

Table S1. Primers.

Primer name	Sequence 5'-3'
SNP_586018 F (<i>glxK</i>)	ATTGAGATTGTACTGCGGGCG
SNP_586018 R	GAAAGATTGCAACCGCGACG
SNP_1633365 F (prophage)	ATCAGAGGACATTTTCAGGAGGTG
SNP_1633365 R	GGCATTCCGGGCATTTTCAG
SNP_3316686 F (prophage)	CAGCTGGCAAGCGAATGTAA
SNP_3316686 R	ACAGTATCAAGGTGAGTGAGATG
SNP_3678749 F (<i>metK</i>)	GTCCGTCTCTGAAGGGCATC
SNP_3678749 R	CCGTCGTCATACTGGAAAGTCA
SNP_3892698 F (<i>fadH</i>)	CAGCTACCAACCGCATCTGG
SNP_3892698 R	TGCGCGTGACCCAAGTAAAT
SNP pO111_3 F	CTTCTTTTTCTGCGCTCTG
SNP pO111_3 R	CCGTATCTCCCCGTCTGTT
95NR1 contig213000000 F (unique sequence of Stx2 phage 3)	ACGATCGTTTCTTGGGTCTTGA
95NR1 contig213000000 R (hypothetical protein encoded at the end of Stx2 phage 2)	TGATGCCGTATGCAGACCAG
Tail fiber F	GTTCGCATAGGGCAGATTACGC
Phage R	ATGAGTACCGCTGGCATTTCG
Stx2 phage 1 cI F	TTGGCGTCTTCCAGTTGCTCAG
Stx 2 phage 2 cI F	TCAAGACCAGCTTCTTTACAGGC
Stx2 phage 3 cI F	ATTCAACCATCTTGCGAGATCCG
Stx2A R	ACCCAGTAACAGGCACAGTACC

Table S2. Prophages and phages used to construct the end sequence profiling index.

Prophage/Phage	Genbank Accession	Reference
ECO111_P01 – ECO111_P16 of O111:H- 11128	AP010960.1	(6)
ECO103_P04 of O103:H2 12009	AP010958.1	(6)
ECO103_P12 of O103:H2 12009	AP010958.1	(6)
ECO103_P14 of O103:H2 12009	AP010958.1	(6)
ECO26_P14 of O26:H11 11368	AP010953.1	(6)
Sp18 of O157:H7 Sakai	BA000007.2	(2)
CRP49 of <i>Citrobacter</i> <i>rodentium</i> ICC168	FN543502.1	(7)
Bacteriophage Mu	AF083977.1	(5)
<i>Escherichia</i> phage TL-2011a	JQ011316.1	(4)
Stx2-converting phage 86	AB255436.1	N/A
Stx2-converting phagevB_EcoP_24B	HM208303.1	(9)
Bacteriophage Min27	EU311208.1	(10)
<i>Escherichia</i> phage TL-2011c	JQ011318.1	(4)
Bacteriophage VT2phi_272	HQ424691.1	N/A
Stx2-converting phage 1717	FJ188381.1	N/A
Phage YYZ-2008	FJ184280.1	N/A
Phage JLK-2012	JQ347801.1	(3)
Stx2-converting phage II	AP005154.1	(8)
Phage BP-4795	AJ556162.1	(1)
nt 1245044-1317701 of O157:H7 Xuzhou21	CP001925.1	(11)

Table S3. *E. coli* genomes used in phylogenetic tree construction.

