

Supplementary Appendix

Genomic investigation of an outbreak of carbapenemase-producing *Enterobacter cloacae*: long-read sequencing reveals the context of *blaIMP4* on a widely distributed IncHI2 plasmid

Leah W. Roberts^{1,2,3}, Patrick N. A. Harris^{4,5}, Nouri L. Ben Zakour^{1,2,3}, Brian M. Forde^{1,2,3}, Elizabeth Catchpoole⁵, Mitchell Stanton-Cook^{1,2}, Minh-Duy Phan^{1,2}, Hanna Sidjabat⁴, Haakon Bergh⁵, Claire Heney⁵, Jayde Gawthorne^{1,2}, Jeffrey Lipman⁶, Anthony Allworth⁷, Kok-Gan Chan⁸, Teik Min Chong⁸, Wai-Fong Yin⁸, Mark A. Schembri^{1,2}, David L. Paterson^{2,4,9}, Scott A. Beatson^{1,2,3}

1. School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD, Australia; 2. Australian Infectious Diseases Research Centre, The University of Queensland, Brisbane, QLD, Australia; 3. Australian Centre for Ecogenomics, The University of Queensland, Brisbane, QLD, Australia; 4. The University of Queensland, UQ Centre for Clinical Research, Brisbane, QLD, Australia; 5. Pathology Queensland, Central Laboratory, Brisbane, QLD, Australia; 6. Burns Trauma and Critical Care Research Centre, The University of Queensland, Brisbane, QLD, Australia; 7. Infectious Disease Unit, Royal Brisbane & Women's Hospital, Brisbane, QLD, Australia; 8. Division of Genetics and Molecular Biology, Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia; 9. Wesley Medical Research, The Wesley Hospital, Auchenflower, QLD, Australia

Appendix contents:

Supplementary Methods	3
<i>Illumina Sequencing and Quality Control (QC)</i>	3
<i>Determining reference for phylogenetic analysis</i>	3
<i>Reassembly of Ecl1 from raw reads</i>	4
<i>Whole genome comparisons and phage analysis</i>	4
<i>Assembly and Annotation of SMRT sequenced isolate MS7884</i>	5
<i>Colony Polymerase chain reaction (PCR) to identify plasmids in MS7884</i>	6
Supplementary Results	8
<i>Table S3: Assembly metrics</i>	8
<i>Table S4: Hospital A and B isolates Resistance gene profile</i>	9
<i>Phenotypic observations may relate to single nucleotide polymorphisms (SNPs)</i>	10
<i>Patient 3 isolates are missing ~25 kb phage region</i>	10
<i>AmpC is unlikely to contribute to carbapenem resistance in the 2015 RBWH isolates</i>	11
<i>Loss of ~34 kb region in MS7889 is mediated by homologous recombination</i>	12
<i>Unique SNPs in aac(6')-Ib could result in increased resistance in MS7889</i>	12
<i>Whole genome sequencing identifies two plasmids carried in RBWH isolates</i>	13
<i>pMS7884A and pMS7884B comparisons to published genomes</i>	14
<i>Summary of methylomes in complete PacBio genomes</i>	14
Table S5: Summary of recognition motifs in MS7884A and MS7884B	14
Table S6: Summary of blastn against Rebase Gold Standard database	16
References	17
Supplementary Figures	20
▪ <i>Figure S1: Whole genome comparison of RBWH draft assemblies using BRIG</i>	
▪ <i>Figure S2: Multiple sequence alignment of Aac(6')-Ib amino acid sequence in MS7884, ARG-ANNOT reference and MS7889</i>	
▪ <i>Figure S3: Nucleotide alignment between MS7884 aac(6')-Ib genes and MS7889 aac(6')-Ib gene</i>	
▪ <i>Figure S4: BRIG comparison of 2015 RBWH draft assemblies to pECL_A from E. cloacae subsp. cloacae ATCC 13047</i>	
▪ <i>Figure S5: Easyfig comparison of pIMP4-SEMI and pMS7884A</i>	

Supplementary Methods

Illumina Sequencing and Quality Control (QC)

Illumina sequencing of the 2015 *bla*IMP+ *E. cloacae* isolates (n=16) and *E. coli* isolate (n=1) was done at the Australian Centre for Ecogenomics (ACE). *E. cloacae* Ecl1 had previously been sequenced on Illumina HiSeq2000 at the Australian Genome Research Facility (Melbourne, Australia). Libraries for isolates MS7884-MS7893 (RBWH isolates) were prepared using 2 x 150 bp Nextera XT V1 chemistry and run on an Illumina NextSeq. Isolates MS7923-MS7926 and MS8077-MS8079 (Hospitals A and B) were prepared using 2 x 300 bp Nextera XT libraries with V3 chemistry and run on an Illumina MiSeq. All libraries were quantified using Vii7 Thermo Fisher qPCR with QC checks with bioanalyzer analysis (Agilent). Contaminant searches on the raw reads were performed using Kraken (v0.10.5-beta). All raw reads were filtered using Nsoni (v0.130) [1] to remove Illumina adaptor sequences, reads shorter than 80 bp and bases below Phred quality 5. MiSeq reads were additionally hard trimmed to 150 bp due to low quality bases between 1-10 and 160-300 using Nsoni (v0.130).

Determining reference for phylogenetic analysis

Kraken (v0.10.5-beta) analysis using raw reads from the ten 2015 RBWH strains determined *E. cloacae* subsp. *cloacae* NCTC 9394 (GenBank: FP929040.1) to be the closest match in the complete genome division of GenBank. Subsequent comparison of the ten 2015 RBWH draft genome assemblies to assemblies within the publicly available whole-genome shotgun (WGS) database revealed a closer match to the IMP-4-producing *E. cloacae* strain Ecl1 from

2013 [2] (Accession: JRFQ01000001). Both were subsequently used to confirm single nucleotide variations (SNVs) between the 2015 RBWH strains.

