Supplementary Appendix for

Multispecies outbreak of VIM-producing multidrug resistant bacteria driven by a promiscuous IncA/C2 plasmid

Table of contents:

Other Supplementary Materials for this manuscript include the following:

Data S1 to S3*

* Data files S1-S3 can be retrieved from **https://github.com/tomdeman-bio/Data_storage/tree/master/Multispecies_outbreak_of_VIM-producing_multidrug_resistant_bacteria**

Molecular Materials and Methods

Isolate preparation and whole genome sequencing

Nineteen isolates, including 14 isolates from 12 case-patients, two environmental isolates, and three historical isolates from the patient reported with VIM-producing CRE from the same hospital in 2013, underwent short-read Illumina sequencing. Specimens were subcultured onto TSA II with a 10 µg meropenem disk (BD, Sparks, Maryland) and incubated for 24 hours at 37°C. Genomic DNA (gDNA) was extracted using the Maxwell® 16 MDx instrument (Promega, Madison, WI) with the Maxwell® 16 Cell LEV DNA Purification Kit according to the manufacturer's protocol. High quality input gDNA, as indicated by an absorbance ratio of 1.8-2.0 (Nanodrop 2000, ThermoFisher Scientific, Waltham, MA), were sheared by ultrasonic fragmentation (M220, Covaris, Woburn, MA). Using a Zephyr® G3 NGS Workstation (Perkin Elmer, Waltham, MA), sequencing libraries were prepared using the Ovation Ultralow DR Multiplex System 1-96 kit (NuGen, San Carlos, CA) according to the manufacturer's protocol. Prepared libraries were pooled, normalized, and sequenced with the MiSeq platform (Illumina, San Diego, CA), yielding 250 base pair (bp) paired-end reads.
For PacBio sequencing, gDNA was extracted and purified using MagAttract HMW DNA (Qiagen, Germantown, MD) following the manufacturer's recommended protocol. Input gDNA was made into 10-kb libraries with the SMRTbell Template Prep Kit 1.0. Libraries were then bound to polymerase using the DNA/Polymerase Binding Kit P6v2 and sequence with C4v2 chemistry for 240 min movies on the RSII instrument (Pacific Biosciences, Menlo Park, CA). Raw sequencing reads and PacBio assemblies were placed in the NCBI Sequence Read Archive (SRA) under BioProject PRJNA302185.

Illumina genome assembly, strain classification, and gene annotation

PhiX174 sequences were removed from raw Illumina reads using BBDuk 35.82 [1]; PhiX-free reads were subsequently trimmed and filtered for high quality, and reads shorter than 50 bp were discarded from the dataset with Trimmomatic 0.35 [2]. Multilocus Sequence Typing (MLST) was performed *in silico* from the cleaned reads prior to assembly using SRST2 [3]. Clean reads were then assembled into contigs using SPAdes 3.9.0 [4] and six k-mer sizes (K21, K33, K55, K77, K99 and K127). In order to correct assembly errors, clean reads were mapped back to the assembled contigs with BWA 0.7.12 (using the "careful" option in SPAdes) [5].
Bacterial species identification was conducted for each isolate with Kraken 0.10.4 [6] and the RefSeq database of the National Center for Biotechnology Information (NCBI). Known antimicrobial resistance (AR) genes were detected using SSTAR [7] and the ResFinder database [8]. AR genes identified on contigs that shared $\geq$ 99% sequence similarity and 100% coverage compared to the ResFinder reference genes were reported. Additional open reading frames were detected and annotated using PGAP from NCBI, and signal peptides were predicted by means of SignalP 4.1 [9]. Confirmed and putative CRISPR regions were detected on the assembled contigs using CRISPRFinder [10].

PacBio genome assembly, strain classification, and gene annotation

PacBio reads for six isolates were initially assembled into unitigs using HGAP and subsequently polished with Quiver [11]. MLST strain types were predicted from the assembled unitigs using MLST [12]. Known resistance mechanisms were determined as described for Illumina contigs. Additionally, Prokka 1.8 [13] was used to detect coding regions on the IncA/C2 plasmid assemblies; regions smaller than 300 bp were excluded from downstream analysis. Additional sequences, non-IncA/C2 plasmids and the chromosome, were functionally annotated with PGAP from NCBI. A functional IncA/C2 plasmid core genome was predicted with Roary [14] using coding regions of $\geq$ 300 bp identified by Prokka. All six PacBio IncA/C2 plasmid assemblies were aligned using BRIG [15] to determine

relatedness and identify sub-clusters. Integron structures were identified using IntegronFinder [16] and annotated by means of INTEGRALL software [17].

