

Supplementary Information

Title

Elucidation of quantitative structural diversity of remarkable rearrangement regions, shufflons, in IncI2 plasmids

Author

Tsuyoshi Sekizuka^{1*}, Michiko Kawanishi², Mamoru Ohnishi³, Ayaka Shima⁴, Kengo Kato¹, Akifumi Yamashita¹, Mari Matsui⁴, Satowa Suzuki⁴, Makoto Kuroda¹

Affiliation

¹Pathogen Genomics Center, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjyuku-ku, Tokyo 162-8640, Japan

²Assay Division II, Bacterial Assay Section, National Veterinary Assay Laboratory, Ministry of Agriculture, Forestry and Fisheries, 1-15-1 Tokura Kokubunji, Tokyo 185-8511 Japan

³Ohnishi Laboratory of Veterinary Microbiology, Japan

⁴Department of Bacteriology II, National Institute of Infectious Diseases, Musashimurayama, Tokyo 208-0011, Japan

* Corresponding author

Dr. Tsuyoshi Sekizuka

Laboratory of Bacterial Genomics, Pathogen Genomic Center,
National Institute of Infectious Diseases

1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan

Tel: +[81]-3-5285-1111 ext. 2525

Fax: +[81]-3-5285-1166

E-mail: sekizuka@nih.go.jp

Supplementary material

Table S1. Summary of PacBio RSII run and HGAP3 assembly information

Section	Protocol / Job Metric	Sample name			
		MRY16-002	MRY15-117	MRY15-131	
Methods	PacBio RSII run status	Average length of Input DNA library (bp)	31,247	34,876	27,218
		Number of SMRT Cells	1	1	1
		Movie length (min)	240	240	240
	HGAP3 Filtering paramter	Minimum Subread Length	500	500	500
		Minimum Polymerase Read Quality	0.8	0.8	0.8
		Minimum Polymerase Read Length	100	100	100
	HGAP3 assembly paramter	Minimum Seed Read Length	6,000	6,000	6,000
		Number of Seed Read Chunks	6	6	6
		Alignment Candidates Per Chunk	10	10	10
		Total Alignment Candidates	24	24	24
		Minimum Coverage For Correction	6	6	6
		Genome Size (Bp)	5,000,000	5,000,000	5,000,000
		Target Coverage	25	25	25
		Overlapper Error Rate	0.06	0.06	0.06
		Overlapper Min Length	40	40	40
		Overlapper K-mer	40	40	40
	HGAP3 mapping paramter	Maximum Divergence (%)	30	30	30
		Minimum Anchor size	12	12	12
Results of HGAP3	Filtering (Post-Filter)	Polymerase Read Bases	900,108,489	1,194,986,182	907,755,639
		Polymerase Reads	62,740	80,504	61,608
		Polymerase Read N50	19,893	20,216	20,230
		Polymerase Read Length	14,346	14,843	14,734
		Polymerase Read Quality	0.857	0.852	0.853
	Subread Filtering	Mean Subread length	11,893	12,011	12,205
		N50	17,237	16,717	17,679
		Total Number of Bases	898,815,238	1,193,475,427	906,550,170
	Adapters	Number of Reads	75,575	99,357	74,273
		Adapter Dimers (0-10bp)	0.01%	0.01%	0.00%
	Mapping	Short Inserts (11-100bp)	0.00%	0.00%	0.00%
		Mapped Subread Length N50 (bp)	16,483	14,897	16,037
		Mapped Polymerase Read Length 95% (bp)	28,320	28,360	26,850
		Mapped Subread Length Mean (bp)	10,855	10,109	10,399
		Mapped Polymerase Read Length Max (bp)	44,886	41,773	37,369
		Mapped Read	59,358	73,857	57,446