Reassembly of Ecl1 from raw reads

In order to determine the relationship between *E. cloacae* Ecl1 and the 2015 RBWH *E. cloacae* strains, core SNP distances were determined using a mapping approach as implemented through Nasoni v0.130 [1] against the publicly available Ecl1 genome (Accession: JRFQ01000001). Manual inspection of the Nasoni output identified a number of clustered SNPs that we suspected were erroneous (Supplementary dataset S1). Reassembly of the raw reads using Spades v3.6.0 [3] under default parameters (without careful flag) was unable to correct the erroneous SNPs (Supplementary dataset S1). Closer examination of the read pileups for Ecl1 identified strand-specific clusters of SNPs, possibly due to technical problems at the time of sequencing. To identify genuine variants, the read pileup for every SNP as reported by Nasoni was manually curated using Artemis and Bamview. SNPs that were due to strand-specific base call errors in the assembly, resided close to contig edges, or resided in repetitive region (e.g. insertion sequences) were omitted from downstream analyses.

Whole genome comparisons and phage analysis

Whole genome comparisons were performed using BRIG [4]. Regions of difference were further investigated using the Artemis Comparison Tool (ACT) [5]. A ~25 kb region, missing in all patient 3 strains, was analysed using BLASTn to determine predicted gene content. As the region was predicted to contain phage-like proteins, Ecl1, MS7884 (RBWH patient 1), MS7886 (RBWH patient 2) and MS7890 (RBWH patient 3) draft assemblies were analysed using PHAST [6] to further characterise the phage-like region.

Assembly and Annotation of SMRT sequenced isolate MS7884

MS7884 was first sequenced on a PacBio RSII sequencer using the P6-C4 sequencing chemistry at the University of Malaya and assembled using the SMRT Analysis suite (version 2.3.0) to give a single chromosome (MS7884B: 4,810,853 bp) and one untypeable plasmid (pMS7884B: 126,208 bp). However, neither pMS7884B nor the chromosome contained any of the previously identified antibiotic resistance genes, including the carbapenemase *bla*_{IMP-4} (previously identified using Illumina and PCR). To ensure that the suspected multidrug resistant (MDR) plasmid was retained, we re-extracted DNA from MS7884 grown with 2 µg/ml meropenem (in lysogeny broth, LB). Sequencing was carried out using a different PacBio RSII Sequencer (P6-C4 sequencing chemistry) at the Doherty Institute, University of Melbourne. Assembly of this genome using the SMRT Analysis suite (version 2.3.0) gave an identical chromosome (4,810,853 bp) and a single IncHI2 plasmid (330,060 bp) carrying *bla*_{IMP-4} (hereafter referred to as pMS7884A). To avoid duplication in the complete genome databases we have compiled the MS7884A chromosome, pMS7884A plasmid and pMS7884B plasmid as a single strain genome, consistent with our original observation that both plasmids were present in the original patient isolate. Individual PacBio Raw data files are available to enable comparisons of the methylome between the two spontaneously cured MS7884 derivative isolates (see 'Summary of methylomes in complete PacBio genomes' below).

Table S1: Complete chromosome and plasmid list

Lab number	Parent strain	Chromosome/ plasmid	Size (bp)	Isolate name	Feature
MS7969	MS7884 (single colony)	Chromosome	4,810,853	MS7884B	
		Plasmid	126,208	pMS7884B	No IMP carbapenemase gene
MS8407	MS7884 (single colony grown with 2 µg/ml meropenem)	Chromosome	4,810,853	MS7884A	
		Plasmid	330,060	pMS7884A	Contains multiple antibiotic resistance genes, including IMP carbapenemase

All PacBio SMRT sequences were manually closed using the Artemis Comparison Tool (ACT) [5]. The chromosome and plasmids were polished using both the PacBio reads (with Quiver through the SMRT Analysis suite) and the Illumina reads (using BWA v0.7.5a-r405 with consensus calling through Pilon v1.21) to remove erroneous indels. The final genome was annotated using Prokka (v1.12-beta). The integron regions were manually annotated using a combination of RAC [7] and Integrall [8]. Insertion sequences (IS) were manually annotated using ISSaga [9]. Modified bases and associated motifs were detected using the SMRT analysis suite (version 2.3.0) to determine genome-wide methylation.

Colony Polymerase chain reaction (PCR) to identify plasmids in MS7884

MS7884 was grown on LB agar overnight at 37°C. Single colonies were resuspended in 50 µl sterile dH₂O, boiled for 15 minutes and centrifuged at 280x g for 2 minutes. 5 µl of supernatant containing colony DNA was used in 25 µl OneTaq PCR reactions, as per manufacturers recommendations (New England Biolabs Inc.). PCR primers were designed to identify both plasmids (Table S2). PCR was run as follows: Annealing 30 seconds at 55°C, extension 1 minute at 68°C, denaturing 30 seconds at 94°C, 29 cycles.