IncA/C2 plasmid classification

All six PacBio IncA/C2 plasmid assemblies were screened against the PlasmidFinder database using BLASTN to confirm the presence of the IncA/C2 replicon. IncA/C2 plasmid sequences were further classified using a recently published pMLST scheme for IncA/C2 plasmids and BLASTN [18]. Lastly, an *rhs* gene sequence (GenBank accession KM670336) was aligned using BLASTN to each PacBio IncA/C2 plasmid assembly to discern between type 1 and type 2 IncA/C2 plasmids.

Phylogenetic analysis

Prior to phylogenetic analysis we constructed a core IncA/C2 plasmid genome sequence from all six PacBio IncA/C2 plasmid assemblies using progressiveMauve 2.4 [19] and a custom Perl script. For each PacBio IncA/C2 plasmid assembly, Locally Collinear Blocks (LCBs) of $\geq$ 500 bp shared among all six plasmid assemblies were extracted from the Mauve core alignment, merged into a single LCB block, and converted to a FASTA file. We randomly selected the merged core plasmid LCBs from Sample N3 as the mapping reference.

Clean Illumina reads from all IncA/C2 replicon-containing isolates were mapped to the merged core LCBs from Sample N3 using lyve-SET1.1.4f [20]. Consensus single nucleotide polymorphism (SNP) calls of 70% and at least 10x coverage were considered high quality and used for further analysis. Clustered SNPs located within 5 bp of each other were filtered such that only one SNP would be reported. Indels were disregarded from phylogenetic analyses. A phylogenetic tree was constructed from high quality SNPs with RAxML version 8.1.16 [21] using a Generalized Time Reversible (GTR) substitution model and a GAMMA model of rate heterogeneity.

Plasmid transmission events between patients or patient and environmental sources were estimated based on pairwise comparisons of SNPs. An *a priori* decision was made that zero SNPs between two plasmid sequences would indicate a potential direct transmission event.

Comparative genomics with unrelated ST1.1 IncA/C2 plasmids

Sixteen ST1.1 IncA/C2 plasmid assemblies were publicly available from GenBank at NCBI as of July 2017 and included for comparative genomic analysis. A core plasmid genome size estimation was performed for 22 ST1.1 IncA/C2 plasmid assemblies, including our 6 PacBio IncA/C2 plasmid assemblies, using progressiveMauve 2.4 [19]. For each PacBio IncA/C2 plasmid assembly, LCBs $\geq$ 500 bp shared among all 22 plasmid assemblies were extracted from the Mauve core alignment, merged into one LCB block, and converted to a FASTA file. Open reading frames (ORFs) were predicted on the merged core plasmid LCBs of representative sample N3 using Prokka 1.8 [13]. ORFs < 300 bp were excluded from downstream analysis.

IncA/C2 plasmid conjugation efficiency

Conjugation experiments were performed using two VIM plasmid-carrying donor isolates from the outbreak (colistin susceptible/imipenem resistant) and five recipient strains; *E. coli* J53 (sodium azide resistant) as well as *K. pneumoniae* (NCBI SRA: SAMN04014928), *E. cloacae, C. amalonaticus,* and *R. ornithinolytica* isolates which were all unrelated to this investigation (colistin resistant/imipenem susceptible). In brief, donor and recipient isolates were subcultured in fresh tryptic soy broth (TSB, Difco) following overnight incubation at 37°C. Mating experiments between VIM-donors and recipients were carried out by combining mid-log-phase cultures of the donor and the recipient, at a ratio of 1:10 in a volume of 1000 µl. *E.coli* J53 transconjugant dilutions were plated onto agar containing 1 µg/ml of imipenem and 200 µg/ml of sodium azide. Similarly, *K. pneumoniae, E. cloacae, C. amalonaticus,* and *R. ornithinolytica* transconjugants were plated on agar containing 1 µg/ml of imipenem and 1 µg/ml of colistin. To count the number of donor and recipient cells used in each mating experiment, serial dilutions were also plated on TSA containing either sodium azide (200

µg/ml) or colistin (1 µg/ml), respectively. Colony counts were used to calculate the conjugation efficiency, which was expressed as conjugants/recipient colony forming unit (CFU).

