Legends to supplementary materials

Dataset S1. *Klebsiella* genomic data used in this study. Part A, *Klebsiella* genomes retrieved from the NCBI Genome database. Part B, chromosomes and plasmids in the complete *Klebsiella* genomes. Part C, other *Klebsiella* plasmids retrieved from the NCBI RefSeq database.

Dataset S2. *In silico* IncFII$_K$ typing results in *Klebsiella* genomes.

Dataset S3. *In silico* multilocus sequence typing (MLST) of *Klebsiella pneumoniae*

Dataset S4. The IncFII$_K$ plasmids with full sequences analysed in this study.

Dataset S5. The antibiotic resistance genes identified on IncFII$_K$ plasmids.

Dataset S6. The transfer regions of IncFII$_K$ plasmids.

Table S1. Distribution of IncFII$_K$ replicons found in the *Klebsiella* genomes.

Table S2. Distribution of multi-replicon IncFII$_K$ plasmids.

Figure S1. The IncFII$_K$ alleles and PlasmidFinder probes used for *in silico* IncFII$_K$ plasmid typing. Alleles were downloaded from pMLST. Colours indicated different allele groups. Percentages in brackets are identities in nucleotide levels between the alleles and the best matching probes. Multiple sequence alignment was performed by MUSCLE v3.7. Then the tree was constructed using neighbor-joining (NJ) method (A) and maximum likelihood (ML) method (B) with 1,000 bootstrap replicates in MEGA5. The two trees display similar topology. IncFII$_K$ alleles (K1-K17) from pMLST were used to fish out the IncFII$_K$ replicon probes from PlasmidFinder. The alleles K1-K3, K5-K8, K11, and K13-16 matched to the literally IncFII$_K$ probe “IncFII(K)_1__CP000648” as the best hit with 95%-100% BLASTn identities. In contrast, alleles K4, K9, K10, K12 and K17 matched to probe

"IncFII_1_pKP91_CP000966" as the best hit. Alleles K4 and K9 were 96% - 100% identical to the probe sequence, while K10, K12 and K17 displayed a lower identity (88%-93%). Although there is literal discrepancy in name, the probe "IncFII_1_pKP91_CP000966" was defined by plasmid pKP91 (GenBank accession number: CP000966), which is also an IncFII$_K$ plasmid.

Figure S2. The *in silico* IncFII$_K$ plasmid typing strategy.

Figure S3. Neighbor-joining (NJ) tree and maximum likelihood (ML) trees of all IncFII$_K$ alleles. The tree was constructed using NJ method (left) and ML method (rightB) with 1,000 bootstrap replicates in MEGA5. The two trees display similar topology. Colours indicated different allele groups. Percentages in brackets are identities in nucleotide levels between the alleles and the best matching probes.

Figure S4. Plasmids of same IncFII$_K$ type are widespread in *K. pneumoniae* with different sequence types. Histogram is the count of occurrence of the IncFII$_K$ alleles in *Klebsiella pneumonia* genomes specified by sequence type. In the block is the zoom-in view of types IncFII$_{K3}$-IncFII$_{K33}$ and unclassified alleles.

Figure S5. Comparison of the conjugation modules between plasmid pKPHS2 and plasmid F. GenBank accession numbers are CP003224 for pKPHS2 and AP001918 for plasmid F. Gene organisation is drawn to scale. Homologous counterparts are connected by orange lines. Identities in amino acid level of the components between the two plasmids are shown below.

Figure S5. Phylogenetic tree of the DNA sequences of the transfer regions of the studied plasmids. Plasmids with same names are discriminated by their accession numbers in the

brackets. The letters ‘a’ and ‘b’ are used to distinguish the two transfer regions on plasmid pKPC-727. Multiple sequence alignment was performed by Kalign v2.04. Then the tree was constructed using neighbor-joining method with 1,000 bootstrap replicates in MEGA5.

Figure S7. Phylogenetic trees of the proteins TraV and TrbE and TrbF of the studied plasmids. Plasmids with same names are discriminated by their accession numbers in the brackets. The letters ‘a’ and ‘b’ are used to distinguish the two conjugation modules on plasmid pKPC-727. Multiple sequence alignment was performed by MUSCLE v3.7. Then the tree was constructed using neighbor-joining method with 1,000 bootstrap replicates in MEGA5.

Table S1. Distribution of IncFII$_K$ replicons found in the *Klebsiella* genomes.