Table S2: Primer list

Plasmid Target	Primer Target	Product size	tm	Sequence
~330 kb IncHI2	MS7884_pA0187 BlaIMP-4	200	58	AGGACACACTCCAGATAACC
			67	TGATGCGTCTCCAGCTTAC
	MS7884_pA0259 RepHI2	800	58.88	TAATGGAGAGCGAGGGGTTTC
			58.62	GCGGTAAATCATGGACGGT
~126 kb untypeable	MS7884_pB0001 Putative methyltransferase	420	59.21	GAAATGTACCGCGTGCTGAA
			58.96	TCCTCAAGCATCTCGATCCC
	MS7884_pB0009 Replication initiation protein repE	590	58.42	AGAATAGCCC GCGAATTGTC
			60.04	CAGGAACCTACGGCGAAAGT

Supplementary Results

Table S3: Assembly metrics

Strain	Location	Depth ²	Contigs ³	Largest contig	N50 ⁴	Total Length
Ec11 (JRFQ01000001)	RBWH	30x	83	514 952	175 645	5 199 581
Ec11 (corrected) ¹	RBWH	36x	90	515 066	256 013	5 223 643
MS7884	RBWH	93x	84	556 131	259 415	5 225 529
MS7885	RBWH	126x	87	515 066	208 661	5 226 022
MS7886	RBWH	85x	85	556 628	259 182	5 228 488
MS7887	RBWH	104x	86	385 936	259 128	5 214 118
MS7888	RBWH	99x	86	406 431	259 182	5 227 406
MS7889	RBWH	80x	79	357 241	195 846	5 181 005
MS7890	RBWH	91x	84	406 431	208 661	5 195 590
MS7891	RBWH	124x	83	556 346	283 853	5 195 194
MS7892	RBWH	121x	89	388 759	259 518	5 194 114
MS7893	RBWH	140x	83	515 066	259 182	5 193 144
MS7923	Hospital A	109x	61	899 556	301 561	5 007 334
MS7924	Hospital A	76x	73	1 033 606	327 391	5 150 594
MS7925	Hospital A	91x	96	622 503	143 352	5 137 892
MS7926	Hospital B	98x	83	721 578	282 470	5 203 172
MS8077	Hospital B	98x	82	588 399	307 036	5 209 475
MS8078	Hospital B	71x	116	329 971	146 076	5 028 820
MS8079	Hospital B	103x	83	733 216	261 436	5 161 736

Table S4: Hospital A and B isolates Resistance gene profile

Strain	Species	Date	Location	Source	ST	Plasmid	Beta-lactam resistance					Rif	Aminoglycoside resistance				Quinolone resistance		Phe		MLS	Sul	Trim	Tet	Fos
							<i>ampC</i>	<i>bla_{SHV-12}</i>	<i>bla_{OXA-1}</i>	<i>bla_{IMP-4}</i>	<i>bla_{TEM-1B}</i>		<i>arr3</i>	<i>strB</i>	<i>strA</i>	<i>aac(6')Ib-cr</i>	<i>aac(3)-IIa</i>	<i>qnrA1</i>	<i>qnrB2</i>	<i>catA2</i>					
MS7925	<i>E. coli</i>	17/06/15	A	Urine	162	IncHI2	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	-
MS8078	<i>E. cloacae</i>	8/11/15	B	Swab Wound	831	IncHI2	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+
MS7923	<i>E. cloacae</i>	19/05/15	A	Rectal swab	656	IncHI2	+	-	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+
MS8077	<i>E. cloacae</i>	11/11/15	B	Urine	830	IncHI2	+	-	+	+	+	+	-	-	+	+	+	+	+	+	+	+	+	+	+
MS8079	<i>E. cloacae</i>	20/11/15	B	Urine	830	IncHI2	+	-	+	+	+	+	-	-	+	+	-	+	+	+	+	+	+	+	+
MS7926	<i>E. cloacae</i>	11/07/15	B	Urine	830	IncHI2	+	-	+	+	+	+	-	-	+	+	+	+	+	+	+	+	+	+	+
MS7924	<i>E. cloacae</i>	8/07/15	A	Rectal swab	830	IncHI2	+	-	-	+	+	+	-	-	+	-	-	+	+	+	+	+	+	+	+
MS7884 ¹	<i>E. cloacae</i>	15/06/15	RBWH	ETT	90	IncHI2	+	-	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	-

Abbreviations: Rif = rifampicin, Tet = Tetracycline, Trim = Trimethoprim, Sul = Sulphonamides, MLS = macrolide, Fos = Fosfomycin, Phe = Phenicol

¹MS7884 – IMP+ *E. cloacae* isolated from RBWH 2015

Phenotypic observations may relate to single nucleotide polymorphisms (SNPs)

Two isolates from Patient 3 exhibited *in vitro* phenotypic differences in adherence, suggesting that the respective SNPs may relate to within-host adaptation. Both isolates (MS7892 and MS7893) exhibited reduced pellet formation after centrifugation and resuspended more readily in solution compared to other cultures. This phenotype may have been advantageous for systemic infection, as both of these isolates were retrieved from wound and blood sites. Investigation into genotypic difference attributable to this phenotype found a non-synonymous SNP in *rcsB* (MS7884_3276) (Supplementary dataset S1), a well-known gene encoding a transcriptional regulatory protein that has previously been associated with colanic acid and biofilm production, and has been shown to provide protection from the bactericidal effect of serum during bloodstream infection [10, 11].

Patient 3 isolates are missing ~25 kb phage region

To improve our understanding of the transmission dynamics of this outbreak, we sought to identify differences in the pan-genome of these strains. Whole genome comparisons determined all 10 2015 RBWH isolates to have essentially identical genome content when compared to Ecl1. The exception was a ~25 kb region missing in all strains isolated from patient 3 (MS7890-MS7893), found to contain predicted phage-related proteins (Figure S1). Comparison of this region to the PHAST prophage database [6] found *Escherichia* phage HK639 (NC_016158.1) to be the closest match (35% coverage, 88% identity). This ~25 kb region represents only half of the full length of phage HK639 (~50 kb), and is present in Ecl1 and patient 1 and 2 isolates. It is unclear if the loss of this phage region could increase the fitness or virulence of the *E. cloacae* variant infecting patient 3. Although the prophage region is likely to have been lost during the infection of patient 3, we cannot rule out that

carbapenemase-producing *E. cloacae* with and without the phage co-exist within a common environmental source in the ICU.

