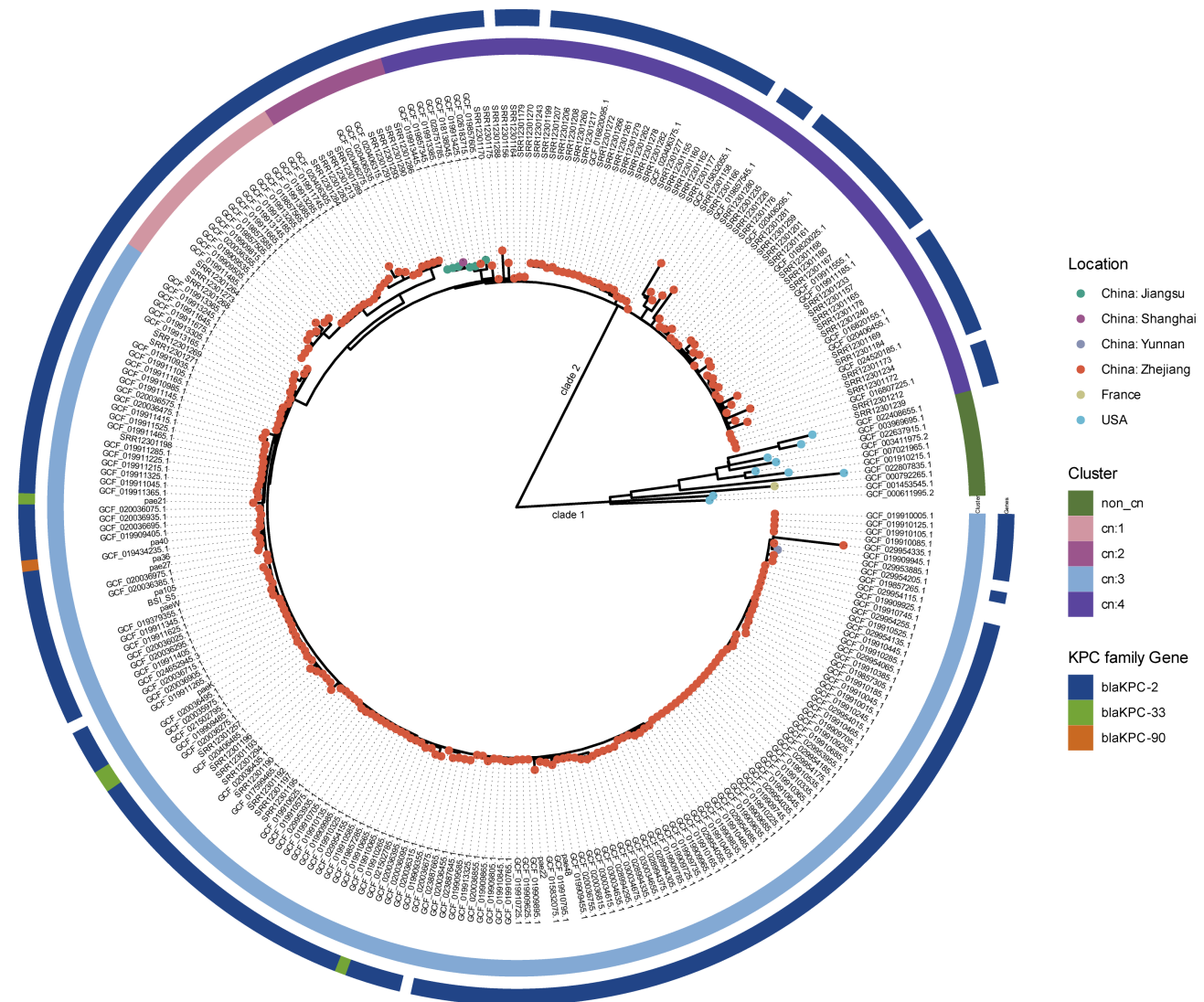


Figure 1. Phylogenetic analysis and geographical distribution of ST463 isolates. The tips of phylogenetic tree of 275 ST463 isolates are colored by their country or city of origin. Clusters are differentiated by colors based on fastbaps analysis. Data on the KPC family genes are mapped onto the tree.

explore the transmission dynamics of ST463 within China, we examined the SNP distribution among the cn clade strains. Our analysis reveals a narrow SNP range among Chinese ST463 strains, spanning from 0 to 116 SNPs with an average of 22 SNPs (Figure S1a). Inferring from the SNP distances, we speculate that the Zhejiang strains are primarily disseminated through hospital networks. Furthermore, our evidence suggests that strains from Jiangsu likely originated from Zhejiang, subsequently experiencing further nosocomial transmissions (Figure S1b).

Temporal dynamics and emergence of ST463 isolates

To elucidate the chronological emergence of the ST463 lineage, a Bayesian framework was applied for analyzing the evolutionary timeline of ST463 isolates. The existence of a significant temporal signal was confirmed by both root-to-tip regression and date randomization tests (Figure S2). Based on this temporal signal, the mean nucleotide substitution rate for the entire ST463 clade was determined to be 2.89×10^{-7} substitutions per site per year [95% highest posterior density (HPD): 2.39×10^{-7} to 3.42×10^{-7}], and we deduce that this clade likely occurred in 1962.8 (95% HPD: 1927.13–1983.37) (Figure 2). Subsequent to the ancestral clade, the non_cn clade was identified, emerging around 1968.08. In contrast, the cn clade exhibited a more recent divergence, with its most recent common ancestor (MRCA) estimated to have appeared around 2007.08 (95% HPD: 2004.63–2008.82). This timeline aligns with prior estimates from an analysis of the evolutionary trajectory of the ST463 clade in China [14]. Notably, the earliest recorded ST463 isolate outside China was identified in 2000 [17], indicating that the Chinese ST463 lineage emerged significantly later than its international counterparts. The internal clustering within cn ST463, specifically cn:4, is believed to have emerged around 2007, with cn:1–3 clusters appearing subsequently around 2013.68, 2014.68, and 2014.89, respectively (Figure 2). To further investigate the population dynamics of this lineage, we employed the skygrowth package in R to infer fluctuations in its effective population size over time. Our analysis revealed that the cn clade underwent an expansion in 2007, reached its peak around 2013, and subsequently maintained a relatively stable population level (Figure S3). The biogeography analysis suggests that North America is the most likely ancestral origin of ST463 isolates, from which these lineages began to diversify (Figure S4). This finding highlights the significantly later emergence of the ST463 lineage in China

compared to the non_cn clade, suggesting a possible scenario of a single introduction event from abroad followed by subsequent domestic dissemination.

Recombination driving the evolution of ST463 in China

Prior research has indicated that recombination played a crucial role in shaping the genomic and phenotypic variability of *P. aeruginosa* during infection [18]. To investigate the reasons behind the emergence and expansion of ST463 in China, we performed recombination analyses and identified a substantial number of recombination events. The extent of recombination differed significantly between the two clades, with the non_cn clade experiencing fewer recombination events (average of 9.4 events) compared to the cn clade, which underwent more recombination events (average of 52.7 events, $p < 0.001$) (Figures 3 and S5a). The overall ratio of base substitutions introduced by homologous recombination relative to point mutations (r/m) was estimated to be 5.31 and 2.56 for the cn clade and non_cn clade, respectively. This suggests that recombination, rather than spontaneous mutation, plays a dominant role in generating diversity within the ST463 population in China.

COG functional analysis of the recombined genes within the cn clade revealed that their involvement across multiple biological processes, from fundamental metabolism to intricate regulatory and signal transduction networks, hinting that recombination events may enhance the adaptability of these strains in China's unique environments (Figure S5b). Notably, the pre-eminent functions observed pertain to transcription, emphasizing the consequential role of genes that regulate gene expression in the recombination process, potentially affecting phenotypic variation and environmental responsiveness of the strains [19]. The presence of numerous genes associated with amino acid transport and metabolism also underscores their significance, which might correlate with bacterial iron metabolism, energy production pathways, and mechanisms of drug resistance. Contrasting with the non-cn clade, where broad-range recombination was detected, cn strains display an exceptionally high recombination density within an integrative conjugative element (ICE) harboring virulence factor *kata* and a genome island (GI) linked to type IV pili (T4P) (Figure 3). KatA, the major catalase, is critical for anti-oxidative defense and virulence [20], while T4P are important for adherence, motility, competence for DNA uptake, and pathogenesis [21]. The concentrated recombination events within these virulence-associated regions suggest that

