Analysis of the features of 105 confirmed CRISPR loci in 487 Klebsiella variicola

Yanyan Xi[1], Jiaxue Zhao[1], Jiangfeng Zhang[1], Yuefei Jin[1], Haiyan Yang[1], Guangcai Duan[2], Shuaiyin Chen[1], and Jinzhao Long[1]

[1]Affiliation not available
[2]Zhengzhou Univ

March 29, 2023

Abstract

Klebsiella variicola (K. variicola) is an emerging human pathogen, which poses a threat to public health. The horizontal gene transfer (HGT) of plasmids is an important driver for the emergence of multiple antibiotic-resistant K. variicola. The clustered regularly interspersed short palindromic repeats coupled with the CRISPR-associated genes (CRISPR/Cas) constitute an adaptive immune system in bacteria, which provide acquired immunity against HGT. However, the information about CRISPR/Cas system in K. variicola is still limited. In this study, a total of 487 genomes from NCBI database were used to analyze the characterization of CRISPR/Cas systems. 105 of the 487 genomes harbored at least one confirmed CRISPR array. Three types of CRISPR/Cas system, including types I-E, I-E*, and IV-A systems, were identified among 105 strains. The distribution of type I system was strongly associated with MLST, whereas type IV system was randomly distributed. Approximately one-third of spacer origins were homologous with plasmids or phages, indicating the role of CRISPR/Cas systems in controlling HGT. Moreover, spacers in K. variicola tended to target mobile genetic elements (MGEs) from Klebsiella pneumoniae, which provides new evidence for their interaction during evolution. Collectively, our results provide valuable insights into the role of CRISPR/Cas systems in K. variicola.

Analysis of the features of 105 confirmed CRISPR loci in 487 *Klebsiella variicola*

Yanyan Xi[1], Jiaxue Zhao[1], Jiangfeng Zhang[3], Yuefei Jin[1], Haiyan Yang[1], Guangcai Duan[1,2], Shuaiyin Chen[1*], Jinzhao Long[1*]

1 College of Public Health, Zhengzhou University, Zhengzhou, Henan, People's Republic of China

2 Henan Key Laboratory of Molecular Medicine, Zhengzhou University, Zhengzhou, Henan, People's Republic of China

3 Henan Provincial People's Hospital, Zhengzhou, Henan, People's Republic of China

*Correspondence:

Correspondence author: Shuaiyin Chen, Ph.D., Department of Epidemiology, College of Public Health, Zhengzhou University, No. 100 Kexue Avenue, Zhengzhou, Henan Province 450001, China, Email: sychen@zzu.edu.cn

Jinzhao Long, Ph.D., Department of Epidemiology, College of Public Health, Zhengzhou University, No. 100 Kexue Avenue, Zhengzhou, Henan Province 450001, China, Email: ljz069@zzu.edu.cn.

Keywords: *Klebsiella variicola* ; CRISPR/Cas system; multi-locus sequence typing; phylogenetic analyses; spacers homology

Abbreviations:

K. variicola : *Klebsiella variicola*

K. pneumoniae :

Klebsiella pneumoniae

MDR: multidrug-resistant

HGT: horizontal gene transfer

ARGs: antibiotic resistance genes

CRISPR: Clustered regularly interspersed short Palindromic repeats

Cas: CRISPR-associated

MGEs: mobile genetic elements

DRs: direct repeats

pre-crRNA: precursor CRISPR RNA

crRNA: CRISPR RNA

MFE: minimum free energy

KW: the non-parametric Kruskal-Wallis

PAM: protospacer adjacent motif

MLST: multiple locus sequence typing

STs: sequence types

Abstract

Klebsiella variicola (*K. variicola*) is an emerging human pathogen, which poses a threat to public health. The horizontal gene transfer (HGT) of plasmids is an important driver for the emergence of multiple antibiotic-resistant *K. variicola* . The clustered regularly interspersed short palindromic repeats coupled with the CRISPR-associated genes (CRISPR/Cas) constitute an adaptive immune system in bacteria, which provide acquired immunity against HGT. However, the information about CRISPR/Cas system in *K. variicola*is still limited. In this study, a total of 487 genomes from NCBI database were used to analyze the characterization of CRISPR/Cas systems. 105 of the 487 genomes harbored at least one confirmed CRISPR array. Three types of CRISPR/Cas system, including types I-E, I-E*, and IV-A systems, were identified among 105 strains. The distribution of type I system was strongly associated with MLST, whereas type IV system was randomly distributed. Approximately one-third of spacer origins were homologous with plasmids or phages, indicating the role of CRISPR/Cas systems in controlling HGT. Moreover, spacers in *K. variicola*tended to target mobile genetic elements (MGEs) from *Klebsiella pneumoniae* , which provides new evidence for their interaction during evolution. Collectively, our results provide valuable insights into the role of CRISPR/Cas systems in *K. variicola* .

