

Supplementary Table S1. Comparative overview of sequenced *E. coli* genomes

Type Strain ECOR group	Non-pathogen						
	Commensal		Laboratory				Environmental
	SE11 B1	HS A	K-12 MG1655 A	K-12 W3110 A	K-12 DH10B A	ATCC 8739 A	SMS-3-5 D
Chromosome size (bp)	4,887,515	4,643,538	4,639,675	4,646,332	4,686,137	4,746,218	5,068,389
GC content (%)	50.8	50.8	50.8	50.0	50.8	50.9	50.5
Protein-coding genes	4,679	4,384	4,133	4,226	4,126	4,200	4,743
tRNA gene	86	88	87	86	86	87	90
rRNA gene	22	22	22	22	22	23	22
Plasmid	6	None	None	None	None	None	4

Pathotype Strain ECOR group	Pathogen						
	EHEC		ETEC	UPEC			APEC
	O157 Sakai E	O157 EDL933 E	E24377A B1	CFT073 B2	UTI89 B2	536 B2	O1 B2
Chromosome size (bp)	5,498,450	5,528,445	4,979,619	5,231,428	5,065,741	4,938,920	5,082,025
GC content (%)	50.5	50.5	50.6	50.5	50.6	50.5	50.5
Protein-coding genes	5,253	5,324	4,755	5,379	5,044	4,629	4,457
tRNA gene	103	100	91	88	88	81	72
rRNA gene	22	22	21	22	22	22	22
Plasmid	2	1	6	None	1	None	4

Supplementary Table S2. IS elements in the SE11 genome

Name	Copy number*
IS1	1 (0)
IS1 (nuxi)	2 (1)
IS2	1 (1)
IS3	2 (0)
IS602	1 (1)
IS629	1 (1)
IS911	1 (1)
IS186A	1 (1)
IS608	1 (1)
IS5	1 (0)
IS26	2 (0)
ISCro1	1 (1)
ISSfl4	1 (0)
IS621	7 (0)
IS609	3 (2)
IS1414	1 (1)
IS604	1 (1)
IS677	10 (8)
ISEc25	2 (1)
IS3 family	4 (4)
IS66 family	2 (2)
IS91 family	1 (1)
IS110 family	1 (1)
IS200/IS605 family	1 (1)
IS 256 family	1 (1)
IS481 family	1 (1)
ISAs1 family	1 (1)
Total	52 (33)

* Numbers of truncated and deleted copies are shown in parentheses.

Supplementary Table S3. The 356 conserved hypothetical genes and 284 genes with assigned function that are present in SE11 but absent in MG1655

Gene ID	Gene product	Presence in E24377A*
ECSE_0018	conserved hypothetical protein	+
ECSE_0019	conserved hypothetical protein	-
ECSE_0020	conserved hypothetical protein	+
ECSE_0030	conserved hypothetical protein	-
ECSE_0031	conserved hypothetical protein	-
ECSE_0034	conserved hypothetical protein	+
ECSE_0081	conserved hypothetical protein	+
ECSE_0150	conserved hypothetical protein	+
ECSE_0194	conserved hypothetical protein	-
ECSE_0239	conserved hypothetical protein	-
ECSE_0249	conserved hypothetical protein	+
ECSE_0251	putative motility protein	+
ECSE_0252	flagellar biosynthetic protein	+
ECSE_0255	conserved hypothetical protein	+
ECSE_0256	peptide chain release factor	+
ECSE_0316	putative invasins	+
ECSE_0317	putative transcription regulator	+
ECSE_0318	putative dehydrogenase	+
ECSE_0327	putative autotransporter	-
ECSE_0328	conserved hypothetical protein	-
ECSE_0329	conserved hypothetical protein	-
ECSE_0330	conserved hypothetical protein	-
ECSE_0345	conserved hypothetical protein	+
ECSE_0346	putative sugar ABC transporter substrate binding component	+
ECSE_0347	putative sugar ABC transporter ATP-binding component	+
ECSE_0348	putative sugar ABC transporter permease component	+
ECSE_0349	putative sugar ABC transporter permease component	+
ECSE_0366	putative truncated cyanate transport protein	-
ECSE_0384	putative transferase	+
ECSE_0392	conserved hypothetical protein	-
ECSE_0393	putative autotransporter	+
ECSE_0396	conserved hypothetical protein	+
ECSE_0402	conserved hypothetical protein	+
ECSE_0413	conserved hypothetical protein	+
ECSE_0421	conserved hypothetical protein	+
ECSE_0432	conserved hypothetical protein	+
ECSE_0450	conserved hypothetical protein	+
ECSE_0451	conserved hypothetical protein	+
ECSE_0452	conserved hypothetical protein	+
ECSE_0461	conserved hypothetical protein	+
ECSE_0525	putative truncated Rhs core protein	-
ECSE_0526	putative truncated Rhs core protein	+
ECSE_0527	conserved hypothetical protein	-
ECSE_0528	conserved hypothetical protein	-
ECSE_0544	conserved hypothetical protein	+
ECSE_0626	conserved hypothetical protein	+
ECSE_0627	conserved hypothetical protein	+
ECSE_0628	conserved hypothetical protein	-
ECSE_0629	putative truncated Rhs core protein	-
ECSE_0630	conserved hypothetical protein	-
ECSE_0631	Rhs core protein	+
ECSE_0632	conserved hypothetical protein	+
ECSE_0633	conserved hypothetical protein	+
ECSE_0646	conserved hypothetical protein	-
ECSE_0677	conserved hypothetical protein	+
ECSE_0693	conserved hypothetical protein	+
ECSE_0700	conserved hypothetical protein	+
ECSE_0719	putative truncated dnaK protein	-
ECSE_0726	conserved hypothetical protein	+
ECSE_0727	conserved hypothetical protein	+
ECSE_0742	conserved hypothetical protein	+
ECSE_0753	conserved hypothetical protein	+
ECSE_0762	putative truncated Rhs core protein	+
ECSE_0765	conserved hypothetical protein	-
ECSE_0779	conserved hypothetical protein	-
ECSE_0792	conserved hypothetical protein	-
ECSE_0841	conserved hypothetical protein	-
ECSE_0853	conserved hypothetical protein	+
ECSE_0871	conserved hypothetical protein	-
ECSE_0873	conserved hypothetical protein	+
ECSE_0941	conserved hypothetical protein	-
ECSE_0956	conserved hypothetical protein	+