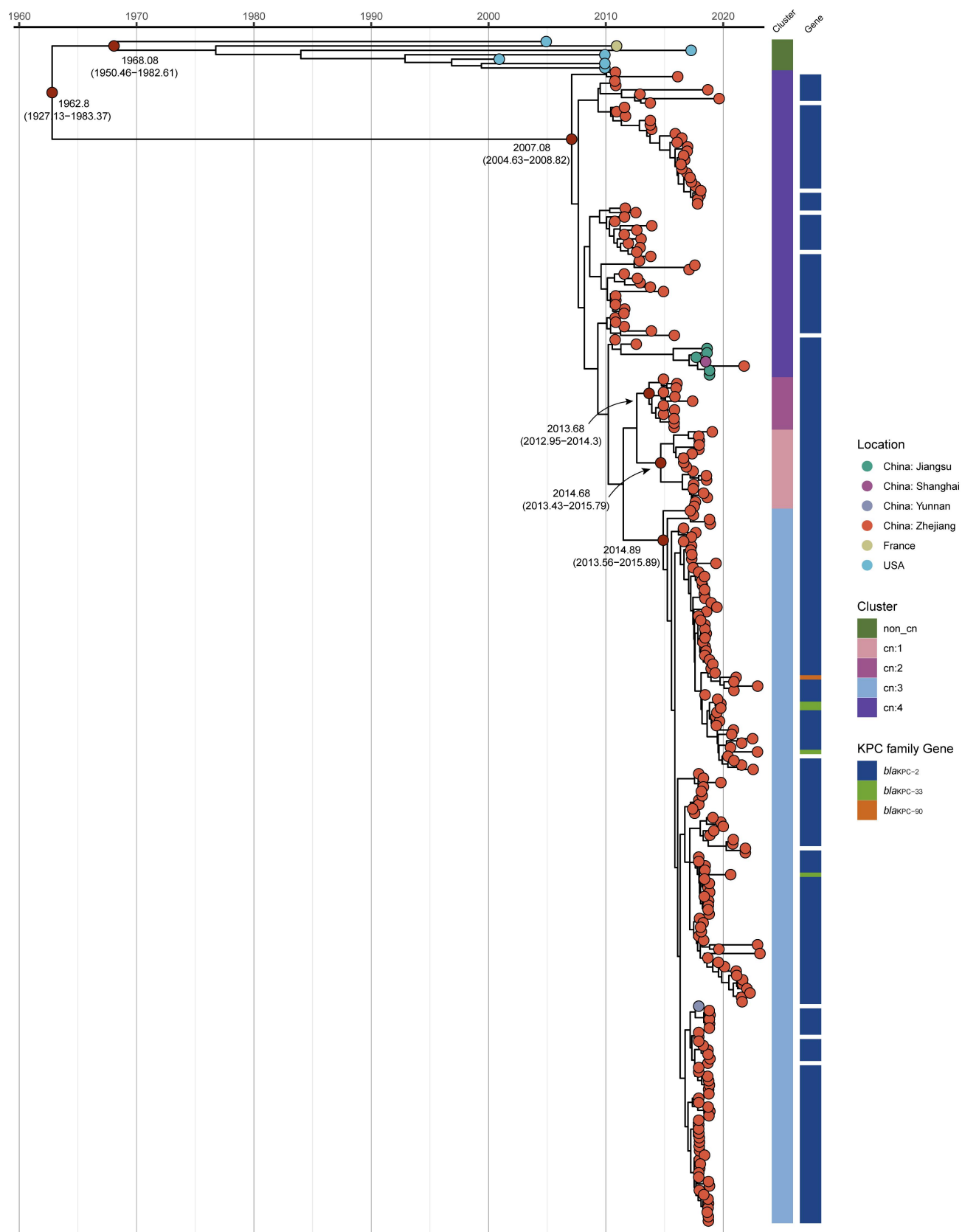


Figure 2. Temporal phylogenetic reconstruction of the ST463 isolates. Characteristics of isolates are shown on the right, including cluster, KPC family genes, and geographical origin (mapped on the tips). The divergence times and corresponding 95% HPD intervals are shown at node of the phylogeny.

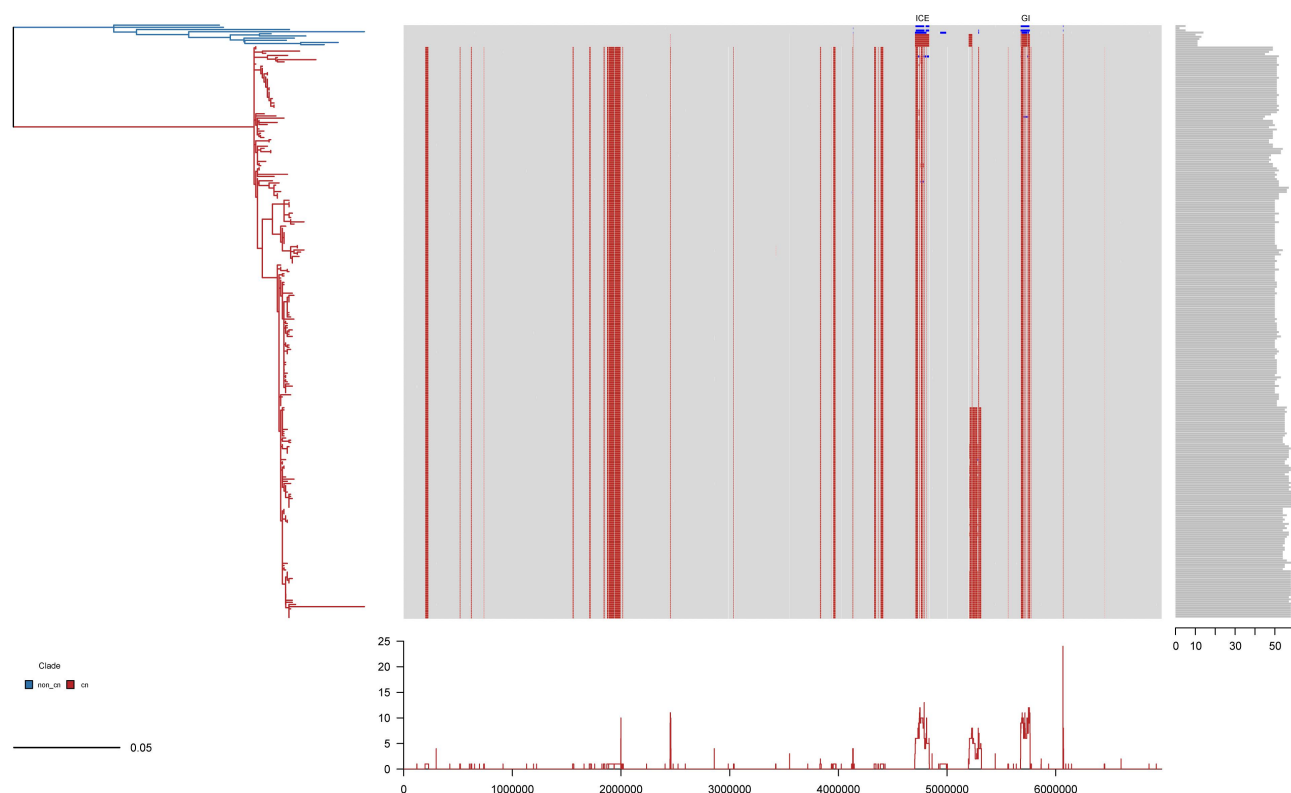


Figure 3. Distribution of recombination events across the ST463. The recombination situation of ST463 is shown on the right side of the phylogenetic tree, which displays the category of each strain's branch and a heatmap of their genomic recombination situations. The heatmap utilizes a color-coded scheme where red indicates recombination detected at non-terminal nodes, and blue signifies events unique to individual isolates. Below the heatmap, a line graph chronicles the recombination event frequency at specific loci along the reference genome. Adjacent to the heatmap, a histogram quantifies the total recombination events recorded for each isolate.

the cn clade may have undergone adaptive evolution to enhance its pathogenic potential and survival fitness in response to specific host-derived selective pressures.

Genomic differences between *cn* and *non-cn* clades

The genomic comparisons between *cn* and *non-cn* clades offer crucial insights into the evolutionary mechanisms that underlie the successful expansion of ST463 in China. Notably, the *cn* clade's loss of a ~64 kb region associated with anti-phage systems (Figure S6) and a *dapA1* (PALES_34361) gene related to the diamminopimelate (DAP) pathway, which is responsible for the biosynthesis of bacterial cell wall and protein components [22]. Intriguingly, the *cn* clade has concurrently acquired three distinct regions, designated here as region 1, region 2, and region 3 (Figure 4a). Region 1 (~11 kb) encompasses genes implicated in fatty acid biosynthesis and has a GC content of 0.61, aligning closely with the genomic GC content (0.66). Region 2 (~61 kb) includes genes associated with the anti-phage ZorAB system [23] and exhibits a markedly lower GC

content of 0.58, suggesting potential acquisition through horizontal gene transfer (HGT). In addition, we identified an OmpA family protein in this region. This protein has an OmpA-like domain, similar to the well-characterized OmpA protein OprF in *Pseudomonas aeruginosa*, suggesting that this gene might be a potential virulence factor. Region 3 (~35 kb), a complete prophage region introduced by a pair of tRNA-Gly genes, harbors the Zeta toxin protein known to disrupt cell wall synthesis [24] (Figure 4b). Its GC content of 0.64, which is closer to the genomic average.