Species	Analysed genome no.	No. of genomes containing IncFII$_K$	Positive ratio	IncFII$_K$ replicons no.	No. of IncFII$_K$ replicons on plasmids	No. of IncFII$_K$ replicons on contigs
K. cf. planticola B43	1					
K. michiganensis	46	12	26.09%	12	1	11
K. oxytoca	84	16	19.05%	16	1	15
K. pneumoniae	**1,258**	**868**	**69.00%**	**1,014**	**40**	**974**
K. quasipneumoniae	15	5	33.33%	5	1	4
Klebsiella sp. 10982	1	1	100.00%	1		1
Klebsiella sp. 1 1 55	1					
Klebsiella sp. 4 1 44FAA	1	1	100.00%	2		2
Klebsiella sp. AA405	1					
Klebsiella sp. AS10	1					

Klebsiella sp. D5A	1	1	100.00%	2		2
Klebsiella sp. G5	1					
Klebsiella sp. KGM-IMP216	1	1	100.00%	1		1
Klebsiella sp. KTE92	1					
Klebsiella sp. MS 92-3	1	1	100.00%	1		1
Klebsiella sp. OBRC7	1					
Klebsiella sp. RIT-PI-d	1					
Klebsiella sp. S1	1					
K. variicola	24	12	50%	15	1	14
Klebsiella variicola CAG:634	0					
Total	1,441	918	63.71%	1,069	44	1,025

Data for *K. pneumoniae* are in bold.

Grey lettering indicates the species not analysed.

IncFII$_K$ replicons are listed in Supplementary Dataset S3.

Table S2. Distribution of multi-replicon IncFII$_K$ plasmids.

	None	IncFIA$_{HI1}$	IncFIB$_K$	IncFIB$_{pQil}$	IncFII$_{Yp}$	IncR	IncFIA$_{HI1}$, IncFIB$_K$	IncFIA$_{HI1}$, IncL/M$_{pOXA-48}$	IncFIB$_K$, IncFII$_{pCRY}$	IncFIB$_K$, IncR	IncFIB$_K$, IncR, IncN	Total
IncFII$_{K1}$			13	3		2				1	1	2
IncFII$_{K2}$	5		1	18		1						25
IncFII$_{K3}$		1										1
IncFII$_{K4}$		2										2
IncFII$_{K5}$	1		4			1			1			7
IncFII$_{K7}$	1	2	1	1				1				6
IncFII$_{K8}$			3									3
IncFII$_{K1}$							2					2
IncFII$_{K12}$		1										1
IncFII$_{K20}$			1									1
IncFII$_{K19}$			1									1

IncFII$_{K13}$			1									1
IncFII$_{K14}$						2						2
IncFII$_{K2}$			2									2
IncFII$_{K21}$					1							1
IncFII$_{K25}$			1									1
IncFII$_{K17}$	13											1
Total	8	6	28	22	1	6	2	1	1	1	1	77

Detail information is listed in Supplementary Dataset S4.