ECSE_1019	conserved hypothetical protein	+
ECSE_1033	conserved hypothetical protein	-
ECSE_1077	conserved hypothetical protein	-
ECSE_1081	putative cytochrome	+
ECSE_1084	conserved hypothetical protein	+
ECSE_1093	conserved hypothetical protein	+
ECSE_1103	conserved hypothetical protein	+
ECSE_1111	conserved hypothetical protein	+
ECSE_1167	conserved hypothetical protein	+
ECSE_1187	conserved hypothetical protein	-
ECSE_1213	conserved hypothetical protein	+
ECSE_1214	conserved hypothetical protein	+
ECSE_1215	putative autotransporter	+
ECSE_1219	putative truncated autotransporter	+
ECSE_1223	conserved hypothetical protein	-
ECSE_1229	putative truncated cytolysin	-
ECSE_1252	conserved hypothetical protein	+
ECSE_1266	conserved hypothetical protein	+
ECSE_1267	conserved hypothetical protein	-
ECSE_1276	conserved hypothetical protein	+
ECSE_1331	conserved hypothetical protein	+
ECSE_1370	putative ABC transporter ATP-binding component	+
ECSE_1379	conserved hypothetical protein	+
ECSE_1380	conserved hypothetical protein	+
ECSE_1462	conserved hypothetical protein	+
ECSE_1492	putative truncated phosphatidate cytidyltransferase	-
ECSE_1499	glyceraldehyde-3-phosphate dehydrogenase	+
ECSE_1502	conserved hypothetical protein	-
ECSE_1518	conserved hypothetical protein	+
ECSE_1519	conserved hypothetical protein	-
ECSE_1527	putative acetyltransferase	-
ECSE_1539	conserved hypothetical protein	+
ECSE_1540	conserved hypothetical protein	+
ECSE_1541	Rhs core protein	+
ECSE_1542	conserved hypothetical protein	+
ECSE_1543	H repeat-associated protein	-
ECSE_1544	H repeat-associated protein	-
ECSE_1545	conserved hypothetical protein	-
ECSE_1547	Rhs core protein	-
ECSE_1548	conserved hypothetical protein	-
ECSE_1549	H repeat-associated protein	-
ECSE_1560	putative outer membrane porin protein	+
ECSE_1562	truncated formate dehydrogenase-N alpha subunit	-
ECSE_1590	conserved hypothetical protein	-
ECSE_1595	putative fimbrial usher protein	-
ECSE_1596	putative fimbrial chaperone	+
ECSE_1597	putative fimbrial protein	-
ECSE_1623	6-phospho-beta-glucosidase	+
ECSE_1624	conserved hypothetical protein	+
ECSE_1625	PTS system cellobiose-specific IIA component	+
ECSE_1626	PTS system cellobiose-specific IIC component	+
ECSE_1627	PTS system cellobiose-specific IIB component	+
ECSE_1628	putative transcriptional regulator	+
ECSE_1639	putative transport protein	+
ECSE_1737	glucuronide transport protein	+
ECSE_1756	conserved hypothetical protein	-
ECSE_1778	conserved hypothetical protein	+
ECSE_1797	conserved hypothetical protein	+
ECSE_1820	conserved hypothetical protein	+
ECSE_1827	conserved hypothetical protein	+
ECSE_1888	conserved hypothetical protein	-
ECSE_1889	conserved hypothetical protein	-
ECSE_1900	conserved hypothetical protein	-
ECSE_1919	conserved hypothetical protein	+
ECSE_1967	conserved hypothetical protein	+
ECSE_2028	conserved hypothetical protein	-
ECSE_2137	conserved hypothetical protein	+
ECSE_2145	conserved hypothetical protein	+
ECSE_2163	conserved hypothetical protein	-
ECSE_2167	conserved hypothetical protein	-
ECSE_2194	putative porin	-
ECSE_2266	conserved hypothetical protein	+
ECSE_2271	conserved hypothetical protein	-
ECSE_2278	conserved hypothetical protein	+
ECSE_2279	putative propanediol utilization protein	+
ECSE_2304	putative glycosyl transferase	-

ECSE_2305	putative glycosyltransferase	-
ECSE_2307	putative glycosyltransferase	-
ECSE_2308	phosphomannomutase	-
ECSE_2309	GDP-mannose pyrophosphorylase	-
ECSE_2310	GDP-mannose mannosyl hydrolase	-
ECSE_2311	GDP-mannose-4-keto-6-D epimerase	-
ECSE_2312	GDP-mannose 4,6-dehydratase	-
ECSE_2313	putative glycosyltransferase	-
ECSE_2354	conserved hypothetical protein	+
ECSE_2355	conserved hypothetical protein	+
ECSE_2359	galactitol utilization operon repressor	+
ECSE_2384	conserved hypothetical protein	+
ECSE_2390	conserved hypothetical protein	+
ECSE_2406	putative channel protein	+
ECSE_2423	conserved hypothetical protein	+
ECSE_2438	conserved hypothetical protein	+
ECSE_2458	putative truncated autotransporter	-
ECSE_2459	putative autotransporter	+
ECSE_2489	conserved hypothetical protein	+
ECSE_2499	conserved hypothetical protein	+
ECSE_2502	conserved hypothetical protein	+
ECSE_2553	conserved hypothetical protein	+
ECSE_2563	conserved hypothetical protein	-
ECSE_2614	conserved hypothetical protein	-
ECSE_2647	putative fimbrial usher protein	+
ECSE_2656	conserved hypothetical protein	+
ECSE_2658	sucrose transport protein	+
ECSE_2659	fructokinase	+
ECSE_2660	sucrose hydrolase	+
ECSE_2661	sucrose operon repressor	+
ECSE_2662	conserved hypothetical protein	+
ECSE_2695	conserved hypothetical protein	-
ECSE_2758	conserved hypothetical protein	-
ECSE_2759	conserved hypothetical protein	+
ECSE_2768	putative hydrogenase 4 membrane subunit	-
ECSE_2790	conserved hypothetical protein	-
ECSE_2865	conserved hypothetical protein	-
ECSE_2877	conserved hypothetical protein	+
ECSE_2906	conserved hypothetical protein	-
ECSE_2907	histidin kinase-like protein	-
ECSE_2910	conserved hypothetical protein	+
ECSE_2911	conserved hypothetical protein	+
ECSE_2933	conserved hypothetical protein	+
ECSE_2934	conserved hypothetical protein	+
ECSE_2989	conserved hypothetical protein	+
ECSE_2990	conserved hypothetical protein	+
ECSE_2991	putative phenylacrylic acid decarboxylase	+
ECSE_2992	transcriptional regulator	+
ECSE_3011	conserved hypothetical protein	+
ECSE_3014	conserved hypothetical protein	+
ECSE_3017	conserved hypothetical protein	-
ECSE_3035	conserved hypothetical protein	+
ECSE_3106	conserved hypothetical protein	+
ECSE_3108	conserved hypothetical protein	-
ECSE_3112	conserved hypothetical protein	+
ECSE_3113	conserved hypothetical protein	-
ECSE_3114	conserved hypothetical protein	+
ECSE_3115	conserved hypothetical protein	-
ECSE_3116	conserved hypothetical protein	+
ECSE_3117	conserved hypothetical protein	-
ECSE_3118	putative lipoprotein of type III secretion apparatus	+
ECSE_3119	putative type III secretion protein	-
ECSE_3120	putative type III secretion protein	+
ECSE_3121	putative type III secretion system regulator	-
ECSE_3122	putative type III secretion protein	-
ECSE_3123	putative type III secretion protein	+
ECSE_3124	putative type III secretion protein	+
ECSE_3125	putative type III secretion protein	+
ECSE_3126	putative type III secretion protein	+
ECSE_3127	putative type III secretion protein	+
ECSE_3128	conserved hypothetical protein	+
ECSE_3177	conserved hypothetical protein	-
ECSE_3192	conserved hypothetical protein	+
ECSE_3193	conserved hypothetical protein	+
ECSE_3194	cobalt ABC transporter permease component	+
ECSE_3195	cobalt ABC transporter ATP-binding component	+