	Mapped Polymerase Read Length	12,992	12,238	12,455
	Mapped Polymerase Read Length n50	18,758	17,843	18,060
	Mapped Subreads	70,846	88,973	68,599
	Mapped Subread Bases	769,056,256	899,432,250	713,389,598
	Mapped Subread Length	10,855	10,109	10,399
	Mean Mapped Subread Concordance	0.863	0.849	0.857
Coverage	Mean Coverage	135.51	159.11	132.57
	Missing Bases (%)	0.04	0.2	0
Pre-Assembler Report	Polymerase Read Bases	898,815,238	1193475427	906550170
	Length Cutoff	23,221	24,800	23,363
	Seed Bases	150,001,201	150,016,858	150,016,570
	Pre-Assembled bases	107,431,279	80,209,190	80,373,405
	Pre-Assembled Yield	0.716	0.535	0.536
	Pre-Assembled Reads	6,108	6,059	5,995
	Pre-Assembled Reads Length	17,588	13,238	13,406
	Pre-Assembled N50	23,981	24,568	23,716
Polished Assembly	Polished Contigs	5	6	3
	Max Contig Length	4,942,963	5,144,467	5,062,492
	N50 Contig Length	4,942,963	5,144,467	5,062,492
	Sum of Contig Lengths	5,497,902	5428266	5297069
	High coverage Contigs ($\geq$ x100)	5	3	3

Table S2. Comparison of drug resistance and virulence factor genes among three isolates