Cn clade carried more prophages

Given the pivotal role of prophages in bacterial evolution and adaptation, we delved into the prophage composition of the two clades. Across the entire ST463 clade, 11 different intact prophages were identified (Table S1), comprising 9 from the *Caudoviricetes* and 2 from the *Inoviridae*. Notably, none of the annotated prophages harbored resistance or virulence genes. In

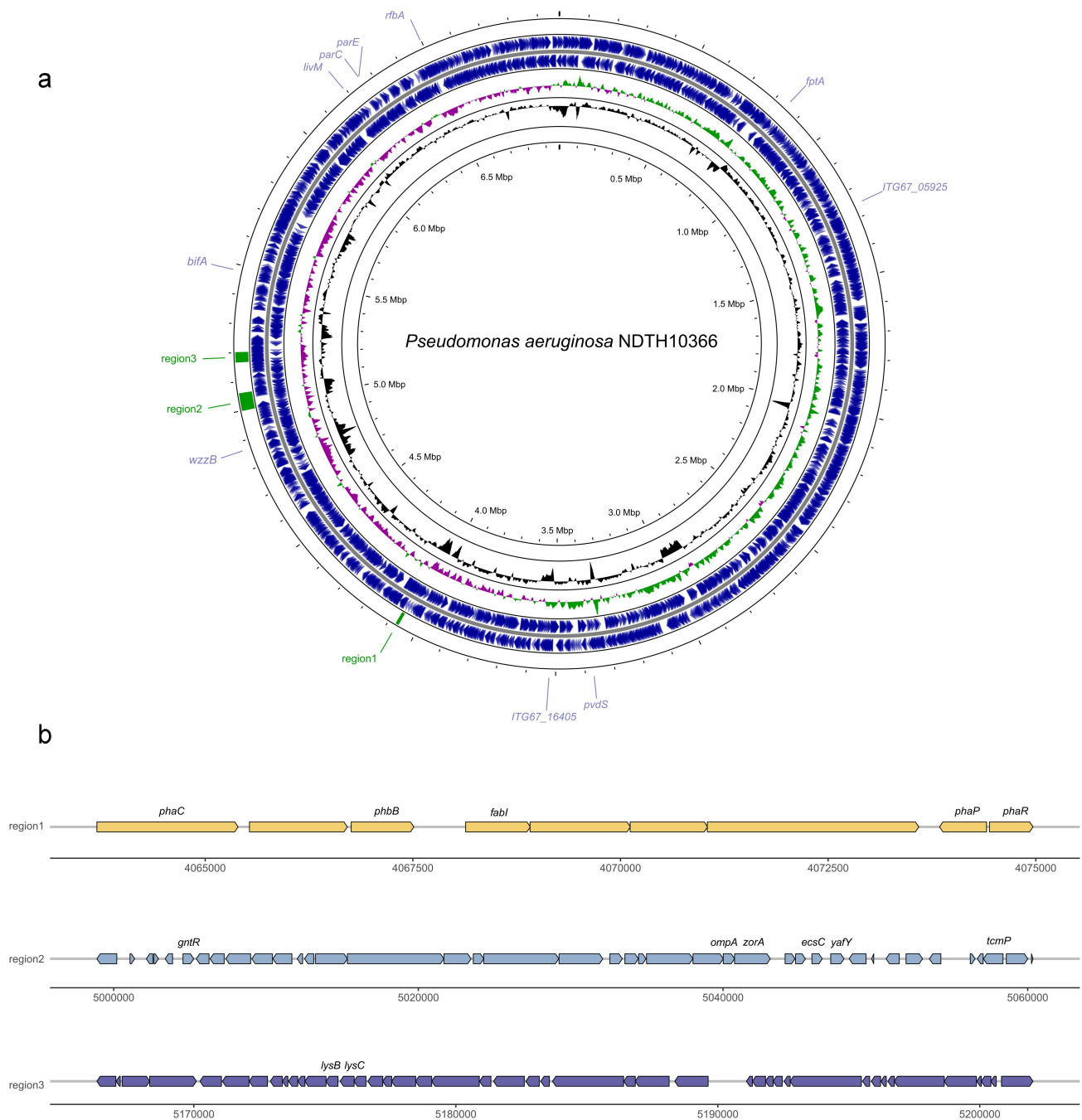


Figure 4. Acquisition regions of the cn clade genome. (a) depicts the positions of the three acquired regions and the positive selection genes on the genome. (b) maps the three regions using gggenes (<https://github.com/wilko/gggenes>), different colors are used to distinguish the three regions.

comparison to the non-cn clade, the cn clade harbored a significantly higher number of prophages (average 4.96 versus 1.1, $p < 0.001$), as illustrated in Figure 5a, which suggests an increased prophage predation burden in Chinese strains. A closer examination into the presence of prophages within the strains indicated that prophages 1–5 were prevalent in most strains, with prophage 3 (i.e.

region 3) found ubiquitously (Figure 5b and Table S1). However, it was observed that some prophages detected in the overall ST463 clade were absent in the cn clade, implying that the cn clade strains, after their introduction into China, faced distinct prophage challenges likely encountered in Chinese hospital settings. In addition, we did not find a known CRISPR/CAS system in ST463.

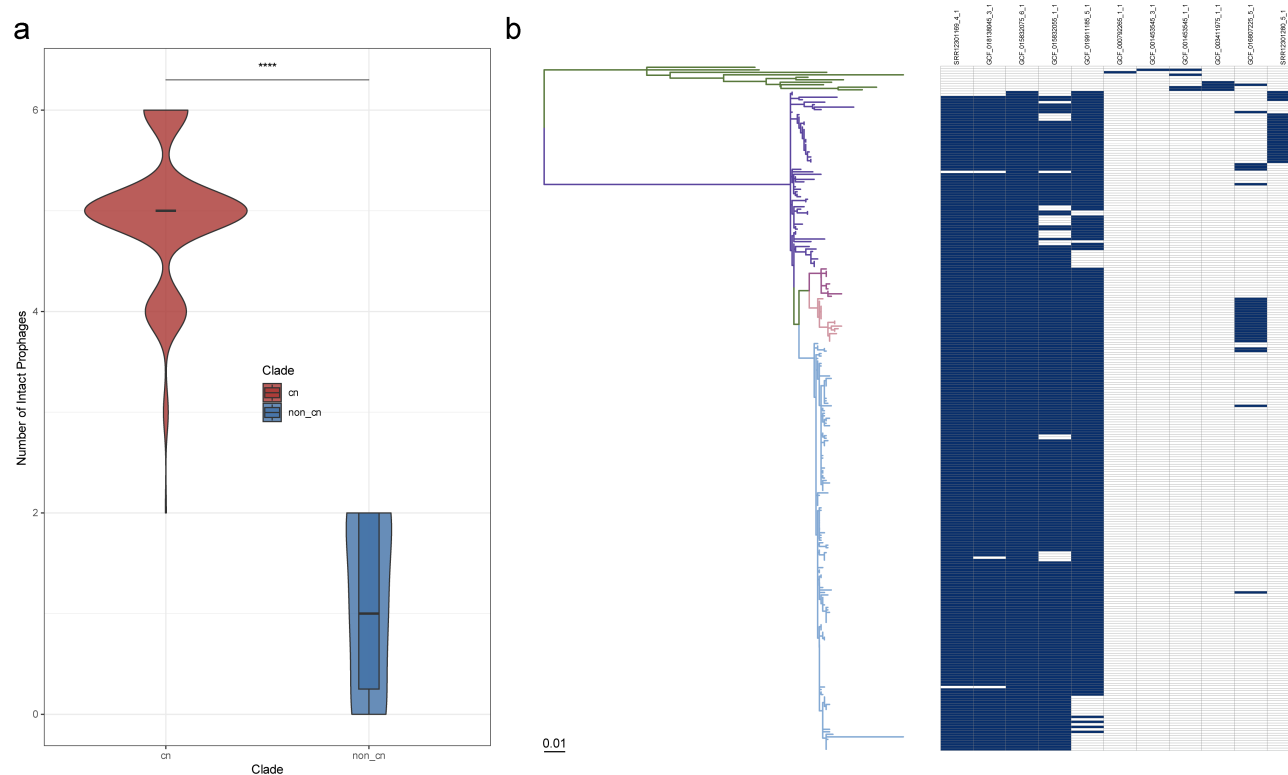


Figure 5. Prophage profiles within ST463 lineages. (a) Comparison of the number of intact prophages between cn and non_cn clades. (b) Maximum-likelihood phylogeny of ST463 (left) juxtaposed with a heatmap depicting the presence (blue) or absence (blank) of intact prophages (right). From left to right are prophages 1–11, respectively. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

Mutation analysis of ST463 in China

To gain more insights into the evolutionary trajectory of ST463 in China, we investigated signals of convergent evolution within the genomes studied. Genes that harbor recurring mutations in phylogenetically distinct lineages serve as hallmarks of convergent evolution, and those exhibiting a high ratio of nonsynonymous to synonymous mutations (dN/dS) are indicative of

positive selection [25]. A total of 10 genes were identified potentially under positive selection (Figure 4a and Table 1), encompassing functions from antimicrobial response (*ITG67_05925*, *parC* and *parE*) to the biosynthesis of vital cellular metabolites and components (*fptA*, *pvdS*, *wzzB*, *livM*, *rfaA*, *bifA* and *ITG67_16045*). Notably, mutations in *parC* and *parE* are linked to fluoroquinolone resistance, highlighting their potential

Table 1. Genes predicted to be under selection in the ST463-cn clade strains.