ECSE_3196	cobalt ABC transporter ATP-binding component	+
ECSE_3206	conserved hypothetical protein	+
ECSE_3227	conserved hypothetical protein	+
ECSE_3237	putative type II secretion protein	+
ECSE_3239	putative type II secretion protein	+
ECSE_3240	putative type II secretion protein	+
ECSE_3241	putative type II secretion protein	+
ECSE_3246	putative type II secretion protein	+
ECSE_3250	conserved hypothetical protein	+
ECSE_3259	conserved hypothetical protein	+
ECSE_3260	conserved hypothetical protein	+
ECSE_3261	conserved hypothetical protein	+
ECSE_3262	conserved hypothetical protein	+
ECSE_3263	conserved hypothetical protein	+
ECSE_3264	conserved hypothetical protein	+
ECSE_3265	conserved hypothetical protein	+
ECSE_3266	conserved hypothetical protein	+
ECSE_3267	conserved hypothetical protein	+
ECSE_3284	conserved hypothetical protein	+
ECSE_3288	conserved hypothetical protein	+
ECSE_3291	conserved hypothetical protein	-
ECSE_3302	conserved hypothetical protein	-
ECSE_3321	conserved hypothetical protein	+
ECSE_3325	putative fimbrial usher protein	+
ECSE_3331	conserved hypothetical protein	-
ECSE_3341	conserved hypothetical protein	+
ECSE_3369	conserved hypothetical protein	+
ECSE_3375	putative fimbrial protein	+
ECSE_3376	putative fimbrial usher protein	+
ECSE_3377	putative fimbrial protein	+
ECSE_3378	putative fimbrial protein	+
ECSE_3406	conserved hypothetical protein	+
ECSE_3407	conserved hypothetical protein	-
ECSE_3418	PTS system N-acetylgalactosamine-specific IIC component	+
ECSE_3419	PTS system N-acetylgalactosamine-specific IID component	+
ECSE_3420	PTS system N-acetylgalactosamine-specific IIA component	+
ECSE_3421	N-acetylgalactosamine-6-phosphate deacetylase	+
ECSE_3496	conserved hypothetical protein	-
ECSE_3524	conserved hypothetical protein	-
ECSE_3531	conserved hypothetical protein	+
ECSE_3534	conserved hypothetical protein	-
ECSE_3550	amino acid ABC transporter substrate binding component	+
ECSE_3659	conserved hypothetical protein	+
ECSE_3665	conserved hypothetical protein	+
ECSE_3667	conserved hypothetical protein	-
ECSE_3668	conserved hypothetical protein	-
ECSE_3673	conserved hypothetical protein	+
ECSE_3677	conserved hypothetical protein	-
ECSE_3693	conserved hypothetical protein	+
ECSE_3694	conserved hypothetical protein	+
ECSE_3695	conserved hypothetical protein	+
ECSE_3696	conserved hypothetical protein	+
ECSE_3711	conserved hypothetical protein	+
ECSE_3725	conserved hypothetical protein	+
ECSE_3750	conserved hypothetical protein	+
ECSE_3755	conserved hypothetical protein	+
ECSE_3770	conserved hypothetical protein	+
ECSE_3771	conserved hypothetical protein	+
ECSE_3772	conserved hypothetical protein	+
ECSE_3773	conserved hypothetical protein	+
ECSE_3803	conserved hypothetical protein	+
ECSE_3808	conserved hypothetical protein	+
ECSE_3809	conserved hypothetical protein	-
ECSE_3811	conserved hypothetical protein	+
ECSE_3818	conserved hypothetical protein	+
ECSE_3844	truncated xylose operon regulator	-
ECSE_3864	conserved hypothetical protein	+
ECSE_3872	putative truncated Rhs core protein	+
ECSE_3878	conserved hypothetical protein	-
ECSE_3883	conserved hypothetical protein	+
ECSE_3884	putative autotransporter	+
ECSE_3894	conserved hypothetical protein	-
ECSE_3905	lipid A-core:surface polymer ligase	-
ECSE_3906	lipopolysaccharide 1,2-N- acetylglucosaminetransferase	-
ECSE_3925	conserved hypothetical protein	+
ECSE_3929	conserved hypothetical protein	-

ECSE_3935	conserved hypothetical protein	+
ECSE_3967	truncated 6-phospho-alpha-glucosidase	-
ECSE_3968	truncated 6-phospho-alpha-glucosidase	-
ECSE_3969	putative PTS system component	+
ECSE_4015	putative fimbrial protein LpfD	+
ECSE_4016	putative fimbrial usher protein LpfC	+
ECSE_4017	putative fimbrial chaperone LpfB	+
ECSE_4018	putative fimbrial protein LpfA	+
ECSE_4049	conserved hypothetical protein	-
ECSE_4051	acetolactate synthase II large subunit	+
ECSE_4059	conserved hypothetical protein	+
ECSE_4064	rho operon leader peptide	-
ECSE_4066	conserved hypothetical protein	+
ECSE_4081	conserved hypothetical protein	+
ECSE_4082	conserved hypothetical protein	+
ECSE_4089	conserved hypothetical protein	+
ECSE_4090	conserved hypothetical protein	+
ECSE_4091	conserved hypothetical protein	-
ECSE_4092	conserved hypothetical protein	-
ECSE_4094	conserved hypothetical protein	+
ECSE_4103	conserved hypothetical protein	+
ECSE_4104	conserved hypothetical protein	+
ECSE_4105	conserved hypothetical protein	-
ECSE_4106	conserved hypothetical protein	-
ECSE_4144	conserved hypothetical protein	+
ECSE_4146	conserved hypothetical protein	-
ECSE_4172	putative lipase	+
ECSE_4173	conserved hypothetical protein	+
ECSE_4174	conserved hypothetical protein	+
ECSE_4181	truncated formate dehydrogenase major subunit	-
ECSE_4193	conserved hypothetical protein	+
ECSE_4205	conserved hypothetical protein	-
ECSE_4226	Rhs core protein	+
ECSE_4254	starvation sensing protein	+
ECSE_4255	putative transport protein	+
ECSE_4266	conserved hypothetical protein	+
ECSE_4291	conserved hypothetical protein	+
ECSE_4296	conserved hypothetical protein	+
ECSE_4310	conserved hypothetical protein	+
ECSE_4311	conserved hypothetical protein	+
ECSE_4312	conserved hypothetical protein	+
ECSE_4319	conserved hypothetical protein	+
ECSE_4320	conserved hypothetical protein	+
ECSE_4329	conserved hypothetical protein	-
ECSE_4330	conserved hypothetical protein	-
ECSE_4358	conserved hypothetical protein	+
ECSE_4372	conserved hypothetical protein	-
ECSE_4376	truncated formate dehydrogenase H	-
ECSE_4402	phosphonate ABC transporter permease component	+
ECSE_4428	conserved hypothetical protein	-
ECSE_4477	conserved hypothetical protein	+
ECSE_4497	conserved hypothetical protein	-
ECSE_4503	putative dicarboxylate-binding periplasmic protein	+
ECSE_4504	putative transport protein	+
ECSE_4505	conserved hypothetical protein	-
ECSE_4506	putative oxidoreductase	+
ECSE_4507	coenzyme A transferase	+
ECSE_4508	conserved hypothetical protein	+
ECSE_4509	conserved hypothetical protein	+
ECSE_4510	conserved hypothetical protein	+
ECSE_4511	conserved hypothetical protein	+
ECSE_4524	conserved hypothetical protein	+
ECSE_4532	conserved hypothetical protein	+
ECSE_4541	cytochrome b562	+
ECSE_4547	conserved hypothetical protein	-
ECSE_4549	conserved hypothetical protein	-
ECSE_4559	conserved hypothetical protein	-
ECSE_4574	putative restriction endonuclease	-
ECSE_4577	conserved hypothetical protein	-
ECSE_4578	putative 3'-phosphoadenosine 5'-phosphosulfate sulfotransferase	-
ECSE_4579	conserved hypothetical protein	-
ECSE_4599	conserved hypothetical protein	-
ECSE_4605	conserved hypothetical protein	+
ECSE_4613	conserved hypothetical protein	-
ECSE_4616	penicillin G acylase	-
ECSE_4620	4-hydroxyphenylacetate 3-monooxygenase small subunit	-