1 | Introduction

Klebsiella variicola (K. variicola), a Gram-negative and facultative anaerobic bacillus, belongs to the*Klebsiella* genus of the *Enterobacteriaceae* family [1]. Such bacteria have been found in human, animals, insects, plants, and environments [2-4]. *K. variicola* infections can cause a series of human diseases, such as bloodstream infections, respiratory tract infections, and neonatal sepsis [5-7]. Moreover, bloodstream infections due to *K. variicola* harbored a higher 30-day mortality rate than its notorious relative,*Klebsiella pneumoniae* (*K. pneumoniae*) [8]. Notably, most cases caused by *K. variicola* infections were often misidentified as being infected with*K. pneumoniae* because of their high similarities in biochemical and phenotypic

features, which leads to the underestimation of the real prevalence of *K. variicola* infections in clinical practice [9]. Additionally, *K. variicola* could cause bovine mastitis and affect the milk production and quality [10]. With the misuse and abuse of antibiotics in clinical and agriculture environments, the last few years has witnessed the rapid increase of multidrug-resistant (MDR) *K. variicola* strains [11-13]. Accumulating evidence has reported that MDR *K. variicola* strains were detected in several countries, such as USA, Australia, Tanzania, Germany and China [3, 14]. Horizontal gene transfer (HGT) is deemed as the main mechanism for the transmission of antibiotic resistance genes (ARGs), in which plasmid serves a vector role. It has been documented that *K. variicola* harbored an open genome and shared similar plasmid types with other members of the*Klebsiella* genus, especially *K. pneumoniae,* thereby indicating frequent plasmid-mediated HGT among these species [1, 15]. In particular, ARGs-related IncFIB$_{\mathrm{k}}$, IncFII$_{\mathrm{k}}$ and IncFII replicons were often shared between*K. variicola* and *K. pneumoniae* [16, 17].

The clustered regularly interspersed short palindromic repeats (CRISPR) and CRISPR-associated genes constitute an RNA-guided adaptive immunity system (CRISPR/Cas system), which protects bacteria against HGT of mobile genetic elements (MGEs), such as plasmids and phages [18]. CRISPR/Cas system is composed of three main parts: i) CRISPR loci that consist of several non-contiguous direct repeats (DRs) separated by stretches of variable sequences called as spacers; ii) a group of *cas* genes essential for adaptive immunity; iii) the leader sequence that acts as the promoter [19]. Generally, the CRISPR/Cas system operates in three stages: adaptation, expression, and interference [20]. During adaptation, a fragment of foreign DNA is captured and integrated into CRISPR array in an ordinal manner, thus forming a new R-S unit. Subsequently, the R-S unit is transcribed into precursor CRISPR RNA (pre-crRNA), which is then processed to produce a mature CRISPR RNA (crRNA). Upon subsequent infection, the crRNA targets and recognizes complementary DNA sequence to degrade exogenous genes with the guide of Cas protein. According to the content of *cas* genes, two classes, six types, and over 45 subtypes of CRISPR/Cas systems have been identified [21]. Class I CRISPR/Cas systems (type I, III, and IV) rely on heteromeric multi-protein effector complexes, whereas class 2 systems (type II, V, and VI) depend on single multi-domain effector proteins, such as Cas9, Cpf1, and C2c2 [22]. In addition to the adaptive immunity, CRISPR/Cas system performs several other biological roles, including regulating gene expression, and participating in DNA repair and genomic evolution [23-25]. Several studies have explored the diversity of CRISPR/Cas systems in multiple species, including *Bifidobacterium*[26], *Escherichia coli* [27], and even the relatives to*K. variicola, K. pneumoniae* [28]. It has been demonstrated that *K. pneumoniae* mainly carried type I-E, I-E* and IV-A CRISPR/Cas systems, which played different roles in the dissemination of antibiotic resistance in *K. pneumoniae* [29]. Although*K. variicola* is becoming another species of concern after*K. pneumoniae* in *Klebsiella* genus due to its ability to acquire and spread ARGs, little is known about the role of CRISPR/Cas system in this species.

Here, we employed a genome mining approach to analyze the characterization of the CRISPR/Cas system in *K. variicola*strains, explore the association between CRISPR evolution and MLST (multiple locus sequence typing), and discussed the potential role of CRISPR/Cas system in ARGs spread.

2 | Methods

2.1 | Data collection

As of September 2021, all *K. variicola* genomes publicly available were retrieved from NCBI (National Center for Biotechnology Information) database (*http://ftp.ncbi.nlm.nih.gov/genomes/all*). A total of 487 available genomes of *K. variicola* were analyzed, including 56 complete and 431 draft genomes (Table S1).

2.2 | CRISPR/Cas identification

Based on the feature of repeat sequences within one single CRISPR locus, CRISPRCasFinder web server (*https://crispr.i2bc.paris-saclay.fr/Server/*) divided CRISPR loci into "confirmed" and "questionable" CRISPR. In all strains containing confirmed CRISPR loci, 10000 bp upstream and downstream of the loci were extracted. These sequences were then uploaded to the CRISPRCasFinder web server to determine the presence and content of *cas* genes. The subtypes of all CRISPR/Cas system were determined by the CRISPRCas typer server (*https://typer.crispr.dk/#/submit*) with default parameters [30]. Type I-E and

type I-E* systems were classified based on the content of *cas* genes, as previously described [31].

2.3 | Repeat structure prediction

The characteristic of repeat sequences was identified by Weblogo (*https://weblogo.berkeley.edu/logo.cgi*) web server. The RNA secondary structures of the repeat sequences were predicted by the RNA fold Web server. Meanwhile, the minimum free energy (MFE) was calculated.

2.4 |MLST and phylogenetic analyses

In silico analysis of MLST was performed by MLST 2.0 available on the CGE website using the seven housekeeping genes (i.e., *leuS, pgi, pgk, phoE, pyrG, rpoB,* and *fusA*) as queries (*https://cge.cbs.dtu.dk/services/MLST/*) [32]. Phylogenetic tree was constructed by Mega v7.0 using neighbor-joining method. Multiple sequence alignment was completed by MUSCLE v3.8.31 [33]. The visualization of the phylogenetic tree was implemented using iTOL v6 (*https://itol.embl.de*).