Locus_tag	Gene name	No. of N ^a SNPs	No. of S ^b SNPs	Product	COG category ^c	GO function/GO process
ITG67_04060	<i>fptA</i>	3	0	Fe ⁽³⁺⁾ -pyochelin receptor FptA	P	Monoatomic cation transport
ITG67_05925	-	3	0	CatB-related O-acetyltransferase	S	Response to antibiotic
ITG67_15905	<i>pvdS</i>	4	0	Pyoverdine signaling pathway sigma factor PvdS	K	DNA biosynthetic process; RNA metabolic process
ITG67_16045	-	4	1	TonB-dependent siderophore receptor	P	Siderophore transport
ITG67_22825	<i>wzzB</i>	6	0	LPS O-antigen chain length determinant protein WzzB	M	Lipopolysaccharide biosynthetic process
ITG67_25645	<i>bifA</i>	5	0	cyclic-di-GMP phosphodiesterase BifA	T	NA
ITG67_29090	<i>livM</i>	3	0	High-affinity branched-chain amino acid ABC transporter permease LivM	E	Amino acid transmembrane transporter
ITG67_29365	<i>parC</i>	6	0	DNA topoisomerase IV subunit A	L	DNA topological change
ITG67_29380	<i>parE</i>	5	0	DNA topoisomerase IV subunit B	L	DNA topological change
ITG67_30385	<i>rfaA</i>	4	0	Glucose-1-phosphate thymidyltransferase RfaA	M	O antigen biosynthetic process

^aNon-synonymous single-nucleotide polymorphisms.

^bSynonymous single-nucleotide polymorphisms.

^cE: Amino acid transport and metabolism; K: Transcription; L: Replication, recombination and repair; M: Cell wall/membrane/envelope biogenesis; P: Inorganic ion transport and metabolism; S: Function unknown; T: Signal transduction mechanisms.

role in antimicrobial resistance. The *fptA* and *ITG67_16045* genes, crucial for siderophore transport, enhance *P. aeruginosa*'s virulence by facilitating iron uptake, a critical factor for bacterial pathogenicity [26]. Furthermore, *pvdS* regulates expression of exotoxin A, a significant virulence factor [27]. The genes *wzzB* and *rfbA* are involved in lipopolysaccharide (LPS) O-antigen synthesis, affecting the bacterial surface characteristics and its interactions with the host's immune system [28]. Lastly, *bifA*'s role in biofilm formation and swarming mobility underscores its importance in bacterial colonization and the spread of infection [29]. This suite of positively selected genes reflects a complex adaptation strategy, enhancing ST463's resilience and pathogenicity in the challenging clinical environment of China.

Patterns of resistance and virulence genes vary between *cn* and *non_cn* clades

Recognizing the clinical significance of antimicrobial resistance genes (ARGs) in *P. aeruginosa*, we compared the distributions of ARGs in two clades. In total, 58 ARGs, spanning 13 classes of antimicrobial agents (β -lactam, aminoglycoside, bleomycin, fluoroquinolone, fosfomycin, macrolide, macrolide/streptogramin, phenicol, quinolone, rifamycin, sulfonamide, tetracycline, and trimethoprim) were identified (Figure 6). Notably, six genes [*aph(3')-IIB*, *bla_{OXA-486}*, *bla_{PDC-8}*, *catB7*, *crpP*, and *fosA*] were inherently prevalent in almost the entire ST463 population, underscoring ST463's multidrug-resistant nature. The *cn* clade was found to have acquired a significant number of ARGs, totaling 47, compared to only 6 in the *non_cn* clade (average 2.26 per isolate versus 1.1, $p < 0.001$, Figure S7a). Of particular interest are the KPC family genes, exclusive to the *cn* clade, predominantly *bla_{KPC-2}* (249/265, 93.96%), *bla_{KPC-33}* (4/265), and *bla_{KPC-90}* (1/265). Mutation analysis further revealed the *cn* clade's propensity for higher chromosomal mutation (Figure 6); for instance, the carbapenem-resistance *ftsI* gene mutation (F533L) was exclusively observed in the *cn* branch across most strains (246/265, 92.83%), and all *cn* strains exhibited the *gyrA* mutation. The high positive rates of *bla_{KPC-2}* and the *ftsI* (F533L) mutation within ST463 in China, marks a pivotal moment in the bacterium's adaptive evolution and resistance development (Figure 2).

A focused examination of all known virulence genes further assessed the pathogenic potential of isolates within both clades, identifying a total of 300 virulence

genes among 275 ST463 isolates (Figure 6). Notably, all ST463 strains possess both the *exoU* and *exoS* genes. The prevalence of virulence genes was comparable between *cn* and *non_cn* isolates (median 293 versus 290, $p > 0.05$, Figure S7b), suggesting that while drug resistance significantly diverges between the clades, virulence potential remains consistent. This observation underscores drug resistance, rather than virulence gene variability, as a key driver of ST463's dominance in China, maintaining its pathogenicity amidst evolving clinical challenges.

Contribution of plasmids to ST463

To elucidate the potential contribution of plasmids to the genomic plasticity of ST463, we investigated plasmid contigs within our dataset. This analysis revealed 781 putative plasmids across 269 genomes (Supplementary Data 2). The identified plasmids exhibited a wide range of sizes (~1–420 kb) and GC content (~40–66%), indicating a diverse plasmid population. To further explore this diversity, we generated a network visualization based on the all-versus-all Mash distances calculated for the ST463 plasmids. The resulting network comprised 16 clusters, with c1 (cluster 1) plasmids being the most prevalent (33.42%, 261/781), followed by c2 (cluster 2, 30.09%, 235/781) and c3 (cluster 3, 28.68%, 224/781) (Figure 7a,b). Intriguingly, we discovered that the c1 cluster was exclusively present in the *cn* clade. Additional analysis of the resistance determinants carried by these plasmids revealed that nearly all KPC genes were harbored within the c1 cluster, implying that c1 serves as a major reservoir for KPC genes (Figure 7c), which aligns with the characteristics of the type 1 plasmid previously described by Zhu et al. [13]. However, all c1 plasmids were predicted to be non-mobilizable, suggesting that the dissemination of *bla_{KPC-2}* in China's ST463 lineage is likely driven by clonal expansion rather than plasmid mobilization. The c2 and c3 plasmids were also highly represented and were commonly found in both the *cn* and *non_cn* clades, indicating that these plasmid types are more likely to be intrinsically associated with ST463.

Discussion

Pseudomonas aeruginosa, a notorious pathogen, is renowned for its complexity in nosocomial infections, particularly in hospital settings. In this study, we investigated the evolution and molecular characteristics of the emerging high-risk lineage ST463 of *P. aeruginosa* in China, which has garnered considerable attention due to its heightened virulence and substantial



Figure 6. Distributions of mutations, ARGs and virulence factors in the ST463 clade. Squares, which are colored by trait category, represent the presence of the trait examined. Virulence genes with a frequency of more than 90% in each clade are considered too conservative, so they are not shown in the figure.

antimicrobial resistance. Unlike the sporadic strains observed abroad, ST463 has achieved clonal success in China, posing an increasing threat within healthcare facilities. This indicates that the ST463 has well adapted to the Chinese healthcare environment. However, the

potential role of ST463 in its prevalence and evolutionary success in China has not been elucidated.