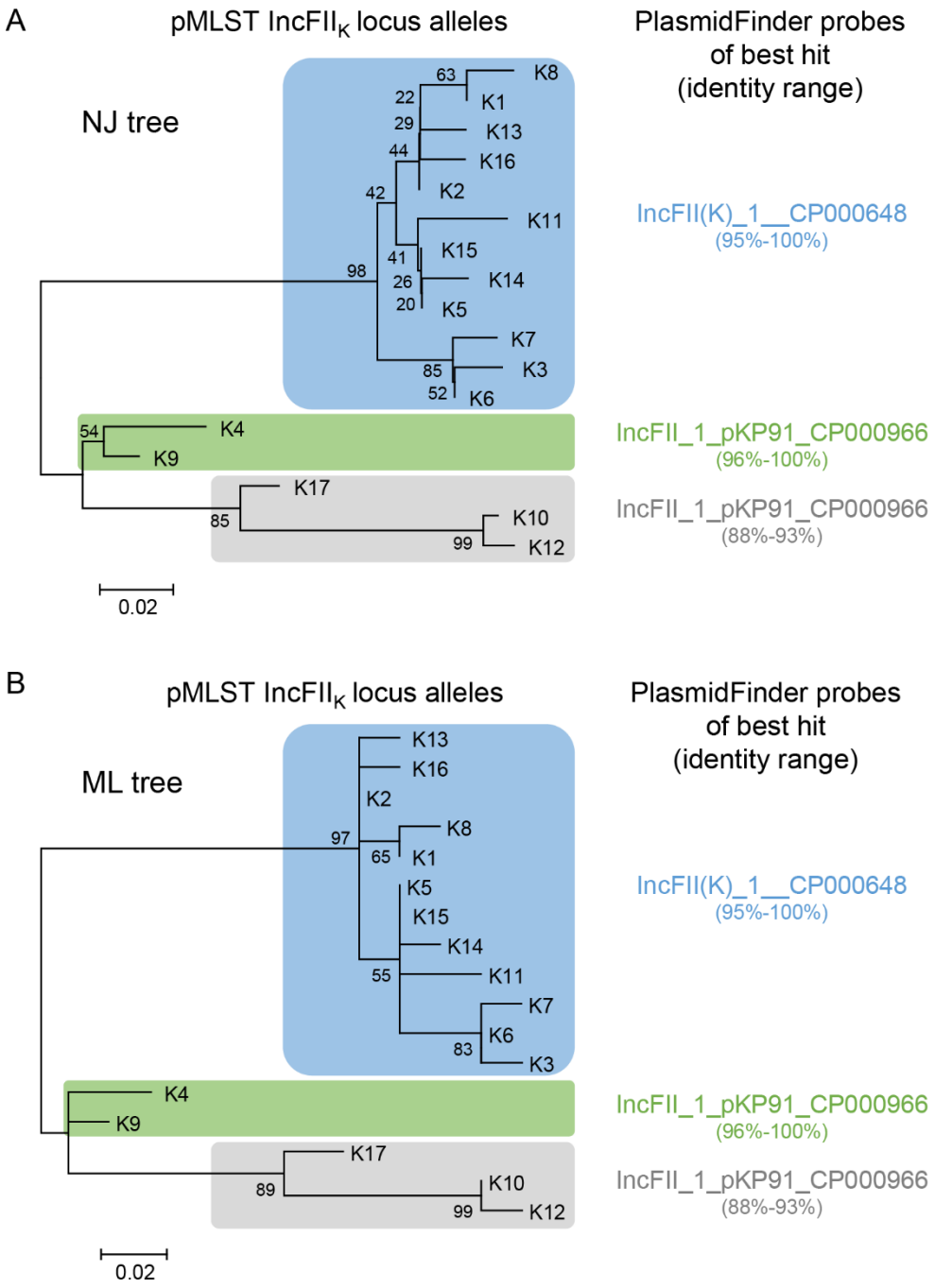


Figure S1. The IncFII$_K$ alleles and PlasmidFinder probes used for *in silico* IncFII$_K$ plasmid typing. Alleles were downloaded from pMLST. Colours indicated different allele groups. Percentages in brackets are identities in nucleotide levels between the alleles and the best matching probes. Multiple sequence alignment was performed by MUSCLE v3.7. Then the tree was constructed using neighbor-joining (NJ) method (A) and maximum likelihood (ML) method (B) with 1,000 bootstrap replicates in MEGA5. The two trees display similar

topology. IncFII$_K$ alleles (K1-K17) from pMLST were used to fish out the IncFII$_K$ replicon probes from PlasmidFinder. The alleles K1-K3, K5-K8, K11, and K13-16 matched to the literally IncFII$_K$ probe “IncFII(K)_1__CP000648” as the best hit with 95%-100% BLASTn identities. In contrast, alleles K4, K9, K10, K12 and K17 matched to probe “IncFII_1_pKP91_CP000966” as the best hit. Alleles K4 and K9 were 96% - 100% identical to the probe sequence, while K10, K12 and K17 displayed a lower identity (88%-93%). Although there is literal discrepancy in name, the probe “IncFII_1_pKP91_CP000966” was defined by plasmid pKP91 (GenBank accession number: CP000966), which is also an IncFII$_K$ plasmid. The IncFII$_K$ alleles fall into three groups according to the probe-mapping profile and the grouping was supported by phylogenetic analysis of those alleles. Therefore, we used the two probes and the 17 alleles to define IncFII$_K$ plasmids and adopted a customised set of thresholds.