Accession	GI	Strain	Pathovar
NC_008563	117622295	<i>Escherichia coli</i> APEC O1	APEC
NC_010468	170018061	<i>Escherichia coli</i> ATCC 8739	Laboratory
NC_012967	254160123	<i>Escherichia coli</i> B str. REL606	Laboratory
NC_010498	170679574	<i>Escherichia coli</i> SMS-3-5	Environmental
NC_010473	170079663	<i>Escherichia coli</i> str. K12 substr. DH10B	Laboratory
NC_011353	209395693	<i>Escherichia coli</i> O157:H7 str. EC4115	EHEC
NC_013008	254791136	<i>Escherichia coli</i> O157:H7 str. TW14359	EHEC
NC_012947	253771435	<i>Escherichia coli</i> 'BL21-Gold(DE3)pLysS AG'	Laboratory
NC_011601	215485161	<i>Escherichia coli</i> O127:H6 E2348/69	EPEC
NC_011741	218552585	<i>Escherichia coli</i> IAI1	Commensal
NC_011745	218687878	<i>Escherichia coli</i> ED1a	Commensal
NC_011750	218698419	<i>Escherichia coli</i> IAI39	UPEC
NC_011748	218693476	<i>Escherichia coli</i> 55989	EAEC
NC_016902	378710836	<i>Escherichia coli</i> KO11	Laboratory
NZ_CM000662	224514698	<i>Escherichia coli</i> O157:H7 str. TW14588	EHEC
NC_000913	49175990	<i>Escherichia coli</i> str. K-12 substr. MG1655	Laboratory
NC_013353	260842239	<i>Escherichia coli</i> O103:H2 str. 12009	EHEC
NC_013361	260853213	<i>Escherichia coli</i> O26:H11 str. 11368	EHEC
NC_013364	260866153	<i>Escherichia coli</i> O111:H- str. 11128	EHEC
NC_013941	291280824	<i>Escherichia coli</i> O55:H7 str. CB9615	EPEC
NC_002695	15829254	<i>Escherichia coli</i> O157:H7 str. Sakai	EHEC
NC_002655	16445223	<i>Escherichia coli</i> O157:H7 EDL933	EHEC
NC_004431	26245917	<i>Escherichia coli</i> CFT073	UPEC
NC_009800	157159467	<i>Escherichia coli</i> HS	Commensal
NC_009801	157154711	<i>Escherichia coli</i> E24377A	ETEC
NC_008253	110640213	<i>Escherichia coli</i> 536	UPEC
NC_007946	91209055	<i>Escherichia coli</i> UTI89	UPEC
NC_012759	238899406	<i>Escherichia coli</i> BW2952	Laboratory
NC_011415	209917191	<i>Escherichia coli</i> SE11	Commensal
NC_011742	218556939	<i>Escherichia coli</i> S88	ExPEC - Meningitis
NC_011751	218703261	<i>Escherichia coli</i> UMN026	UPEC
AP009048	85674274	<i>Escherichia coli</i> str. K12 substr. W3110	Laboratory
AP009378	281177210	<i>Escherichia coli</i> SE15	Commensal
AM946981	313848522	<i>Escherichia coli</i> BL21(DE3)	Laboratory
CP001637	260447279	<i>Escherichia coli</i> DH1	Laboratory
CU651637	222031834	<i>Escherichia coli</i> LF82	AIEC
CP001671	307551844	<i>Escherichia coli</i> ABU 83972	Commensal
FN554766	284919779	<i>Escherichia coli</i> 042	EAEC
CP001855	312944605	<i>Escherichia coli</i> O83:H1 str. NRG 857C	AIEC

CP002729	332341332	<i>Escherichia coli</i> UMNK88	porcine EPEC
FN649414	309700213	<i>Escherichia coli</i> ETEC H10407	EPEC
CP001969	294489418	<i>Escherichia coli</i> IHE3034	ExPEC - Meningitis
CP002167	307625127	<i>Escherichia coli</i> UM146	AIEC
CP002211	355418401	<i>Escherichia coli</i> str. 'clone D i2'	UPEC
CP002212	355423321	<i>Escherichia coli</i> str. 'clone D i14'	UPEC
AP012030	315134697	<i>Escherichia coli</i> DH1 (ME8569)	Laboratory
CP002291	383101383	<i>Escherichia coli</i> P12b	Laboratory
CP002967	383403426	<i>Escherichia coli</i> W	Laboratory
CP003034	349736152	<i>Escherichia coli</i> O7:K1 str. CE10	ExPEC - Meningitis
CP002797	333968017	<i>Escherichia coli</i> NA114	UPEC
CP003109	374356928	<i>Escherichia coli</i> O55:H7 str. RM12579	EPEC