Consistent with the findings of Zhang et al. [30], our analysis delineated ST463 into two main clades: cn and non_cn. All ST463 strains from China were clustered

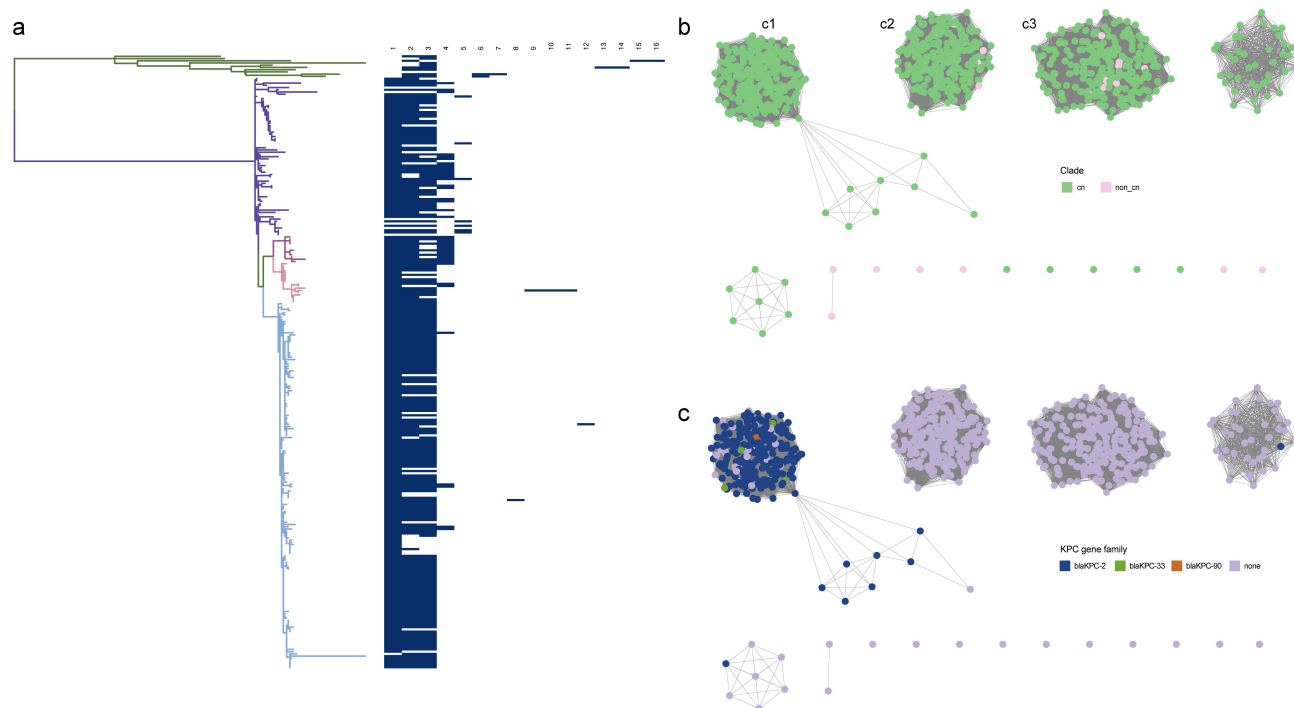


Figure 7. Plasmid clusters distribution within the ST463 lineage. (a) Maximum-likelihood phylogeny of ST463 (left) juxtaposed with a heatmap illustrating the distribution of the 16 identified plasmid clusters (right). (b) and (c) depict distribution patterns of various plasmid clusters within the clade, with specific attention to those harboring KPC family genes, respectively.

within the same cn clade, indicating that the Chinese ST463 originated from a single ancestral lineage. This observation, coupled with the divergence time of the clade and the results of biogeography analysis, supports the hypothesis that ST463 was introduced to China from abroad (North America), followed by subsequent expansion within the country. Notably, ST463 was first identified predominantly in the coastal Zhejiang region of southern China, implicating trade interactions around 2007 as a potential vector for its introduction. SNP analysis indicated that these strains primarily spread within hospitals and gradually extended to Shanghai, Yunnan, and other regions. Interestingly, ST463 is predominantly found in southern China and is relatively rare in Shanghai. This geographic distribution may be attributed to two factors. Firstly, the dry climate of northern China may hinder the growth of ST463, whereas the warm and humid conditions in the south favor its survival [31]. Additionally, the restricted patient movement due to the previous COVID-19 virus outbreak has contributed to the more localized emergence of the *P. aeruginosa* ST463 lineage [32]. Therefore, it is imperative to conduct wider surveillance of ST463 to prevent its widespread spread in the future.

Compared to the non_cn clade, the cn clade underwent significant genomic changes after its introduction

into China, which enhanced the adaptability of ST463 in the Chinese environment. Specifically, the cn clade experienced more extensive recombination events, particularly with *katA*-related ICE and T4P island. The frequent recombination of these virulence-related genetic elements suggests that the cn isolates have adapted to enhance pathogenicity and adherence, reflecting a shift in colonization strategies within the Chinese host environment, likely driven by selective pressures like immune responses and antibiotic treatments. It is noteworthy that strains in the cn clade lost a region related to anti-phage activity and a gene associated with cell wall synthesis, while acquiring three regions, including fatty acid synthesis, a new anti-phage region, and a prophage related to cell wall degradation. Considering that the cn clade harbors more prophages, we speculate that ST463 may have adapted to the Chinese settings by enhancing its energy metabolism pathways and developing novel anti-phage systems in response to the increased prophage burden. Additionally, the *ompA* gene found in region 2 might be a potential virulence factor, as the OmpA protein family typically plays a crucial role in pathogenesis by contributing to biofilm formation, stimulation of proinflammatory cytokines, bacterial adhesion, invasion, and intracellular survival [33,34], which may have facilitated the enhanced virulence of ST463 in China. In addition,

the deletion of the *dapA1* gene and the presence of the zeta toxin family protein in region 3 may result in cell wall modifications, potentially facilitating the acquisition of exogenous genes following the introduction of ST463 into China. Moreover, the absence of a known CRISPR/CAS system in ST463, a feature also observed in the widely prevalent multidrug-resistant clonal complex CC235, may contribute to its ability to incorporate exogenous genetic elements, such as bacteriophages [35]. Furthermore, by searching for convergent evolution events and independent evolution of mutations at the same nucleotide site or gene, we observed additional signatures (i.e. positively selected genes) that may have contributed to the success of ST463 in China. These genes not only include those related to antimicrobial resistance but also those associated with virulence, host immunity, and biofilm formation. The positive selection of these genes highlights the multiple selective pressures faced by ST463 in adapting to the hospital environment in China. These complex selective pressures have collectively driven the rapid evolution and adaptation of ST463 in the Chinese hospital environment, making it a successful nosocomial infection lineage.

The acquisition of drug resistance is also pivotal for the success of bacteria in clinical settings. We found that ST463 intrinsically carries six ARGs, establishing it as inherently multidrug-resistant. Furthermore, the prevalence of the T83I mutation in the *gyrA* gene, which is known to confer resistance to fluoroquinolone antibiotics [36], in Chinese ST463 isolates may have been driven by the extensive use of these antibiotics, where they are the one of most frequently used antibiotics in China [37]. Moreover, the Chinese ST463 has further acquired additional drug resistance genes and mutations, particularly the *bla*_{KPC-2} gene from the KPC family of carbapenemases and the F533L mutation in the *ftsI* gene. It is noteworthy that the acquisition of the *bla*_{KPC-2} gene is closely associated with the emergence and expansion of ST463 in China, suggesting that this gene may have played a crucial role in the adaptation and dissemination of ST463. Carbapenems are essential for treating infections caused by multidrug-resistant bacteria; the KPC-2 enzyme, encoded by *bla*_{KPC-2}, efficiently hydrolyzes carbapenems, conferring resistance to this class of antibiotics [38]. Similarly, the F533L mutation in the *ftsI* gene is closely linked to the carbapenem resistance [39]. Consequently, ST463 strains harboring the *bla*_{KPC-2} gene and/or the *ftsI* mutation possess a significant survival advantage under the selective pressure of carbapenems, facilitating their colonization and spread within hospital environments. Furthermore, we noticed that

the c1-type plasmid carrying *bla*_{KPC-2} in ST463 is exclusive to the cn clade. Although the c1-type plasmids were predicted to be non-mobilizable, limiting their potential for broad epidemiological transmission, they serve as a crucial reservoir of resistance genes, particularly *bla*_{KPC-2}, for the ST463 lineage in China. The acquisition and maintenance of these plasmids, coupled with the clonal expansion of ST463, have likely contributed to the widespread carbapenem resistance observed in this lineage.

Our study has limitations. First, although the research covers ST463 strains from different regions of China, the geographical distribution of the samples may not be sufficiently extensive or balanced. This limitation may impede our comprehensive understanding of the genetic diversity of this ST both nationally and globally. In particular, strain data from countries other than China are relatively scarce. For instance, despite reports of ST463 from South Korea [40], the absence of corresponding strain sequences restricts our ability to thoroughly investigate genetic variations across countries. Additionally, the limited temporal scope of our sample collection might not adequately reflect the long-term evolutionary and transmission dynamics of ST463. Furthermore, while this study provides some insights into the evolutionary mechanisms underlying the successful spread and adaptation of ST463 in China, these speculations require further support from experimental evidence. In particular, the role of three acquired genomic regions in the ST463 strains in facilitating the organisms to better adapt the host and be more virulent for the host remains to be addressed in the future. Finally, the spread and adaptation of ST463 are likely influenced by a myriad of factors, including medical practices, infection control measures, and host immunity, which have not been addressed in our analysis.

In conclusion, this study offers valuable insights into the molecular evolution of the emerging high-risk lineage ST463 of *Pseudomonas aeruginosa* in China, emphasizing the critical roles of recombination, prophages, ARGs and plasmids in shaping the remarkable adaptability of ST463. These factors have collectively contributed to its exceptional ability to adapt and its extensive drug resistance profile. Although this lineage currently exhibits a certain degree of regional specificity, the increasing frequency of global travel and trade raises the possibility of ST463 spreading to wider geographical areas. Therefore, we strongly urge international public health agencies and research institutions to strengthen clinical surveillance of this lineage. Enhanced monitoring will enable timely

detection of its dissemination trends and facilitate the implementation of effective measures to control its spread.

