

Supplementary Information

Identification and Pathogenomic Analysis of an *Escherichia coli* Strain Producing a Novel Shiga Toxin 2 Subtype

Running title: A novel Subtype of Shiga Toxin 2

Xiangning Bai^{1†}, Shanshan Fu^{1†}, Ji Zhang^{2†}, Ruyue Fan¹, Yanmei Xu¹, Hui Sun¹, Xiaohua He³, Jianguo Xu^{1,4}, and Yanwen Xiong^{1,4*}

1. State Key Laboratory of Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Changping, Beijing, China;
2. mEpiLab, Institute of Veterinary, Animal and Biomedical Sciences, Massey University, New Zealand;
3. Western Regional Research Center, U.S. Department of Agriculture, Agricultural Research Service, Albany, California, USA;
4. Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Hangzhou, Zhejiang Province, China

† These authors contributed equally to this work.

* Corresponding author. Tel.: +8610 58900749; Fax: +8610 58900748

E-mail address: xiongyanwen@icdc.cn

Table S1. *E. coli*/*Shigella* genomes used in whole genome analysis.

Isolate	Pathovar	Accession number
EDL933	STEC ^a	NZ_CP008957.1
Sakai	STEC ^a	NC_002695.1
Xuzhou21	STEC ^a	NC_017906.1
EC4115	STEC ^a	NC_011353
TW14359	STEC ^a	NC_013008
2011C-3493	STEC ^b	NC_018658.1
12009	STEC ^c	NC_013353.1
11368	STEC ^d	NC_013361.1
RM13514	STEC ^e	CP006027.1
11128	STEC ^f	NC_013364.1
NRG 857C	AIEC	NC_017634
UM146	AIEC	NC_017632
042	EAEC	NC_017626
55989	EAEC	CU928145
CB9615	EPEC	NC_013941
E2348/69	EPEC	NC_011601
UMNK88	ETEC	NC_017641
E24377A	ETEC	NC_009801
IAI39	UPEC	NC_011750
CFT073	UPEC	NC_004431
S88	NMEC	NC_011742
APEC O1	APEC	NC_008563
SE11	Commensal	NC_011415
HS	Commensal	NC_009800
MG1655	K12	NC_000913
W	Lab	NC_017635
ATCC 8739	Lab	NC_010468
BL21(DE3)	Lab	NC_012971
<i>Shigella sonnei</i> Ss046	<i>Shigella</i>	NC_007384
<i>S. dysenteriae</i> Sd197	<i>Shigella</i>	NC_007606
<i>S. flexneri</i> 301	<i>Shigella</i>	NC_004337
<i>S. boydii</i> 227	<i>Shigella</i>	NC_007613

^aO157:H, ^bO104:H4, ^cO103:H2, ^dO26:H11, ^eO145:H28, ^fO111:H-

Table S2. Virulence traits of strain STEC299

* C, virulence genes on chromosome; p1, genes on plasmid pSTEC299-1; p2, genes on plasmid pSTEC299-2.

Gene	Description	Presence/Postion*
<i>stx</i>	Shiga-toxin 1	Negative
<i>stx2</i>	Shiga-toxin 2	Positive/C
<i>eae</i>	Intimin (adhesin)	Negative
<i>ehxA</i>	Enterohaemolysin A	Negative
<i>saa</i>	Shiga-toxin-producing <i>E. coli</i> autoagglutinationadhesin	Negative
<i>efa1</i>	Enterohaemorrhagic <i>E. coli</i> factor for adherence	Negative
<i>fuyA</i>	Yersiniabactinsiderophore (receptor)	Positive/C
<i>chuA</i>	hemin uptake-related factor	Positive/C
<i>ibeA</i>	Endothelial invasion	Positive/C
<i>tsh</i>	Temperature-sensitive haemagglutinin	Positive/C
<i>fimA</i>	Type I fimbriae	Positive/C
<i>yfcV</i>	Fimbrial protein	negative
<i>paa</i>	Adherence	p1
<i>gsp</i>	Type II secretion protein	p1
<i>pic</i>	Pic involved in the intestinal colonization	p2
<i>draC</i>	Dr family of adhesins	p1
<i>focC</i>	F1C fimbriae	C
<i>ecpA</i>	<i>E. coli</i> common pilus	C
<i>draC</i>	Dr family of adhesins	p1
<i>irp1</i>	Yersiniabactin biosynthetic protein	C
<i>irp2</i>	Yersiniabactin biosynthetic protein	C
<i>entA</i>	Iron chelator	C
<i>espC</i>	Enterotoxin gene	p1
<i>ecpA</i>	<i>E. coli</i> common pilus	C