Statistical Analysis

Descriptive statistical analysis was conducted using Epi Info 7, version 7.0.8.0 (CDC).

Supplementary Results

IncA/C2 Plasmid Functional Core and Accessory Genomes

Gene detection identified approximately 160 open reading frames of at least 300 bp long in each IncA/C2 plasmid, totaling 181 unique genes. 142 genes were shared among all six plasmids and were therefore considered the functional core plasmid genome for this outbreak dataset. This core included, but was not limited to, beta-lactamase gene, bla_{VIM-1}, sulfonamide resistant dihydropteroate synthase gene, *sul1*, aminoglycoside resistant transferase, *aadA1*, genes encoding mercury and chromate resistance, and conjugal transfer genes **(Data S3)**.

The functional accessory IncA/C2 plasmid genomes for the 160 Kb and 164 Kb sub-groups consisted of 13 and 14 genes, respectively, and were dominated by transposable elements, antimicrobial resistance genes, and genes with unknown function.

One *K. pneumoniae* isolate contained an IncA/C2 plasmid, belonging to the 164 Kb plasmid sub-group, which carried a bla_{TEM-1B} presumed non-functional because it lacked the first 88 nucleotides compared to the ResFinder reference (Genbank: JF910132) due to a disruption from an IS6 element. This insertional inactivation event was also independently confirmed by assessing the same bla_{TEM-1B} variant identified on one of the contiguous Illumina sequences from the same isolate. No differences in MIC values were observed between the isolates carrying an active or inactive bla_{TEM-1B} gene, likely due to the presence of other beta-lactamases like VIM.

Comparative genomics with other ST1.1 IncA/C2 plasmids

All six IncA/C2 plasmids sequenced with PacBio technology belonged to IncA/C2 pMLST type ST1 and sub-group ST.1.1. ST1.1 is highly associated with carriage of bla_{NDM} and bla_{CMY} and has been isolated in many countries. However, five publicly available ST1.1 IncA/C2 plasmid assemblies were also isolated in the US and two of them lack bla_{NDM} and bla_{CMY}, similar to our outbreak plasmids. Moreover, the older members of this group KF976462.2 (Australia, 1997) and JX141473.1 (Thailand, 2007) also lack bla_{CMY} and bla_{NDM}, which might suggest a possible origin of these plasmids (Table S2). All six IncA/C2 plasmids from this outbreak were most closely related to the historical IncA/C2 plasmid pRMH760 from Australia isolated in 1997 (KF976462.2), and plasmid pKPC_CAV1344 isolated in the US in 2010 (CP011622), which contained a bla_{KPC} but lacked a bla_{VIM} (Figure S1). Our outbreak IncA/C2 plasmids and plasmids pRMH760 and pKPC_CAV1344 contained a conserved backbone, and differed predominately in the region harboring the resistance mechanisms (Figure S1).

The estimated core plasmid genome of all publicly available IncA/C2 ST1.1 plasmid assemblies (n=16) and our six PacBio VIM plasmids was 57,484 bp, representing 35-36% of our VIM IncA/C2 plasmids. This core plasmid genome contained 65 genes of $\geq$300 bp and the majority exhibited unknown function (n=49, 75%). This core genome also harbored genes involved in conjugative transfer (*traG*, *traH, traF*), plasmid stability and maintenance (*parA*, *parB*), and toxin-antitoxin system genes (*higA, higB)*. No known genes encoding virulence factors, holotoxins or antimicrobial resistance genes were detected within this core plasmid genome.