AmpC is unlikely to contribute to carbapenem resistance in the 2015 RBWH isolates

Chromosomally encoded *ampC* was detected in all 10 2015 RBWH *E. cloacae* isolates (Table 1). While *ampC* is usually repressed and present at low levels in cells, SNPs resulting in derepression of *ampC* have been shown to increase resistance to beta-lactams [12, 13]. As such, we compared the nucleotide and amino acid sequence of our strains against several previously described SNPs in the regulatory genes *ampD*, *ampR* and *ampG* to predict *ampC* overexpression within our isolates.

All isolates contained 100% identical AmpC genes at both nucleotide and amino acid levels. The conserved AmpC residues Ser-64, Lys-67, Tyr-150, Asn-152, Lys-315 and Ala-318 [12] were all found to be present. Nucleotide and amino acid comparisons of AmpD between all 10 isolates were found to be 100% identical, with none of the previously described SNPs [14-16]. Amino acid comparison of AmpR to *E. cloacae* MNH1 [17] was 92% identical, with the highly conserved residues Arg-86, Gly-102, Ser-35, Tyr-264 and Asp-135 present in our strains. Comparison of AmpG to previous literature again found no obvious evidence to suggest heightened AmpC production [18-20]. We also checked the region upstream of *ampC* for the presence of insertion sequences (IS), as IS-driven overexpression of *ampC* has been previously reported [21-23]. Again, we found no evidence of IS upstream of *ampC*. We also found no evidence of plasmid-encoded *ampC*.

Additionally, observation of isolates grown with and without 2 ug/ml meropenem selection in LB found that colonies that lost the *bla*_{IMP-4} carbapenemase were unable to grow under the meropenem selective pressure, further suggesting that the *bla*_{IMP-4} carbapenemase is the major

driver of carbapenem resistance in these strains, and that *ampC* expression has little or no effect.

Loss of ~34 kb region in MS7889 is mediated by homologous recombination

Compared to all other *E. cloacae* from the 2015 RBWH isolates, MS7889 (patient 2, isolate 4) was found to have lost a number of antibiotic resistance genes, including *bla*_{IMP-4}. SMRT sequencing of the *E. cloacae* isolate MS7884 (patient 1, isolate 1) enabled full resolution of a 330,060 bp IncHI2 plasmid harbouring a ~55 kb MDR region that encompassed the majority of antibiotic resistance genes. Analysis of this region and comparison to the draft assembly for MS7889 determined that the most likely mechanism of loss of *bla*_{IMP-4} and other antibiotic resistance genes is homologous recombination between two near identical *aac(6')-Ib* genes, resulting in the loss of a ~34 kb region. Nucleotide comparison of the two *aac(6')-Ib* gene from MS7884 and the single *aac(6')-Ib* gene from MS7889 identified possible SNP evidence for homologous recombination, with the *aac(6')-Ib* gene from MS7889 appearing as a hybrid between the two *aac(6')-Ib* genes from MS7884 (Figure S3).

Unique SNPs in *aac(6')-Ib* could result in increased resistance in MS7889

The carbapenem-sensitive isolate MS7889 was found by Vitek2 to have increased resistance to the aminoglycosides tobramycin and amikacin, despite the loss of several other antibiotic resistance genes. Comparison of the aminoglycoside resistance gene *aac(6')-Ib* between the SMRT sequenced MS7884 reference and the MS7889 draft assembly identified 3 non-synonymous SNPs (Figure S2), which could contribute to the observed increase in resistance to tobramycin and amikacin.

Whole genome sequencing identifies two plasmids carried in RBWH isolates

Illumina whole genome sequencing of the 10 RBWH *E. cloacae* strains directly from clinical agar plates identified two large plasmids. Initial sequencing of isolate MS7884 (grown from a single colony) using PacBio SMRT sequencing was only able to identify one plasmid; a 126,208 bp untypeable plasmid (pMS7884B) that lacked all of the previously identified antibiotic resistance genes determined from the Illumina sequencing data, including the *bla*_{IMP-4} carbapenemase. Repeat PacBio SMRT sequencing of MS7884 (grown from a single colony with 2 µg/ml meropenem selection to avoid loss of *bla*_{IMP-4}) enabled the successful resolution of a large, 330,060 bp IncHI2 plasmid (pMS7884A) that was found to harbour all previously identified antibiotic resistance genes. However, pMS7884B was not present in this subsequent sequencing run. We therefore sought to determine whether these two plasmids were incompatible, and whether our clinical samples were comprised of a mixed population of carbapenemase-producing and non-producing *E. cloacae*.

Glycerol stock of the original MS7884 clinical sample was streaked onto Mueller-Hinton (MH) agar and grown overnight at 37°C. 100 colonies were pick and patched onto both plain MH agar and MH agar with 1 µg/ml meropenem and grown overnight at 37°C. 15 of 100 colonies did not grow on the meropenem plate, suggesting loss of pMS7884A or *bla*_{IMP-4}. Colony PCR determined 8 of the 15 colonies retained pMS7884A, but had lost *bla*_{IMP-4}. The remaining 7 colonies appeared to have lost pMS7884A entirely.

Colony PCR of the 100 single colonies was performed using primers designed to target pMS7884B. Of 100 colonies, only 1 appeared to have lost the pMS7884B. This confirmed that these two plasmids are not incompatible and that loss of one of the two plasmids in MS7884 is not uncommon and independent plasmid loss likely occurred prior to PacBio SMRT sequencing.

pMS7884A and pMS7884B comparisons to published genomes

The complete ~55kb MS7884 MDR region shares the most similarity to a region identified in pEI1573, an IncL/M plasmid carrying *bla*_{IMP-4} (isolated in Sydney 2012), with 99% nucleotide identity across 71% of the ~55 kb MDR region (Figure 4B) [24]. pMS7884A shares 86% of its IncHI2 backbone (with 99% nucleotide identity) to the previously reported IMP-producing pEC-IMPQ (GenBank: EU855788.1) plasmid isolated from China before 2009 [25]. The most closely related plasmid to pMS7884B in the NCBI database was the IncFII plasmid pECL_A (Genbank: CP001919.1) from *E. cloacae* subsp. *cloacae* ATCC 13047 [26], with 64% query coverage at 99% identity (Figure S4).

