SUPPLEMENTAL MATERIAL

Next-generation sequencing based hospital outbreak investigation yields insight into *Klebsiella aerogenes* population structure and determinants of carbapenem resistance and pathogenicity

Adel Malek[1], Kelly McGlynn[1], Samantha Taffner[1], Lynn Fine[3], Brenda Tesini[4], Jun Wang[1], Heba Mostafa[1], Sharon Petry[1], Archibald Perkins[1], Paul Graman[4], Dwight Hardy[1,2], and Nicole Pecora[1,#]

Author affiliations:
[1]Department of Pathology and Laboratory Medicine, University of Rochester Medical Center, New York, USA
[2]Department of Microbiology and Immunology, University of Rochester Medical Center, New York, USA
[3]Infection Prevention Program, University of Rochester Medical Center, New York
[4]Department of Medicine, Infectious Diseases Division, University of Rochester Medical Center, New York, USA

[#]**Correspondence to:**
Nicole Pecora, MD, PhD, Email: nicole_pecora@urmc.rochester.edu, Phone: (585) 276-4674
University of Rochester Medical Center, 601 Elmwood Avenue,
Rochester, NY-14642, USA

INDEX:

I. SUPP. METHODS

Sequencing library preparation and raw data acquisition

Single colony genomic DNA extractions were performed using MagNA Pure Compact instrument (Roche Diagnostics, Indianapolis, IN); DNA quality was analyzed using QuantiFluor dsDNA system (Promega). Nextera XT kit (Illumina, San Diego, CA) was used for preparing dual-indexed WGS libraries. For QC dsDNA quality of individual samples was analyzed using the 4200 TapeStation system (Agilent). Following library cleanup, individual libraries were pooled at equimolar ratios and denatured; DNA concentration was determined using Qubit ssDNA kit (Thermo fisher Scientific). The libraries were combined with 20ρM PhiX control and sequenced on Illumina Miseq^{TM} using V3 kit and 2X300 bp paired-end protocol.

Genomic analyses

A) Quality control: For each sample, the read quality scores across all bases were assessed using FastQC v0.11.5 (1) and low quality reads were trimmed using Trimmomatic v0.36 (2). Genomic coverage of the sequencing reads relative to the reference *K. aerogenes* KCTC 2190 genome (ATCC 13048^{T}; Refseq accession number: NC_015663.1) were determined using Bowtie 2 v2.2.9 (3). Across the study isolates, alignments showed average genomic coverage of 90.84% (minimum 89%) with an average depth of coverage: 100X (excluding regions below 12X). Quality metrics are detailed in Dataset 2.

B) Core-genome SNP calling workflow: A read mapping approach was used to assess SNPs in the genomes of the study CR-KA strains relative to the reference genome *K. aerogenes* KCTC 2190 genome (ATCC 13048^{T}; Refseq accession number: NC_015663.1). Mapping, variant calling and phylogenetic analysis were performed by locally installed CFSAN SNP Pipeline v1.0.0 (4), an analysis workflow developed by the U.S Food and Drug Administration (FDA). CFSAN pipeline employs a 2-phase variant calling workflow and the 'optimized' version of the pipeline with criteria as applied by Saltykova et al (higher stringency in allele frequency thresholds and coverage) (5) was applied. In the first phase, variants were called based on *mpileup function* of SAMTools and *mpileup2snp tool* from VarScan (minimum average base quality=20, minimum read depth of coverage at site=12, minimum allele frequency=90%). Densely clustered SNPs that could arise due to recombination were excluded (>/=3 SNPs in a 50 bp window). High-confidence SNP variants meeting the criteria were composed to a list. In the 2nd phase, nucleotide sites at the listed positions were determined for all sites (minimum allele frequency threshold for SNP filtering=90%). SNPs located in mobile elements as annotated by RAST were excluded. The final nucleotide sites post filtering corresponding to the listed positions were used to build a SNP matrix. A multi FASTA file with concatenated SNP matrix entries were used for inferring phylogeny by FastTree (6), which were visualized using ITOL (7).