ECSE_4621	4-hydroxyphenylacetate 3-monooxygenase large subunit	-
ECSE_4622	4-hydroxyphenylacetate 3-monooxygenase operon regulator	-
ECSE_4623	putative 4-hydroxyphenylacetate transport protein	-
ECSE_4624	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	-
ECSE_4625	2-oxo-hept-3-ene-1,7-dioate hydratase	-
ECSE_4626	5-carboxymethyl-2-hydroxymuconate isomerase	-
ECSE_4627	3,4-dihydroxyphenylacetate 2,3-dioxygenase	-
ECSE_4628	5-carboxy-2-hydroxymuconate semialdehyde dehydrogenase	-
ECSE_4629	5-oxo-1,2,5-tricarboxilic-3-penten acid decarboxilase/isomerase	-
ECSE_4630	homoprotocatechuate degradative operon repressor	-
ECSE_4660	conserved hypothetical protein	-
ECSE_P1-0001	replication initiation protein	+
ECSE_P1-0002	conserved hypothetical protein	+
ECSE_P1-0003	conserved hypothetical protein	+
ECSE_P1-0004	conserved hypothetical protein	-
ECSE_P1-0005	conserved hypothetical protein	-
ECSE_P1-0006	conserved hypothetical protein	-
ECSE_P1-0007	truncated YafB protein	-
ECSE_P1-0010	tetracycline repressor protein TetR	-
ECSE_P1-0011	tetracycline resistance structural protein TetA	-
ECSE_P1-0014	putative resolvase	-
ECSE_P1-0017	truncated YafB protein	-
ECSE_P1-0018	conserved hypothetical protein	-
ECSE_P1-0019	colicin protein	-
ECSE_P1-0020	colicin immunity protein	-
ECSE_P1-0021	conserved hypothetical protein	+
ECSE_P1-0022	conserved hypothetical protein	+
ECSE_P1-0023	conserved hypothetical protein	-
ECSE_P1-0025	conserved hypothetical protein	-
ECSE_P1-0026	conserved hypothetical protein	-
ECSE_P1-0028	conserved hypothetical protein	-
ECSE_P1-0029	conserved hypothetical protein	-
ECSE_P1-0030	conserved hypothetical protein	-
ECSE_P1-0031	abortive infection protein	-
ECSE_P1-0032	conserved hypothetical protein	-
ECSE_P1-0033	conserved hypothetical protein	-
ECSE_P1-0034	conserved hypothetical protein	-
ECSE_P1-0035	conserved hypothetical protein	-
ECSE_P1-0036	resolvase	+
ECSE_P1-0037	conserved hypothetical protein	+
ECSE_P1-0038	conserved hypothetical protein	+
ECSE_P1-0039	ParA protein	+
ECSE_P1-0040	ParB protein	+
ECSE_P1-0041	ImpB protein	-
ECSE_P1-0042	ImpA protein	-
ECSE_P1-0043	ImpC protein	-
ECSE_P1-0044	conserved hypothetical protein	-
ECSE_P1-0045	conserved hypothetical protein	+
ECSE_P1-0046	conserved hypothetical protein	-
ECSE_P1-0047	conserved hypothetical protein	+
ECSE_P1-0048	conserved hypothetical protein	-
ECSE_P1-0049	conserved hypothetical protein	-
ECSE_P1-0050	conserved hypothetical protein	+
ECSE_P1-0051	conserved hypothetical protein	-
ECSE_P1-0052	conserved hypothetical protein	+
ECSE_P1-0053	conserved hypothetical protein	+
ECSE_P1-0054	conserved hypothetical protein	-
ECSE_P1-0055	conserved hypothetical protein	-
ECSE_P1-0056	single-stranded DNA binding protein Ssb	-
ECSE_P1-0057	conserved hypothetical protein	-
ECSE_P1-0058	conserved hypothetical protein	-
ECSE_P1-0059	plasmid SOS inhibition protein PsiB	+
ECSE_P1-0060	plasmid SOS inhibition protein PsiA	+
ECSE_P1-0061	conserved hypothetical protein	+
ECSE_P1-0062	anti-restriction protein	+
ECSE_P1-0063	conserved hypothetical protein	-
ECSE_P1-0064	conserved hypothetical protein	-
ECSE_P1-0065	conserved hypothetical protein	+
ECSE_P1-0066	conserved hypothetical protein	+
ECSE_P1-0067	conserved hypothetical protein	+
ECSE_P1-0068	conserved hypothetical protein	+
ECSE_P1-0069	conserved hypothetical protein	+
ECSE_P1-0070	conserved hypothetical protein	-
ECSE_P1-0071	conserved hypothetical protein	+
ECSE_P1-0072	conserved hypothetical protein	+
ECSE_P1-0073	relaxosome component NikA	-