Summary of methylomes in complete PacBio genomes

Sequencing two single plasmid derivatives of the same isolate with PacBio SMRT sequencing enabled us to analyse the contribution of each plasmid to the genome-wide methylation status of every nucleotide (i.e. the methylome). Table S4 summarises the methylated motifs in MS7884A (with the ~330 kb IncHI2 plasmid) and MS7884B (with the ~126 kb untypeable plasmid), as determined using the SMRT Analysis Suite. Only one difference was determined between the two strains: an additional m6A methylated motif in MS7884B (ACCTRGCA).

Table S5: Summary of recognition motifs¹ in MS7884A and MS7884B

Motif string	Modification type	Methylated (%)	Number Detected	Number in Genome	Mean Score (Qmod)	Mean IPD Ratio
MS7884A						
GATC	m6A	98.2	50733	51688	123.40524	4.0679536
CCYAN ₉ TGAY	m6A	98.0	450	459	122.58444	4.87911
RTCAN ₉ TRGG	m6A	97.4	447	459	121.434006	4.6271815
CAGCNAC	m6A	97.0	4521	4662	112.2818	3.8938332
MS7884B						
GATC	m6A	99.1	49412	49856	238.59088	4.6236324
RTCAN ₉ TRGG	m6A	99.3	446	449	235.08296	5.354307
CCYAN ₉ TGAY	m6A	99.3	446	449	237.71748	5.4802275
CAGCNAC	m6A	98.1	4445	4530	218.76625	4.31024
ACCTRGCA	m6A	55.9	133	238	97.1203	1.9882706

¹ Probable false positives (modified bases and motifs with a Qmod < 80) are not shown

Searching this motif against REBASE PacBio motifs found 9 matches to previously identified PacBio motifs from *Salmonella* and *Enterobacter* species (CP015024, CP017087, CP017186, CP017187, CP017180, CP017179, CP017181, CP016012, CP016357, CP015923). The authors of many of these submissions have speculated that this motif could be a miscall of CCWGG and would be predicted to be methylated by an m5C methylase. As of yet an enzyme has not been determined for this motif.

To identify the causative methyltransferase (MTase) for this motif, sequences for the MS7884 chromosome, pMS7884A plasmid and pMS7884B plasmid were compared against the REBASE Gold Standard Database using BLASTn to determine MTase genes carried in the genome (Table S5). No previously identified MTase was found in pMS7884B when compared to the Rebase Gold Standard database suggesting the presence of a novel MTase on pMS7884B. Only one putative MTase gene was annotated on the plasmid: MS7884_pB0001 – a putative DNA adenine MTase (position 109-801 bp). Searches of the Conserved Domain Database at NCBI revealed that the encoded protein matched the Pfam DNA methylase domain (pfam01555; residues 21-203, E-value = 1.21e-41) and a more specific CDD domain with a putative methylase function (PRK13699; 1-219, E-value = 4.36e-44). As this is the only annotated MTase encoded on pMS7884B it is possible that MS7884_pB0001 encodes a novel MTase that accounts for the additional methylation detected at ACCTRGCA. However, we note that methylation was relatively low frequency (55.9% of sites), the mean score and IPD ratios are also low and the motif is not typical of adenine DNA MTases of any type, supporting the contention that MS7884_pB0001 may encode a m5C methylase that actually methylates CCWGG. Further work is required to elucidate the MTase properties (if any) of MS7884_pB0001.

Table S6: Summary of methyltransferases

Strain	Hit ¹	Percentage identity (nt)	Query coverage	Locus Tag	Location
MS7884A (chromosome)²					
	M.EclNIH2 Dam GATC 813nt	97%	789/813	MS7884_2116	2249369..2250181
	M.Csa8155I GAANNNNNNntAAA 1563nt	90%	672/744	MS7884_1344	1405871..1406629
pMS7884A (IncHI2)					
	M.Sen 2050ORF235P GATC 813nt	99%	806/813	MS7884_pA0268	227341..228153
pMS7884B (untypeable)					
	MS7884_pB0001 ³	-	-	MS7884_pB0001	109..801

¹ Hits > 50 bp coverage, \geq 90% identity with blastn against REBASE gold standard database (redundant hits excluded) downloaded 30/12/16

² MS7884A and MS7884B have identical gene content on the chromosome

³ Putative methyltransferase identified on pMS7884B untypeable plasmid

References

1. Victorian-Bioinformatics-Consortium. Nsoni. Available at: <https://github.com/Victorian-Bioinformatics-Consortium/nsoni>. Accessed June 2016.
2. Sidjabat HE, Heney C, George NM, Nimmo GR, Paterson DL. Interspecies transfer of blaIMP-4 in a patient with prolonged colonization by IMP-4-producing Enterobacteriaceae. *J Clin Microbiol* **2014**; 52(10): 3816-8.
3. Bankevich A, Nurk S, Antipov D, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* **2012**; 19(5): 455-77.
4. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **2011**; 12: 402.
5. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. ACT: the Artemis Comparison Tool. *Bioinformatics* **2005**; 21(16): 3422-3.
6. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. *Nucleic Acids Res* **2011**; 39(Web Server issue): W347-52.
7. Tsafnat G, Coptly J, Partridge SR. RAC: Repository of Antibiotic resistance Cassettes. Database **2011**; 2011(0): bar054-bar.
8. 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(8): 1096-8.
9. Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. ISSaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biology* **2011**; 12(3): R30.
10. Miajlovic H, Cooke NM, Moran GP, Rogers TRF, Smith SG. Response of Extraintestinal Pathogenic *Escherichia coli* to Human Serum Reveals a Protective