IncA/C2 conjugation efficiencies

We observed variation in conjugation efficiency for both IncA/C2 sub-groups during lab experiments when transferring the plasmids to *E. coli* J53, *K. pneumoniae* (NCBI's SRA isolate: SAMN04014928), *E. cloacae*, *C. amalonaticus,* and *R. ornithinolytica*. In *E. coli* and *K. pneumoniae*, higher conjugation efficiencies were observed for the smaller 160 Kb variant (*E. coli*: 1.3 x $10^{-3,}$ conjugants/recipient CFU, *Klebsiella*: 7.8 x 10^{-2}) compared to the larger 164 Kb variants (*E. coli*: 1.4 x 10^{-5}, *Klebsiella*: 7.1 x 10^{-3}). We were not able to determine conjugation efficiencies for the 160 Kb plasmid variant in *E. cloacae*, *C. amalonaticus,* and *R. ornithinolytica*, however we observed conjugation efficiencies for the 164 Kb variant in those three species of 2.1 x 10^{-5}, 1.3 x 10^{-2}, and 2.9 x 10^{-4}, respectively (Table S3).

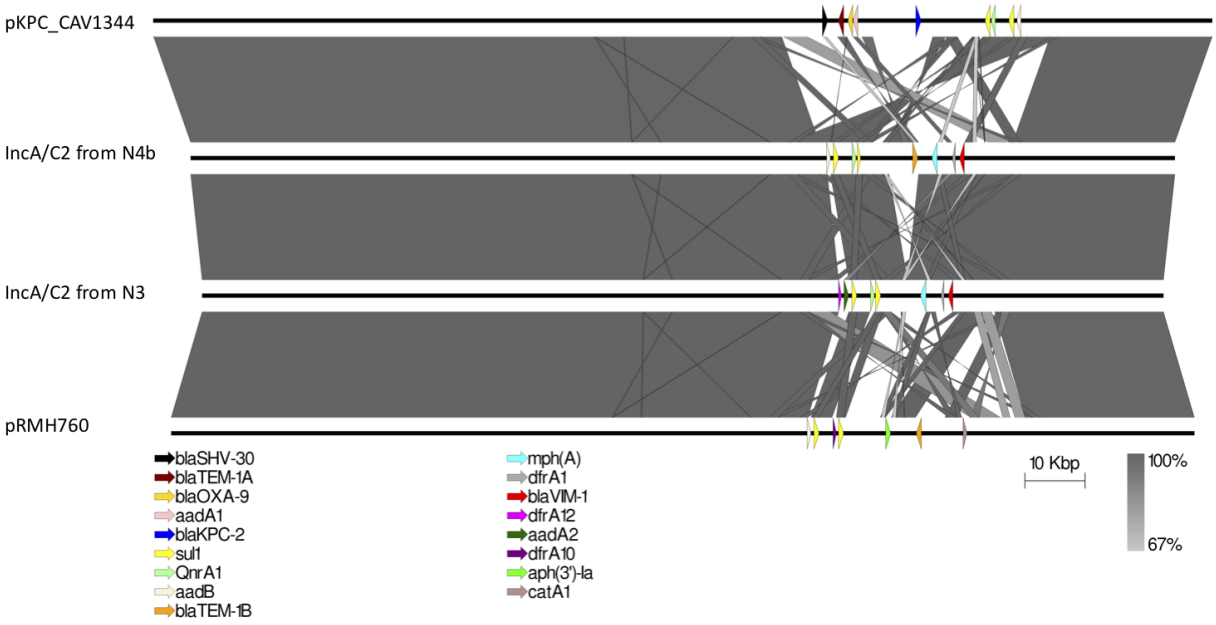

Figure S1.

The IncA/C2 plasmids isolated and characterized during this outbreak were similar to IncA/C2 plasmids previously isolated from the US and Australia. 1. IncA/C2 plasmid pKPC_CAV1344 isolated in the US. 2. 164 Kb IncA/C2 plasmid from isolate N4b. 3. 160 Kb IncA/C2 plasmid from isolate N3. 4. IncA/C2 plasmid pRMH760 isolated in Australia.