ECSE_P1-0074	relaxase NikB	+
ECSE_P1-0075	TrbC protein	+
ECSE_P1-0076	TrbB protein	-
ECSE_P1-0077	TrbA protein	-
ECSE_P1-0078	plasmid stability protein	-
ECSE_P1-0079	conserved hypothetical protein	-
ECSE_P1-0080	conserved hypothetical protein	-
ECSE_P1-0081	conserved hypothetical protein	-
ECSE_P1-0082	putative regulator protein	+
ECSE_P1-0083	surface exclusion protein ExcA	-
ECSE_P1-0084	TraY protein	-
ECSE_P1-0085	TraX protein	-
ECSE_P1-0086	TraW protein	-
ECSE_P1-0087	TraV protein	-
ECSE_P1-0088	TraU protein	-
ECSE_P1-0089	TraT protein	-
ECSE_P1-0090	TraS protein	-
ECSE_P1-0091	TraR protein	-
ECSE_P1-0092	TraQ protein	-
ECSE_P1-0093	TraP protein	-
ECSE_P1-0094	TraO protein	-
ECSE_P1-0095	TraN protein	-
ECSE_P1-0096	TraM protein	-
ECSE_P1-0097	TraL protein	-
ECSE_P1-0098	DNA primase SogL	-
ECSE_P1-0099	endonuclease	-
ECSE_P1-0100	TraK protein	-
ECSE_P1-0101	TraJ protein	-
ECSE_P1-0102	TraI protein	-
ECSE_P1-0103	TraH protein	-
ECSE_P1-0104	TraG protein	-
ECSE_P1-0105	TraF protein	-
ECSE_P1-0106	TraE protein	-
ECSE_P1-0107	shufflon protein	-
ECSE_P1-0108	prepilin peptidase PilU	-
ECSE_P1-0109	PilT protein	-
ECSE_P1-0110	type IV prepilin PilS	-
ECSE_P1-0111	PilR protein	-
ECSE_P1-0112	ATP-binding protein PilQ	-
ECSE_P1-0113	pilus assembly protein PilP	-
ECSE_P1-0114	PilO protein	-
ECSE_P1-0115	lipoprotein PilN	-
ECSE_P1-0116	PilM protein	+
ECSE_P1-0117	lipoprotein PilL	+
ECSE_P1-0118	PilK protein	-
ECSE_P1-0119	PilJ protein	-
ECSE_P1-0120	PilI protein	-
ECSE_P1-0121	TraD protein	-
ECSE_P1-0122	TraC protein	+
ECSE_P1-0123	TraB protein	+
ECSE_P1-0124	TraA protein	-
ECSE_P2-0001	putative F1 capsule chaperone	+
ECSE_P2-0002	putative F1 capsule anchoring protein	-
ECSE_P2-0003	putative F1 capsule antigen protein	-
ECSE_P2-0004	putative F1 capsule antigen protein	-
ECSE_P2-0005	putative F1 capsule antigen protein	-
ECSE_P2-0006	conserved hypothetical protein	-
ECSE_P2-0008	conserved hypothetical protein	+
ECSE_P2-0010	conserved hypothetical protein	+
ECSE_P2-0011	plasmid maintenance protein CcdB	+
ECSE_P2-0012	resolvase	+
ECSE_P2-0013	replication protein RepB	-
ECSE_P2-0014	partitioning protein ParA	+
ECSE_P2-0015	partitioning protein ParB	+
ECSE_P2-0016	truncated UV protection protein ImpB	-
ECSE_P2-0020	conserved hypothetical protein	-
ECSE_P2-0021	conserved hypothetical protein	-
ECSE_P2-0022	conserved hypothetical protein	-
ECSE_P2-0023	truncated UV protection protein ImpB	-
ECSE_P2-0024	UV protection protein ImpA	-
ECSE_P2-0025	conserved hypothetical protein	-
ECSE_P2-0026	conserved hypothetical protein	-
ECSE_P2-0027	conserved hypothetical protein	-
ECSE_P2-0028	conserved hypothetical protein	+
ECSE_P2-0029	conserved hypothetical protein	+
ECSE_P2-0030	conserved hypothetical protein	+

ECSE_P2-0031	conserved hypothetical protein	-
ECSE_P2-0032	conserved hypothetical protein	+
ECSE_P2-0033	conserved hypothetical protein	+
ECSE_P2-0034	conserved hypothetical protein	-
ECSE_P2-0035	conserved hypothetical protein	-
ECSE_P2-0036	conserved hypothetical protein	-
ECSE_P2-0037	conserved hypothetical protein	+
ECSE_P2-0038	conserved hypothetical protein	+
ECSE_P2-0041	single-stranded DNA binding protein	+
ECSE_P2-0042	conserved hypothetical protein	+
ECSE_P2-0043	conserved hypothetical protein	+
ECSE_P2-0044	plasmid SOS inhibition protein PsiB	+
ECSE_P2-0045	plasmid SOS inhibition protein PsiA	+
ECSE_P2-0046	conserved hypothetical protein	+
ECSE_P2-0049	conserved hypothetical protein	+
ECSE_P2-0050	conserved hypothetical protein	+
ECSE_P2-0051	conserved hypothetical protein	+
ECSE_P2-0053	conserved hypothetical protein	+
ECSE_P2-0054	putative TraM protein	-
ECSE_P2-0058	putative TraA protein	-
ECSE_P2-0059	TraL protein	+
ECSE_P2-0060	TraE protein	+
ECSE_P2-0061	TraK protein	-
ECSE_P2-0062	TraB protein	-
ECSE_P2-0063	TraV protein	-
ECSE_P2-0064	TraC protein	-
ECSE_P2-0065	TrbI protein	-
ECSE_P2-0066	TraR protein	+
ECSE_P2-0067	TraW protein	-
ECSE_P2-0068	TraU protein	-
ECSE_P2-0070	TrbD protein	-
ECSE_P2-0071	TrbC protein	-
ECSE_P2-0072	TraN protein	-
ECSE_P2-0073	TraF protein	-
ECSE_P2-0075	TrbB protein	-
ECSE_P2-0076	TraH protein	+
ECSE_P2-0077	TraG protein	+
ECSE_P2-0079	TraT protein	+
ECSE_P2-0081	conserved hypothetical protein	+
ECSE_P2-0082	TraD protein	+
ECSE_P2-0083	TraI protein	-
ECSE_P2-0084	TraX protein	+
ECSE_P2-0085	FinO protein	+
ECSE_P2-0086	conserved hypothetical protein	-
ECSE_P2-0087	conserved hypothetical protein	+
ECSE_P2-0089	replication protein	-
ECSE_P2-0090	conserved hypothetical protein	-
ECSE_P2-0091	replication initiation protein	-
ECSE_P2-0092	conserved hypothetical protein	+
ECSE_P2-0093	conserved hypothetical protein	+
ECSE_P2-0105	conserved hypothetical protein	-
ECSE_P2-0106	putative transcriptional regulator	-
ECSE_P2-0107	conserved hypothetical protein	-
ECSE_P2-0110	conserved hypothetical protein	-
ECSE_P2-0111	conserved hypothetical protein	-
ECSE_P2-0112	conserved hypothetical protein	-
ECSE_P3-0001	anaerobic sulfite reductase subunit A	-
ECSE_P3-0002	anaerobic sulfite reductase subunit B	-
ECSE_P3-0003	anaerobic sulfite reductase subunit C	-
ECSE_P3-0004	putative truncated xylanase/chitin deacetylase	+
ECSE_P3-0008	conserved hypothetical protein	-
ECSE_P3-0009	conserved hypothetical protein	-
ECSE_P3-0010	conserved hypothetical protein	-
ECSE_P3-0011	integrase	-
ECSE_P3-0012	replication protein RepA	-
ECSE_P3-0013	conserved hypothetical protein	-
ECSE_P3-0016	conserved hypothetical protein	+
ECSE_P3-0017	conserved hypothetical protein	-
ECSE_P3-0021	conserved hypothetical protein	-
ECSE_P3-0023	putative cation transport protein	-
ECSE_P3-0024	thiosulfate reductase subunit A	-
ECSE_P3-0025	thiosulfate reductase iron-sulfur subunit	-
ECSE_P3-0026	thiosulfate reductase cytochrome b subunit	-
ECSE_P3-0031	fimbrial usher protein FaeD	-
ECSE_P3-0032	fimbrial chaperone FaeE	-
ECSE_P3-0033	fimbrial protein FaeF	-