Table S4. SNPs unique to 95NR1 and 95JB1

Strain	11128 site	Base change	Amino acid change	11128 gene number	Annotation
95JB1	586018	c-t	P-L	ECO111_0546	<i>glxK</i> (glycerate kinase II)
95NR1	1633365 (3328268)	t-g	Synonymous	ECO111_1622	putative tail length tape measure protein
95JB1	3316686	g-t	V-F	-	End of P16 prophage
95JB1	3678749	g-a	E-K	ECO111_3690	<i>metK</i> (methionine adenosyltransferase 1)
95JB1	3892698	g-c	Stop codon- Y (but next codon is also stop)	ECO111_3903	<i>fadH</i> (2,4-dienoyl-CoA reductase)
95JB1	35489 (pO111_3)	g-c	P-A	ECO111_p3-43	pseudogene (conserved predicted protein, N-terminal part)

Table S5. Genes outside of annotated prophages of 11128 to which reads of 95NR1 and 95JB1 do not align.

11128 gene number	Annotation	BLAST-p		Description
ECO111_0321	Hypothetical protein	Hypothetical protein	No Pfam-A match	IE01 (end)
0322	Hypothetical protein	Hypothetical protein, plasmid pRiA4b ORF-3 like protein	PRiA4_ORF3 (Plasmid pRiA4b ORF-3-like protein)	
0323	Hypothetical protein	Hypothetical protein, putative phage protein	No Pfam-A match	
0324	Hypothetical protein	Hypothetical, prophage protein	No Pfam-A match	
0325	Hypothetical protein	Hypothetical protein, putative prophage protein	No Pfam-A match	
0326	Hypothetical protein	Hypothetical protein, putative prophage protein	No Pfam-A match	
0327	Hypothetical protein	Hypothetical protein	No Pfam-A match	
0328	Hypothetical protein	Hypothetical protein	No Pfam-A match	
0880	Hypothetical protein	Hypothetical protein	No Pfam-A match	Region in 12009 and 11128 only in the NCBI nucleotide collection
0881	Predicted methyltransferase	Methyltransferase, modification methylase, DNA methylase N-4/N-6 domain containing protein, restriction endonuclease M subunit	N6_N4_Mtase (DNA methylase)	
0882	Site specific recombinase	Hypothetical protein, site	Resolvase (resolvase, N	

		specific recombinase, resolvase	terminal domain)	
0883	Endonuclease-like protein	Endonuclease, HNH endonuclease	HNH (HNH endonuclease)	
0884	Hypothetical protein	Hypothetical protein	No Pfam-A match	
0885	Putative transcriptional regulator	Transcriptional regulator, DNA-binding protein, XRE family transcriptional regulator	HTH 31 (helix-turn-helix domain)	
0886	Hypothetical protein	Hypothetical protein	No Pfam-A match	
0887	Hypothetical protein	Hypothetical protein, resolvase	No Pfam-A match	
1286	Hypothetical protein	Hypothetical protein, transcriptional regulator, parB-like nuclease domain protein, immunoglobulin-binding regulator IbrB	ParBc (ParB-like nuclease domain)	IE02
1287	Transcriptional regulator PchD-homolog	Hypothetical protein, transcriptional regulator, perC transcriptional activator family protein	PerC (PerC transcriptional activator)	
1288	Hypothetical protein	Hypothetical protein	No Pfam-A match	
1289	Conserved predicted protein	Hypothetical protein	No Pfam-A match	
1290	Putative IS629 transposase	IS629 transposase OrfB, transposase IS629, integrase core domain protein	HTH 21 (HTH-like domain), rve (integrase core domain)	