EPI ID	Species	MLST	A/C2 variant	Beta-lactamase genes
E1	*Citrobacter amalonaticus*	No MLST scheme	160 Kb	bla_{VIM-1}
A1	*Enterobacter hormaechei*	ST-110*	160 Kb	bla_{VIM-1}, bla_{TEM-1B}, bla_{SHV-12} bla_{ACT-15}
2013_1a	*Enterobacter hormaechei*	ST-1015	160 Kb	bla_{VIM-1}, bla_{TEM-1B}, bla_{SHV-12}, bla_{ACT-7}
2013_1b	*Enterobacter hormaechei*	ST-1015	160 Kb	bla_{VIM-1}, bla_{TEM-1B}, bla_{SHV-12}, bla_{ACT-7}
A7	*Enterobacter hormaechei*	ST-135*	164 Kb	bla_{VIM-1}, bla_{TEM-1B}, bla_{ACT-14}
N1	*Enterobacter hormaechei*	ST-254	164 Kb	bla_{VIM-1}, bla_{TEM-1B}, bla_{ACT-7}
N4a	*Enterobacter hormaechei*	ST-254	164 Kb	bla_{VIM-1}, bla_{TEM-1B}, bla_{ACT-7}
A3	*Enterobacter hormaechei*	ST-428	164 Kb	bla_{VIM-1}, bla_{TEM-1B}, bla_{SHV-12}
A4	*Enterobacter hormaechei*	ST-428	164 Kb	bla_{VIM-1}, bla_{TEM-1B}, bla_{SHV-12}
A5	*Enterobacter hormaechei*	ST-428	164 Kb	bla_{VIM-1}, bla_{TEM-1B}, bla_{SHV-12}
N5	*Enterobacter hormaechei*	ST-363*	160 Kb	bla_{VIM-1}, bla_{TEM-1B}, bla_{SHV-12}, bla_{ACT-14}
A6	*Enterobacter hormaechei*	ST-51	160 Kb	bla_{VIM-1}, bla_{TEM-1B}, bla_{SHV-12} bla_{ACT-7}
A2	*Enterobacter hormaechei*	ST-656*	160 Kb	bla_{VIM-1}, bla_{TEM-1B}, bla_{SHV-12} bla_{ACT-15}
N3	*Escherichia coli*	ST-6140	160 Kb	bla_{VIM-1}
2013_1c	*Klebsiella pneumoniae*	ST-1717*	160 Kb	bla_{VIM-1}, bla_{SHV-12}
N4b	*Klebsiella pneumoniae*	ST-422	164 Kb	bla_{VIM-1}, bla_{TEM-1B}, bla_{SHV-1}
N2b	*Klebsiella pneumoniae*	ST-643	164 Kb	bla_{VIM-1}, bla_{TEM-1B}**

N2a	*Raoultella ornithinolytica*	No MLST scheme	160 Kb	*bla*$_{VIM-1}$

Table S1.
Carbapenem-resistant Enterobacteriaceae (CRE) species and Multilocus sequence typing (MLST) genotypes identified during the outbreak. Associated IncA/C2 plasmid variant and beta-lactam resistance genes are listed for each isolate.
* Closely related to this MLST but not an exact match; ** Lacking a signal peptide sequence due to IS6 disruption

Genbank	IncA/C2 Plasmid	Origin	Beta-lactamase genes
KF976462	pRMH760	Australia	*bla*$_{TEM-1B}$
KU997026	pDGO100	Australia	*bla*$_{TEM-1B}$
KX458222	pB2-1	Australia	*bla*$_{TEM-1B}$
CP012902	pNDM15-1078	Canada	*bla*$_{NDM-1}$, *bla*$_{OXA-10}$
KU302801	pNDM1_SZ1	China	*bla*$_{NDM-1}$
KU302802	pNDM1_SZ2	China	*bla*$_{CTX-M-15}$, *bla*$_{NDM-1}$
KP742988	pNDM-SAL	India	*bla*$_{CMY-4}$, *bla*$_{NDM-1}$
KR827393	pKAZ4	India	*bla*$_{OXA-10}$, *bla*$_{SHV-12}$
LN831185	pNDM-116-17	India	*bla*$_{OXA-10}$, *bla*$_{NDM-1}$
AP012208	pNDM-1_Dok01	Japan	*bla*$_{CMY-4}$, *bla*$_{NDM-1}$, *bla*$_{TEM-1A}$
JX141473	pR148	Thailand	*bla*$_{OXA-10}$
CP011622	pKPC_CAV1344	USA	*bla*$_{KPC-2}$, *bla*$_{OXA-9}$, *bla*$_{TEM-1A}$, *bla*$_{SHV-30}$
CP013324	pCAV1193-166	USA	*bla*$_{OXA-9}$, *bla*$_{TEM-1A}$, *bla*$_{SHV-30}$
CP020049	AR_0118_unitig_1	USA	*bla*$_{NDM-1}$, *bla*$_{CMY-6}$
CP021551	AR_0159_tig00000137	USA	*bla*$_{NDM-1}$
CP021719	AR_0128_tig00000792	USA	*bla*$_{CMY-6}$

Table S2.
Publicly available ST.1.1 IncA/C2 plasmid sequences and associated beta-lactam resistance genes.

