

Table S1. 16S rDNA sequences of four isolates 114, 318, 71 and 166.

Table S2. Comparison of the virulence genes among O157, K88, EcN, K12 and the *E. coli* 166 genomes.

Figure S1. Genome circle of *E. coli* 166 and LASTZ dot plots comparison. (A) The concentric circles show from outside to inside: 1st circle, Genomic sequence information. 2nd circle, Coding sequences (CDS) and non-coding RNA regions (rRNA, tRNA) from the reference genome. The outer layer represents the positive strand, and the inner layer represents the negative strand. 3rd circle, GC skew curve with a 2,000 bp sliding window. The dashed line represents GC skew = 0. 4th circle, GC content curve with a 2,000 bp sliding window. The dashed line represents the average GC content of the reference genome. **(B)** LASTZ (Large-Scale Genome Alignment Tool) analysis compare the consensus in genomes among *E. coli* 166 and other four *E. coli* reference strains.

Table S1. 16S rDNA sequences of four isolates 114, 318, 71 and 166.

>71

ACGGCCTGGGCGGCAGCTACACATGCAGTCGAACGGTAACAGGAAGCAGCTTGCTGCTTCGCTGACG
AGTGGCGGACGGGTGAGTAATGTCTGGGAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTA
GCTAATACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCGGATGTGCCC
AGATGGGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGA
TGACCAGCCACACTGGAAGTGAACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGC
ACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCCTTCGGGTTGTAAAGTACTT
TCAGCGGGGAGGAAGGGAGTAAAGTTAATACCTTTGCTCATTGACGTTACCCGCAGAAGAAGCACCG
GCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAA
AGCGCACGCAGGCGGTTTGTAAAGTCAGATGTGAAATCCCCGGGCTCAACCTGGGAAGTGCATCTGAT
ACTGGCAAGCTTGAGTCTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCT
GGAGGAATACCGTGGCGAAGGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGG
GAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGTGCACTTGAGAGTTGTGCCCTT
GAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAAC
TCAAATGAATTGACGGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGA
ACCTTACCTGGTCTTGACATCCACAGAACTTTCCAGAGATGGATTGGTGCCTTCGGGAAGTGTGAGAC
AGGTGCTGCATGGCTGTCGTCAGCTCGTGTGTGAAATGTTGGGTAAAGTCCCGCACGAGCGCACCCCT
ATCCTTTGTTGCCAGCGGTCCGCGGGACTCAAGAGACTGCCAGTGATAACCTGGAAGAAGGTGGGGA
AGACCGTCAGTCATCATGGCCTTACAACACAGGGCCTACACACTGCTAATGGGCATACAAAGAGAAGC
GCACCTCCGAAAGAAGCGGACCTCATTATGCGTCCTCTATAATCCCGATGA

>318

GCGGGCTGGGCGGGAGGCTACACATGCAGTCGAACGGTAACAGGAAGCAGCTTGCTGCTTCGCTGAC
GAGTGGCGGACGGGTGAGTAATGTCTGGGAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGT
AGCTAATACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCGGATGTGCC
CAGATGGGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGG
ATGACCAGCCACACTGGAAGTGAACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTG
CACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCCTTCGGGTTGTAAAGTACT
TTCAGCGGGGAGGAAGGGAGTAAAGTTAATACCTTTGCTCATTGACGTTACCCGCAGAAGAAGCACCC
GGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTA
AAGCGCACGCAGGCGGTTTGTAAAGTCAGATGTGAAATCCCCGGGCTCAACCTGGGAAGTGCATCTG
ATACTGGCAAGCTTGAGTCTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGAT
CTGGAGGAATACCGTGGCGAAGGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTG
GGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGTGCACTTGAGAGTTGTGCCC
TTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAA
ACTCAAATGAATTGACGGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAA
GAACCTTACCTGGTCTTGACATCCACAGAACTTTCCAGAGATGGATATGTGCCTTCGGGACTGTGAGA
CAGTGCTGCATGGCTGTCGTCAGCTCGTGTGTGAAATGTGGGTAAAGTCCGCACGAGCGCACCCCTTATC
TTTGTGTCAGCGGTCCGCGGGACTCAAGGGACTGCATGATAACTGGAGGAGGTGGGGATGACGTCAG
TCATCATGCTTACAACAGGCTACACATTGCTCAATGGCCATACAAGAGAGCGCACCCCTCGCGGGAGCA
CGGAACCTTTAAGGCGCGGATCCGAATGATCGCACTGATCCTGAATCGAATCTGTAATCGATACAAGTA
CAGCGGAATCGTTCCGGGCGTTGTAATCAG