1291	Hypothetical protein	Hypothetical protein	No Pfam-A match	
1292	Putative membrane transport protein	Hypothetical protein, membrane transport protein, major facilitator superfamily transport protein	MFS 1 (major facilitator family protein)	
1293	Siderophore biosynthesis protein (<i>iucA</i>)	IucA protein, Aerobactin siderophore biosynthesis protein, siderophore biosynthesis protein	IucA_IucC (IucC/IucA family protein), FhuF (Ferric iron reductase FhuF-like protein)	
1294	Siderophore biosynthesis protein (<i>iucB</i>)	IucB, N(6)-hydroxylysine O-acetyltransferase	Acetyltransf_8 (Acetyltransferase [GNAT] domain)	
1295	Siderophore biosynthesis protein (<i>iucC</i>)	IucC, Aerobactin siderophore biosynthesis protein, siderophore biosynthesis protein	IucA_IucC (IucC/IucA family protein), FhuF (Ferric iron reductase FhuF-like protein)	
1296	Siderophore biosynthesis protein (<i>iucD</i>)	Siderophore biosynthesis protein, IucD, lysine 6-monooxygenase, lysine:N6-hydroxylase	K_oxygenase (L-lysine 6-monooxygenase [NADPH-requiring])	
1297	Putative ferric siderophore receptor (<i>iutA</i>)	IutA, ligand-gated channel protein	Plug (TonB-dependent Receptor Plug Domain) TonB_Dep_Rec, (TonB dependent receptor)	
2137	Conserved hypothetical protein (<i>ydhS</i>) [half deleted]	Hypothetical protein	NAD_binding_9 (FAD-NAD[P]-binding)	Region found in multiple <i>E. coli</i> in NCBI nucleotide collection

2138	Conserved hypothetical (<i>ydhT</i>)	Hypothetical protein	No Pfam-A match	
2139	Predicted cytochrome (<i>ydhU</i>)	Nickel-dependent hydrogenase b-type cytochrome subunit, thiosulfate reductase cytochrome B, prokaryotic cytochrome b561 family protein, PhsC	Ni_hydr_CYTB (prokaryotic cytochrome b561)	
2140	Predicted 4Fe-4S ferridoxin-type protein (<i>ydhX</i>)			
2141	Conserved predicted protein (<i>ydhW</i>)			

Table S6. Identification of phage insertion sites of 95NR1 and 95JB1 and associated phage terminal regions using paired-end reads.

Location of the phage insertion site in relation to the 11128 reference genome	Location of the mate reads of the reads aligned to the 5' end of the phage insertion site	Location of the mate reads of the reads aligned to the 3' end of the phage insertion site
ECO111_tRNA006: tRNA-Arg	ECO111_P01 (5' end)	ECO111_P01 (3' end)
Adjacent <i>ybhC</i> (0783)	ECO111_P02 (5' end)	ECO111_P02 (3' end)
Adjacent <i>yccA</i> (1038)	ECO111_P03 (5' end)	ECO111_P03 (3' end)
Adjacent <i>potC</i> (1401)	ECO111_P05 (5' end)	ECO111_P05 (3' end)
Adjacent <i>yefD</i> (1476)	O103:H2 str. 12009's ECO103_P04 (5' end)	O103:H2 str. 12009's ECO103_P04 (3' end)
Adjacent <i>yciI</i> (1579)	ECO111_P06 (5' end)	ECO111_P06 (3' end)
Adjacent <i>ydfJ</i> (1946)	ECO111_P08 (3' end)	ECO111_P08 (5' end)
Adjacent <i>tonB</i> (1986)	ECO111_P09 (3' end)	ECO111_P09 (5' end)
Adjacent <i>btuC</i> (2181)	ECO111_P10 (5' end)	ECO111_P10 (3' end)
Adjacent <i>pphA</i> (2346)	11368 ECO26_P14 (3' end)	11368 ECO26_P14 (5' end)
Adjacent <i>yobB</i> (2553)	ECO111_P12 (3' end)	ECO111_P12 (5' end)
<i>yfaT</i> (2981)	ECO111_P15 (3' end)	ECO111_P15 (5' end)
ECO111_tRNA056: tRNA-Arg	ECO103_P12 (3' end), VB_ECOP_24B(3' end), VT2phi_272(3' end)	ECO103_P12 (5' end), VT2phi_272(5' end)
tmRNA <i>ssrA</i> (adjacent 3340)	ECO111_P16 (3' end)	ECO111_P16 (5' end)
Adjacent <i>yicC</i> (4466)	O103:H2 str. 12009's ECO103_P14 (5' end)	O103:H2 str. 12009's ECO103_P14 (3' end)
Hypothetical protein (4843)	Sakai phage (5' end)	Sakai phage (3' end)
<i>yjbN</i> (4923)	ECO111_P17 (5' end)	ECO111_P17 (3' end)
Hypothetical protein(5095) [95NR1 only]	ECO103_P12 (5' end), VT2phi_272 (5' end)	VT2phi_272 (3' end)