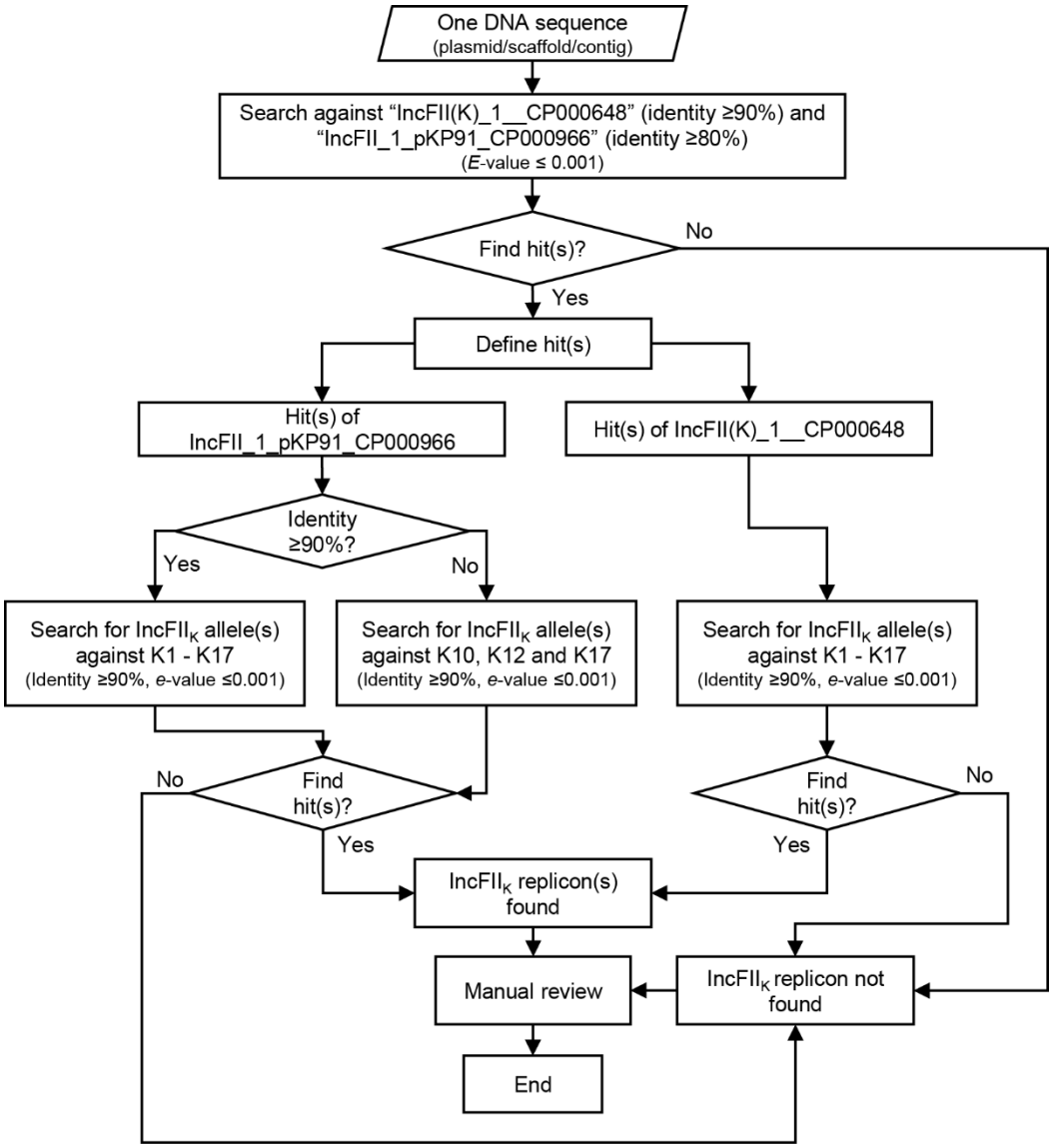


Figure S2. The *in silico* IncFII$_K$ plasmid typing strategy.