Methods and materials

Strain collection and genome assembly

All *Pseudomonas aeruginosa* genomes were retrieved from the NCBI GenBank database as of 1 October 2023 and MLST v2.23.0 (<https://github.com/tseemann/mlst>) was used to determine their multi-locus sequence type (ST) against the PubMLST database (<https://pubmlst.org/>) [41], subsequently focusing on those identified as ST463. We then searched for publications on *Pseudomonas aeruginosa* ST463 and downloaded the associated genomes using the NCBI Datasets tool (<https://github.com/ncbi/datasets>) and SRA Toolkit (<https://github.com/ncbi/sra-tools>). Furthermore, we collected 10 strains identified as ST463 from human sputum, blood, or stool samples of patients diagnosed with bloodstream or lung infections between 2021 and 2023 as part of routine clinical management at the First Affiliated Hospital, Zhejiang University School of Medicine, and subsequently used them for Illumina sequencing (2×150 -bp paired-end libraries). The sequencing raw data were quality controlled using fastp v0.23.2 [42] and then genome assembly was carried out using shovill v1.1.0 (<https://github.com/tseemann/shovill>), with contigs shorter than 200 bp being filtered out. Genome quality and completeness were evaluated using QUAST v5.2.0 [43] and Checkm2 v1.0.2 [44]. Assemblies with N50 < 50 kb or contigs > 300 were considered low-quality genomes and subsequently excluded from further analysis.

Population structure and phylogenetic construction

We chose strain NDTH10366 (CP064401.1; GCF_019857345.1) [13] with the highest completeness in Checkm2 as the reference strain of ST463. Reference-based mapping and single nucleotide polymorphisms (SNPs) identification within the ST463 core genome were performed using Snippy v4.6.0 (<https://github.com/tseemann/snippy>). Recombined regions in the core genome were detected and excluded using Gubbins v3.2.1 [45]. ModelTest-NG v0.1.7 [46] determined the optimal phylogeny model, and a genetic tree was constructed from the recombination-free core-genome alignment using RAxML v8.2.12 [47] with a GTR model and gamma correction, supported by 1,000 bootstrap replications. The geographic origins of ST463 isolates were deduced using the Bayesian binary MCMC method in

RASP v4.2 [48], employing 10 parallel chains each running for 50 million cycles. The population structure of all ST463 genomes and the China clade was inferred using fastbaps v1.0.8 [49]. A minimum spanning tree was constructed to visualize the genetic relationships among all ST463 isolates using PHYLOViZ v2.0 [50].

Bayesian phylogenetic inference

To deduce the origin of the ST463 ancestor and to pinpoint the timeframe of its emergence in China, isolates without dates of collection were removed from the phylogeny. We assessed the linear correlation between root-to-tip divergence and sampling dates, achieving a moderate correlation ($R^2 = 0.529$) using TempEst v1.5.3 [51], and one isolate displaying significant deviation from the temporal signal was eliminated. To further corroborate the temporal signal within our dataset, we utilized the TipDatingBeast package [52] in R, conducting date randomization tests across multiple replicates. The sampling dates of strains were randomly shuffled 20 times, adhering to the same parameters established in the original dataset. BEAST v1.8.4 [53] was employed to estimate both the substitution rate and the timing of the most recent common ancestor (MRCA) for the ST463 evolutionary lineage. An uncorrelated log-normal distribution was employed for the substitution rate, and a constant population size was applied to the tree priors following path sampling [54]. The analysis entailed running three chains of 2×10^8 generations each, sampling every 1,000 generations, and discarding the initial 10% of each chain as burn-in to ensure independent convergence. We evaluated convergence using Tracer v1.7.2 [55], ensuring that all essential parameters achieved effective sample sizes (ESS) greater than 200. The estimation of effective population size dynamics was performed with the sky-growth package in R [56].

Pan-genome analysis

Annotation of assemblies was used Bakta v1.8.2 [57]. The pan-genome analysis was performed using Panaroo v1.3.2 [58] with the default parameters, the GFF3 file generated by Batka as the input file. Significantly higher substitution accumulation in core genes (FDR-adjusted $p < 0.05$) compared to the expectations of a neutral evolutionary model (Poisson process), or the presence of homoplastic SNPs (independently evolved at least three times), was indicative of the presence of positive selection [59]. The

reconstruction of SNP ancestral states was constructed using PAML [60].

Bacteriophage prediction

Bacteriophage regions within each genome were predicted using the online tool PHASTEST (<https://phastest.ca>) [61]. These regions were clustered using cd-hit v4.8.1 [62], and CheckV v1.0.1 [63] was used to assess the integrity and quality of these clusters. Only those regions with a completeness score exceeding 90% were considered as intact bacteriophages. Finally, to investigate the distribution of these intact bacteriophage regions across different contigs in the genome, the sequences were aligned using BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) and manually examined.

Identification of antimicrobial resistance genes, virulence factors and CRISPR-Cas systems

Antimicrobial resistance genes (ARGs) were annotated using AMRFinderPlus v3.11.20 [64], and virulence factors were scanned using ABRicate v1.0.0 (<https://github.com/tseemann/abricate>) with a minimum sequence similarity threshold of 90% and a minimum coverage threshold of 90% based on VFDB database [65]. The putative CRISPR-Cas systems were detected using CRISPRCasTyper v1.8.0 [66]. Integrative conjugative element (ICE) and genome island (GI) were predicted by VRprofile2 [67].

Plasmid analysis

As described previously [68], plasmids were identified and reconstructed using the mob-recon tool from the MOB-suite toolkit v3.1.2 [69]. Plasmids clustering was performed using mash v2.3 [70] with a distance of 0.05 and the result was visualized in Cystoscope v3.10.1 [71].

Statistical analysis

The chi-square test or Fisher's exact test, alongside the Wilcoxon rank-sum test, were utilized to perform the statistical analyses. A P-value of less than 0.05 ($p < 0.05$) was considered statistically significant. All data processing and visualization were conducted in R version 4.2.1 (<https://www.r-project.org/>).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by National Infectious Disease Medical Center startup fund (Y.Z.) [B2022011-1], Jinan Microecological Biomedicine Shandong Laboratory project [JNL-2022050B], and Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang [No. 2021R01012] and Chinese Postdoctoral Science Foundation [2021TQ0278].

Authors contributions

Xu Dong, Tiantian Wu, and Ying Zhang conceptualized and designed the study. Xu Dong and Yanghui Xiang performed the data analysis and interpretation, and drafted the manuscript. Lanjuan Li, Tiantian Wu, and Ying Zhang reviewed and revised the manuscript. All authors have read and approved the final version of the manuscript.

Data availability statement

The genome sequences of *Pseudomonas aeruginosa* isolates collected in this study have been deposited in GenBank under the accession number PRJNA1102836. The data supporting the findings of this study, along with Supplementary Data1 and Supplementary Data2, are openly available on Figshare at doi.org/10.6084/m9.figshare.27144330.

Ethics statement

Ethical approval for this study was approved by the Ethics Committee of the First Affiliated Hospital of Zhejiang University in accordance with the Declaration of Helsinki. The requirement for informed consent was waived by the Ethics Committee as this was a retrospective study and all bacterial samples from patients were used anonymously for scientific purposes, in compliance with the Decree No. 11 of the National Health Commission of the People's Republic of China (Operational Guideline for the Ethics Review of Biomedical Research Involving Human Subjects, issued on 30 September 2016 and implemented from). The data used in this study, other than those generated by the study itself, were obtained from publicly available NCBI databases, which are openly accessible and permit unrestricted re-use under an open license, exempting the study from requiring additional ethical approval.