Genome References

1. **Creuzburg, K., J. Recktenwald, V. Kuhle, S. Herold, M. Hensel, and H. Schmidt.** 2005. The Shiga toxin 1-converting bacteriophage BP-4795 encodes an NleA-like type III effector protein. *J Bacteriol* **187**:8494-8498.
2. **Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C. G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori, and H. Shinagawa.** 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* **8**:11-22.
3. **Kyle, J. L., C. A. Cummings, C. T. Parker, B. Quinones, P. Vatta, E. Newton, S. Huynh, M. Swimley, L. Degoricija, M. Barker, S. Fontanoz, K. Nguyen, R. Patel, R. Fang, R. Tebbs, O. Petrauskene, M. Furtado, and R. E. Mandrell.** 2012. *Escherichia coli* serotype O55:H7 diversity supports parallel acquisition of bacteriophage at Shiga toxin phage insertion sites during evolution of the O157:H7 lineage. *J Bacteriol* **194**:1885-1896.
4. **L'Abée-Lund, T. M., H. J. Jorgensen, K. O'Sullivan, J. Bohlin, G. Ligard, P. E. Granum, and T. Lindback.** 2012. The highly virulent 2006 Norwegian EHEC O103:H25 outbreak strain is related to the 2011 German O104:H4 outbreak strain. *PLoS One* **7**:e31413.
5. **Morgan, G. J., G. F. Hatfull, S. Casjens, and R. W. Hendrix.** 2002. Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus*, *Neisseria* and *Deinococcus*. *J Mol Biol* **317**:337-359.
6. **Ogura, Y., T. Ooka, A. Iguchi, H. Toh, M. Asadulghani, K. Oshima, T. Kodama, H. Abe, K. Nakayama, K. Kurokawa, T. Tobe, M. Hattori, and T. Hayashi.** 2009. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *Proc Natl Acad Sci U S A* **106**:17939-17944.
7. **Petty, N. K., R. Bulgin, V. F. Crepin, A. M. Cerdeno-Tarraga, G. N. Schroeder, M. A. Quail, N. Lennard, C. Corton, A. Barron, L. Clark, A. L. Toribio, J. Parkhill, G. Dougan, G. Frankel, and N. R. Thomson.** 2010. The *Citrobacter rodentium* genome sequence reveals convergent evolution with human pathogenic *Escherichia coli*. *J Bacteriol* **192**:525-538.
8. **Sato, T., T. Shimizu, M. Watarai, M. Kobayashi, S. Kano, T. Hamabata, Y. Takeda, and S. Yamasaki.** 2003. Genome analysis of a novel Shiga toxin 1 (Stx1)-converting phage which is closely related to Stx2-converting phages but not to other Stx1-converting phages. *J Bacteriol* **185**:3966-3971.
9. **Smith, D. L., D. J. Rooks, P. C. Fogg, A. C. Darby, N. R. Thomson, A. J. McCarthy, and H. E. Allison.** 2012. Comparative genomics of Shiga toxin encoding bacteriophages. *BMC Genomics* **13**:311.
10. **Su, L. K., C. P. Lu, Y. Wang, D. M. Cao, J. H. Sun, and Y. X. Yan.** 2010. [Lysogenic infection of a Shiga toxin 2-converting bacteriophage changes host gene expression, enhances host acid resistance and motility]. *Mol Biol (Mosk)* **44**:60-73.
11. **Xiong, Y., P. Wang, R. Lan, C. Ye, H. Wang, J. Ren, H. Jing, Y. Wang, Z. Zhou, X. Bai, Z. Cui, X. Luo, A. Zhao, S. Zhang, H. Sun, L. Wang, and J. Xu.** 2012. A novel *Escherichia coli* O157:H7 clone causing a major hemolytic uremic syndrome outbreak in China. *PLoS One* **7**:e36144.