2.5 | Spacer analysis, protospacer target identification andprotospacer adjacent motif (PAM) determination

The putative origin of CRISPR spacers was acquired by the CRISPRtarget web server (*http://crispr.otago.ac.nz/CRISPRTarget/crispr analysis.html*). The 8bp nucleotide sequences from upstream of the predicted protospacers were extract to predict PAM using Weblogo (*https://weblogo.berkeley.edu/logo.cgi*) web server. The hierarchical clustering analysis of spacers was performed by the "seaborn" module in python script. The network of *K. variicola*spacers and MGEs from other species were visualized in Gephi with the layout generated by a combination of Fruchterman Reingold and Noverlap algorithms (*https://github.com/gephi/gephi*). Each pair of species was connected by at least one spacer-protospacer match.

2.6 | Statistical analysis

Statistical analyses were conducted using SPSS 21.0. The non-parametric Kruskal-Wallis (KW) test was used to compare spacer numbers in three types of CRISPR/Cas systems. p [?] 0.05 was considered to be statistically significant.

3 | Results

3.1 | Occurrence and diversity of CRISPR/Cas systems in*K. variicola*

The 487 genomes of *K. variicola* strains from the NCBI database were analyzed the occurrence and diversity of CRISPR/Cas system. Based on CRISPRCasFinder search results, 21.56% of *K. variicola*strains (105/487) harbored CRISPR/Cas systems, which was lower than the estimated occurrence rate of other bacteria (45%) [34]. Based on signature Cas proteins and repeat sequence, three different types of CRISPR/Cas systems (i.e., type I-E, I-E* and IV-A) were identified, which were located on the chromosome or the plasmids. The type I system was the most prevalent CRISPR/Cas system in *K. variicola* . Specifically, 72 type I-E, 10 type I-E* and 29 IV-A systems were identified among 105 strains (Figure 1A). As shown in Table S2, most strains (94.29%, 99/105) only had one type of CRISPR/Cas system, but 6 strains were found to contain two different types of systems. Type I-E and IV-A systems coexisted in five strains (strain HUMB 14438, K5-12018, 7041, CRE843, and TUM14103), and type I-E* and IV-A system coexisted in strain AS012291. Furthermore, most of CRISPR/Cas systems (91.89%, 102/111) harbored a series of intact *cas* gene clusters, but the partial absence of *cas* genes were observed in 9 type I-E CRISPR/Cas systems. As described in Figure S1, 6 strains were absent of*cas2* , 2 strains were absent of *cas3* , and 1 strain was absent of both *cas2* and *cas3* . Interestingly, we observed that the order of type I-E *cas* genes was slightly different from that of type I-E* *cas* genes. Regarding the type IV-A, *csf1, cas6, csf4, csf3* and *csf2* genes were encoded by the strains (Figure 1B).

3.2 | Analysis of CRISPR repeat-spacer arrays

The CRISPR array was composed of discontinuous DR sequences and intervening spacer sequences. As presented in Figure 2A, the size and base arrangement of the repeat sequence were relatively conserved within CRISPR types. The length of the repeat sequences was 29 nucleotides for type I-E and IV-A systems, 28 nucleotides for I-E* system, respectively (Figure 2D). In addition, the prediction results of RNA secondary structure showed that all types of repeats formed stable 'stem-loops' structures in the middle (Figure 2C). According to the predictions, the RNA secondary structure included 9, 7, and 6 bp stem lengths for types I-E, I-E*, and IV-A, respectively (Figure 2C). Based on the structure diagram and MFE value, the secondary structures of repeat sequences can be analyzed for conservation and stability. A small MFE value indicated high structural stability, the length of stem was proportional to structural stability. As shown in Figure 2C, the secondary structure of type I-E repeat sequence had the least MFE values and the longest stem colored in red, suggesting that the secondary structure of the type I-E repeat was the most stable.

Spacer sequences are captured into CRISPR array with the aid of Cas proteins. Spacer number reflects the activity of CRISPR/Cas system. As seen in Figure 2E, the number of the spacers was diverse. Among 105 strains carrying CRISPR/Cas system, strain WUSM_KV_47 had the largest number of spacers (41 type I-E spacers). The smallest spacer number was identified in strain EuSCAPE_TR218 (3 type IV-A spacers). Moreover, type I-E system (26.5, 17-37) had more spacers than type I-E* (13, 8.5-15; $p < 0.001$) and type IV-A (16, 13-20; $p < 0.001$). However, there was no significant difference in spacer numbers between type I-E* and type IV-A systems (13, 8.5-15 vs 16, 13-20; $p = 0.128$).

PAM plays an important role in the acquisition of spacer sequences. As shown in Figure 2B, PAM sequences for type I-E system, I-E* system was inferred to be 5'-AAG-3' and 5'-(C) GAA-3', respectively. Considering that PAM was essential elements for Cas protein to recognize and degrading foreign DNA, diverse PAM represented different Cas protein variants. The difference of PAM in type I-E and I-E* system further supported the evolutionary and functional divergence. Notably, PAM predicted for type IV-A system (5'-AAG-3') was identical to that predicted for type I-E.

3.3 | Relationship between CRISPR distribution and MLST

Based on MLST results, 243 different sequence types (STs) were identified in 427 *K. variicola* strains, but 60 strains were not assigned to a defined ST due to the limited information in PubMLST database. The most prevalent ST was ST20 (18/427, 4.22%), followed by ST60 (14/427, 3.28%) and ST10 (11/427, 2.58%). Further analysis found that the distributions of type I-E and type I-E* systems were strong associated with MLST, but type IV-A system was scattered throughout the whole genetic lineage (Figure 3). For example, once one strain within one ST harbor type I-E or I-E* system, all strains within the same ST were type I-E-positive (e.g., ST20, ST92, ST108, ST137) or I-E*-positive (e.g., ST188). However, this phenomenon was not found in strains containing type IV-A system.

