

**Table S1. General genomic features of O26, O111 and O103 EHECs, O157 EHEC and 21 other, fully sequenced *E. coli*/Shigella strains**

Strain	Phylogroup	Pathotype	Serotype	Chromosome				Plasmid			No. of CDS family	Reference
				Size (kb)	CDS*	tRNA	Accession number	Size (kb)	CDS in total <sup>a</sup>	Accession number		
O157 Sakai†	E	EHEC	O157:H7	5,498	5,363 (247)	105	BA000007	93/3	95 (11)	AB011548-9	4,807	(1)
O26_11368	B1	EHEC	O26:H11	5,697	5,609 (254)	101	AP010953	85/63/6/4	186 (31)	AP010954-7	4,946	This study
O111_11128	B1	EHEC	O111:H-	5,371	5,264 (291)	106	AP010960	205/98/78/8/7	468 (33)	AP010961-5	4,965	This study
O103_12009	B1	EHEC	O103:H2	5,449	5,264 (209)	98	AP010958	72	90 (21)	AP010959	4,705	This study
E24377A	B1	EPEC	O139:H28	4,979	4,873 (118)	91	CP000800	79/74/70/34/6/5	242 (34)	CP000795-9, 801	4,760	(2)
SE11	B1	commensal	O152:H28	4,888	4,679	86	AP009240	100/91/61/7/5/4	323	AP009241-6	4,821	(3)
IAI1	B1	commensal	O8	4,701	4,491 (51)	94	CU928160	-	-	-	4,267	(4)
55989	B1	EAEC	-	5,155	4,969 (79)	86	CU928145	72	100 (7)	CU928159	4,635	(4)
MG1655	A	laboratory	-	4,639	4,294 (101)	88	U00096	-	-	-	4,033	(5)
HS	A	commensal	O9	4,643	4,478 (94)	88	CP000802	-	-	-	4,248	(2)
E2348/69	B2	EPEC	O127:H6	4,965	4,703 (145)	92	FM180568	97/6	121 (23)	FM180569-70	4,497	(6)
CFT073	B2	ExPEC	O6:K2:H1	5,231	5,473 (94)	89	AE14075	-	-	-	5,101	(7)
536	B2	ExPEC	O6:K15:H7	4,938	4,685	81	CP000247	-	-	-	4,457	(8)
UTI89	B2	ExPEC	-	5,065	5,066	88	CP000243	114	145	CP000244	5,073	(9)
APEC O1	B2	ExPEC	O1:K1:H7	5,082	4,467	94	CP000468	241/174	423	DQ517526, DQ381420	4,730	(10)
S88	B2	ExPEC	O45:K1:H7	5,032	4,859 (90)	91	CU928161	134	144 (9)	CU928146	4,666	(4)
ED1a	B2	commensal	O81	5,210	5,129 (95)	91	CU928162	120	150 (11)	CU928147	4,634	(4)
IAI39	D	ExPEC	O7:K1	5,132	4,906 (80)	88	CU928164	-	-	-	4,432	(4)
UMN026	D	ExPEC	O17:K52:H18	5,202	4,918 (45)	88	CU928163	122/34	198 (8)	CU928148-9	4,867	(4)
<i>S. boydii</i> Sb227	S1	shigella	4	4,520	4,353 (217)	91	CP000036	127	149	CP000037	3,550	(11)
<i>S. sonnei</i> Ss046	SS	shigella	-	4,825	4,434 (210)	97	CP000038	214/8/5/2	252	CP000039, CP000641-3	3,819	(11)
<i>S. flexneri</i> Sf301	S3	shigella	2a	4,607	4,434 (254)	97	AE005674	222	267	AF386526	3,729	(12)
<i>S. flexneri</i> Sf2457T	S3	shigella	2a	4,599	4,073 (372)	98	AE014073	-	-	-	3,514	(13)
<i>S. flexneri</i> Sf8401	S3	shigella	5b	4,574	4,316 (198)	97	CP000266	-	-	-	3,577	(14)
<i>S. dysenteriae</i> Sd197	SD1	shigella	1	4,369	4,557 (285)	85	CP000034	183/9	231	CP000035, CP000640	3,371	(11)

\*Total numbers of CDSs including pseudogenes are shown. Numbers of pseudogenes are indicated in parentheses if available.

†Another O157:H7 strain (EDL933) has been sequenced (15).

Reference list:

- Hayashi T, et al. (2001) Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* 8:11-22.
- Rasko DA, et al. (2008) The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J Bacteriol* 190:6881-6893.
- Oshima K, et al. (2008) Complete genome sequence and comparative analysis of the wild-type commensal *Escherichia coli* strain SE11 isolated from a healthy adult. *DNA Res* 15:375-386.
- Touchon M, et al. (2009) Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* 5:e1000344.

5. Blattner FR, et al. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1453-1474.
6. Iguchi A, et al. (2008) The complete genome sequence and comparative genome analysis of enteropathogenic *E. coli* O127:H6 strain E2348/69. *J Bacteriol* 191:347-354.
7. Welch RA, et al. (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* 99:17020-17024.
8. Brzuszkiewicz E, et al. (2006) How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains. *Proc Natl Acad Sci U S A* 103:12879-12884.
9. Chen SL, et al. (2006) Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. *Proc Natl Acad Sci U S A* 103:5977-5982.
10. Johnson TJ, et al. (2007) The genome sequence of avian pathogenic *Escherichia coli* strain O1:K1:H7 shares strong similarities with human extraintestinal pathogenic *E. coli* genomes. *J Bacteriol* 189:3228-3236.
11. Yang F, et al. (2005) Genome dynamics and diversity of *Shigella* species, the etiologic agents of bacillary dysentery. *Nucleic Acids Res* 33:6445-6458.
12. Jin Q, et al. (2002) Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K-12 and O157. *Nucleic Acids Res* 30:4432-4441.
13. Wei J, et al. (2003) Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. *Infect Immun* 71:2775-2786.
14. Nie H, et al. (2006) Complete genome sequence of *Shigella flexneri* 5b and comparison with *Shigella flexneri* 2a. *BMC Genomics* 7:173.
15. Perna NT, et al. (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409:529-533.