ORCID

Xu Dong  <http://orcid.org/0000-0001-8568-5106>

Tiantian Wu  <http://orcid.org/0000-0002-8147-4077>

References

- [1] Silby MW, Winstanley C, Godfrey SA, et al. *Pseudomonas* genomes: diverse and adaptable. FEMS Microbiol Rev. 2011;35(4):652-680.
- [2] Barbier F, Andremont A, Wolff M, et al. Hospital-acquired pneumonia and ventilator-associated pneumonia: recent advances in epidemiology and

- management. *Curr Opin Pulm Med.* 2013;19(3):216–228. doi: [10.1097/MCP.0b013e32835f27be](https://doi.org/10.1097/MCP.0b013e32835f27be)
- [3] De Oliveira DMP, Forde BM, Kidd TJ, et al. Antimicrobial resistance in ESKAPE pathogens. *Clin Microbiol Rev.* 2020;33(3):3. doi: [10.1128/CMR.00181-19](https://doi.org/10.1128/CMR.00181-19)
 - [4] Tacconelli E, Carrara E, Savoldi A, et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect Dis.* 2018;18(3):318–327. doi: [10.1016/S1473-3099\(17\)30753-3](https://doi.org/10.1016/S1473-3099(17)30753-3)
 - [5] Tenover FC, Nicolau DP, Gill CM. Carbapenemase-producing *Pseudomonas aeruginosa* – An emerging challenge. *Emerg Microbes Infect.* 2022;11(1):811–814.
 - [6] Pang Z, Raudonis R, Glick BR, et al. Antibiotic resistance in *Pseudomonas aeruginosa*: mechanisms and alternative therapeutic strategies. *Biotechnol Adv.* 2019;37(1):177–192. doi: [10.1016/j.biotechadv.2018.11.013](https://doi.org/10.1016/j.biotechadv.2018.11.013)
 - [7] Reyes J, Komarow L, Chen L, et al. Global epidemiology and clinical outcomes of carbapenem-resistant *Pseudomonas aeruginosa* and associated carbapenemases (POP): a prospective cohort study. *Lancet Microbe.* 2023;4(3):e159–e170.
 - [8] Ge C, Wei Z, Jiang Y, et al. Identification of KPC-2-producing *Pseudomonas aeruginosa* isolates in China. *J Antimicrob Chemother.* 2011;66(5):1184–1186. doi: [10.1093/jac/dkr060](https://doi.org/10.1093/jac/dkr060)
 - [9] Hu YY, Gu DX, Cai JC, et al. Emergence of KPC-2-producing *Pseudomonas aeruginosa* sequence type 463 isolates in Hangzhou, China. *Antimicrob Agents Chemother.* 2015;59(5):2914–2917. doi: [10.1128/AAC.04903-14](https://doi.org/10.1128/AAC.04903-14)
 - [10] Hu YY, Wang Q, Sun QL, et al. A novel plasmid carrying carbapenem-resistant gene *bla*_{KPC-2} in *Pseudomonas aeruginosa*. *Infect Drug Resist.* 2019;12:1285–1288.
 - [11] Hu Y, Peng W, Wu Y, et al. A potential high-risk clone of *Pseudomonas aeruginosa* ST463. *Front Microbiol.* 2021;12:670202. doi: [10.3389/fmicb.2021.670202](https://doi.org/10.3389/fmicb.2021.670202)
 - [12] Wu T, Zhang Z, Li T, et al. The type III secretion system facilitates systemic infections of *Pseudomonas aeruginosa* in the clinic. *Microbiol Spectr.* 2023;12(1):e0222423.
 - [13] Zhu Y, Chen J, Shen H, et al. Emergence of ceftazidime- and avibactam-resistant *Klebsiella pneumoniae* carbapenemase-producing *Pseudomonas aeruginosa* in China. *mSystems.* 2021;6(6):e0078721. doi: [10.1128/mSystems.00787-21](https://doi.org/10.1128/mSystems.00787-21)
 - [14] Hu Y, Liu C, Wang Q, et al. Emergence and expansion of a carbapenem-resistant *Pseudomonas aeruginosa* clone are associated with plasmid-borne *bla*_{KPC-2} and virulence-related genes. *mSystems.* 2021;6(3):13
 - [15] Zhu Y, Kang Y, Zhang H, et al. Emergence of ST463 *exoU*-Positive, Imipenem-Nonsusceptible *Pseudomonas aeruginosa* Isolates in China. *Microbiol Spectr.* 2023;11(4):e0010523. doi: [10.1128/spectrum.00105-23](https://doi.org/10.1128/spectrum.00105-23)
 - [16] Sun L, Guo H, Chen Y, et al. Genomic and phylogenetic analysis of a multidrug-resistant *Pseudomonas aeruginosa* ST463 strain co-carrying *bla*_{KPC-2}, *bla*_{OXA-246}, and *bla*_{OXA-486} in China. *J Glob Antimicrob Resist.* 2023;33:301–303. doi: [10.1016/j.jgar.2023.04.011](https://doi.org/10.1016/j.jgar.2023.04.011)
 - [17] Allen JP, Ozer EA, Minasov G, et al. A comparative genomics approach identifies contact-dependent growth inhibition as a virulence determinant. *Proc Natl Acad Sci USA.* 2020;117(12):6811–6821.
 - [18] Darch SE, McNally A, Harrison F, et al. Recombination is a key driver of genomic and phenotypic diversity in a *Pseudomonas aeruginosa* population during cystic fibrosis infection. *Sci Rep.* 2015;5(1):7649. doi: [10.1038/srep07649](https://doi.org/10.1038/srep07649)
 - [19] Moradali MF, Ghods S, Rehm BH. *Pseudomonas aeruginosa* lifestyle: a paradigm for adaptation, survival, and persistence. *Front Cell Infect Microbiol.* 2017;7:39.
 - [20] Lee JS, Heo YJ, Lee JK, et al. KatA, the major catalase, is critical for osmoprotection and virulence in *Pseudomonas aeruginosa* PA14. *Infect Immun.* 2005;73(7):4399–4403. doi: [10.1128/IAI.73.7.4399-4403.2005](https://doi.org/10.1128/IAI.73.7.4399-4403.2005)
 - [21] Burrows LL. *Pseudomonas aeruginosa* twitching motility: type IV pili in action. *Annu Rev Microbiol.* 2012;66(1):493–520. doi: [10.1146/annurev-micro-092611-150055](https://doi.org/10.1146/annurev-micro-092611-150055)
 - [22] Impey RE, Panjikar S, Hall CJ, et al. Identification of two dihydrodipicolinate synthase isoforms from *Pseudomonas aeruginosa* that differ in allosteric regulation. *FEBS J.* 2020;287(2):386–400. doi: [10.1111/febs.15014](https://doi.org/10.1111/febs.15014)
 - [23] Doron S, Melamed S, Ofir G, et al. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science.* 2018;359(6379). doi: [10.1126/science.aar4120](https://doi.org/10.1126/science.aar4120)
 - [24] Mutschler H, Gebhardt M, Shoeman RL, et al. A novel mechanism of programmed cell death in bacteria by toxin-antitoxin systems corrupts peptidoglycan synthesis. *PLOS Biol.* 2011;9(3):e1001033.
 - [25] Chung the H, Boinett C, Pham Thanh D, et al. Dissecting the molecular evolution of fluoroquinolone-resistant *Shigella sonnei*. *Nat Commun.* 2019;10(1):4828. doi: [10.1038/s41467-019-12823-0](https://doi.org/10.1038/s41467-019-12823-0)
 - [26] Mislin GL, Hoegy F, Cobessi D, et al. Binding properties of pyochelin and structurally related molecules to FptA of *Pseudomonas aeruginosa*. *J Mol Biol.* 2006;357(5):1437–1448.
 - [27] Hunt TA, Peng WT, Loubens I, et al. The *Pseudomonas aeruginosa* alternative sigma factor PvdS controls exotoxin a expression and is expressed in lung infections associated with cystic fibrosis. *Microbiol (Read).* 2002;148(Pt 10):3183–3193.
 - [28] Dasgupta T, Lam JS. Identification of *rfaB*, involved in B-band lipopolysaccharide biosynthesis in *Pseudomonas aeruginosa* serotype O5. *Infect Immun.* 1995;63(5):1674–1680. doi: [10.1128/iai.63.5.1674-1680.1995](https://doi.org/10.1128/iai.63.5.1674-1680.1995)
 - [29] Kuchma SL, Brothers KM, Merritt JH, et al. BifA, a cyclic-Di-GMP phosphodiesterase, inversely regulates biofilm formation and swarming motility by *Pseudomonas aeruginosa* PA14. *J Bacteriol.* 2007;189(22):8165–8178. doi: [10.1128/JB.00586-07](https://doi.org/10.1128/JB.00586-07)
 - [30] Zhang P, Wu W, Wang N, et al. *Pseudomonas aeruginosa* High-risk sequence type 463 Co-producing KPC-2 and AFM-1 Carbapenemases, China, 2020–2022. *Emerg Infect Dis.* 2023;29(10):2136–2140.
 - [31] Bedard E, Prevost M, Deziel E. *Pseudomonas aeruginosa* in premise plumbing of large buildings. *Microbiologyopen.* 2016;5(6):937–956. doi: [10.1002/mbo3.391](https://doi.org/10.1002/mbo3.391)