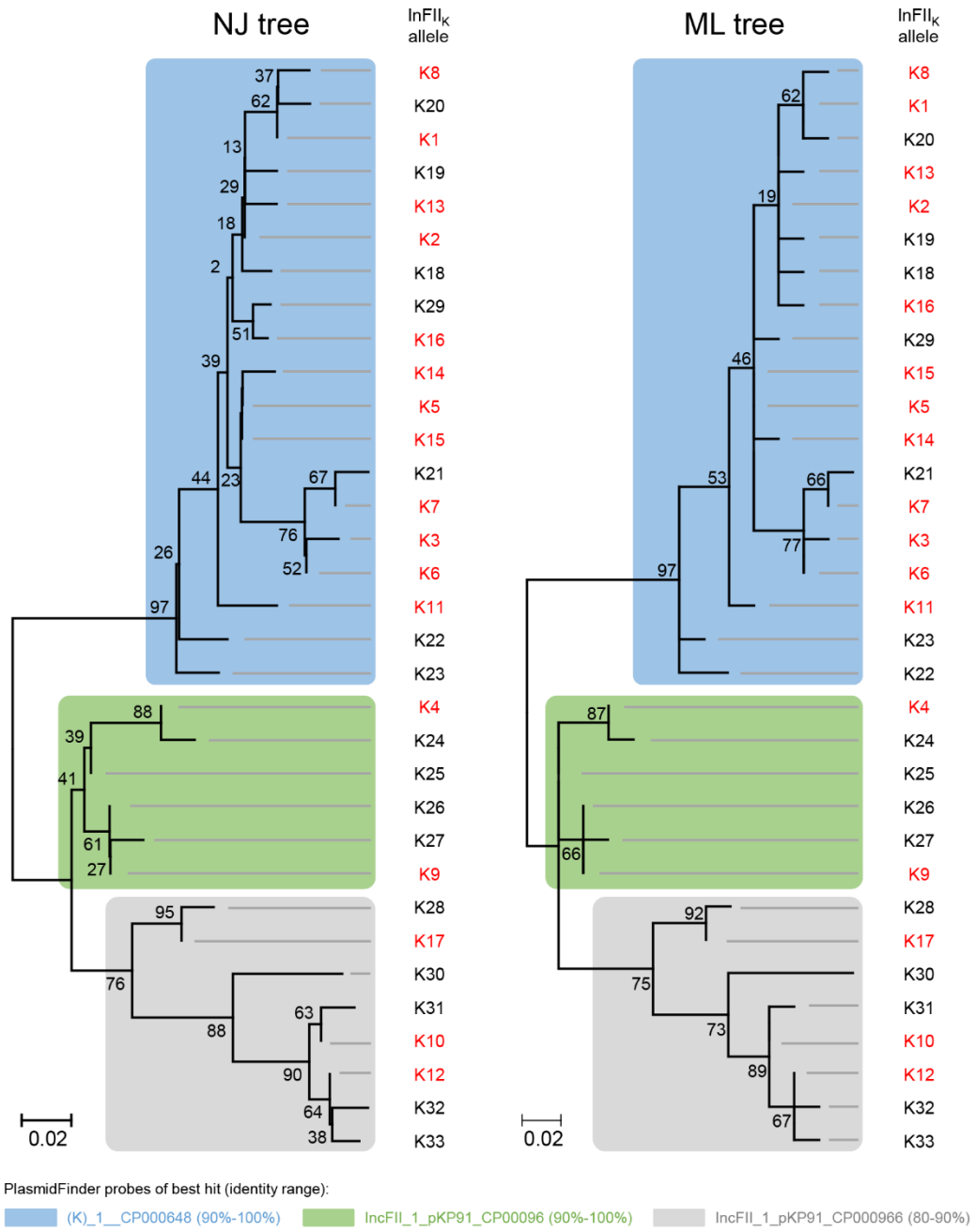


Figure S3. Neighbor-joining (NJ) tree and maximum likelihood (ML) trees of all IncFII$_K$ alleles. The tree was constructed using NJ method (left) and ML method (rightB) with 1,000 bootstrap replicates in MEGA5. The two trees display similar topology. Colours indicated different allele groups. Percentages in brackets are identities in nucleotide levels between the alleles and the best matching probes.