- Role for Rcs-Regulated Exopolysaccharide Colanic Acid. *Infection and Immunity* **2014**; 82(1): 298-305.
11. Lacour S, Bechet E, Cozzone AJ, Mijakovic I, Grangeasse C. Tyrosine Phosphorylation of the UDP-Glucose Dehydrogenase of *Escherichia coli* Is at the Crossroads of Colanic Acid Synthesis and Polymyxin Resistance. *PLOS ONE* **2008**; 3(8): e3053.
 12. Jacoby GA. AmpC beta-lactamases. *Clin Microbiol Rev* **2009**; 22(1): 161-82, Table of Contents.
 13. Babouee Flury B, Ellington MJ, Hopkins KL, et al. Association of Novel Nonsynonymous Single Nucleotide Polymorphisms in ampD with Cephalosporin Resistance and Phylogenetic Variations in ampC, ampR, ompF, and ompC in *Enterobacter cloacae* Isolates That Are Highly Resistant to Carbapenems. *Antimicrob Agents Chemother* **2016**; 60(4): 2383-90.
 14. Stapleton P, Shannon K, Phillips I. DNA sequence differences of ampD mutants of *Citrobacter freundii*. *Antimicrob Agents Chemother* **1995**; 39(11): 2494-8.
 15. Kopp U, Wiedemann B, Lindquist S, Normark S. Sequences of wild-type and mutant ampD genes of *Citrobacter freundii* and *Enterobacter cloacae*. *Antimicrob Agents Chemother* **1993**; 37(2): 224-8.
 16. Ehrhardt AF, Sanders CC, Romero JR, Leser JS. Sequencing and analysis of four new *Enterobacter* ampD Alleles. *Antimicrob Agents Chemother* **1996**; 40(8): 1953-6.
 17. Kuga A, Okamoto R, Inoue M. ampR gene mutations that greatly increase class C beta-lactamase activity in *Enterobacter cloacae*. *Antimicrob Agents Chemother* **2000**; 44(3): 561-7.
 18. Korfmann G, Sanders CC. ampG is essential for high-level expression of AmpC beta-lactamase in *Enterobacter cloacae*. *Antimicrob Agents Chemother* **1989**; 33(11): 1946-51.

19. Lindquist S, Weston-Hafer K, Schmidt H, et al. AmpG, a signal transducer in chromosomal beta-lactamase induction. *Mol Microbiol* **1993**; 9(4): 703-15.
20. Zhang Y, Bao Q, Gagnon LA, et al. ampG gene of *Pseudomonas aeruginosa* and its role in beta-lactamase expression. *Antimicrob Agents Chemother* **2010**; 54(11): 4772-9.
21. Bagge N, Ciofu O, Hentzer M, Campbell JI, Givskov M, Hoiby N. Constitutive high expression of chromosomal beta-lactamase in *Pseudomonas aeruginosa* caused by a new insertion sequence (IS1669) located in ampD. *Antimicrob Agents Chemother* **2002**; 46(11): 3406-11.
22. Lindberg F, Lindquist S, Normark S. Inactivation of the ampD gene causes semiconstitutive overproduction of the inducible *Citrobacter freundii* beta-lactamase. *J Bacteriol* **1987**; 169(5): 1923-8.
23. Hilty M, Sendi P, Seiffert SN, et al. Characterisation and clinical features of *Enterobacter cloacae* bloodstream infections occurring at a tertiary care university hospital in Switzerland: is cefepime adequate therapy? *Int J Antimicrob Agents* **2013**; 41(3): 236-49.
24. Partridge SR, Ginn AN, Paulsen IT, Iredell JR. pEI1573 Carrying blaIMP-4, from Sydney, Australia, is closely related to other IncL/M plasmids. *Antimicrob Agents Chemother* **2012**; 56(11): 6029-32.
25. Chen YT, Liao TL, Liu YM, Lauderdale TL, Yan JJ, Tsai SF. Mobilization of qnrB2 and ISCR1 in plasmids. *Antimicrob Agents Chemother* **2009**; 53(3): 1235-7.
26. Ren Y, Ren Y, Zhou Z, et al. Complete genome sequence of *Enterobacter cloacae* subsp. *cloacae* type strain ATCC 13047. *J Bacteriol* **2010**; 192(9): 2463-4.
27. Gupta SK, Padmanabhan BR, Diene SM, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother* **2014**; 58(1): 212-20.