ECSE_P3-0034	fimbrial protein FaeG	-
ECSE_P3-0035	fimbrial protein FaeH	-
ECSE_P3-0036	fimbrial protein FaeI	-
ECSE_P3-0037	fimbrial protein FaeJ	-
ECSE_P3-0038	conserved hypothetical protein	-
ECSE_P3-0044	conserved hypothetical protein	-
ECSE_P3-0045	NAD-linked malate dehydrogenase	-
ECSE_P3-0046	choloyglycine hydrolase	-
ECSE_P3-0048	putative acetoin dehydrogenase	-
ECSE_P3-0051	conserved hypothetical protein	+
ECSE_P3-0052	conserved hypothetical protein	+
ECSE_P3-0053	conserved hypothetical protein	-
ECSE_P3-0057	putative transcription regulator	-
ECSE_P3-0058	putative transcription regulator	-
ECSE_P3-0059	conserved hypothetical protein	-
ECSE_P3-0060	putative fimbrial usher protein	-
ECSE_P3-0061	putative fimbrial chaperone	-
ECSE_P3-0062	putative fimbrial chaperone	-
ECSE_P3-0063	putative fimbrial protein	-
ECSE_P3-0064	putative fimbrial protein	-
ECSE_P3-0065	putative fimbrial chaperone	-
ECSE_P3-0066	putative fimbrial protein	-
ECSE_P4-0004	mobilization protein MobB	-
ECSE_P4-0005	mobilization protein MobA	+
ECSE_P4-0006	mobilization protein MobC	-
ECSE_P4-0007	RNA I inhibition modulator protein Rom	-
ECSE_P4-0009	conserved hypothetical protein	-
ECSE_P4-0010	conserved hypothetical protein	-
ECSE_P5-0001	mobilization protein MobA	-
ECSE_P5-0002	mobilization protein MobB	-
ECSE_P5-0004	conserved hypothetical protein	-
ECSE_P5-0005	conserved hypothetical protein	-
ECSE_P5-0007	conserved hypothetical protein	-
ECSE_P6-0001	replication protein	-
ECSE_P6-0002	conserved hypothetical protein	-
ECSE_P6-0003	conserved hypothetical protein	-

* "+" indicates genes that are shared with *E. coli* E24377A and may be specific to phylogenetic B1 group; "-" indicates genes not shared with *E. coli* E24377A.

Supplementary Table S4. The 118 genes present in MG1655 but absent in SE11

Gene ID	Gene product
b4412	toxic membrane protein, small
b0164	predicted phosphatase
b4586	hypothetical protein
b0228	conserved protein
b0232	predicted antitoxin of the YafO-YafN toxin-antitoxin system
b0233	predicted toxin of the YafO-YafN toxin-antitoxin system
b4504	predicted protein
b0295	predicted protein
b0342	thiogalactoside acetyltransferase
b4637	ryhB-regulated fur leader peptide
b0689	predicted protein
b4590	hypothetical protein
b4513	potassium ion accessory transporter subunit
b0702	predicted inner membrane protein
b0706	conserved protein
b4515	conserved protein
b0816	predicted protein
b4592	hypothetical protein
b4594	hypothetical protein
b4419	toxic polypeptide, small
b4423	toxic polypeptide, small
b1228	predicted protein
b1229	predicted protamine-like protein
b1268	predicted inner membrane protein
b4428	toxic polypeptide, small
b1420	regulatory peptide
b4532	predicted protein
b4598	hypothetical protein
b1460	conserved protein
b1509	conserved protein
b4599	hypothetical protein
b4601	hypothetical protein
b4602	hypothetical protein
b1715	phenylalanyl-tRNA synthetase operon leader peptide
b1932	predicted acyltransferase
b2032	lipopolysaccharide biosynthesis protein
b2033	predicted acyl transferase
b2034	conserved protein
b2035	O-antigen polymerase
b2036	UDP-galactopyranose mutase, FAD/NAD(P)-binding
b2038	dTDP-4-deoxyrhamnose-3,5-epimerase
b2039	glucose-1-phosphate thymidyltransferase
b2040	dTDP-4-dehydrorhamnose reductase subunit, NAD(P)-binding, of dTDP-L-rhamnose synthase
b2041	dTDP-glucose 4,6 dehydratase, NAD(P)-binding
b2071	predicted protein
b4542	predicted protein
b4502	conserved protein
b4604	hypothetical protein
b2219	sensory histidine kinase in two-component regulatory system with AtoC
b2220	fused response regulator of ato operon, in two-component system with AtoS
b2221	acetyl-CoA:acetoacetyl-CoA transferase, alpha subunit
b2222	acetyl-CoA:acetoacetyl-CoA transferase, beta subunit
b2223	short chain fatty acid transporter
b2224	acetyl-CoA acetyltransferase
b4605	hypothetical protein
b4543	predicted protein
b2246	predicted transporter
b2247	predicted enolase
b2248	predicted DNA-binding transcriptional regulator
b2270	conserved protein
b2273	predicted protein
b2274	predicted protein
b2275	predicted protein
b2305	conserved protein
b4501	response regulator inhibitor for tor operon

b2364 DNA-binding transcriptional dual regulator
b2365 predicted transporter
b2420 conserved protein
b4606 hypothetical protein
b2504 predicted protein
b2598 pheA gene leader peptide
b4548 predicted protein
b2647 adhesin-like autotransporter
b2756 predicted protein
b2759 predicted protein
b4610 hypothetical protein
b2940 predicted protein
b3020 predicted transporter subunit: periplasmic-binding component of ABC superfamily
b3021 predicted DNA-binding transcriptional regulator
b3022 predicted cyanide hydratase
b3083 conserved protein
b3214 periplasmic protein
b3215 predicted periplasmic chaperone protein
b3216 predicted outer membrane protein
b3219 predicted transcriptional regulator
b3322 part of gsp divergon involved in type II protein secretion
b3323 general secretory pathway component, cryptic
b3324 general secretory pathway component, cryptic
b3329 predicted general secretory pathway component, cryptic
b3330 general secretory pathway component, cryptic
b3331 predicted general secretory pathway component, cryptic
b3333 general secretory pathway component, cryptic
b3334 general secretory pathway component, cryptic
b3338 periplasmic endochitinase
b4612 hypothetical protein
b3483 predicted protein
b3593 rhsA element core protein RshA
b3622 O-antigen ligase
b3623 lipopolysaccharide core biosynthesis
b3624 lipopolysaccharide core biosynthesis protein
b3628 UDP-D-galactose:(glucosyl)lipopolysaccharide-1,6-D-galactosyltransferase
b3629 lipopolysaccharide core biosynthesis protein
b4617 lexA-regulated predicted protein
b4618 lexA-regulated toxic peptide
b4557 predicted protein
b3801 acrylsulfatase-like enzyme
b3817 conserved inner membrane protein
b3818 predicted inner membrane protein
b4011 conserved protein
b4620 hypothetical protein
b4157 predicted protein
b4158 conserved inner membrane protein
b4204 predicted protein
b4215 predicted protein
b4264 DNA-binding transcriptional repressor, 5-gluconate-binding
b4265 L-idonate and D-gluconate transporter
b4266 5-keto-D-gluconate-5-reductase
b4267 L-idonate 5-dehydrogenase, NAD-binding
b4268 D-gluconate kinase, thermosensitive
b4269 predicted alcohol dehydrogenase, Zn-dependent and NAD(P)-binding
b4655 expressed protein
b4565 predicted enzyme IIB component of PTS
b4309 conserved protein
b4345 5-methylcytosine-specific restriction enzyme McrBC, subunit McrC
b4346 5-methylcytosine-specific restriction enzyme McrBC, subunit McrB
b4347 conserved protein
b4348 specificity determinant for hsdM and hsdR
b4349 DNA methylase M
b4350 endonuclease R
b4351 methylated adenine and cytosine restriction protein
b4385 predicted DNA-binding transcriptional regulator