To further clarify the relationship between CRISPR evolution and MLST, a hierarchical clustering analysis was performed based on the presence of spacers. Likewise, there was a strong association between the spacer contents of type I system and MLST (Figure 4A and 4B). For example, all ST20 strains harbored relatively conserved type I-E spacer contents. Likewise, type I-E* spacers also showed obvious aggregation (e.g., ST188). Differently, type IV-A spacer contents were random across MLST. As shown in Figure 4C, type IV-A spacers compositions of ST271, ST115, and ST148 were highly similar though they were phylogenetically unrelated.

3.4 | Analysis of spacer sequences and homology to foreign DNA

Spacers are the products of exogenous genetic elements, which records the encounters of host with invading DNA. There was a total of 536 type I-E, 168 I-E*, and 128 IV-A spacers present in 111 CRISPR/Cas systems. The homology search revealed that more than one-third of the spacers (39.30%, 327/832) were homologous to plasmids or phages. Specially, 173 type I-E spacers (32.28%, 173/536), 59 type I-E* spacers (35.12%, 59/168), and 95 type IV-A spacers (74.22%, 95/128) were homologous to plasmids or phages (Figure 5A). There were cases where one spacer sequence targeted both the plasmid and the phage. Besides, we observed that type I-E and IV-A systems displayed a target bias towards plasmids, whereas type I-E*

systems exhibited preference for phage targets (Figure 5A).

Through BLASTp analyses against the NCBI database, the function of proteins targeted by spacers was further investigated. If the hypothetical proteins, unknown and non-coding regions were not considered, conjugation transfer proteins were the most commonly targeted, such as TrbI, TrbD, and TrbH (Figure 5B). Notably, spacers from *K. variicola* targeted MGEs from multiple species, including*K. pnenumoniae* , *Escherichia coli, Salmonella enterica,* and other *species* (Figure 6 and Table S6-S8). Moreover, 103 strains have at least one spacer targeted MGEs from *K. pneumoniae* .

4 | Discussion

K. variicola is an emerging human pathogen with increasing antimicrobial resistance and virulence, which poses a threat to public health [1]. CRISPR/Cas system is an acquired immunity system, the most abundant form of innate immunity in prokaryotes. However, the diversity and characterization about CRISPR/Cas system in *K. variicola* is not well known. Similar to *K. pneumoniae* [28],*K. variicola* strains harbored three types of CRISPR/Cas systems, including type I-E, type I-E*, and type IV-A. Indeed, *K. variicola* was a member of the*K. pneumoniae* complex, which phylogenetically related to *K. pneumoniae* [6, 16, 35]. The consistency of CRISPR types provides new evidence for their evolutionary relationship.

Multiple types of CRISPR/Cas systems coexisted in the same strain in*K. variicola* . The coexistence of type I-E and IV-A system was found in five strains, and type I-E* and IV-A system coexisted in one strain. An explanation for this phenomenon was that the coexistence of different types of CRISPR/Cas systems contribute to exploiting mutual *cas* genes to perform corresponding functions [36]. In this study, all type IV-A systems lacked *cas1* and *cas2* genes. *cas1* and *cas2*genes are indispensable in the acquisition of spacers during adaptation stage [37]. It could be speculated that type IV-A system can utilize*cas* genes from type I system to perform adaptation roles. Meanwhile, we also noted that the PAM sequences of type I-E and type IV-A systems were identical, which further affirms their functional relationships. Additionally, it has been demonstrated that the synergy between coexisting subtypes was more conducive to avoiding immune evasion of MGEs [38, 39]. One study has found that type III CRISPR/Cas systems could provide redundancy to overcome phage escape from type I systems [38]. Accordingly, the coexistence of type I and type IV-A system in *K. variicola* reinforces the immune defenses against MGEs. In addition, the absence of *cas2* and *cas3*were found in type I-E CRISPR/Cas system. During CRISPR adaptation, the Cas2 dimer functions as an adaptor protein and forms a binding surface for the protospacer DNA combined with two Cas1 dimers [40]. The*cas3* harbors both nuclease and helicase activities, which is responsible for cutting and degrading invading DNA [41]. The different degree absence of *cas2* and *cas3* genes in type I-E system may represent a degenerated adaptation immunity role.

MLST has been applied to infer the phylogenetic relationship of strains [42]. Here, we observed the strong association between the distribution of chromosome-encoded type I CRISPR systems and MLST in*K. variicola* , which was also found in *K. pneumoniae*[43]. Nevertheless, plasmid-borne type IV-A systems in *K. variicola* were randomly distributed across MLST. This phenomenon could be attributed to the stable inheritance of chromosome-encoded CRISPR systems to the offspring and the HGT of CRISPR/Cas system mediated by plasmid. Moreover, it has been demonstrated that the scarcity of chromosome-encoded type I systems contributes to the global expansion of multidrug-resistant ST11 and ST258 clones in *K. pneumoniae* [44]. It can be speculated that the uneven distribution of CRISPR systems may be a contributor to the accumulation of ARGs in specific lineages of *K. variicola* . It has been reported that CRISPR-positive plasmids were rich in ARGs in*Klebsiella* genus [29]. The adaptive immune function carried by CRISPR might confer drug-resistant plasmids more competitive advantage in survival.