>166

GCGGCGGGCGGCGGACTACACGTGCAGTCGAACGGTAACAGGGAAGCAGCTTGCTGCTTCGCTGACG
AGTGGCGGACGGGTGAGTAATGTCTGGGAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTA
GCTAATACCGCATAACGTGCGAAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCGGATGTGCCC
AGATGGGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGA
TGACCAGCCACACTGGAAGTGAACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGC
ACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTTGTAAAGTACTT
TCAGCGGGGAGGAAGGGAGTAAAGTTAATACCTTTGCTCATTGACGTTACCCGCGAGAAGAAGCACCG
GCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAA
AGCGCACGCAGGCGGTTTGTAAAGTCAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCTGAT
ACTGGCAAGCTTGAGTCTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCT
GGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGG
GAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGTCGACTTGGAGGTTGTGCCCTT
GAGGCGTGGCTTCGGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAAAAC
TCAATGAATTGACGGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTAATTTCGATGCAACGCGAAGA
ACCTTACCTGGTCTTGACATCCACAGAACTTTCCAGAGATGGATATGTGCCTTCGGGAACTGTGAGAC
AGTGCTGCATGGCTGTCGTCAGCTCGTGTGTGAAATGTTGGGTTAGTCCGCACGAGCGCACCCCTATC
CTTTGTTGCCAGCGGTCCGGCCGGGACTCAAGGAGACTGCCAGTGATAACTGGAGGAAGGGGGGGAT
GACGTCAGTCATCATGACCTACGAACAGGCTACCACACGTGCTACATGGCGCATACAAAGAGAAGCG
ACCTTCGCGGAAGACACGGACTTAAGTGCGCTGTATTCCGAAGTGGAAGTCTGCACTCGACCCTGATC
CGAATCGCTAAATCGAATCGAATGTACACGCGAATATGCTTTCCCGGCGCCGTGTTATATATACGCCGCT
>114

GCGGCTGGGCGGCGGCTACACATGCAGTCGAACGGTAACAGGAAGCAGCTTGCTGCTTCGCTGACGA
GTGGCGGACGGGTGAGTAATGTCTGGGAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAG
CTAATACCGCATAACGTGCGAAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCGGATGTGCCCCA
GATGGGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGAT
GACCAGCCACACTGGAAGTGAACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCA
CAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTTGTAAAGTACTTT
CAGCGGGGAGGAAGGGAGTAAAGTTAATACCTTTGCTCATTGACGTTACCCGCGAGAAGAAGCACCGG
CTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAA
GCGCACGCAGGCGGTTTGTAAAGTCAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCTGATA
CTGGCAAGCTTGAGTCTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTG
GAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGG
AGCAAACAGGATTAGATACCTGGTAGTCCACGCCGTAAACGATGTCGACTTGGAGGTTGTGCCCTTG
AGGCGTGGCTTCGGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAAACT
CAAATGAATTGACGGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTAATTTCGATGCAACGCGAAGAA
CCTTACCTGGTCTTGACATCCACAGAACTTTCCAGAGATGGATTGGTGCCTTCGGGACTGTGAGACAG
GTGCTGCATGGCTGTCGTCAGCTCGTGTGTGAAATGTTGGGTTAAGTCCCGCACGAGCGCACCCCTAT
CTTTGTTGCAGCGGTGCGGGGACTCAAGGAACTGCATGATAAACTGGAGGAAGGGTGGGGGATGACG
TCAGTCATCATGCTTACGAACAGGCTACACACTGCTACATGGCCATACAAAGAGAGCGACCTCGCGAG
ACACGAACCTTAATGGCGCTGGATCCGAGTGAGTGCACCTATCTCTGATCGGAGCTCTGTAATCGTACA
AGTGACCGGGTATACTATATCCCGCCTGTTATATCACGCGCGTCA