Category	Gene	Product	MRY16-002		MRY15-117		MRY15-131	
			Chromosome	Plasmid	Chromosome	Plasmid	Chromosome	Plasmid
Drug resistance	<i>sul2</i>	sulfonamide-resistant dihydropteroate synthase Sul2	-	-	+	-	+	-
	<i>strA</i>	streptomycin phosphotransferase StrA	-	-	+	-	+	-
	<i>strB</i>	streptomycin phosphotransferase StrB	-	-	+	-	+	-
	<i>tet(A)</i>	tetracycline resistance protein TetA	-	+	+	-	+	-
	<i>floR</i>	florfenicol/ chloramphenicol export protein FloR	-	-	+	-	+	-
	<i>dfrA14</i>	dihydrofolate reductase DfrA14	-	-	-	+	-	+
	<i>mph(A)</i>	macrolide 2'-phosphotransferase	-	-	-	+	-	+
	<i>erm(B)</i>	rRNA adenine N-6-methyltransferase	-	-	-	+	-	+
	<i>aac(3)-IIa</i>	aminoglycoside N(3')-acetyltransferase	-	-	-	+	-	+
	<i>bla</i> _{CTX-M-27}	beta-lactamase CTX-M-27 precursor	-	-	-	+	-	+
Virulence factor	<i>mcr-1</i>	phosphoethanolamine transferase	-	+	-	+	-	+
	<i>aph(3')-Ia</i>	aminoglycoside 3'-phosphotransferase	-	+	-	-	-	-
	<i>afaA-VIII</i>	AfaA-VIII protein	-	-	-	-	+	-
	<i>afaB-VIII</i>	chaperone protein AfaB-VIII	-	-	-	-	+	-
	<i>afaC-VIII</i>	AfaC-VIII usher protein	-	-	-	-	+	-
	<i>afaD-VIII</i>	AfaD-VIII protein	-	-	-	-	+	-
	<i>afaE-VIII</i>	AfaE-VIII adhesin	-	-	-	-	+	-
	<i>aslA</i>	putative arylsulfatase	+	-	+	-	+	-
	<i>cdtA</i>	cytolethal distending toxin subunit A precursor	+	-	-	-	-	-
	<i>cdtB</i>	cytolethal distending toxin subunit B precursor	+	-	-	-	-	-
	<i>cdtC</i>	cytolethal distending toxin subunit C precursor	+	-	-	-	-	-
	<i>chuA</i>	outer membrane heme/hemoglobin receptor ChuA	+	-	+	-	+	-
	<i>chuS</i>	heme oxygenase ChuS	+	-	+	-	+	-
	<i>chuT</i>	periplasmic heme-binding protein ChuT	+	-	+	-	+	-
	<i>chuU</i>	heme permease protein ChuU	+	-	+	-	+	-
	<i>chuV</i>	ATP-binding hydrophilic protein ChuV	+	-	+	-	+	-
	<i>chuW</i>	putative oxygen independent coproporphyrinogen III oxidase	+	-	+	-	+	-
	<i>chuX</i>	putative heme-binding protein ChuX	+	-	+	-	+	-
	<i>chuY</i>	ChuY protein	+	-	+	-	+	-
	<i>ecpA</i>	pilus structural protein subunit EcpA	+	-	+	-	+	-
	<i>ecpB</i>	pilus chaperone protein EcpB	+	-	+	-	+	-
	<i>ecpC</i>	pilus usher protein EcpC	+	-	+	-	+	-
	<i>ecpD</i>	polymerized tip adhesin	+	-	+	-	+	-
	<i>ecpE</i>	pilus chaperone protein EcpE	+	-	+	-	+	-
	<i>ecpR</i>	regulator protein EcpR	+	-	+	-	+	-
	<i>entA</i>	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase	+	-	+	-	+	-
	<i>entB</i>	isochorismatase	+	-	+	-	+	-
	<i>entC</i>	isochorismate synthase 1	+	-	+	-	+	-
	<i>entD</i>	phosphopantetheinyl transferase component	+	-	-	-	-	-
	<i>entE</i>	2,3-dihydroxybenzoate-AMP ligase component	+	-	+	-	+	-
	<i>entF</i>	enterobactin synthase multienzyme complex component	+	-	+	-	+	-
	<i>entS</i>	enterobactin exporter	+	-	+	-	+	-
	<i>espL1</i>	type III secretion system effector EspL1	+	-	+	-	-	-
	<i>espR1</i>	type III secretion system effector EspR1	-	-	+	-	+	-
	<i>espX4</i>	type III secretion system effector EspX4	-	-	+	-	+	-
	<i>espY2</i>	type III secretion system effector EspY2	-	-	+	-	+	-
	<i>fepA</i>	ferrienterobactin outer membrane transporter	+	-	+	-	+	-
	<i>fepB</i>	ferrienterobactin ABC transporter periplasmic binding protein	+	-	+	-	+	-
	<i>fepC</i>	ferrienterobactin ABC transporter ATPase	+	-	+	-	+	-
	<i>fepD</i>	ferrienterobactin ABC transporter permease	+	-	+	-	+	-
	<i>fepG</i>	iron-enterobactin ABC transporter permease	+	-	+	-	+	-
	<i>fes</i>	enterobactin/ferric enterobactin esterase	+	-	+	-	+	-
	<i>fimA</i>	type I fimbriae major pilin FimA	+	-	+	-	+	-
	<i>fimB</i>	type I fimbriae chaperone protein FimB	+	-	+	-	+	-
	<i>fimC</i>	type I fimbriae outer membrane usher protein precursor FimC	+	-	+	-	+	-
	<i>fimD</i>	type I fimbriae fimbrial adhesin FimD	+	-	+	-	+	-
	<i>fimE</i>	type I fimbriae regulatory protein FimE	+	-	+	-	+	-
	<i>fimF</i>	type I fimbriae adaptor protein FimF	+	-	+	-	+	-
	<i>fimG</i>	type I fimbriae FimG protein precursor	+	-	+	-	+	-
	<i>fimH</i>	type I fimbriae minor fimbrial subunit FimH, adhesin	+	-	+	-	+	-
	<i>fimI</i>	type I fimbriae fimbrial protein internal segment FimI	+	-	+	-	+	-
	<i>gad</i>	glutamate decarboxylase	+	-	+	-	+	-
	<i>gspC</i>	general secretion pathway protein C	+	-	+	-	+	-
	<i>gspD</i>	general secretion pathway protein D	+	-	+	-	+	-
	<i>gspE</i>	general secretion pathway protein E	+	-	+	-	+	-
	<i>gspF</i>	general secretion pathway protein F	+	-	+	-	+	-
	<i>gspG</i>	general secretion pathway protein G	+	-	+	-	+	-
	<i>gspH</i>	general secretion pathway protein H	+	-	+	-	+	-
	<i>gspI</i>	general secretion pathway protein I	+	-	+	-	+	-
	<i>gspJ</i>	general secretion pathway protein J	+	-	+	-	+	-
	<i>gspK</i>	general secretion pathway protein K	+	-	+	-	+	-
	<i>gspL</i>	general secretion pathway protein L	+	-	+	-	+	-