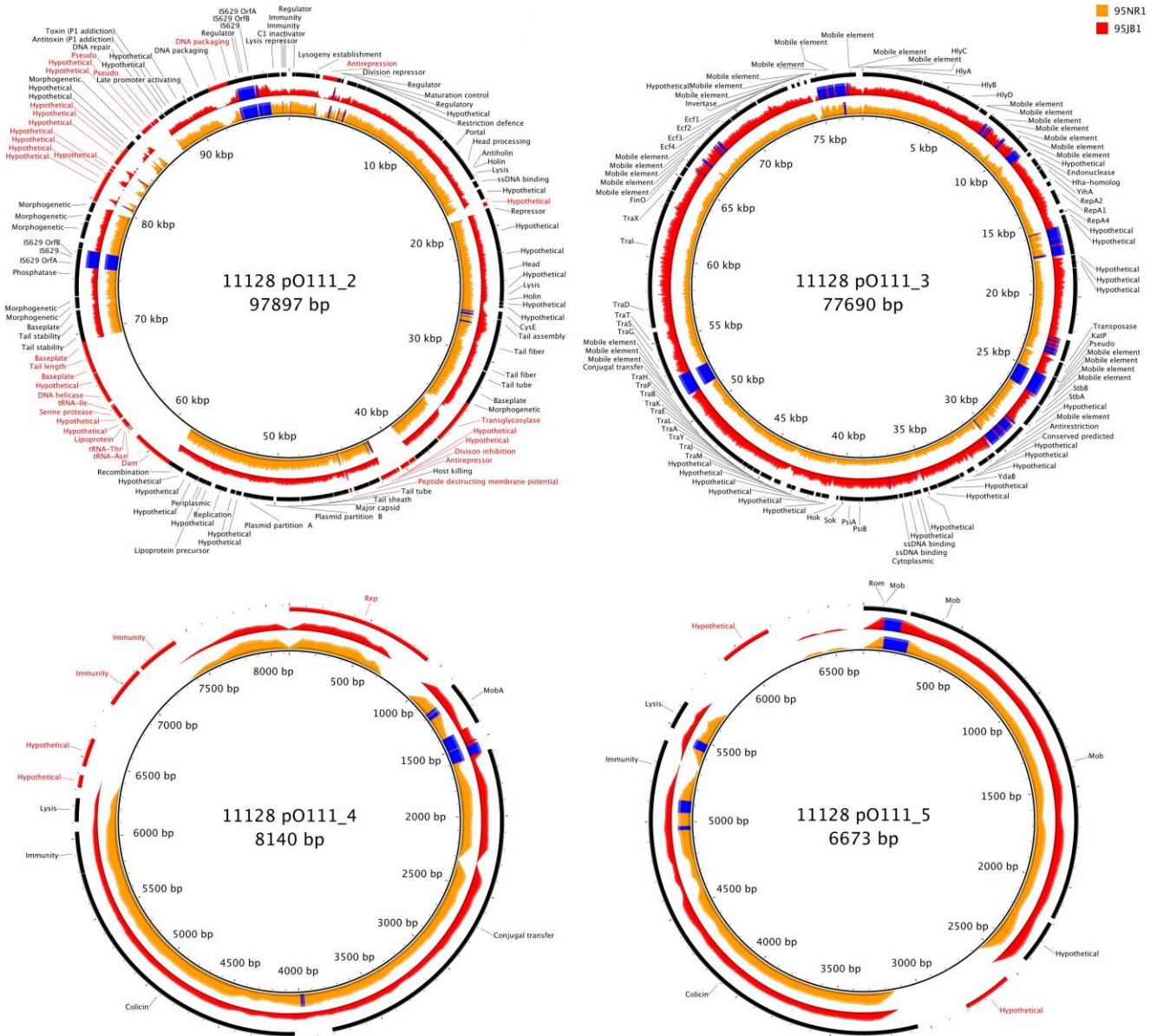


Figure S1. Mapping of 95NR1 and 95JB1 reads to plasmids p0111_2 - p0111_5 of 11128. Reads of 95NR1 and 95JB1 were aligned to the plasmids using Bowtie2. The inner (orange) rings represent 95NR1 coverage and the outer (red) rings represent 95JB1 coverage. Areas of high coverage are shown in royal blue. Genes without alignment are annotated in red.

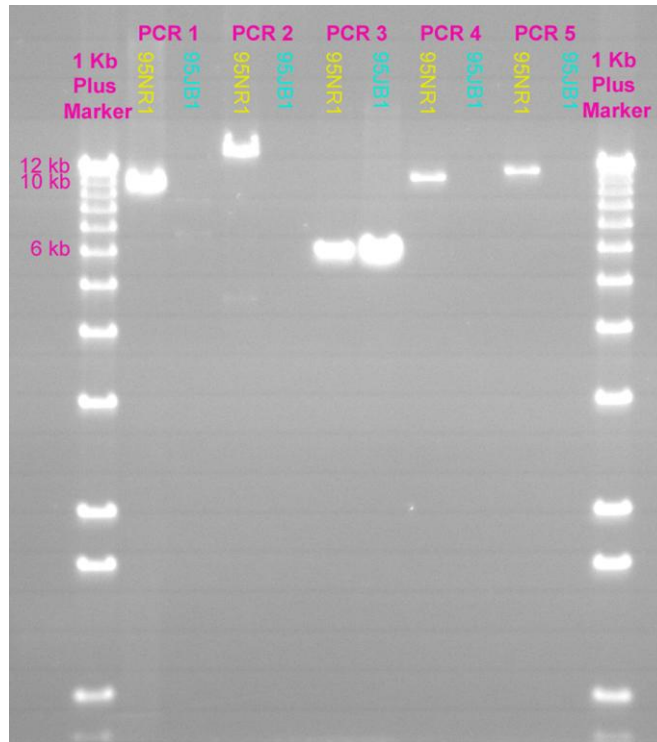


Figure S2. Stx2-converting phage PCRs of this study.

PCRs were performed using 95NR1 and 95JB1 genomic DNA with the following primers:

PCR 1: 95NR1 contig213000000 F and R

PCR 2: Tail fiber F and Hypothetical R

PCR 3: Stx2 phage 1 cI F and Stx2A R

PCR 4: Stx2 phage 2 cI F and Stx2A R

PCR 4: Stx2 phage 3 cI F and Stx2A R

Samples were electrophoresed with 1 Kb Plus DNA Ladder (Life Technologies, Victoria, Australia) on a 0.8% agarose gel in TBE for 6 hours at 60 V and post-stained using GelRed Nucleic Acid Gel Stain (Biotium, California, USA).