Virulence factors		Related genes	Table S2. Comparison of the virulence genes among O157, K88, EeN, K12 and the <i>E. coli</i> 166 genomes.				<i>E. coli</i> 166
			<i>E. coli</i> O157:H7 str. EC4115 (EHEC) chromosome NC_011353 (5572075 bp)	<i>E. coli</i> UMNK88 (EPEC) chromosome NC_017641 (5186416 bp)	<i>E. coli</i> str. K-12 substrate MG1655 chromosome NC_000913 (4639675 bp)	<i>E. coli</i> Nissle 1917 chromosome NZ_CP007799 (5441200 bp)	
Adherence							
AIDA-1 type	tfaA					Y	
Curli fibers	csgG					Y	
	csgH					Y	
	csgD					Y	
	csgB					Y	
	csgA					Y	
	csgC					Y	
E.coli hemin-binding fimbriae (HBF)	eHA	Y	Y	Y		Y	
	eHD	Y	Y	Y		Y	
	eHC		Y	Y		Y	
	eHG		Y	Y		Y	
EACP	eacH	Y	Y				
	yagK/csgR	Y	Y	Y	Y		
	yagP/wspA	Y	Y	Y	Y	Y	
	yagV/csgB	Y	Y	Y	Y	Y	
	yagX/csgC	Y	Y	Y	Y	Y	
	yagW/csgD	Y	Y	Y	Y	Y	
	yagV/csgE	Y		Y			
EhaA, AIDA-1 type	ehaA	Y					
EhaB, AIDA-1 type	ehaB		Y				
F1C fimbriae	focA				Y	Y	
	focG				Y		
	focH				Y		
Hemorrhagic E.coli pilus (HEP)	hepA	Y	Y	Y			
	hepB	Y	Y	Y			
	hepC	Y	Y	Y			
	papX				Y		
P fimbriae	papD					Y	
	papC					Y	
	papB				Y		
	papF				Y		
Paa	paa	Y					
	paaC						
S fimbriae	sfaD					Y	
	sfaF					Y	
	sfaG					Y	
Type 1 fimbriae	fimB	Y		Y	Y		
	fimD	Y		Y	Y		
	fimA	Y	Y	Y	Y		
	fimT	Y		Y	Y		
	fimC	Y		Y	Y	Y	
	fimD	Y		Y	Y	Y	
	fimF	Y		Y	Y	Y	
	fimG	Y		Y	Y	Y	
	fimH	Y		Y	Y	Y	
	UspG adhesin, trimeric AT	uspG/chaG	Y				
Invasion							
IbcA	ibcB	Y	Y				
	ibcC	Y	Y				
Effector delivery system							
ACE: T8SS	ace32		Y				
	ace31	Y	Y				
	ace30		Y				
	ace29	Y	Y				
	ace28	Y	Y				
	ace27/espV	Y	Y				
	ace26	Y	Y				
	ace25	Y	Y				
	ace24	Y	Y				
	ace23	Y	Y				
	ace22	Y	Y				
	-	Y	Y				
	ace19	Y	Y				
	ace18	Y	Y				
	ace17	Y	Y				
	ace16	Y	Y				
	ace15	Y	Y				
	Intimin	esc	Y				
escF		Y					
escD2		Y					
espB		Y					
espD		Y					
espA		Y					
espE		Y					
escD		Y					
escT		Y					
escF		Y					
espQ/escQ		Y					
escO		Y					
escN		Y					
escV		Y					
escL		Y					
escJ		Y					
espD		Y					
escC		Y					
T1SS		ghrR	Y				
	etgA	Y			Y	Y	
	escU	Y					
	escT	Y					
	escS	Y					
	escR	Y					
	escL	Y					
	escK	Y					
	ceaAB	Y					
	escE	Y					
	roaT	Y					
	espB	Y					
	espG	Y					
	espH	Y					
	espF	Y					
	tr	Y					
	mmp	Y			Y	Y	
	espA/escP	Y					
	espJ	Y					
nleA/espI	Y						
nleC	Y						
T2SS secreted effectors	espK	Y					
	espL1	Y					
	espL2	Y					
	espL4	Y					
	espM1	Y					
	espM	Y					
	espR1	Y					
	espR3	Y					
	espR4	Y					
	espW	Y					
	espX1	Y					
	espX2	Y					
	espX4	Y					
	espX5	Y					
	espX6	Y					
	espX7/nleL	Y					
	espY2	Y					
	espY3	Y					
	espY4	Y					
	nleB1	Y					
	nleB2	Y					
	nleE	Y					
	nleG-1	Y					
	nleG2-3	Y					
	nleG2-4	Y					
	nleG-3	Y					
	nleG5-1	Y					
	nleG5-2	Y					
	nleG6-2	Y					
nleG6-3	Y						
nleG7	Y						
nleH1	Y						
nleH2	Y						
Vacuolating autotransporter gene Vat	vat				Y		
Toxins							
α-hemolysin	hlyD				Y		
β-hemolysin	hlyE/hlyA	Y			Y	Y	
Stx	stx2A	Y					
	stx2B	Y					
Biofilms							
Antigen 43, AIDA-1 type	aga43			Y	Y		
CalA, AIDA-1 type	calA	Y	Y		Y		
Neutrophil/leukocyte factor							
Actobactin	batA				Y		
	chsS	Y			Y		
	chsA	Y			Y		
	chsT	Y			Y		
	chsW	Y					
	chsX	Y					
	chsV	Y			Y		
Chu	chuL	Y					
	chuA				Y		
Iron/manganese transport	itfB				Y		
	itfC				Y		
	itfD				Y		
	itfN				Y		
Salmonella siderophore	iroB				Y		
	iroD				Y		
Yersinia siderophore	iroC				Y		
	iroB				Y		
	ybaS				Y		
	ybaX				Y		
	ybaQ				Y		
	ybaP				Y		
	ybaA				Y		
	irp2				Y		
	irp1				Y		
	ybaU				Y		
Ler	ybaT				Y		
	ybaE				Y		
	fyxA				Y		
	ler	Y					
Others							