C) Sequence assembly, scaffolding, annotation and analyses: De novo genome assemblies (average N_{50}: 344943 across all isolates) were generated by SPAdes genome Assembler v3.11.1 (8), and the assembly quality was assessed by QUAST v4.5 (9). Draft genomes were annotated with RAST (10) and Prokka (11). MLST on the *K. aerogenes* genomes was performed using a newly developed publicly available scheme (12). Alleles and genetic markers corresponding to acquired antibiotic resistance, virulence and plasmid

replicons were identified using PlasmidFinder (13), ResFinder (14), and Virulence Factor Database (15). Typing of genetic loci associated with mobilizable yersiniabactin siderophore and genotoxin colibactin systems was performed using Kleborate (16), and the loci were visualized with MacVector software (MacVector, Cary, NC). Sequence variations in chromosomal genes associated with carbapenem resistance in the study strains were assessed relative to alleles in control strains (URMC 201, URMC 223) and 'wild-type' alleles in carbapenem susceptible (17) reference genome KCTC 2190 [Genbank ID-*ampD*: 10792472 (EAE_11350), *ampG*: 10792757 (EAE_12735), *ampR*: 10792632 (EAE_12115), *omp35*: 10793271 (EAE_15245), *omp36*: 10795060 (EAE_24205) and *ompR*: 10791222 (EAE_05245)]. Multiple sequence alignments of gene alleles and corresponding proteins was performed using Vector NTI software (Invitrogen, Carlsbad, CA) and the alignments were manually inspected to identify substitutions that would likely impact function.

D) Comparative genomics of publicly available global *K. aerogenes* assembled genomes with URMC CR-KA assemblies: Using a custom shell script, publicly available *K. aerogenes* genomes available as of July 2018 in the NCBI genome database were downloaded from the NCBI FTP site. Harvest genomics suite was used to perform intraspecific core-genome alignments as described before (18); phylogenies were visualized using ITOL (7).

II. SUPP. FIGURES

Fig. S1. Patient A timeline following CICU admit highlighting diagnostic workup, microbiological findings and antibiotic administration prior to development of CR-KA respiratory infection.