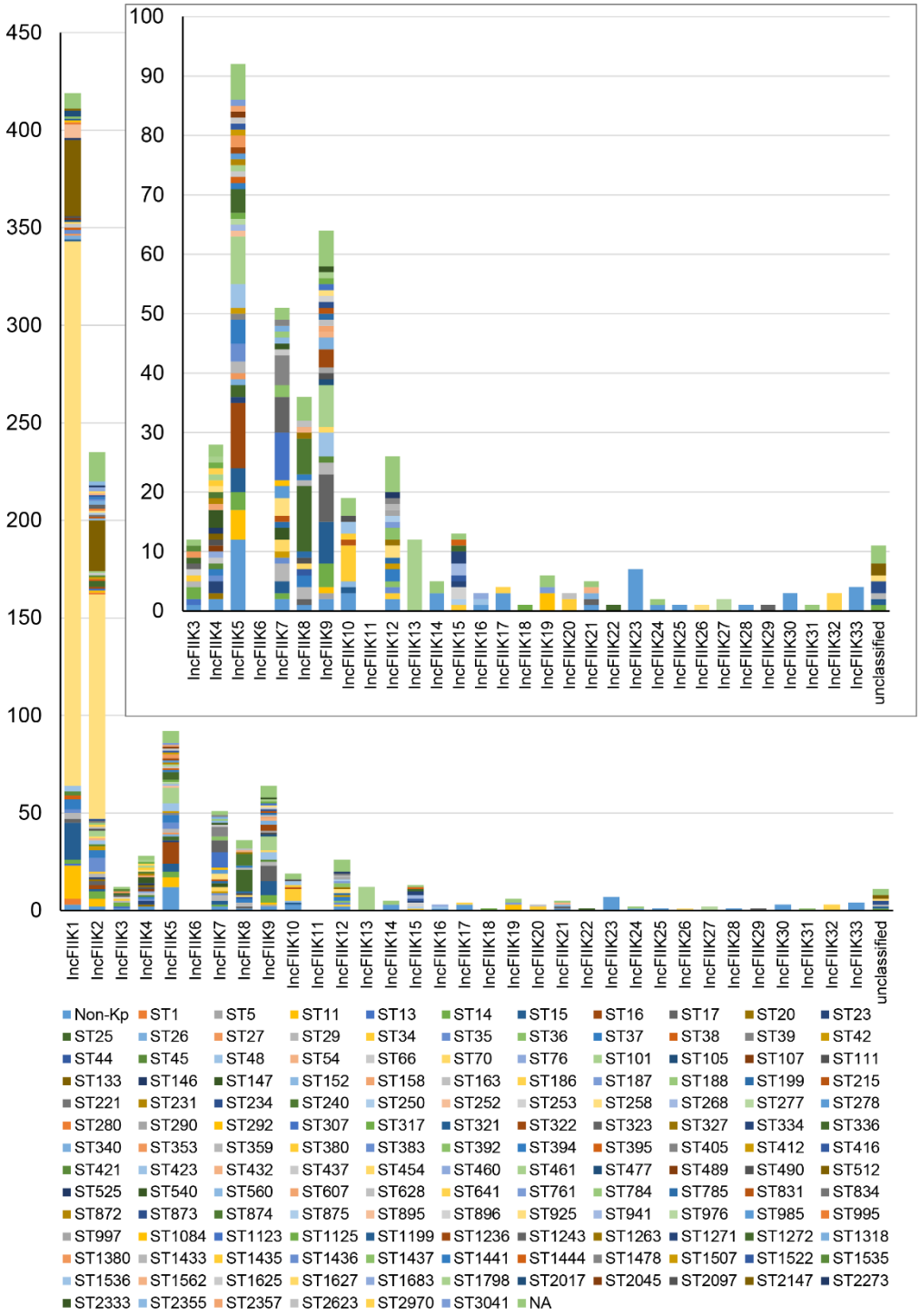
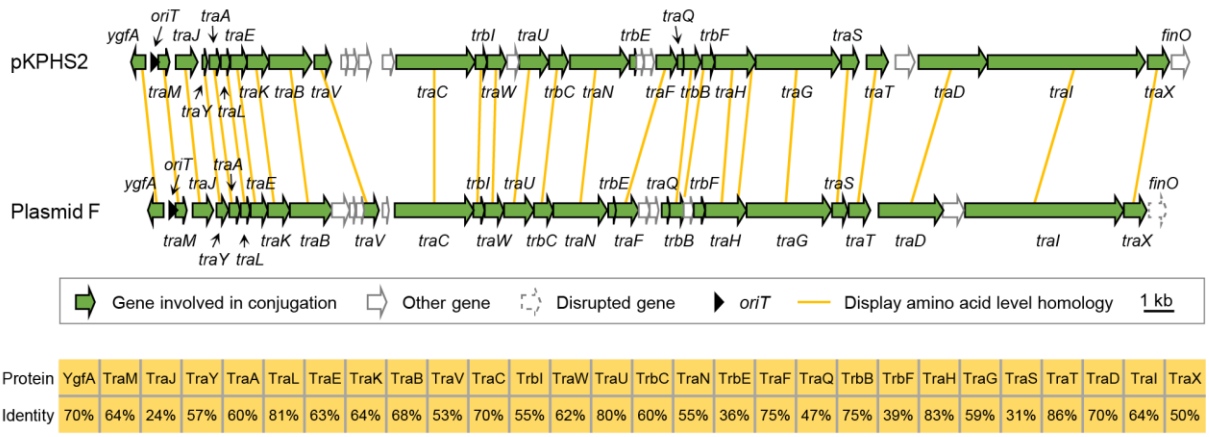


Figure S4. Plasmids of same IncFII$_K$ type are widespread in *K. pneumoniae* with different sequence types. Histogram is the count of occurrence of the IncFII$_K$ alleles in *Klebsiella pneumonia* genomes specified by sequence type. In the block is the zoom-in view of types IncFII$_{K3}$-IncFII$_{K33}$ and unclassified alleles.