Supplementary Figures

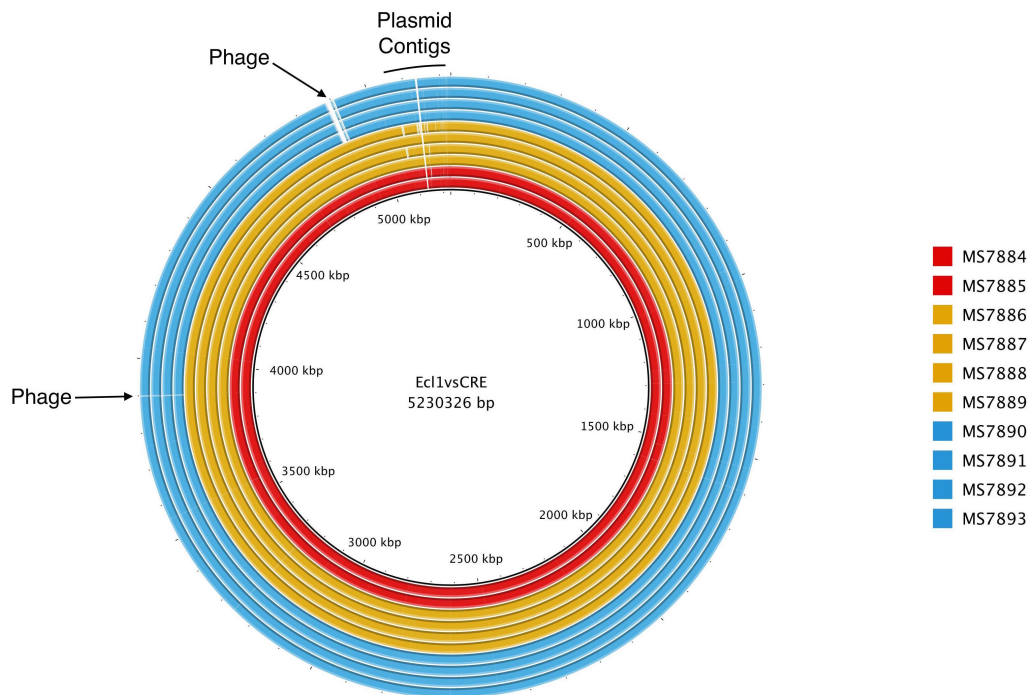


Figure S1: Whole genome comparison of RBWH draft assemblies using BRIG:

Draft assemblies of all RBWH isolates were compared against the Ecl1 draft reassembly (*E. cloacae*, 2013 RBWH) using the blast comparison tool BRIG [4]. Whole genome comparison identified two main differences between the patient 3 isolates (blue) and the patient 1 and 2 isolates (red and gold, respectively). All patient 3 isolates were missing a ~25 kb region identified as containing phage-related genes.

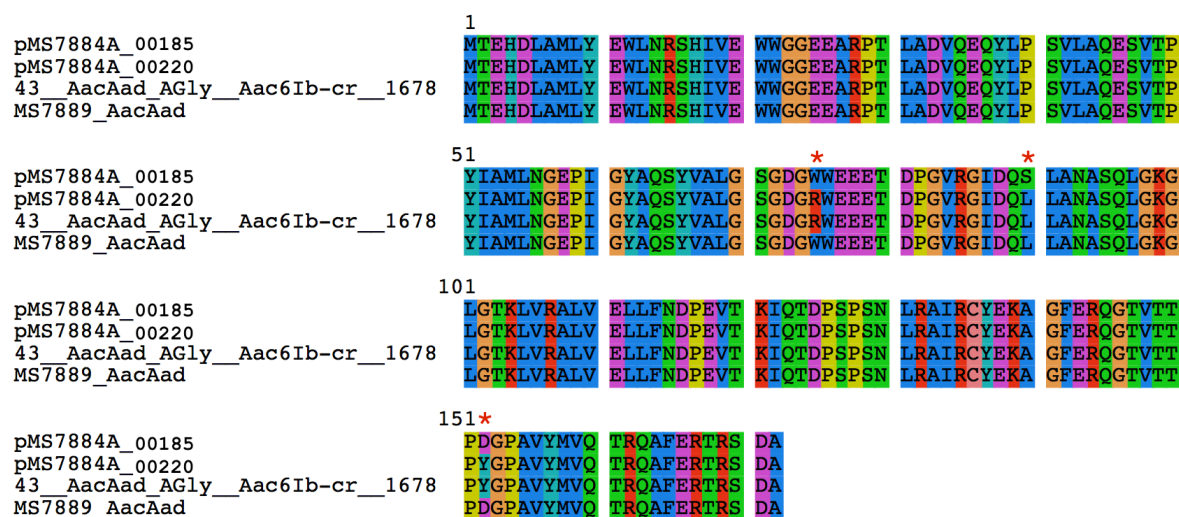


Figure S2: Multiple sequence alignment of Aac(6')-Ib amino acid sequence in MS7884, ARG-ANNOT reference and MS7889: Amino acid alignment of *aac(6')-Ib* genes in the complete plasmid pMS7884A (MS7884A_0185, MS7884A_0220), the reference *aac(6')-Ib* from ARG-ANNOT [27] (43__AacAad_AGly__Aac6Ib-cr__1678, Genbank: EF636461.1) and the single *aac(6')-Ib* from MS7889 (MS7889_AacAad) as viewed in Seaview (v4.5.4). MS7884A_0220 was identical at the amino acid level to the ARG-ANNOT reference for *aac(6')-Ib*, while MS7884A_0185 differed by three amino acids (W75R, S90L, D152Y, marked with red asterisks). MS7889_AacAad differed from the ARG-ANNOT reference by two amino acids (W75R, D152Y).



Figure S3: Nucleotide alignment between MS7884 *aac(6')*-Ib genes and MS7889

***aac(6')*-Ib gene:** The *aac(6')*-Ib gene in MS7889 was found to be a hybrid of both *aac(6')*-Ib genes in MS7884 as determined by SNP analysis (highlighted in red boxes). Positions 1-131 of the *aac(6')*-Ib gene in MS7889 were found to be more similar to one variant of *aac(6')*-Ib in MS7884 (pMS7884A_2). However, positions 311 onwards are more similar to the other MS7884 *aac(6')*-Ib variant (pMS7884A_1), suggesting a hybrid *aac(6')*-Ib formed during recombination between these two near identical genes.