- [32] Ji H, Tong H, Wang J, et al. The effectiveness of travel restriction measures in alleviating the COVID-19 epidemic: evidence from Shenzhen, China. *Environ Geochem Health*. 2022;44(9):3115–3132. doi: [10.1007/s10653-021-00920-3](https://doi.org/10.1007/s10653-021-00920-3)
- [33] Confer AW, Ayalew S. The OmpA family of proteins: roles in bacterial pathogenesis and immunity. *Vet Microbiol*. 2013;163(3–4):207–222. doi: [10.1016/j.vetmic.2012.08.019](https://doi.org/10.1016/j.vetmic.2012.08.019)
- [34] Vila-Farres X, Parra-Millan R, Sanchez-Encinales V, et al. Combating virulence of gram-negative bacilli by OmpA inhibition. *Sci Rep*. 2017;7(1):14683. doi: [10.1038/s41598-017-14972-y](https://doi.org/10.1038/s41598-017-14972-y)
- [35] Miyoshi-Akiyama T, Tada T, Ohmagari N, et al. Emergence and spread of epidemic multidrug-resistant *Pseudomonas aeruginosa*. *Genome Biol Evol*. 2017;9(12):3238–3245. doi: [10.1093/gbe/evx243](https://doi.org/10.1093/gbe/evx243)
- [36] Agnello M, Finkel SE, Wong-Beringer A. Fitness cost of fluoroquinolone resistance in clinical isolates of *Pseudomonas aeruginosa* differs by type III secretion genotype. *Front Microbiol*. 2016;7:1591. doi: [10.3389/fmicb.2016.01591](https://doi.org/10.3389/fmicb.2016.01591)
- [37] Xu W, Zhang G, Li X, et al. Occurrence and elimination of antibiotics at four sewage treatment plants in the Pearl River Delta (PRD), South China. *Water Res*. 2007;41(19):4526–4534.
- [38] Barnes MD, Winkler ML, Taracila MA, et al. *Klebsiella pneumoniae* carbapenemase-2 (KPC-2), substitutions at ambler position Asp179, and resistance to ceftazidime-avibactam: unique antibiotic-resistant phenotypes emerge from beta-lactamase protein engineering. *mBio*. 2017;8(5).17
- [39] Cabot G, Zamorano L, Moya B, et al. Evolution of *Pseudomonas aeruginosa* antimicrobial resistance and fitness under low and high mutation rates. *Antimicrob Agents Chemother*. 2016;60(3):1767–1778. doi: [10.1128/AAC.02676-15](https://doi.org/10.1128/AAC.02676-15)
- [40] Hong JS, Kim JO, Lee H, et al. Characteristics of metallo-beta-lactamase-producing *Pseudomonas aeruginosa* in Korea. *Infect Chemother*. 2015;47(1):33–40.
- [41] Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST. Org website and their applications. *Wellcome Open Res*. 2018;3:124. doi: [10.12688/wellcomeopenres.14826.1](https://doi.org/10.12688/wellcomeopenres.14826.1)
- [42] Chen S, Zhou Y, Chen Y, et al. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884–i890. doi: [10.1093/bioinformatics/bty560](https://doi.org/10.1093/bioinformatics/bty560)
- [43] 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](https://doi.org/10.1093/bioinformatics/btt086)
- [44] Chklovski A, Parks DH, Woodcroft BJ, et al. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods*. 2023;20(8):1203–1212. doi: [10.1038/s41592-023-01940-w](https://doi.org/10.1038/s41592-023-01940-w)
- [45] Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res*. 2015;43(3):e15. doi: [10.1093/nar/gku1196](https://doi.org/10.1093/nar/gku1196)
- [46] Darriba D, Posada D, Kozlov AM, et al. ModelTest-NG: a new and scalable tool for the selection of DNA and Protein evolutionary models. *Mol Biol Evol*. 2020;37(1):291–294. doi: [10.1093/molbev/msz189](https://doi.org/10.1093/molbev/msz189)
- [47] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–1313. doi: [10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033)
- [48] Yu Y, Blair C, He X, et al. RASP 4: ancestral state reconstruction tool for multiple genes and characters. *Mol Biol Evol*. 2020;37(2):604–606. doi: [10.1093/molbev/msz257](https://doi.org/10.1093/molbev/msz257)
- [49] Tonkin-Hill G, Lees JA, Bentley SD, et al. Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Res*. 2019;47(11):5539–5549. doi: [10.1093/nar/gkz361](https://doi.org/10.1093/nar/gkz361)
- [50] Nascimento M, Sousa A, Ramirez M, et al. PHYLOViZ 2.0: providing scalable data integration and visualization for multiple phylogenetic inference methods. *Bioinformatics*. 2017;33(1):128–129. doi: [10.1093/bioinformatics/btw582](https://doi.org/10.1093/bioinformatics/btw582)
- [51] Rambaut A, Lam TT, Max Carvalho L, et al. Exploring the temporal structure of heterochronous sequences using TempEst (formerly path-O-Gen). *Virus Evol*. 2016;2(1):vew007. doi: [10.1093/ve/vew007](https://doi.org/10.1093/ve/vew007)
- [52] Rieux A, Khatchikian CE. Tipdatingbeast: an r package to assist the implementation of phylogenetic tip-dating tests using beast. *Mol Ecol Resour*. 2017;17(4):608–613. doi: [10.1111/1755-0998.12603](https://doi.org/10.1111/1755-0998.12603)
- [53] Baele G, Li WL, Drummond AJ, et al. Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. *Mol Biol Evol*. 2013;30(2):239–243.
- [54] Baele G, Lemey P, Bedford T, et al. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol*. 2012;29(9):2157–2167. doi: [10.1093/molbev/mss084](https://doi.org/10.1093/molbev/mss084)
- [55] Rambaut A, Drummond AJ, Xie D, et al. Posterior summarization in bayesian phylogenetics using tracer 1.7. *Syst Biol*. 2018;67(5):901–904. doi: [10.1093/sysbio/syy032](https://doi.org/10.1093/sysbio/syy032)
- [56] Volz EM, Didelot X, Townsend J. Modeling the growth and decline of pathogen effective population size provides insight into epidemic dynamics and drivers of antimicrobial resistance. *Syst Biol*. 2018;67(4):719–728. doi: [10.1093/sysbio/syy007](https://doi.org/10.1093/sysbio/syy007)
- [57] Schwengers O, Jelonek L, Dieckmann MA, et al. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb Genom*. 2021;7(11). doi: [10.1099/mgen.0.000685](https://doi.org/10.1099/mgen.0.000685)
- [58] Tonkin-Hill G, MacAlasdair N, Ruis C, et al. Producing polished prokaryotic pangenomes with the panaroo pipeline. *Genome Biol*. 2020;21(1):180. doi: [10.1186/s13059-020-02090-4](https://doi.org/10.1186/s13059-020-02090-4)
- [59] Sun Z, Yang F, Ji J, et al. Dissecting the genotypic features of a fluoroquinolone-resistant *Pseudomonas aeruginosa* ST316 sublineage causing ear infections in Shanghai, China. *Microb Genom*. 2023;9(4). doi: [10.1099/mgen.0.000989](https://doi.org/10.1099/mgen.0.000989)

- [60] Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24(8):1586–1591. doi: [10.1093/molbev/msm088](https://doi.org/10.1093/molbev/msm088)
- [61] Wishart DS, Han S, Saha S, et al. PHASTEST: faster than PHASTER, better than PHAST. *Nucleic Acids Res.* 2023;51(W1):W443–W450. doi: [10.1093/nar/gkad382](https://doi.org/10.1093/nar/gkad382)
- [62] Fu L, Niu B, Zhu Z, et al. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 2012;28(23):3150–3152. doi: [10.1093/bioinformatics/bts565](https://doi.org/10.1093/bioinformatics/bts565)
- [63] Nayfach S, Camargo AP, Schulz F, et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol.* 2021;39(5):578–585. doi: [10.1038/s41587-020-00774-7](https://doi.org/10.1038/s41587-020-00774-7)
- [64] Feldgarden M, Brover V, Gonzalez-Escalona N, et al. AMRFinderPlus and the reference gene catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci Rep.* 2021;11(1):12728. doi: [10.1038/s41598-021-91456-0](https://doi.org/10.1038/s41598-021-91456-0)
- [65] Liu B, Zheng D, Zhou S, et al. VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res.* 2022;50(D1):D912–D917. doi: [10.1093/nar/gkab1107](https://doi.org/10.1093/nar/gkab1107)
- [66] Russel J, Pinilla-Redondo R, Mayo-Munoz D, et al. CRISPRCasTyper: automated identification, annotation, and classification of CRISPR-Cas Loci. *CRISPR J.* 2020;3(6):462–469. doi: [10.1089/crispr.2020.0059](https://doi.org/10.1089/crispr.2020.0059)
- [67] Wang M, Goh YX, Tai C, et al. VRprofile2: detection of antibiotic resistance-associated mobilome in bacterial pathogens. *Nucleic Acids Res.* 2022;50(W1):W768–W773. doi: [10.1093/nar/gkac321](https://doi.org/10.1093/nar/gkac321)
- [68] Dong X, Jia H, Yu Y, et al. Genomic revisitation and reclassification of the genus *Providencia*. *mSphere.* 2024;9(3):e0073123. doi: [10.1128/msphere.00731-23](https://doi.org/10.1128/msphere.00731-23)
- [69] Robertson J, Nash JHE. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb Genom.* 2018;4(8). doi: [10.1099/mgen.0.000206](https://doi.org/10.1099/mgen.0.000206)
- [70] Ondov BD, Treangen TJ, Melsted P, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 2016;17(1):132. doi: [10.1186/s13059-016-0997-x](https://doi.org/10.1186/s13059-016-0997-x)
- [71] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–2504. doi: [10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303)