Table S5. Identification of EHEC/EPEC-specific genes and their locational and functional classifications

Category	No. of CDS families*										
	Total	Location			Predicted function						
		Backbone	Prophages and integrative elements	Plasmids	Virulence-related	Phage function	Plasmid function	O antigen	Fimbria	Others	Unknown
EHEC/EPEC-specific†	1,761 (427)										
Present in all the EHECs‡	87 (34)	0	79 (34)	8	51 (28)	8 (2)	0	0	0	4	24 (4)
O157-specific	314	118	172	24	29	65	2	10	27	29	152
O26-specific	175	21	120	34	7	29	15	8	0	7	109
O111-specific	430	18	93	319	7	19	108	17	0	29	250
O103-specific	174	16	138	20	20	37	2	11	0	6	98
E2348/69-specific	353	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
Others	228 (40)	22 (2)	197 (37)	9 (1)	59 (18)	47 (5)	2 (1)	0	0	13 (1)	107 (15)
Present in all the *E. coli* strains	1,919	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
Absent in EHEC strains	6,318	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
Others	2,942	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
Total	12,940	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd

* All CDSs from the 25 fully sequenced *E. coli* and *Shigella* strains were grouped into 12,940 CDS families based on the results from BLASTP analysis. The numbers of CDS families of each strain were indicated in Table S1.

† CDS families that are present in at least one of the four EHEC and the EPEC (E2348/69) strains. The numbers of CDS families conserved in EPEC E2348/69 are indicated in parentheses.

‡ ECs numbers of the 87 genes (CDS families) that are present in all EHECs (O157, O26, O111 and O103) are listed below. Predicted functions of each CDS family are shown in parentheses. Best-hit homologues of each CDS family are also shown in parentheses (bacterial species and strains and amino acid sequence identities) if homologues showing ≥50% amino acid sequence identity (alignment length; ≥60% of the length of query sequence) were found in bacerial species/strains other than EHEC/EPEC strains in the nr database search.

Virlence-related: ECs0814 (systemic factor protein SfpA; *Salmonella* Typhi CT18, 72%) ECs0848 (T3SS secreated effector NleH1-1; *Citrobacter rodentium*, 83%) ECs1091/2182/2737 (transcriptional regulator PchABC family; no hit) ECs1322 (urease gamma subunit UreA; *Klebsiella pneumoniae* MGH 78578, 96%)□□ECs1323 (urease beta subunit UreB; *K. pneumoniae* 342, 82%) ECs1324 (urease alpha subunit UreC; *K. pneumoniae* MGH 78578, 90%) ECs1325 (urease accessory protein UreE; *K. pneumoniae* 342, 84%) ECs1326 (urease accessory protein UreF; *K. pneumoniae* MGH 78578, 79%) ECs1327 (urease accessory protein UreG; *Citrobacter* sp. 30_2, 91%) ECs1560 (T3SS secreated effector EspX7; no hit) ECs1561 (T3SS secreated effector EspN; no hit) ECs1568 (T3SS secreated effector EspK; no hit) ECs1772 (porcine attaching-effacing associated protein Paa/adherence factor AdfO; *E. coli* ECL8389, 93%) ECs1995/2155 (T3SS secreated effector NleG family; no hit) ECs2226 (T3SS secreated effector NleG7; no hit) ECs2715 (T3SS secreated effector TccP/EspFU / /EspF2-2; no hit) ECs3486 (T3SS secreated effector NleG8-2; no hit) ECs3487 (T3SS secreated effector EspW; no hit) ECs3855 (T3SS secreated effector EspL2; no hit) ECs3857 (T3SS secreated effector NleB1; *C. rodentium*, 89%) ECs3858 (T3SS secreated effector NleE, *C. rodentium*, 86%) ECs4552 (T3SS structure protein EscF; *E. albertii* TW07627, 98%) ECs4553 (chaperone CesD2; *E. albertii* TW07627, 91%) ECs4557 (secretion switching protein SepL; *C. rodentium*, 89%)□□ECs4558 (T3SS structure protein EscD; *E. albertii* TW07627, 94%) ECs4560 (chaperone CesT; *E. albertii* TW07627, 95%) ECs4568 (translocator EscN; *E. albertii* TW07627, 95%) ECs4569 (translocator EscV; *E. albertii* TW07627, 95%) ECs4570 (regulator Mpc; *E. albertii* TW07627, 96%) ECs4572 (chaperone of T3SS Rorf8; *C. rodentium*, 77%) ECs4573 (T3SS structure protein EscJ; *E. albertii* TW07627, 97%) ECs4574 (secretion switching protein SepD; *C. rodentium*, 90%) ECs4575 (T3SS structure protein EscC; *E. albertii* TW07627, 97%) ECs4576 (chaperone CesD; *E. albertii* TW07627, 96%) ECs4577 (positive regulator GrlA; *C. rodentium*, 91%) ECs4578 (negative regulator GrlR; *C. rodentium*, 87%) ECs4580 (T3SS structure protein EscU; *E. albertii* TW07627, 94%) ECs4581 (T3SS structure protein EscT; *E. albertii* TW07627, 93%) ECs4582 (T3SS structure protein EscS; *E. albertii* TW07627, 95%) ECs4583 (T3SS structure protein EscR; *E. albertii* TW07627, 96%) ECs4584 (T3SS machinery component; *C. rodentium*, 76%) ECs4585 (T3SS machinery component; *E. albertii* TW07627, 92%) ECs4587 (T3SS machinery component; *E. albertii* TW07627, 93%) ECs4588 (transcription regulator Ler; *C. rodentium*, 89%)□□ECsL17 (Hemolysin C; *E. coli* 536, 67%) ECsL18 (Hemolysin A; *E. coli* UTI89, 62%) ECsL19 (Hemolysin B; *E. coli* UTI89, 72%) ECsL20 (Hemolysin D; *E. coli* CFT073, 57%) ECsL88 (polysaccharide deacetylase; *E. fergusonii* ATCC 35469, 70%) ECsL89 (LPS -1,7-N-acetylglucosamine transferase; *E. fergusonii* ATCC 35469, 59%) ECsL91 (lipid A myristoyl transferase; *Citrobacter* sp. 30_2, 67%)