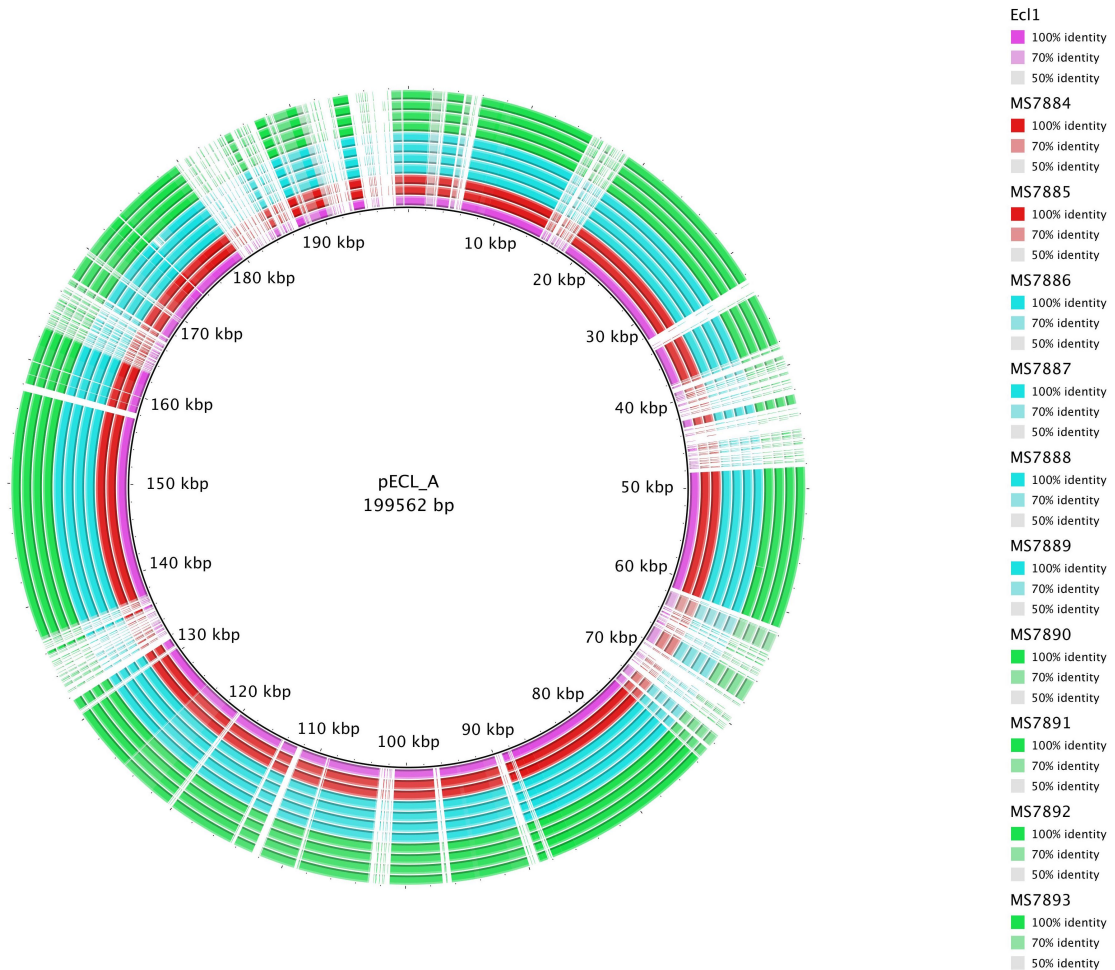


Figure S4: BRIG comparison of 2015 RBWH draft assemblies to pECL_A (Genbank: CP001919.1) from *E. cloacae* subsp. *cloacae* ATCC 13047: pMS7884B was found to be most closely related to pECL_A from *E. cloacae* subsp. *cloacae* ATCC 13047 (BLASTn; 64% query coverage at 99% identity). Draft assemblies for the 2013 strain Ecl1 (pink), 2015 Patient 1 strains (red), 2015 Patient 2 strains (light blue), 2015 Patient 3 strains (green) were compared against pECL_A using BRIG [4].

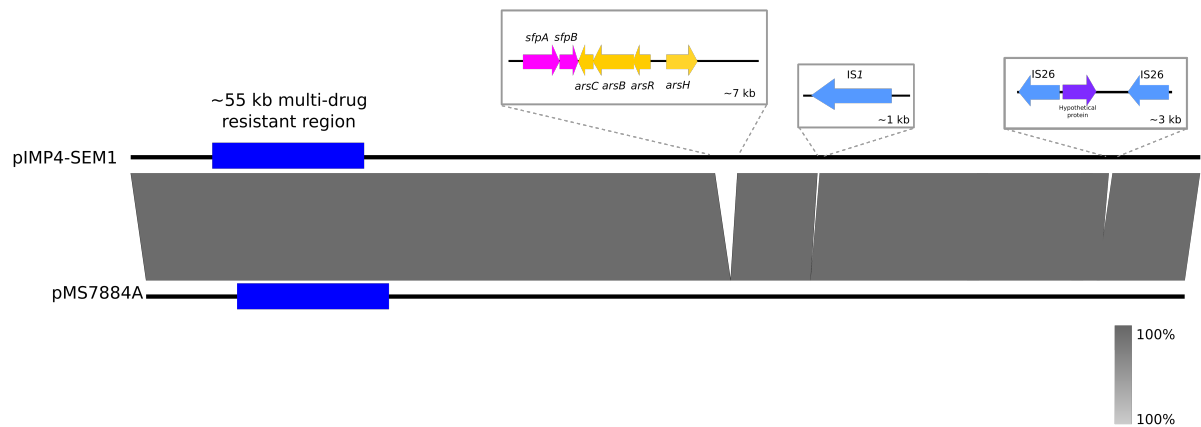


Figure S5: Easyfig comparison of pIMP4-SEM1 and pMS7884A: Black lines represent the complete plasmids pIMP4-SEM1 and pMS7884A. Grey blocks indicate nucleotide identity between the two plasmids. Both plasmids were found to be near identical, carrying the same large ~55 kb multi-drug resistant region. Three additional regions were found in pIMP4-SEM1: a ~7 kb region carrying *sfpAB* and *arsCBRH*, an *IS1* inserted upstream of *trhR*, and an *IS26*-bound transposon carrying a hypothetical protein.