Although different CRISPR loci contain diverse repeat sequences, the base composition and arrangement of repeat sequence are conserved in the same CRISPR/Cas types [45]. We also noticed that the repeat sequences showed conservation within the same subtypes in *K. variicola* . However, there were a small number of base differences in the repeat sequences of the same CRISPR/Cas system. Although the repeat sequences were diverse during evolution, the function was conserved. The repeat sequences could be transcribed during

expression stage, thereby forming a stable RNA secondary structure [45, 46]. In this study, all types of repeat sequence formed relatively stable secondary structure. In secondary structures, conserved motifs interact to generate "stem-loop" that act as processing points for pre-crRNA through the *cas6* endonuclease [45-47]. According to MFE value theory, stems with higher GC content and larger match base numbers tend to harbor lower MFE, suggesting more stable secondary structure. Here, the MFE of the type I-E repeat sequence was the smallest, representing the most stable RNA secondary structure.

The adaptation activity of CRISPR/Cas system can be measured by the number of the spacer [22]. The number of the spacers in *K. variicola* was diverse for different CRISPR/Cas types, ranging from 3 to 41 spacers, which indicated their different abilities to obtain spacers. Type I-E system contained the most spacers, thereby demonstrating its strong adaptation ability. The protospacer sequence was the specific sequence on the foreign gene that was consistent with spacers [48, 49]. About one-third of spacers in *K. variicola* were homologous to plasmids or phages. The proportion of *K. variicola* spacers targeting foreign genetic elements was significantly higher than other bacteria, such as *Staphylococcus* [50] and *Clostridium perfringens* [51]. It might be owing to that the sequencing of*K. variicola* -related MGEs was overrepresented in the current databases. In this study, type I-E spacers showed homology with plasmids more than phages. Previous studies have shown that type I-E CRISPR/Cas systems in *K. pneumoniae* affected the dissemination of IncF antibiotic-resistant plasmids [44, 52]. Considering the evolution relationship between *K. variicola* and *K. pneumoniae* , it is possible that type I-E systems in *K. variicola* play a similar role in limiting the uptake and survival of antibiotic-resistant plasmids. Different from type I-E system, type I-E* system biasedly targeted phages, suggesting their different responsiveness to diverse MGEs. Uniquely, plasmid-borne type IV-A spacers target plasmids more effectively than phages. Moreover, the target preference for plasmids in type IV-A system was more obvious than chromosome-encoded type I system (type I-E and I-E*). The strong association between type IV-A system and plasmids can be attributed to the fierce battle between similar entities competing for overlapping niches and resources [53-55].

The spacer sequence reflects the history of interaction between bacteria and MGEs [56]. We observed that *K. variicola* spacers prefer to target MGEs from *K. pneumoniae,* which suggested that *K. varricolla* were frequently challenged by *K. penumoniae* MGEs. Previous study has shown that *K. variicola* displayed similar biological features to *K. pneumoniae* [57]. The strong interaction between *K. varricola* and *K. pneumoniae* that are implied by spacer target events provides new evidence for their evolutionary affinity. In addition, we analyzed the genes-targeted characteristics. The diversity of coding products indicated the wide range of spacer sources and the diversity of CRISPR functions [58]. Conjugative transfer-related proteins were the most frequently targeted proteins, such as TrbI, TrbD, and TrbH. These Trb proteins were required for pilus extension and conjugal DNA transfer, which were relatively active [59-61]. Conjugative transfer was the main mechanism in the spread of ARGs in which conjugative transfer proteins played an important role in mediating plasmid DNA exchange [62]. In our study, we observed that 15.15% (35/231) spacers targeted conjugative transfer proteins, which further highlight the important role of CRISPR/Cas system in influencing the dissemination of antibiotic resistance. In addition, conserved genes encoding phage tail protein, methyltransferase, kinase, terminase were also frequently targeted by CRISPR spacers. These direct targets to conserved gene regions is beneficial for improving the defense efficiency of CRISPR systems and avoiding immunity escape of MGEs.

In conclusion, our results discovers that *K. variicola* harbors diverse and complex CRISPR/Cas systems, including three types. The distribution of chromosome-encoded type I CRISPR/Cas system is associated with MLST whereas plasmid-encoded IV-A system is randomly distributed across MLST. Spacer homology analysis reveal that CRISPR/Cas systems targeted diverse MGEs, especially conjugative transfer-related proteins related to antibiotic-resistant plasmids, thereby suggesting the critical role of CRISPR/Cas system in control ARG spread in *K. variicola* . Overall, this study provides new insights into the evolution of CRISPR/Cas systems in this specie.

Acknowledgements

The work was supported by the China Postdoctoral Science Foundation (2022M712859) and National Sci-

ence and Technology Specific Projects (2018ZX10301407). We thank National Supercomputing Center in Zhengzhou for its technical supports.

Author contributions statements

JL and SC designed the study. JZ, HY and YJ helped collected some data. YX and JL analyzed data and wrote the paper. The paper was written, reviewed, and edited by all of the authors. All authors read and approved the final manuscript.

Conflict of interest statement

The authors declare no conflict of interest.

Figure Legend

Figure 1 CRISPR/Cas systems in *K. variicola* . (A) The pie chart shows the proportion of 105 strains with and without CRISPR/Cas systems. The stacking plot represents the number of each CRISPR/Cas subtype. (B) Schematic diagram of each CRISPR/Cas subtype. Different colors are used to represent the CRISPR/Cas system and its gene neighbor. The *K. variicola* type I-E system is located between the*cysH* and *iap* genes, and the *cas* genes are shown in blue. The type I-E* system is located on the downstream of the ABC transporter system, and the *cas* genes are shown in green. The type IV-A system is adjacent to the *umuD* and *umuC* , and the*cas* genes are shown in orange. The repeats and spacers of CRISPR are represented by white diamonds and black rectangles, respectively. All genes are drawn to scale.