Protein	YgfA	TraM	TraJ	TraY	TraA	TraL	TraE	TraK	TraB	TraV	TraC	TrbI	TraW	TraU	TrbC	TraN	TrbE	TraF	TraQ	TrbB	TrbF	TraH	TraG	TraS	TraT	TraD	TraI	TraX
Identity	70%	64%	24%	57%	60%	81%	63%	64%	68%	53%	70%	55%	62%	80%	60%	55%	36%	75%	47%	75%	39%	83%	59%	31%	86%	70%	64%	50%

Figure S5. Comparison of the conjugation modules between plasmid pKPHS2 and plasmid F. GenBank accession numbers are CP003224 for pKPHS2 and AP001918 for plasmid F. Gene organisation is drawn to scale. Homologous counterparts are connected by orange lines. Identities in amino acid level of the components between the two plasmids are shown below.

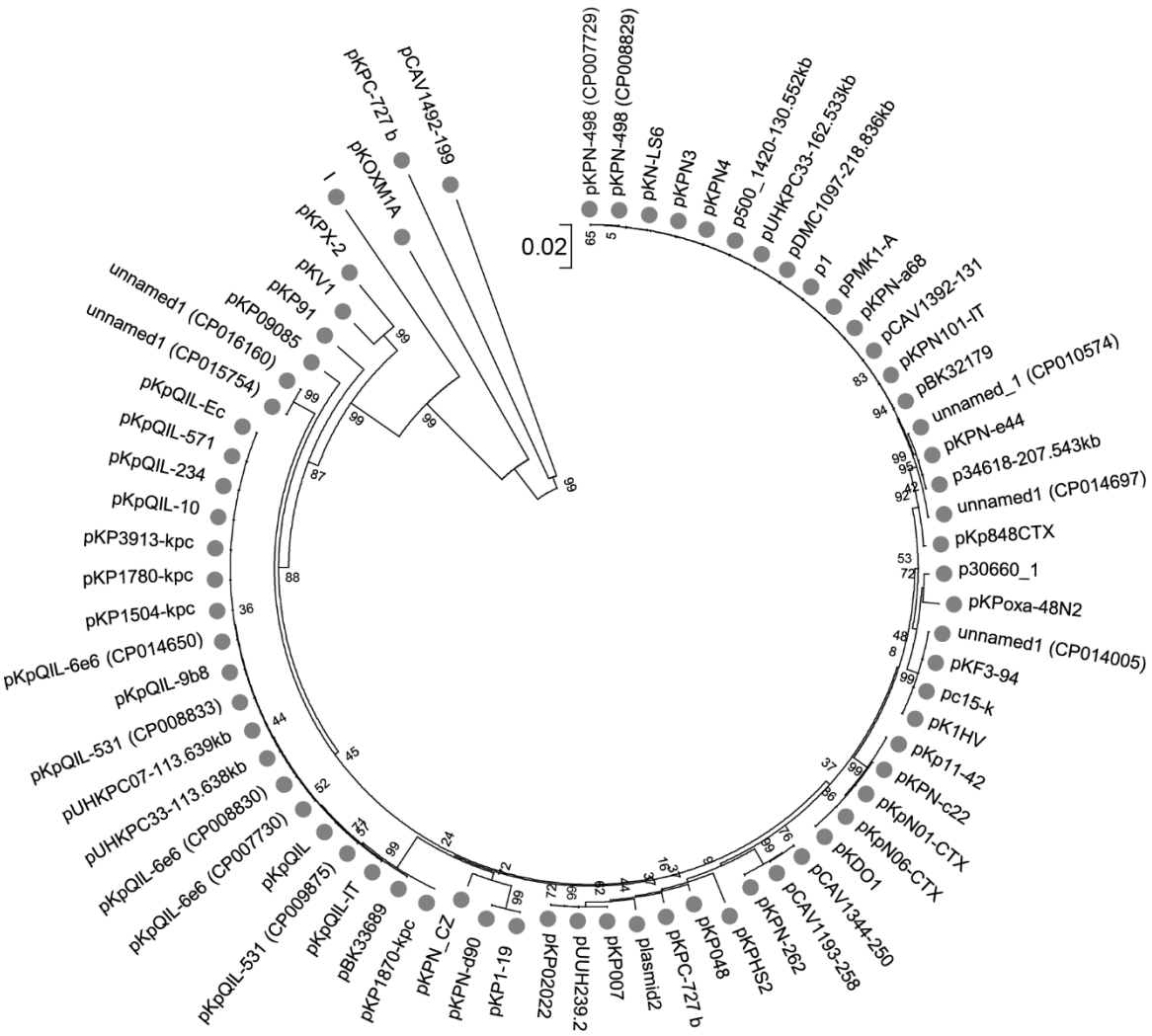


Figure S6. Phylogenetic tree of the DNA sequences of the transfer regions of the studied plasmids. Plasmids with same names are discriminated by their accession numbers in the brackets. The letters ‘a’ and ‘b’ are used to distinguish the two transfer regions on plasmid pKPC-727. Multiple sequence alignment was performed by Kalign v2.04. Then the tree was constructed using neighbor-joining method with 1,000 bootstrap replicates in MEGA5.