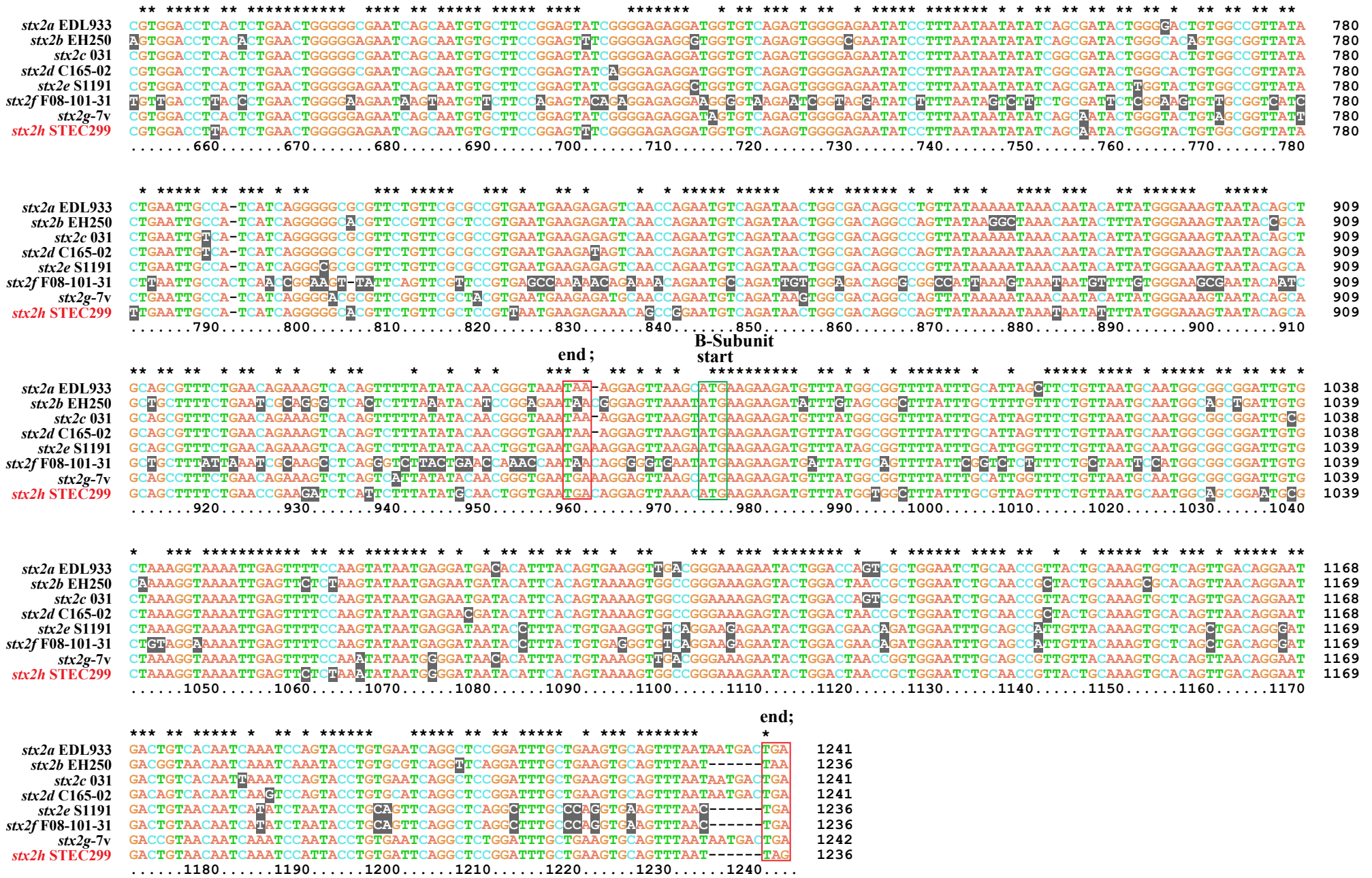


Figure S1. Nucleotide comparison of stx2h with other stx2 subtypes.
Nucleotide conserved with the stx2 sequences are indicated with an asterisk. Shaded letters represent nucleotide differences between the sequences of different stx2 subtypes. Nucleotides in green and red boxes indicate start and stop codon respectively for subunit A and B.



Figure S2. Phylogenetic analysis of Stx2 subtypes and variants by the neighbor-joining method

The neighbor-joining tree was inferred from comparison of combined A and B holotoxin amino acid sequences of Stx2h and 93 representatives of other known Stx2 subtypes and variants. Bar, 0.02 substitutions per site. Stx2 subtypes are indicated by different colors.