Figure 2 (A) Visualization and alignment of nucleotides of repeat sequences in each CRISPR/Cas subtype. (B) PAM prediction of each CRISPR/Cas subtype. (C) Secondary structure prediction and MFE value of each CRISPR/Cas subtype. Nucleotide bases are colored by base-pair probability. (D) The repeat length of three types of CRISPR/Cas system. (E) The number of three types of CRISPR/Cas system. The box plot represents the number of spacers detected in each subtype. *: $p< 0.05$, **: $p < 0.01$, ***: $p< 0.001$, ns: non-significance.

Figure 3 Phylogenetic tree based on MLST. From inside to outside, the colored rings indicate the type I-E, type I-E*, type IV-A CRISPR/Cas systems, respectively. The short band indicates that the*cas* genes in the system is incomplete, and the long band indicates that the CRISPR/Cas system is complete.

Figure 4 Hierarchical clustering analysis of spacer arrangements in type I-E (A), type I-E* (B), type IV-A systems (C). The various colors correspond to different STs, and the STs are marked on the left. The blue and gray squares denote the presence and absence of spacers, respectively.

Figure 5 (A) Comparison of detected spacer-protospacer matches for type I-E, type I-E*, and type IV-A systems. (B) Spacers target specific exogenous genes. Nearly one-quarter of *K. variicola*spacers have a match to genes with informative annotations, and the proportion of gene matching is presented.

Figure 6 The network of*K. variicola* and MGEs from other species based on protospacer-spacer matches in type I-E (A), type I-E* (B), type IV-A (C) systems. Nodes indicate individual spacers and edges represent CRISPR spacer targeting based on spacer-protospacer matches.

Figure S1 Comparison of *cas* genes between *K. variicola* strain WCHKP19 and 9 *K. variicola* strains. The genes (*cas3, cse1, cse2, cas7* , and *cas5*) responsible for the interference are shown in blue. The gene (*cas6*) involve in expression is highlighted in red. The adaptation genes (*cas1* and*cas2*) are depicted in orange. The repeats and spacers are represented by white diamonds and black rectangles, respectively. All genes are drawn to scale.

References

[1] N. Rodriguez-Medina, H. Barrios-Camacho, J. Duran-Bedolla, U. Garza-Ramos, *Klebsiella variicola* : an emerging pathogen in humans, Emerg Microbes Infect, 8 (2019) 973-988.

[2] R.F. Potter, W. Lainhart, J. Twentyman, M.A. Wallace, B. Wang, C.A. Burnham, et al., Population Structure, Antibiotic Resistance, and Uropathogenicity of *Klebsiella variicola* , mBio, 9 (2018).

[3] H. Barrios-Camacho, A. Aguilar-Vera, M. Beltran-Rojel, E. Aguilar-Vera, J. Duran-Bedolla, N. Rodriguez-Medina, et al., Molecular epidemiology of *Klebsiella variicola* obtained from different sources, Sci Rep, 9 (2019) 10610.

[4] E. Martinez-Romero, J. Silva-Sanchez, H. Barrios, N. Rodriguez-Medina, J. Martinez-Barnetche, J. Tellez-Sosa, et al., Draft Genome Sequences of *Klebsiella variicola* Plant Isolates, Genome Announc, 3 (2015).

[5] K. Imai, N. Ishibashi, M. Kodana, N. Tarumoto, J. Sakai, T. Kawamura, et al., Clinical characteristics in blood stream infections caused by *Klebsiella pneumoniae* , *Klebsiella variicola* , and*Klebsiella quasipneumoniae* : a comparative study, Japan, 2014-2017, BMC Infect Dis, 19 (2019) 946.

[6] M. Rosenblueth, L. Martinez, J. Silva, E. Martinez-Romero,*Klebsiella variicola* , a novel species with clinical and plant-associated isolates, Syst Appl Microbiol, 27 (2004) 27-35.

[7] E. Piepenbrock, P.G. Higgins, J. Wille, K. Xanthopoulou, J. Zweigner, P. Jahn, et al., *Klebsiella variicola* causing nosocomial transmission among neonates - an emerging pathogen?, J Med Microbiol, 69 (2020) 396-401.

[8] M. Maatallah, M. Vading, M.H. Kabir, A. Bakhrouf, M. Kalin, P. Naucler, et al., *Klebsiella variicola* is a frequent cause of bloodstream infection in the stockholm area, and associated with higher mortality compared to *K. pneumoniae* , PLoS One, 9 (2014) e113539.

[9] S. Brisse, V. Passet, P.A.D. Grimont, Description of*Klebsiella quasipneumoniae* sp. nov., isolated from human infections, with two subspecies, *Klebsiella quasipneumoniae*subsp. *quasipneumoniae* subsp. nov. and *Klebsiella quasipneumoniae* subsp. *similipneumoniae* subsp. nov., and demonstration that *Klebsiella singaporensis* is a junior heterotypic synonym of *Klebsiella variicola* , Int J Syst Evol Microbiol, 64 (2014) 3146-3152.

[10] M.P. Podder, L. Rogers, P.K. Daley, G.P. Keefe, H.G. Whitney, K. Tahlan, *Klebsiella* species associated with bovine mastitis in Newfoundland, PLoS One, 9 (2014) e106518.

[11] Y. Lu, Y. Feng, A. McNally, Z. Zong, Occurrence of colistin-resistant hypervirulent *Klebsiella variicola* , J Antimicrob Chemother, 73 (2018) 3001-3004.

[12] K.L. Hopkins, J. Findlay, M. Doumith, B. Mather, D. Meunier, S. D'Arcy, et al., IMI-2 carbapenemase in a clinical *Klebsiella variicola* isolated in the UK, J Antimicrob Chemother, 72 (2017) 2129-2131.