Phage function: ECs0815 (late gene regulator of lambda-like phage; *E. coli* SMS-3-5, 74%) ECs1084 (antitermination protein of lambda-like phage; *E. coli* 536, 86%) ECs1778 (late gene regulator of lambda-like phage; *E. coli* ED1a, 88%) ECs2183/2738 (lipoprotein precursor of lambda-like phage; *E. coli* MG1655, 72%) ECs2166/2725/2949 (tail length tape measure protein of lambda-like phage; *S. boydii* CDC 3083-94, 97%) ECs2168/2951 (minor tail protein of lambda-like phage; *S. boydii* CDC 3083-94, 96%) ECs1092/1216/1787/2256/2965/5623 (lipoprotein precursor of lambda-like phage; *E. coli* BL21, 72%) ECs1214/1533/1785/1965/2258/2967 (antirepressor of lambda-like phage; Acyrthosiphon pisum, T5A, 69%)

Others: ECs1316 (diacylglycerol kinase; *C. koseri*, 75%) ECs1332 (colicin immunity protein; no hit) ECs2153/2218 (DNA damage-inducible protein; E. coli ED1a, 76%) ECs0813/3502 (serine/threonin protein phosphatase; *S. flexneri* phage Sf6, 98%)

Unknown: ECs1085 (Providencia rettgeri DSM 1131, 50%) ECs1097 (no hit) ECs1315 (no hit) ECs1319 (no hit) ECs1320 (no hit) ECs1331 (no hit) ECs1333 (*E. coli* ED1a, 83%) ECs1342 (*E. coli* CFT073, 68%)□□ECs1774 (*E. coli* SMS-3-5, 97%) ECs2196 (*S.* Typhi 404ty, 53%) ECs1244/2197 (Enterobacter sakazakii ATCC BAA-894, 53%) ECs2199 (*E. coli* SE11, 55%) ECs2200 (*E. coli* SE11, 76%) ECs2202 (*S.* Typhimurium LT2, 56%) ECs2203 (*E. coli* 101-1, 78%) ECs2213 (no hit) ECs1816/2221/2255 (Edwardsiella ictaluri 93-146, 96%) ECs1210/2971 (no hit) ECs3489/2716/1809/2230/2157/2940/1229/0845/1993/1124 (*E. albertii* TW07627, 93%) ECs3500/0816 (*E. albertii* TW07627, 97%) ECs4567 (*E. albertii* TW07627, 92%) ECs4579 (*C. rodentium*, 90%) ECs5637/5627/5625/1960/2748/5614 (*E. albertii* TW07627, 56%) ECsL90 (*E. coli* SMS-3-5, 81%)