<i>gspM</i>	general secretion pathway protein M	+	-	+	-	+	-
<i>ireA</i>	siderophore receptor	+	-	+	-	+	-
<i>irp1</i>	yersiniabactin biosynthetic protein Irp1	+	-	-	-	-	-
<i>irp2</i>	yersiniabactin biosynthetic protein Irp2	+	-	-	-	-	-
<i>irp3</i>	yersiniabactin biosynthetic protein Irp3	+	-	-	-	-	-
<i>irp4</i>	yersiniabactin biosynthetic protein Irp4	+	-	-	-	-	-
<i>irp5</i>	yersiniabactin siderophore biosynthetic protein Irp5	+	-	-	-	-	-
<i>irp6</i>	lipoprotein inner membrane ABC-transporter Irp6	+	-	-	-	-	-
<i>irp7</i>	inner membrane ABC-transporter Irp7	+	-	-	-	-	-
<i>irp8</i>	putative signal transducer Irp8	+	-	-	-	-	-
<i>irp9</i>	salicylate synthase Irp9	+	-	-	-	-	-
<i>fyuA</i>	yersiniabactin receptor protein	+	-	-	-	-	-
<i>iss</i>	increased serum survival protein	+	+	+	-	+	-
<i>kpsD</i>	KpsD protein	-	-	+	-	+	-
<i>lpfA</i>	long polar fimbria protein LpfA	+	-	+	-	+	-
<i>mchF</i>	ABC transporter protein MchF	-	-	+	-	+	-
<i>ompA</i>	outer membrane protein A	+	-	+	-	+	-
<i>papB</i>	regulatory protein PapB	-	-	+	-	+	-
<i>papC</i>	usher protein PapC	-	-	+	-	-	-
<i>papD</i>	chaperone protein PapD	-	-	+	-	-	-
<i>papH</i>	P pilus termination subunit PapH	-	-	+	-	-	-
<i>papI</i>	regulatory protein PapI	-	-	+	-	+	-
<i>papJ</i>	P pilus assembly protein PapJ	-	-	+	-	-	-
<i>papK</i>	P pilus minor subunit PapK	-	-	+	-	-	-
<i>papX</i>	flagellum synthesis regulation protein PapX	-	-	+	-	+	-
<i>pic</i>	serine protease autotransporter Pic	+	-	-	-	-	-
<i>tsh</i>	haemoglobin protease	+	-	-	-	-	-
<i>ybtA</i>	yersiniabactin-related transcriptional regulator YbtA	+	-	-	-	-	-
<i>iroB</i>	glucosyltransferase IroB	-	+	-	-	-	-
<i>iroC</i>	ATP binding cassette transporter IroC	-	+	-	-	-	-
<i>iroD</i>	esterase IroD	-	+	-	-	-	-
<i>iroE</i>	esterase IroE	-	+	-	-	-	-
<i>iroN</i>	salmochelins receptor IroN	-	+	-	-	-	-
<i>iucA</i>	aerobactin synthesis protein IucA	-	+	-	-	-	-
<i>iucB</i>	aerobactin synthesis protein IucB	-	+	-	-	-	-
<i>iucC</i>	aerobactin synthesis protein IucC	-	+	-	-	-	-
<i>iucD</i>	aerobactin synthesis protein IucC, lysine 6-monooxygenase	-	+	-	-	-	-
<i>katP</i>	plasmid-encoded catalase peroxidase	-	+	-	-	-	-

+, presence; -, absence.