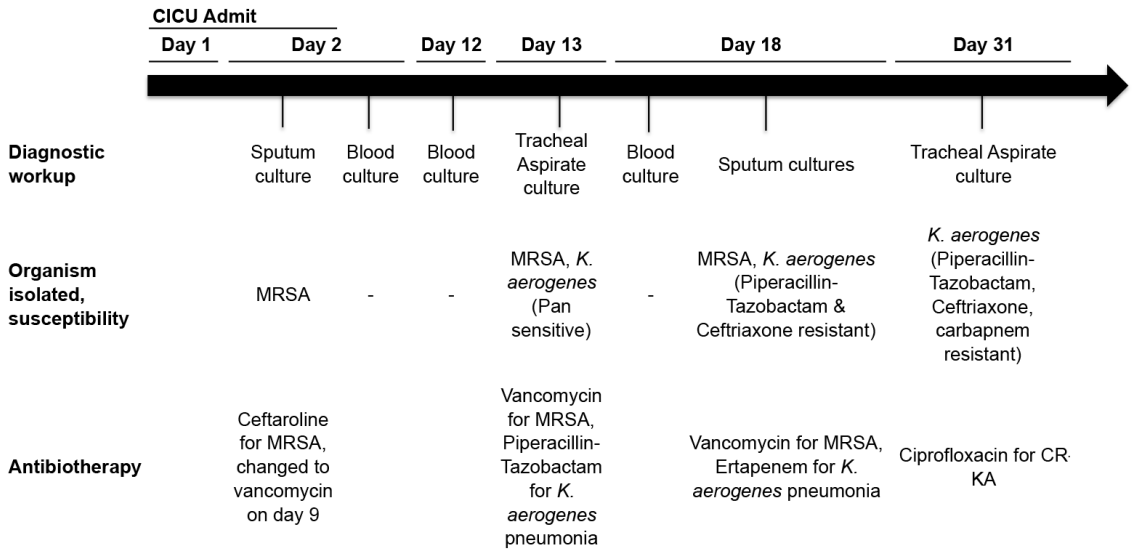
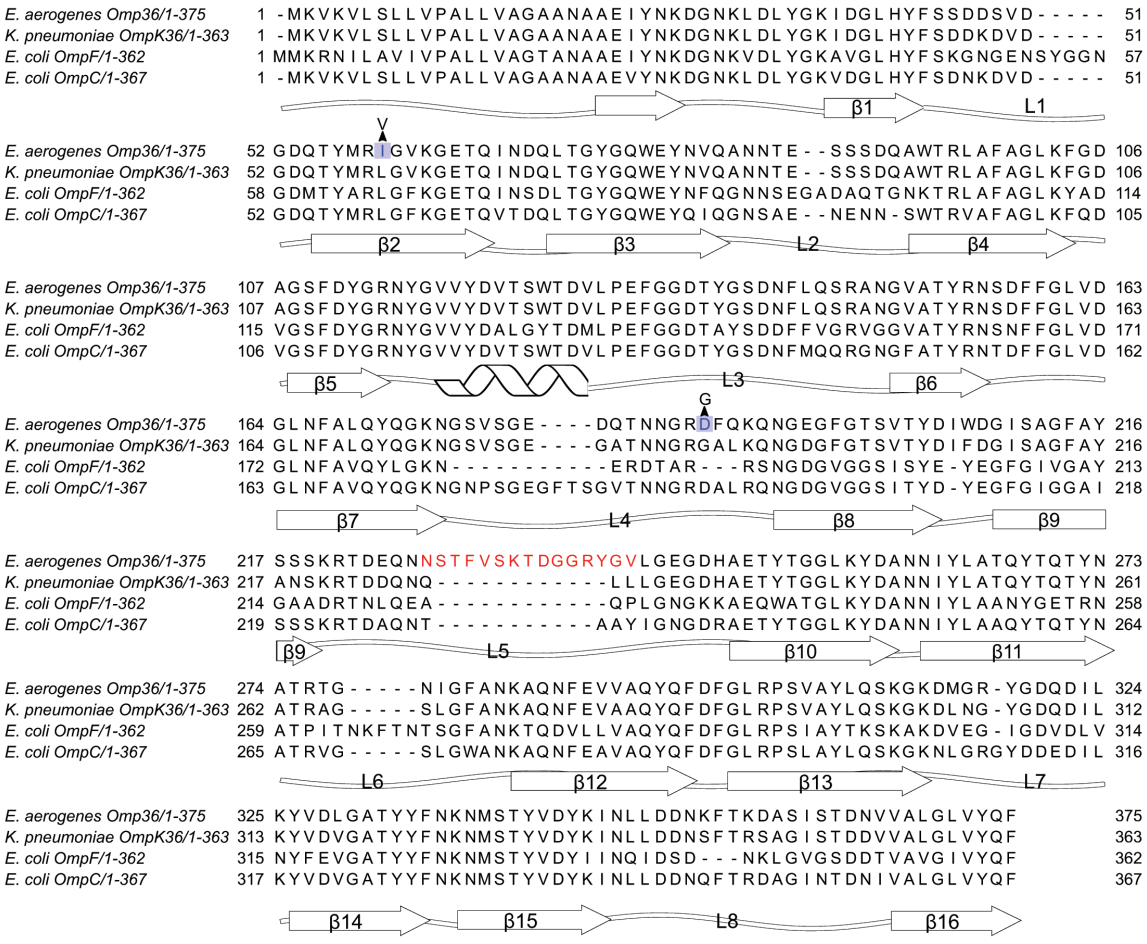


Fig. S2. Comparison of *K. aerogenes* Omp36 with representative porins of the OmpC and OmpF families. Predicted β-sheets, loops and helixes are shown. Non-synonymous SNP mutations in CICU outbreak CR-KA strains resulting in substitutions relative to reference highlighted in blue.