Donors	Recipients				
E. hormaechei **(IncA/C2** ***bla$_{VIM-1}$*** **plasmid)**	***E. coli*** **J53**	***K. pneumoniae*** **2013732991**	***E. cloacae*** **3000072915**	***Citrobacter amalonaticus*** **2008015970**	***Raoultella ornithinolytica*** **2014005534**
2013_1a (160 kb variant)	1.3×10^{-3}	7.8×10^{-2}	ND	ND	ND
N1 (164 kb variant)	1.4×10^{-5}	7.1×10^{-3}	2.1×10^{-5}	1.3×10^{-2}	2.9×10^{-4}

Table S3.
Conjugation efficiencies of IncA/C2 VIM plasmids from two *E. hormaechei* isolates to five different recipients. The efficiency of conjugation is the number of transconjugants per recipient cell.

Data S1.
Carbapenem-resistant Enterobacteriacea (CRE) species and Multilocus sequence typing (MLST) genotypes identified during the outbreak. Alternative isolate identifier used at the NCBI repository, associated IncA/C2 plasmid variant, associated beta-lactam resistance genes, and CRISPR mechanisms are listed for each isolate.

Data S2.
Six IncA/C2 plasmid-harboring isolates sequenced using long-read technology (Pacific Biosciences, Menlo Park, CA). Resistance genes and their cellular location are listed. Resistance genes unique to each IncA/C2 plasmid, which were used as marker genes, are highlighted.

Data S3.
IncA/C2 plasmid functional core genome for this outbreak dataset.

References:

1. Bushnell B. BBDuk: filters or trims reads for adapters and contaminants using *k*-mers. 2014;Available from: http://jgi.doe.gov/data-and-tools/bbtools/

2. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–20.

3. Inouye M, Dashnow H, Raven L-A, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Medicine 2014;6.

4. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 2012;19:455–77.

5. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–60.

6. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biology 2014;15:R46.

7. de Man TJB, Limbago BM. SSTAR, a Stand-Alone Easy-To-Use Antimicrobial Resistance Gene Predictor. mSphere 2016;1(1).

8. Zankari E, Hasman H, Cosentino S, et al. Identification of acquired antimicrobial resistance genes. Journal of antimicrobial Chemotherapy 2012;67:2640–4.

9. Petersen TN, Brunak S, Heijne von G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods 2011;8:785–6.

10. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Research 2007;35:W52–7.

11. Chin C-S, Alexander DH, Marks P, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nature Methods 2013;10:563–9.

12. Seemann T. Scan contig files against PubMLST typing schemes. Available from: https://github.com/tseemann/mlst

13. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014;30:2068–9.

14. Page AJ, Cummins CA, Hunt M, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 2015;31:3691–3.

15. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 2011;12:402.

16. Cury J, Jové T, Touchon M, Néron B, Rocha EP. Identification and analysis of integrons and cassette arrays in bacterial genomes. Nucleic Acids Research 2016;44:4539–50.

17. Moura A, Soares M, Pereira C, Leitao N, Henriques I, Correia A. INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. Bioinformatics 2009;25:1096–8.

18. Hancock SJ, Phan M-D, Peters KM, et al. Identification of IncA/C Plasmid Replication and Maintenance Genes and Development of a Plasmid Multilocus Sequence Typing Scheme. Antimicrobial Agents and Chemotherapy 2017;61.

19. Darling AE, Mau B, Perna NT. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. Plos one 2010;5:e11147.

20. Katz LS, Griswold T, Williams-Newkirk AJ, et al. A Comparative Analysis of the Lyve-SET Phylogenomics Pipeline for Genomic Epidemiology of Foodborne Pathogens. Frontiers in Microbiology 2017;8.

21. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30:1312–3.