Supplementary Table S5. The 62 poorly conserved genes between SE11 and MG1655

SE11 gene	MG1655 gene	BLAST score ratio	Gene product
ECSE_3015	b2760	0.07	conserved hypothetical protein
ECSE_3016	b2761	0.15	conserved hypothetical protein
ECSE_3012	b2757	0.16	conserved hypothetical protein
ECSE_3013	b2758	0.16	conserved hypothetical protein
ECSE_3238	b3332	0.23	putative type II secretion protein
ECSE_1559	b1470	0.26	conserved hypothetical protein
ECSE_1432	b1365	0.32	phage DNA methylase/nuclease
ECSE_3597	b3335	0.34	putative leader peptidase
ECSE_2314	b2037	0.34	conserved hypothetical protein
ECSE_3907	b3626	0.39	putative UDP-glucose:(galactosyl) LPS alpha1,2-glucosyltransferase
ECSE_2642	b2332	0.39	conserved hypothetical protein
ECSE_3245	b3325	0.40	putative type II secretion protein
ECSE_3243	b3327	0.42	putative type II secretion protein
ECSE_2154	b1923	0.44	flagellar protein FlhC
ECSE_0264	b2349	0.45	putative phage integrase
ECSE_0290	b2634	0.46	hypothetical phage protein
ECSE_2904	b2622	0.47	phage integrase
ECSE_0292	b0251	0.49	hypothetical phage protein
ECSE_4339	b4047	0.49	conserved hypothetical protein
ECSE_2644	b2334	0.49	putative fimbrial protein
ECSE_2645	b2335	0.50	putative fimbrial protein
ECSE_0297	b2644	0.52	hypothetical phage protein
ECSE_0289	b0253	0.52	conserved hypothetical protein
ECSE_2254	b1570	0.54	putative phage regulator
ECSE_3909	b3627	0.54	UDP-D-galactose:(glucosyl)lipopolysaccharide-alpha-1,3-D-galactosyltransferase
ECSE_3244	b3326	0.56	putative type II secretion protein
ECSE_3908	b3625	0.57	lipopolysaccharide core biosynthesis protein
ECSE_2643	b2333	0.57	putative fimbrial protein
ECSE_1696	b1575	0.58	hypothetical phage protein
ECSE_0294	b0249	0.59	conserved hypothetical protein
ECSE_3242	b3328	0.59	putative type II secretion protein
ECSE_3403	b3119	0.60	threonine dehydratase operon activator protein
ECSE_0135	b0135	0.61	putative fimbrial protein
ECSE_0862	b0806	0.62	conserved hypothetical protein
ECSE_1692	b1571	0.62	hypothetical phage protein
ECSE_0400	b0380	0.64	conserved hypothetical protein
ECSE_4558	b4253	0.65	conserved hypothetical protein
ECSE_4450	b4148	0.67	suppressor of groEL
ECSE_1959	b1788	0.67	conserved hypothetical protein
ECSE_3831	b4455	0.67	conserved hypothetical protein
ECSE_1676	b1560	0.68	hypothetical phage protein
ECSE_2012	b1837	0.68	conserved hypothetical protein
ECSE_0296	b2643	0.69	hypothetical phage protein
ECSE_0299	b2646	0.69	hypothetical phage protein
ECSE_2440	b2171	0.69	putative elongation factor
ECSE_1528	b1445	0.69	conserved hypothetical protein
ECSE_1600	b1510	0.72	putative autotransporter
ECSE_2531	b2271	0.72	conserved hypothetical protein
ECSE_1697	b1576	0.72	hypothetical phage protein
ECSE_3554	b3279	0.73	putative transferase
ECSE_3912	b3632	0.73	lipopolysaccharide core biosynthesis protein
ECSE_2565	b1158	0.74	phage DNA invertase
ECSE_4354	b4060	0.74	conserved hypothetical protein
ECSE_0206	b0210	0.75	putative biotin synthesis protein
ECSE_0122	b0122	0.75	conserved hypothetical protein
ECSE_3521	b3242	0.75	conserved hypothetical protein
ECSE_1188	b1122	0.76	conserved hypothetical protein
ECSE_1409	b1355	0.76	conserved hypothetical protein
ECSE_0298	b2645	0.77	hypothetical phage protein
ECSE_0295	b0252	0.78	hypothetical phage protein
ECSE_3651	b3390	0.78	shikimate kinase I
ECSE_3268	b2983	0.79	conserved hypothetical protein