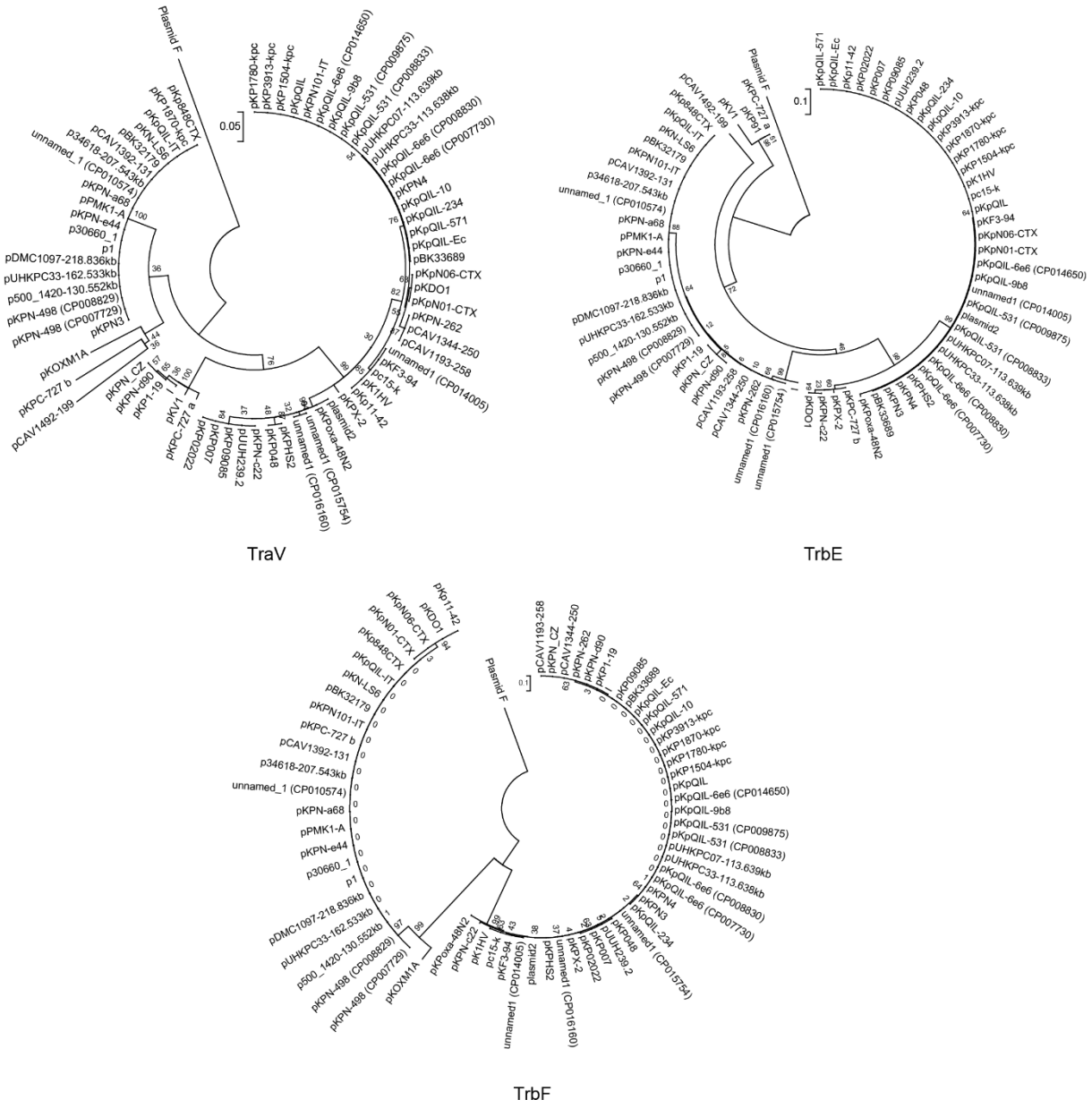


Figure S7. Phylogenetic trees of the proteins TraV and TrbE and TrbF of the studied plasmids. Plasmids with same names are discriminated by their accession numbers in the brackets. The letters ‘a’ and ‘b’ are used to distinguish the two conjugation modules on plasmid pKPC-727. Multiple sequence alignment was performed by MUSCLE v3.7. Then the tree was constructed using neighbor-joining method with 1,000 bootstrap replicates in MEGA5.