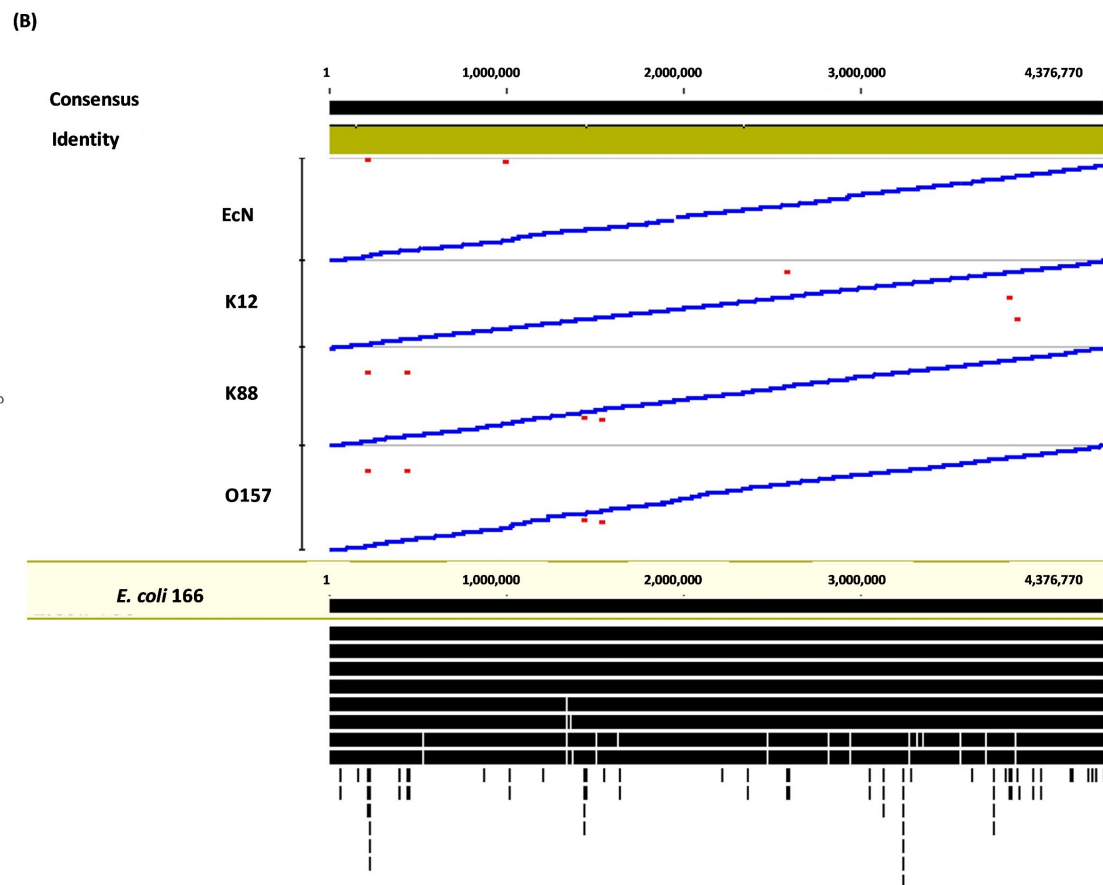
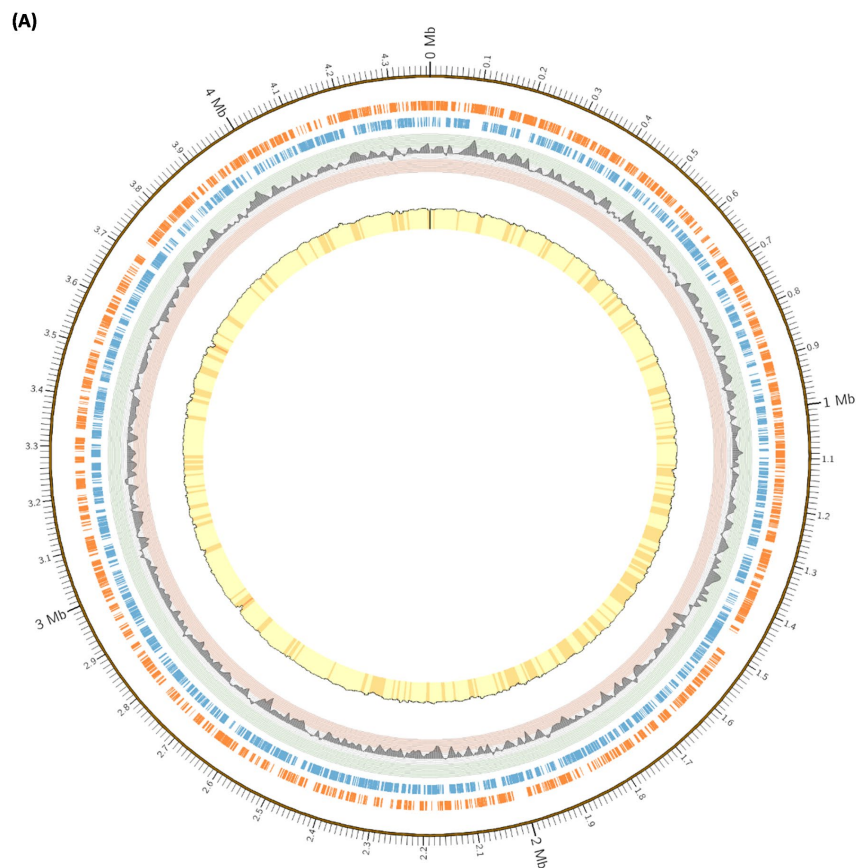


Figure S1. Genome circle of *E. coli* 166 and LASTZ dot plots comparison. (A) The concentric circles show from outside to inside: 1st circle, Genomic sequence information. 2nd circle, Coding sequences (CDS) and non-coding RNA regions (rRNA, tRNA) from the reference genome. The outer layer represents the positive strand, and the inner layer represents the negative strand. 3rd circle, GC skew curve with a 2,000 bp sliding window. The dashed line represents GC skew = 0. 4th circle, GC content curve with a 2,000 bp sliding window. The dashed line represents the average GC content of the reference genome. (B) LASTZ (Large-Scale Genome Alignment Tool) analysis compare the consensus in genomes among *E. coli* 166 and other four *E. coli* reference strains.