Table S3. Summary of the differences in 20 or 23 bp conserved region among shufflon inverted repeat regions

		Number of mismatch in 23 bp conserved region (idenitiy)					
		ORF A	ORF A'	ORF B'	ORF D'	ORF C	ORF C'
Number of mismatch in 20 bp conserved region (idenitiy)	ORF A		2 (91.3%)	3 (87.0%)	2 (91.3%)	3 (87.0%)	6 (73.9%)
	ORF A'	0 (100%)		5 (78.3%)	4 (82.6%)	5 (78.3%)	5 (78.3%)
	ORF B'	2 (90.0%)	2 (90.0%)		1 (95.7%)	3 (87.0%)	6 (73.9%)
	ORF D'	2 (90.0%)	2 (90.0%)	0 (100%)		3 (87.0%)	6 (73.9%)
	ORF C	2 (90.0%)	2 (90.0%)	2 (90.0%)	2 (90.0%)		6 (73.9%)
	ORF C'	3 (85.0%)	3 (85.0%)	3 (85.0%)	3 (85.0%)	4 (80.0%)	

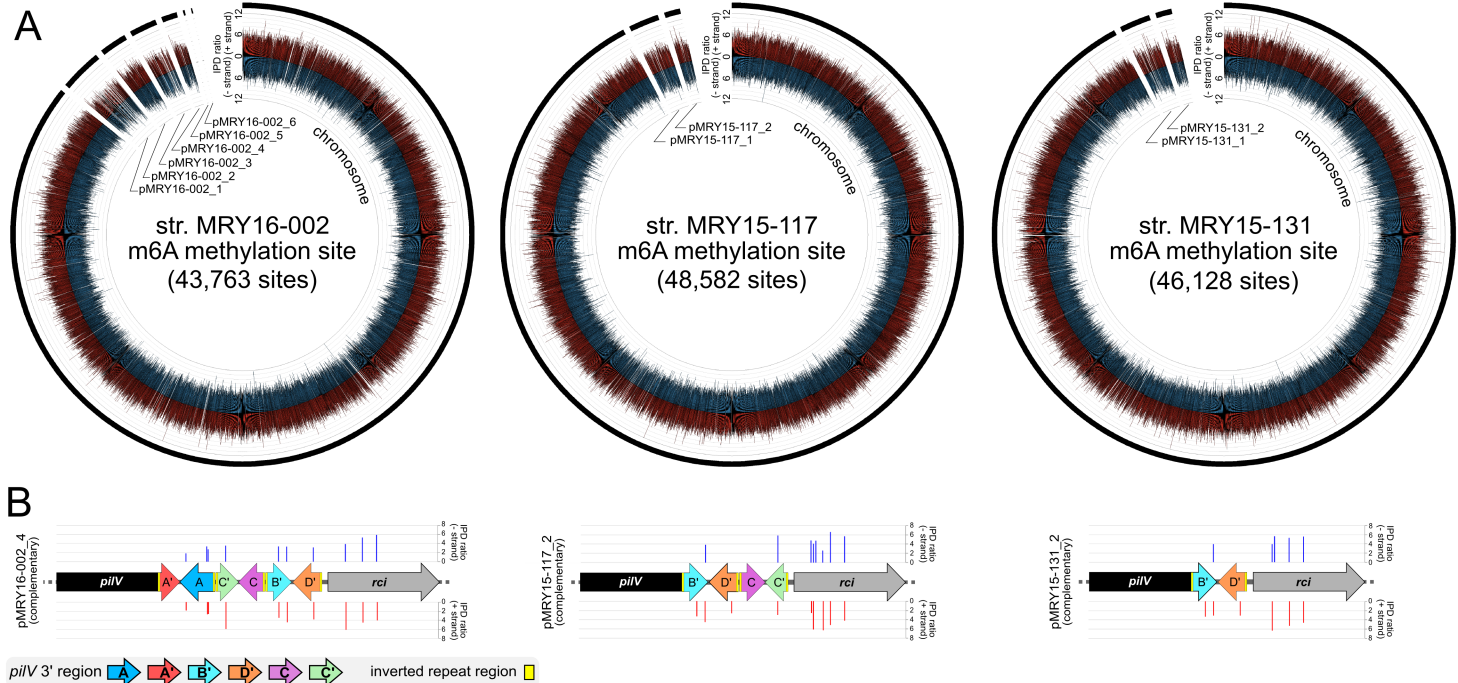


Figure S1 N⁶-Methyladenosine (m6A) methylation site detected in three *Escherichia coli* strains. (A) Circos plot displaying the genome-wide detections of m6A methylation site. The interpulse duration (IPD) ratio on the positive and negative strands are plotted as the red and blue bars, respectively. (B) Schematic representation of m6A methylation sites on shufflon regions and adjacent gene loci in three *E. coli* strains. The m6A methylation site is not detected in all shufflon inverted repeat regions.

Table S4. Summary of discovered methylation motifs in three *Escherichia coli* strains

Motif	Modified Position	Modification Type	MRY16-002				MRY15-117				MRY15-131			