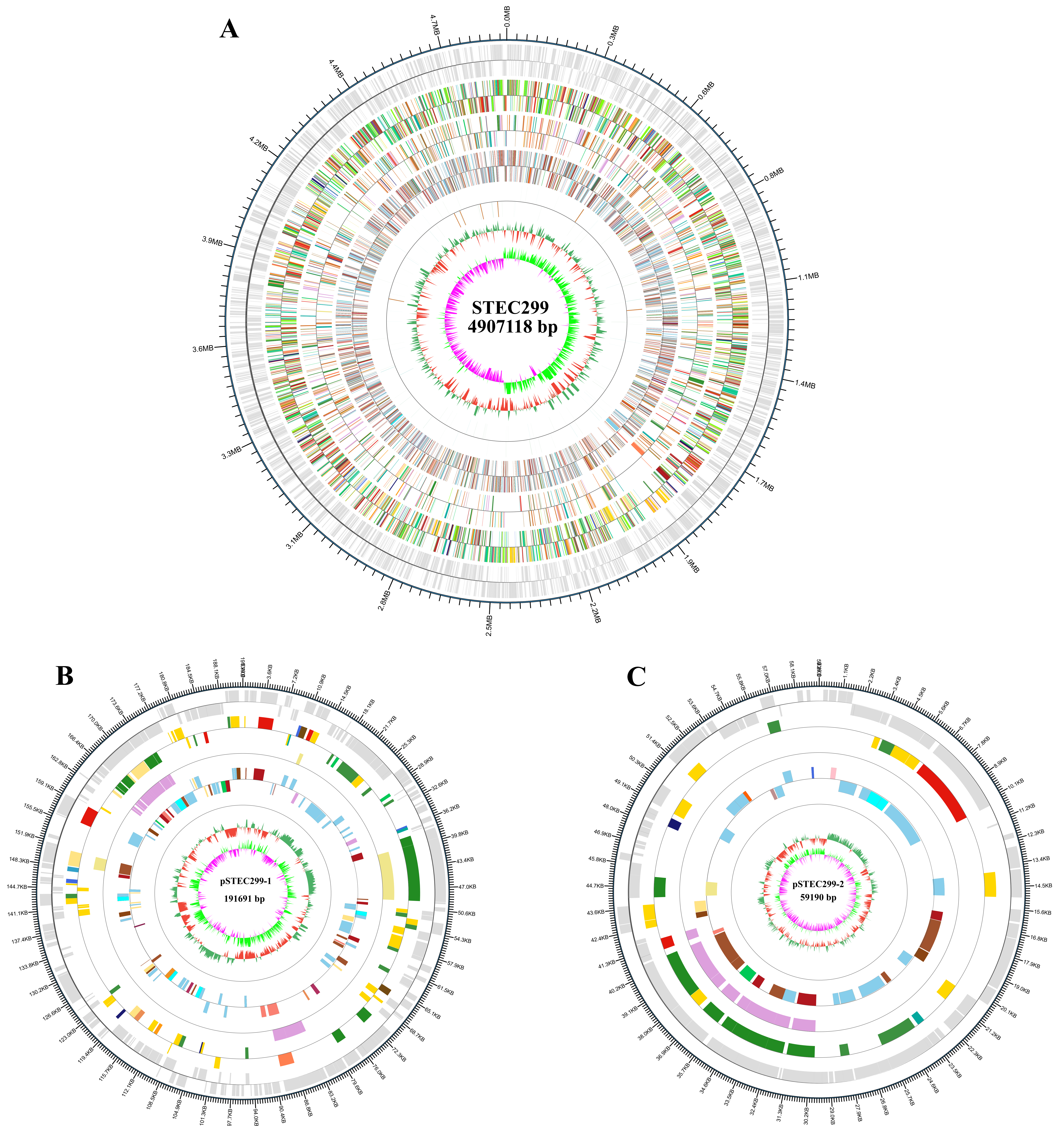


Figure S3. Genome features of STEC299.

Circular maps of the chromosome (A), and the two plasmids, pSTEC299-1 (B) and pSTEC299-2 (C). For both chromosome and plasmids, from outer circle to inner circle, each represents CDS (circle 1 and 2), COG annotation (circle 3 and 4), KEGG annotation (circle 5 and 6), GO annotation (circle 7 and 8), ncRNA (circle 9 and 10), GC Content (circle 11), GC Skew (circle 12). The two circles in one parentheses contains positive and negative strands respectively.