Supplementary Figure S1

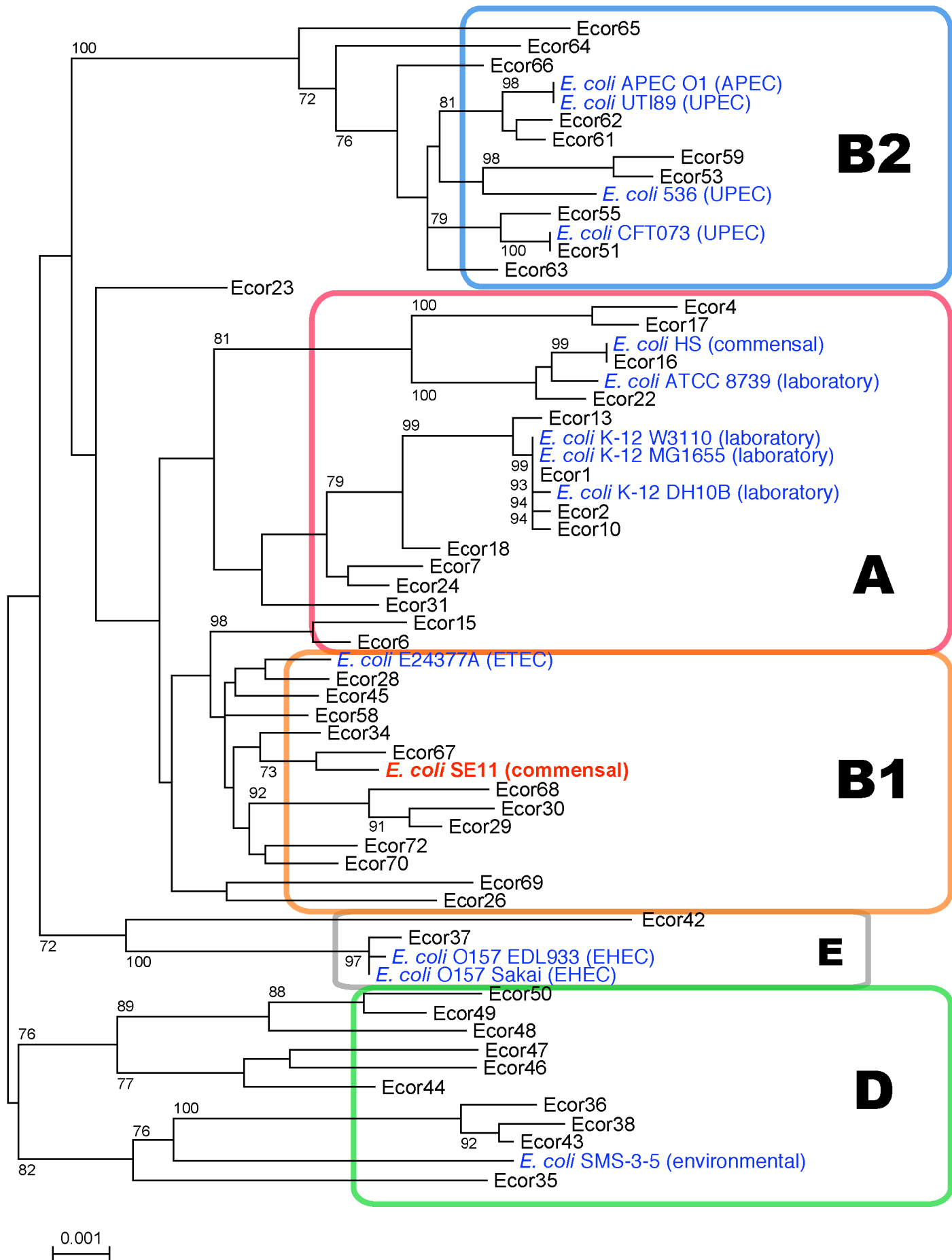
Phylogenetic analysis of SE11

Phylogenetic tree of members of *E. coli* was built using the neighbor-joining method based on partial nucleotide sequences of seven housekeeping genes (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA*). Multiple alignments of concatenated sequences (3,423 nt) were carried out using ClustalW (a bootstrap trial of 1,000). Bootstrap values greater than 70% are indicated at the nodes. SE11 is shown by red, and other completely sequenced *E. coli* strains by blue. Scale bar represents the number of substitutions per site. Proposed *E. coli* phylogroups (four major groups, A, B1, B2, and D, and a minor group E) are indicated. Abbreviations: ETEC, enterotoxigenic *E. coli*; EHEC, enterohemorrhagic *E. coli*; UPEC, uropathogenic *E. coli*; APEC, avian pathogenic *E. coli*.

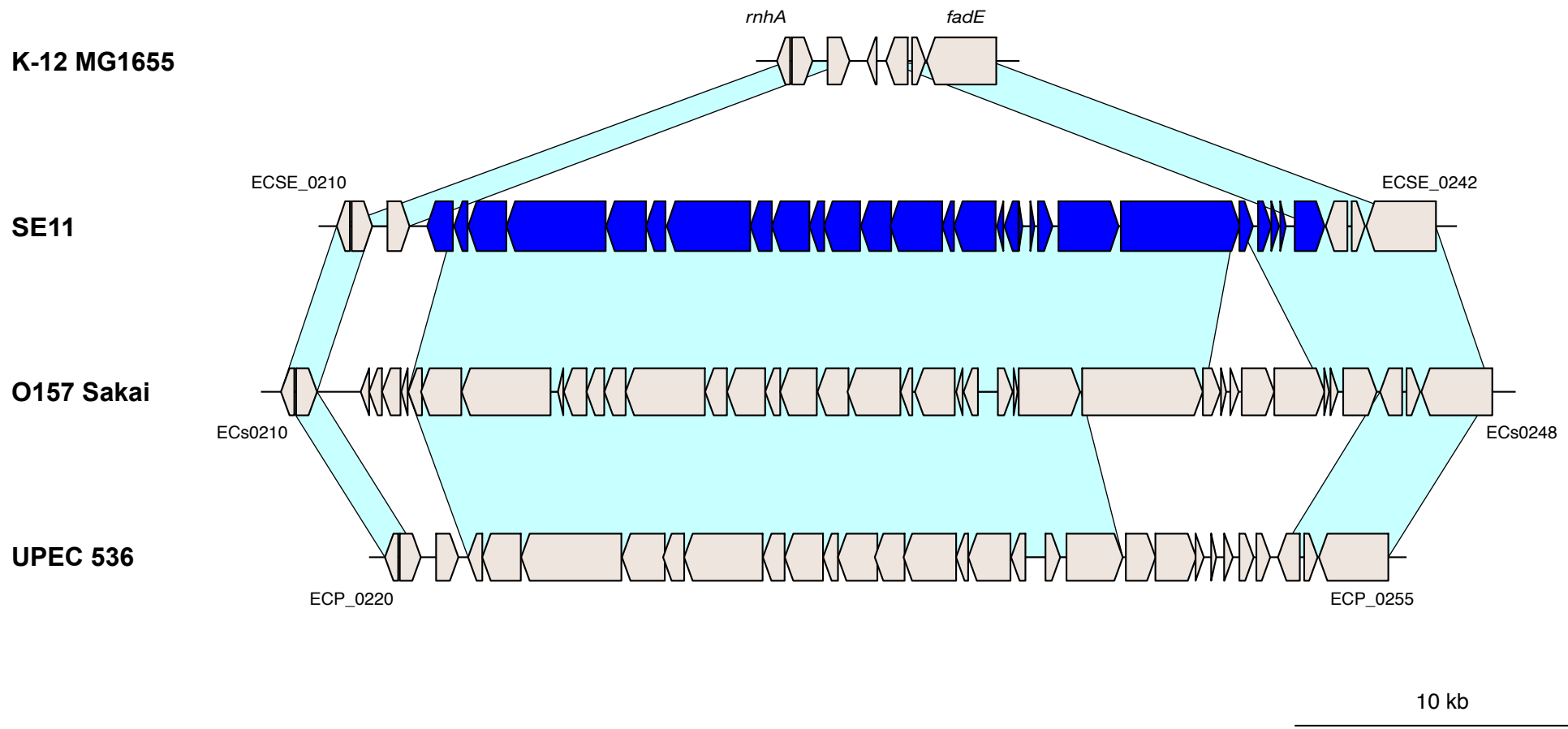
Supplementary Figure S2

Comparisons of the genomic location of the large segment near the *aspV* in SE11 with the corresponding location of MG1655, O157, and UPEC 536 strains

Genes and their orientations are depicted with arrows. Genes within the large segment in SE11 are shown by blue. Light blue bars indicate orthologous regions.



Supplementary Figure S1



Supplementary Figure S2