			% Detected methylations (# detected methylations)	Mean Score	Mean IPD Ratio	Mean Coverage	% Detected methylations (# detected methylations)	Mean Score	Mean IPD Ratio	Mean Coverage	% Detected methylations (# detected methylations)	Mean Score	Mean IPD Ratio	Mean Coverage
GATC	2	m6A	98.96% (42,145)	103.63	4.65	67.16	99.16% (41,899)	119.00	4.60	79.88	99.04% (41,499)	100.73	4.61	64.67
GAGNNNNRTAC	2	m6A	98.67% (519)	106.13	6.01	67.31	99.43% (521)	120.11	5.88	79.32	99.42% (512)	101.63	5.94	64.34
GTAYNNNNCTC	3	m6A	99.24% (522)	102.56	5.25	68.54	99.43% (521)	117.29	5.11	81.10	99.42% (512)	97.99	5.17	65.65
CACNNNNNGGG	2	m6A	N.D.	N.D.	N.D.	N.D.	99.10% (1,535)	106.89	4.23	79.83	98.28% (1,488)	90.39	4.28	64.67
CTANNNNNNTATC	3	m6A	98.63% (289)	100.26	5.16	69.59	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
GATANNNNNNTAG	4	m6A	98.29% (288)	101.05	5.49	69.42	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
ATAYNNNNNGTG	3	m6A	N.D.	N.D.	N.D.	N.D.	89.11% (1,039)	95.62	3.82	80.13	N.D.	N.D.	N.D.	N.D.
CACNNNNNRTAT	2	m6A	N.D.	N.D.	N.D.	N.D.	99.57% (1,161)	114.81	4.73	80.79	N.D.	N.D.	N.D.	N.D.
GAGNNNNNNNTCA	3	m6A	N.D.	N.D.	N.D.	N.D.	99.58% (957)	116.52	5.72	78.10	N.D.	N.D.	N.D.	N.D.
TGANNNNNNCTTC	3	m6A	N.D.	N.D.	N.D.	N.D.	98.75% (949)	118.23	4.81	80.54	N.D.	N.D.	N.D.	N.D.
AAAGNNNNNNNTCA	3	m6A	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	99.91% (1,062)	97.79	6.19	62.07
TGANNNNNNCTTT	3	m6A	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	99.25% (1,055)	100.16	5.07	65.90

N.D., Not detected.