III. SUPP. TABLES

Table S1: Attributes of patients associated with the 2017 URMC CICU CR- KA cluster

Patient ID	Isolate ID	Underlying condition/disease and indications for admission	Age range	Culture site and Type
A	URMC 205*	Cardiogenic shock, endocarditis, sepsis, respiratory failure, history of ECMO	20-30	Tracheal aspirate
B	URMC 206	Cardiac tamponade, cardiogenic shock, history of ECMO, biventricular failure, LVAD present	70-80	Sputum
C	URMC 207	Ischemic mitral valve dysfunction, rapid atrial fibrillation, COPD, acute respiratory failure, mitral stenosis	70-80	Sputum
D	URMC 208	Heart failure, LVAD present, hemorrhagic stroke and recurrent infections	60-70	Sputum
E	URMC 209	Congestive heart failure, elevated mycocardial infraction, sarcoidosis	50-60	Sputum
F	URMC 211	History of fibromyalgia, cardiogenic shock, history of ECMO	40-50	Sputum
G	URMC 212	Cardiac tamponade, critical aaortic stenosis	60-70	Rectal Surveillance
H	URMC 213	Chronic systolic heart failure, ischemic cardiomyopathy	60-70	Rectal Surveillance
I	URMC 215	Ascending aortic aneurysm repair	60-70	Rectal Surveillance
J	URMC 216	Coronary artery disease, aortic valve insufficiency	70-80	Rectal Surveillance
K	URMC 218	Congestive heart failure, history of ECMO, acute heart transplant rejection	70-80	Rectal Surveillance
L	URMC 219	Paroxysmal atrial firbrillation, ventricular dysfunction	60-70	Rectal Surveillance
M	URMC 224	Endocarditis, end-stage renal disease	40-50	Blood
N	URMC 225	Chronix systolic heart failure, LVAD complication, cardiac tamponade	50-60	Rectal Surveillance
O	URMC 226	Coronary artery disease, end stage renal disease	70-80	Rectal Surveillance

*URMC 205, first patient diagnosed with clinical CR-KA infection in the URMC CICU in 2017

Abbr: ECMO: Extracorporeal membrane oxygenation, LVAD: left ventricular assist device present, COPD: chronic obstructive pulmonary disease

Table S2: Univariate analysis of risk-factors associated with infection/colonization by CR-KA in the CICU

Characteristic	CRE cases (n=15)	Controls (n= 30)	p-value
Mean age in years	63.6	55.1	NS
No. (%) of patients with the following specific risk factors:			
Male	40	61	NS
Unit location	73	55	NS
ECMO[a]	40	32	NS
Transfer from another facility	53	42	NS
Bronchoscopy	13	13	NS
Surgery History	100	61	p=.004
Carbapenem exposure	13	13	NS
Ventilator ≥5 days	60	58	NS
CRRT[b]/CVVH[c]	20	20	NS

[a]ECMO: Extracorporal membrane oxygenation
[b]CRRT: Continuous renal replacement therapy
[c]CVVH: Continuous veno-venous hemofiltration

Table S3: Summary of additional hospital *K. aerogenes* strains in this study (isolates not associated with CICU CR-KA cluster)

Isolate ID	Year of positive culture	Gender	Age range	Culture site and type	Patient's unit/ward
URMC 200	2015	M	20-30	Catheter urine	General Medicine
URMC 201	2015	M	30-40	Blood	Acute Care Medicine
URMC 202	2016	M	60-70	Broncholaveolar lavage	Burn/Trauma
URMC 203	2016	M	70-80	Biliary drain fluid	Solid Organ Transplant
URMC 204	2016	M	70-80	Catheter urine	Cardiovascular Progressive Care
URMC 210	2017	F	80-90	Urine	Outpatient
URMC 214	2017	F	90-100	Urine	Outpatient
URMC 217	2017	F	70-80	Urine, cytoscopy	Acute Medicine
URMC 221	2017	F	90-100	Urine	Outpatient
URMC 222	2017	F	40-50	Abdominal ascites fluid	Emergency Department
URMC 223*	2017	F	70-80	Urine	Cardiovascular Step Down

* All strains except URMC 223 and URMC 201 were deemed carbapenem resistant by CDC and CLSI definitions

URMC 223 was a pan-sensitive *K. aerogenes* isolated from the urine of a patient in the CICU in Oct 2017 (within the outbreak period)