[13] K. Zurfluh, L. Poirel, P. Nordmann, J. Klumpp, R. Stephan, First detection of *Klebsiella variicola* producing OXA-181 carbapenemase in fresh vegetable imported from Asia to Switzerland, Antimicrob Resist Infect Control, 4 (2015) 38.

[14] J. Osei Sekyere, M.A. Reta, Global evolutionary epidemiology and resistome dynamics of *Citrobacter* species, *Enterobacter hormaechei* , *Klebsiella variicola* , and *Proteeae* clones, Environ Microbiol, 23 (2021) 7412-7431.

[15] K.E. Holt, H. Wertheim, R.N. Zadoks, S. Baker, C.A. Whitehouse, D. Dance, et al., Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae* , an urgent threat to public health, Proc Natl Acad Sci U S A, 112 (2015) E3574-3581.

[16] S.W. Long, S.E. Linson, M. Ojeda Saavedra, C. Cantu, J.J. Davis, T. Brettin, et al., Whole-Genome Sequencing of Human Clinical*Klebsiella pneumoniae* Isolates Reveals Misidentification and Misunderstandings of *Klebsiella pneumoniae* , *Klebsiella variicola* , and *Klebsiella quasipneumoniae* , mSphere, 2 (2017).

[17] K.L. Wyres, K.E. Holt, *Klebsiella pneumoniae* Population Genomics and Antimicrobial-Resistant Clones, Trends Microbiol, 24 (2016) 944-956.

[18] J.E. Garneau, M.E. Dupuis, M. Villion, D.A. Romero, R. Barrangou, P. Boyaval, et al., The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA, Nature, 468 (2010) 67-71.

[19] R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, et al., CRISPR provides acquired resistance against viruses in prokaryotes, Science, 315 (2007) 1709-1712.

[20] F. Hille, H. Richter, S.P. Wong, M. Bratovic, S. Ressel, E. Charpentier, The Biology of CRISPR-Cas: Backward and Forward, Cell, 172 (2018) 1239-1259.

[21] K.S. Makarova, Y.I. Wolf, O.S. Alkhnbashi, F. Costa, S.A. Shah, S.J. Saunders, et al., An updated evolutionary classification of CRISPR-Cas systems, Nat Rev Microbiol, 13 (2015) 722-736.

[22] P. Mohanraju, K.S. Makarova, B. Zetsche, F. Zhang, E.V. Koonin, J. van der Oost, Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems, Science, 353 (2016) aad5147.

[23] E.R. Westra, A. Buckling, P.C. Fineran, CRISPR-Cas systems: beyond adaptive immunity, Nat Rev Microbiol, 12 (2014) 317-326.

[24] B. Wiedenheft, S.H. Sternberg, J.A. Doudna, RNA-guided genetic silencing systems in bacteria and archaea, Nature, 482 (2012) 331-338.

[25] A.F. Andersson, J.F. Banfield, Virus population dynamics and acquired virus resistance in natural microbial communities, Science, 320 (2008) 1047-1050.

[26] L. Ou, J. Long, Y. Teng, H. Yang, Y. Xi, G. Duan, et al., Diversity of the type I-U CRISPR-Cas system in *Bifidobacterium* , Arch Microbiol, 203 (2021) 3235-3243.

[27] J. Long, Y. Xu, L. Ou, H. Yang, Y. Xi, S. Chen, et al., Polymorphism of Type I-F CRISPR/Cas system in *Escherichia* coli of phylogenetic group B2 and its application in genotyping, Infect Genet Evol, 74 (2019) 103916.

[28] M. Kamruzzaman, J.R. Iredell, CRISPR-Cas System in Antibiotic Resistance Plasmids in *Klebsiella pneumoniae* , Front Microbiol, 10 (2019) 2934.

[29] J. Long, J. Zhang, Y. Xi, J. Zhao, Y. Jin, H. Yang, et al., Genomic Insights into CRISPR-Harboring Plasmids in the *Klebsiella*Genus: Distribution, Backbone Structures, Antibiotic Resistance, and Virulence Determinant Profiles, Antimicrob Agents Chemother, 67 (2023) e0118922.

[30] J. Russel, R. Pinilla-Redondo, D. Mayo-Munoz, S.A. Shah, S.J. Sorensen, CRISPRCasTyper: Automated Identification, Annotation, and Classification of CRISPR-Cas Loci, CRISPR J, 3 (2020) 462-469.

[31] J. Shen, L. Lv, X. Wang, Z. Xiu, G. Chen, Comparative analysis of CRISPR-Cas systems in *Klebsiella* genomes, J Basic Microbiol, 57 (2017) 325-336.

[32] Y. Matsumura, [Multilocus sequence typing (MLST) analysis], Rinsho Byori, 61 (2013) 1116-1122.

[33] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res, 32 (2004) 1792-1797.

[34] I. Grissa, G. Vergnaud, C. Pourcel, The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats, BMC Bioinformatics, 8 (2007) 172.

[35] E. Mondo, R. Rinnovati, A. Spadari, F. Giacometti, A. Serraino, F. Savini, et al., First isolation of *Klebsiella variicola* from a horse pleural effusion, BMC Vet Res, 17 (2021) 75.

[36] G. Yadav, R. Singh, In silico analysis reveals the co-existence of CRISPR-Cas type I-F1 and type I-F2 systems and its association with restricted phage invasion in *Acinetobacter baumannii* , Front Microbiol, 13 (2022) 909886.

[37] O. Musharova, S. Medvedeva, E. Klimuk, N.M. Guzman, D. Titova, V. Zgoda, et al., Prespacers formed during primed adaptation associate with the Cas1-Cas2 adaptation complex and the Cas3 interference nuclease-helicase, Proc Natl Acad Sci U S A, 118 (2021).