Table S5. NGS platform information of complete Incl2 plasmids

Organim	Complete plasmid name	NGS platform	Sequence Read Archive (SRA) ID	Sequence accession number	Ref. (PMID)	PilV 3' end region serch with assembled sequences						PilV 3' end region serch with NGS raw reads					
						ORF A	ORF A'	ORF B'	ORF D'	ORF C	ORF C'	ORF A	ORF A'	ORF B'	ORF D'	ORF C	ORF C'
<i>Escherichia coli</i>	pMRY15-131_2	Illumina MiSeq, PacBio	DRR065948, DRR065951	AP017622	this study	-	-	+	+	-	-	-	-	+	+	-	-
<i>Escherichia coli</i>	pMRY15-117_2	Illumina MiSeq, PacBio	DRR065947, DRR065950	AP017619	this study	-	-	+	+	+	+	-	-	+	+	+	+
<i>Escherichia coli</i>	pMRY16-002_4	Illumina MiSeq, PacBio	DRR065946, DRR065949	AP017614	this study	+	+	+	+	+	+	+	+	+	+	+	+
<i>Escherichia coli</i>	pHNSHP45	Illumina MiSeq	N.A.	KP347127	26603172	+	+	+	+	-	-	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
<i>Salmonella enterica</i>	pSH146_65	454 GS-FLX	N.A.	JN983044	23896467	+	+	+	+	+	pseudo	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
<i>Escherichia coli</i>	unnamed 3	PacBio	N.A.	CP009581	N.A.	+	+	+	pseudo	+	+	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
<i>Escherichia coli</i>	R721	Sanger	N.A.	AP002527	1400257	+	+	+	+	+	+	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
<i>Escherichia coli</i>	pChi7122-3	N.A.	N.A.	FR851304	22238616	+	+	+	+	-	-	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
<i>Escherichia coli</i>	pRM12761	Sanger, 454 GS-FLX, Illumina, PacBio	N.A.	CP007134	24855308	+	+	+	+	-	-	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
<i>Escherichia coli</i>	pRM13516	454 GS-FLX, Illumina HiSeq, PacBio	N.A.	CP006264	24410921	+	+	+	+	-	-	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
<i>Escherichia coli</i>	pHN1122-1	Sanger	N.A.	JN797501	23478963	+	+	+	+	-	-	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
<i>Escherichia coli</i>	pHNY2	454 GS-FLX	N.A.	KF601686	25987615	+	+	+	+	-	-	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
<i>Escherichia coli</i>	pCTXM132_P0421	IonTorrent	N.A.	KP198615	25861872	pseudo	pseudo	+	+	-	-	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
<i>Escherichia coli</i>	pHNLDH19	454 GS-FLX	N.A.	KM207012	25987615	+	+	+	+	-	-	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
<i>Escherichia coli</i>	pHNAH46-1	454 GS-FLX	N.A.	KJ020576	25987615	+	+	+	+	-	-	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
<i>Escherichia coli</i>	pCTXM64_C0967	IonTorrent	N.A.	KP091735	25861872	+	+	+	+	-	-	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
<i>Shigella sonnei</i>	p1081-CTXM	N.A.	N.A.	KJ460501	25405884	+	+	+	+	-	-	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
<i>Klebsiella pneumoniae</i>	pUHKPC45-77	PacBio, Illumina HiSeq	SRX283875, SRX277000	JMSX01000003	N.A.	+	+	+	+	-	-	+	+	+	+	-	-
<i>Klebsiella pneumoniae</i>	pBK15692	454 GS-FLX	N.A.	KC845573	23896467	+	+	+	+	-	-	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
<i>Klebsiella pneumoniae</i>	pDMC1097-77.775kb	PacBio, Illumina HiSeq	SRX276936, SRX305269, SRX1075361-4	CP011978	N.A.	+	+	+	+	-	-	+	+	+	+	-	-
<i>Klebsiella pneumoniae</i>	pKPC_CAV1596-78	HiSeq	SRR1582868	CP011645	25561339	+	+	+	+	-	-	+	+	+	+	-	-

N.A., not available; +, presence; -, absence; pseudo, pseudogene.