Table S4: Specific SNP substitutions in URMC CICU CR-KA strains relative to patient A strain (URMC 205)

Region (RAST based annotation)	Gene length	Gene location	URMC 206	URMC 207	URMC 208	URMC 209	URMC 211	URMC 212	URMC 213	URMC 215	URMC 216	URMC 218	URMC 219	URMC 224	URMC 225	URMC 226
ABC transporter involved in cytochrome c biogenesis ATPase component CcmA	623	227	-	-	-	-	-	-	G>T	-	-	-	-	-	-	-
FIG00731541: hypothetical protein	1358	188	-	-	-	-	-	-	-	-	-	-	-	-	-	C>A
Glutamate-ammonia-ligase adenylyltransferase	2834	1750	-	-	-	-	-	-	G>T	-	-	-	-	G>T	G>T	G>T
Glutamate-ammonia-ligase adenylyltransferase	2834	1751	-	-	-	-	-	-	G>T	-	-	-	-	G>T	G>T	G>T
Phosphoenolpyruvate-dihydroxyacetone phosphotransferase regulatory protein DhaR	1931	1795	T>A	-	T>A	T>A	T>A	T>A	T>A	T>A	T>A	T>A	T>A	T>A	T>A	T>A
Non-coding region	None	None	-	-	-	-	-	-	-	-	-	-	-	-	-	T>A
Non-coding region	None	None	A>T	-	-	-	-	-	-	-	-	-	-	-	-	-
N-acetylmuramoyl-L-alanine amidase AmpD	563	283	G>T	-	G>T	-	-	-	-	-	G>T	-	-	-	-	-
N-acetylmuramoyl-L-alanine amidase AmpD	563	481	A>G	-	A>G	-	-	-	-	-	A>G	-	-	-	-	-
Sugar/maltose fermentation stimulation protein homolog	716	105	-	-	-	-	-	-	-	-	-	-	-	-	-	C>T
Xanthine-guanine phosphoribosyltransferase	458	284	-	-	-	-	-	-	-	C>A	-	-	C>A	-	-	-
Permeases of the major facilitator superfamily	1193	58	-	C>T	-	-	-	-	-	-	-	-	-	-	-	-
Transcriptional regulator TetR family	566	461	-	-	-	C>T	C>T	-	-	C>T	-	C>T	C>T	-	-	-
Transport ATP-binding protein CydD	1766	272	-	-	-	-	C>G	-	-	-	-	-	-	-	-	-
Nitrate/nitrite transporter	1232	156	-	-	-	-	-	-	A>G	-	-	-	-	-	-	-
Tyrosyl-tRNA synthetase	1274	394	-	-	-	G>T	-	-	-	-	-	-	-	-	-	-
Malonate decarboxylase beta subunit	833	427	-	-	-	-	T>G	-	-	-	-	-	-	-	-	-
Candidate type III effector Hop protein	1235	1065	-	-	-	C>T	C>T	-	-	C>T	-	C>T	C>T	-	-	-
Dipeptide transport system permease protein DppC	851	456	-	-	-	G>T	-	-	-	-	-	-	-	-	-	-
hypothetical protein	791	618	-	-	-	-	-	T>C	-	-	-	-	-	-	-	-
FIG00644701: hypothetical protein	2036	1437	-	-	-	-	-	C>G	-	-	-	-	-	-	-	-
Membrane-bound lytic murein transglycosylase E	611	457	-	-	-	-	-	-	-	-	A>C	-	-	-	-	-
Outer membrane protein C precursor	305	122	-	-	-	T>A	T>A	-	-	T>A	-	T>A	T>A	-	-	-
Non-coding region	None	None	-	-	-	-	-	-	-	-	-	-	-	C>T	-	-
SNPs vs URMC 205			4	1	3	6	6	3	5	5	4	4	5	4	3	6

Table S5: Acquired antibiotic resistance markers, plasmid replicon types and plasmid multilocus sequence typing of the study *K. aerogenes* strains