[38] S. Silas, P. Lucas-Elio, S.A. Jackson, A. Aroca-Crevillen, L.L. Hansen, P.C. Fineran, et al., Type III CRISPR-Cas systems can provide redundancy to counteract viral escape from type I systems, Elife, 6 (2017).

[39] L. Deng, R.A. Garrett, S.A. Shah, X. Peng, Q. She, A novel interference mechanism by a type IIIB CRISPR-Cmr module in*Sulfolobus* , Mol Microbiol, 87 (2013) 1088-1099.

[40] J. Wang, J. Li, H. Zhao, G. Sheng, M. Wang, M. Yin, et al., Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems, Cell, 163 (2015) 840-853.

[41] Y. Huo, K.H. Nam, F. Ding, H. Lee, L. Wu, Y. Xiao, et al., Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation, Nat Struct Mol Biol, 21 (2014) 771-777.

[42] M.C. Maiden, Multilocus sequence typing of bacteria, Annu Rev Microbiol, 60 (2006) 561-588.

[43] H.Y. Li, C.Y. Kao, W.H. Lin, P.X. Zheng, J.J. Yan, M.C. Wang, et al., Characterization of CRISPR-Cas Systems in Clinical*Klebsiella pneumoniae* Isolates Uncovers Its Potential Association With Antibiotic Susceptibility, Front Microbiol, 9 (2018) 1595.

[44] Y. Zhou, Y. Tang, P. Fu, D. Tian, L. Yu, Y. Huang, et al., The type I-E CRISPR-Cas system influences the acquisition of bla(KPC)-IncF plasmid in *Klebsiella pneumonia* , Emerg Microbes Infect, 9 (2020) 1011-1022.

[45] F.J. Mojica, C. Diez-Villasenor, E. Soria, G. Juez, Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria, Mol Microbiol, 36 (2000) 244-246.

[46] V. Kunin, R. Sorek, P. Hugenholtz, Evolutionary conservation of sequence and secondary structures in CRISPR repeats, Genome Biol, 8 (2007) R61.

[47] N. Wakefield, R. Rajan, E.J. Sontheimer, Primary processing of CRISPR RNA by the endonuclease Cas6 in *Staphylococcus epidermidis* , FEBS Lett, 589 (2015) 3197-3204.

[48] C. Pourcel, G. Salvignol, G. Vergnaud, CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies, Microbiology (Reading), 151 (2005) 653-663.

[49] A. Bolotin, B. Quinquis, A. Sorokin, S.D. Ehrlich, Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin, Microbiology (Reading), 151 (2005) 2551-2561.

[50] M. Zhang, C. Bi, M. Wang, H. Fu, Z. Mu, Y. Zhu, et al., Analysis of the structures of confirmed and questionable CRISPR loci in 325 *Staphylococcus* genomes, J Basic Microbiol, 59 (2019) 901-913.

[51] J. Long, Y. Xu, L. Ou, H. Yang, Y. Xi, S. Chen, et al., Diversity of CRISPR/Cas system in *Clostridium perfringens* , Mol Genet Genomics, 294 (2019) 1263-1275.

[52] Y. Tang, P. Fu, Y. Zhou, Y. Xie, J. Jin, B. Wang, et al., Absence of the type I-E CRISPR-Cas system in *Klebsiella pneumoniae* clonal complex 258 is associated with dissemination of IncF epidemic resistance plasmids in this clonal complex, J Antimicrob Chemother, 75 (2020) 890-895.

[53] E.V. Koonin, K.S. Makarova, Mobile Genetic Elements and Evolution of CRISPR-Cas Systems: All the Way There and Back, Genome Biol Evol, 9 (2017) 2812-2825.

[54] J.E. Peters, K.S. Makarova, S. Shmakov, E.V. Koonin, Recruitment of CRISPR-Cas systems by Tn7-like transposons, Proc Natl Acad Sci U S A, 114 (2017) E7358-E7366.

[55] S.M. Deane, D.E. Rawlings, Plasmid evolution and interaction between the plasmid addiction stability systems of two related broad-host-range IncQ-like plasmids, J Bacteriol, 186 (2004) 2123-2133.

[56] E.V. Koonin, Y.I. Wolf, Evolution of the CRISPR-Cas adaptive immunity systems in prokaryotes: models and observations on virus-host coevolution, Mol Biosyst, 11 (2015) 20-27.

[57] C. Rodrigues, V. Passet, A. Rakotondrasoa, S. Brisse, Identification of *Klebsiella pneumoniae* , *Klebsiella quasipneumoniae* , *Klebsiella variicola* and Related Phylogroups by MALDI-TOF Mass Spectrometry, Front Microbiol, 9 (2018) 3000.

[58] M. Touchon, E.P. Rocha, The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella* , PLoS One, 5 (2010) e11126.

[59] R.L. Harris, P.M. Silverman, Tra proteins characteristic of F-like type IV secretion systems constitute an interaction group by yeast two-hybrid analysis, J Bacteriol, 186 (2004) 5480-5485.

[60] R.L. Harris, V. Hombs, P.M. Silverman, Evidence that F-plasmid proteins TraV, TraK and TraB assemble into an envelope-spanning structure in *Escherichia* coli, Mol Microbiol, 42 (2001) 757-766.

[61] J.A. Heinemann, G.F. Sprague, Jr., Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast, Nature, 340 (1989) 205-209.

[62] M. Llosa, F.X. Gomis-Ruth, M. Coll, F. de la Cruz Fd, Bacterial conjugation: a two-step mechanism for DNA transport, Mol Microbiol, 45 (2002) 1-8.