Study Isolates		Antimicrobial resistance marker	Replicon types	pMLST summary
2017 URMC CICU associated CR-KA strains	URMC 205		IncFIB(pENTAS01),IncFII(Y),ColRNAI	IncF[Y2*:A-:B-]
	URMC 206		IncFIB(pENTAS01),IncFII(Y),ColRNAI	IncF[Y2*:A-:B-]
	URMC 207		IncFIB(pENTAS01),IncFII(Y),ColRNAI	IncF[Y2*:A-:B-]
	URMC 208		IncFIB(pENTAS01),IncFII(Y),ColRNAI	IncF[Y2*:A-:B-]
	URMC 209		IncFIB(pENTAS01),IncFII(Y),ColRNAI	IncF[Y2*:A-:B-]
	URMC 211		IncFIB(pENTAS01),IncFII(Y),ColRNAI	IncF[Y2*:A-:B-]
	URMC 212		IncFIB(pENTAS01),IncFII(Y),ColRNAI	IncF[Y2*:A-:B-]
	URMC 213		IncFIB(pENTAS01),IncFII(Y),ColRNAI	IncF[Y2*:A-:B-]
	URMC 215		IncFIB(pENTAS01),IncFII(Y),ColRNAI	IncF[Y2*:A-:B-]
	URMC 216		IncFIB(pENTAS01),IncFII(Y),ColRNAI	IncF[Y2*:A-:B-]
	URMC 218		IncFIB(pENTAS01),IncFII(Y),ColRNAI	IncF[Y2*:A-:B-]
	URMC 219		IncFIB(pENTAS01),IncFII(Y),ColRNAI	IncF[Y2*:A-:B-]
	URMC 224		IncFIB(pENTAS01),IncFII(Y),ColRNAI	IncF[Y2*:A-:B-]
	URMC 225		IncFIB(pENTAS01),IncFII(Y),ColRNAI	IncF[Y2*:A-:B-]
	URMC 226		IncFIB(pENTAS01),IncFII(Y),ColRNAI	IncF[Y2*:A-:B-]
Other URMC KA strains	URMC 200		ColRNAI	NA
	URMC 201		No replicons found	NA
	URMC 202		ColRNAI	NA
	URMC 203	bla_{NMC-A}	IncFIB(pENTAS01),ColRNAI	IncF(unknown ST)
	URMC 204		No replicons found	NA
	URMC 210		IncFIB(pKPHS1),IncFIB(K),IncFII(Yp)	IncF[S3*:A-:B-]
	URMC 214		ColRNAI	NA
	URMC 217		IncFIB(pENTAS01),ColRNAI	IncF(unknown ST)
	URMC 221		No replicons found	NA
	URMC 222		IncFII(SARC14),IncFII(29),IncFIB(K)	IncF[S3*:A-:B-]
	URMC 223		IncFIB(K),IncHI1B	IncHI1(unknown ST); IncF(unknown ST)

Table S6:

Typing yersiniabactin and colibactin encoding gene clusters in URMC *K. aerogenes* strains using Kleborate

Study Isolates		Yersiniabactin	YbST	Colibactin	CbST
2017 URMC CICU associated CR-KA strains	URMC 205	ybt 17; ICEKp10	289-1LV	clb 3	17-3LV
	URMC 206	ybt 17; ICEKp10	289-1LV	clb 3	17-3LV
	URMC 207	ybt 17; ICEKp10	289-1LV	clb 3	17-3LV
	URMC 208	ybt 17; ICEKp10	289-1LV	clb 3	17-3LV
	URMC 209	ybt 17; ICEKp10	289-1LV	clb 3	17-3LV
	URMC 211	ybt 17; ICEKp10	289-1LV	clb 3	17-3LV
	URMC 212	ybt 17; ICEKp10	289-1LV	clb 3	17-3LV
	URMC 213	ybt 17; ICEKp10	289-1LV	clb 3	17-3LV
	URMC 215	ybt 17; ICEKp10	289-1LV	clb 3	17-3LV
	URMC 216	ybt 17; ICEKp10	289-1LV	clb 3	17-3LV
	URMC 218	ybt 17; ICEKp10	289-1LV	clb 3	17-3LV
	URMC 219	ybt 17; ICEKp10	289-1LV	clb 3	17-3LV
	URMC 224	ybt 17; ICEKp10	289-1LV	clb 3	17-3LV
	URMC 225	ybt 17; ICEKp10	289-1LV	clb 3	17-3LV
	URMC 226	ybt 17; ICEKp10	289-1LV	clb 3	17-3LV
Other URMC KA strains	URMC 200	ybt 0; ICEKp10	220-2LV	clb 3	13-2LV
	URMC 201	ybt 17; ICEKp10	289-1LV	clb 3	17-1LV
	URMC 202	ybt 17; ICEKp10	267	clb 3	19-1LV
	URMC 203	-	0	-	0
	URMC 204	ybt 16; ICEKp12	277-3LV	clb 3	17-1LV
	URMC 210	-	0	-	0
	URMC 214	-	0	-	0
	URMC 217	-	0	-	0
	URMC 221	-	0	-	0
	URMC 222	-	0	-	0
	URMC 223	ybt 4; plasmid	221-3LV	-	0

Abbr: YbST: Yersiniabactin locus sequence type, CbST: Colibactin locus sequence type

IV. REFERENCES

1. **Andrews S.** Babraham Institute Bioinformatics. FastQC. 2010, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

2. **Bolger AM, Lohse M, Usadel B.** 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30:**2114-2120.

3. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9:**357-359.

4. **Davis S, Pettengill JB, Luo Y, Payne J, Shpuntoff A, Rand H, Strain E.** 2015. CFSAN SNP Pipeline: an automated method for constructing SNP matrices from next-generation sequence data. PeerJ Computer Science **1:**e20.

5. **Saltykova A, Wuyts V, Mattheus W, Bertrand S, Roosens NHC, Marchal K, De Keersmaecker SCJ.** 2018. Comparison of SNP-based subtyping workflows for bacterial isolates using WGS data, applied to *Salmonella enterica* serotype Typhimurium and serotype 1,4,[5],12:i. PLOS One **13:**e0192504.

6. **Price MN, Dehal PS, Arkin AP.** 2010. FastTree 2-approximately maximum-likelihood trees for large alignments. PLOS One **5:**e9490.

7. **Letunic I, Bork P.** 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res **44:**W242-245.

8. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19:**455-477.

9. **Gurevich A, Saveliev V, Vyahhi N, Tesler G.** 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics **29:**1072-1075.

10. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O.** 2008. The RAST Server: rapid annotations using subsystems technology. BMC genomics **9:**75.

11. **Seemann T.** 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30:**2068-2069.

12. *Klebsiella aerogenes* MLST website, https://pubmlst.org/kaerogenes/, sited at the University of Oxford (**Jolley et al.** Wellcome Open Res 2018, 3:124 [version 1; referees: 2 approved]); MLST scheme developed by **Fengjun Sun** and colleagues at the Department of Pharmacy, Army Medical University, Chongqing, China. Database curated by **Qian Yuan**.

13. **Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, Moller Aarestrup F, Hasman H.** 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother **58:**3895-3903.

14. **Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV.** 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother **67:**2640-2644.

15. **Chen L, Zheng D, Liu B, Yang J, Jin Q.** 2016. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. Nucleic Acids Res **44:**D694-697.

16. **Lam MMC, Wick RR, Wyres KL, Gorrie CL, Judd LM, Jenney AWJ, Brisse S, Holt KE.** 2018. Genetic diversity, mobilisation and spread of the yersiniabactin-encoding mobile element ICE*Kp* in *Klebsiella pneumoniae* populations. Microb Genom.

17. **Lavigne JP, Sotto A, Nicolas-Chanoine MH, Bouziges N, Bourg G, Davin-Regli A, Pages JM.** 2012. Membrane permeability, a pivotal function involved in antibiotic resistance and virulence in *Enterobacter aerogenes* clinical isolates. Clin Microbiol Infect **18:**539-545.

18. **Treangen TJ, Ondov BD, Koren S, Phillippy AM.** 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol **15:**524.