Supplementary Table 1. Strains containing RegA homologues

Organism	**Strain notes**	**RegA accession**	**Reference**
C. rodentium: **ICC168/ICC169**	Mouse pathogenic strain	emb|CBG90830.1; gb|ACI29349.1	(1)
E. coli: **4_1_47FAA**	Human faecal strain, reference genome for human microbiome project	gb|EHP63452	Taxonomy ID: 1127356
E. coli: **97.0259**	Cow commensal strain, O11:H- encoding *stx1*.	gb|EIH11864.1	Taxonomy ID: 869672
E. coli: **AA86**	Cow faecal strain, B2 group (Korea)	gb|EGH36988.1	(2)
E. coli:**B088**	Bird strain (*Circus assimilis*, NSW, Australia)	gb|EFE60329.1, ref|ZP_06664842.1	David Gordon, Taxonomy ID: 550672
E. coli: **B093**	Bird strain (*Strepera graculina*, NSW, Australia)	gb|EHN96312.1	Taxonomy ID: 550674
E. coli**: E22**	Rabbit enteropathogenic *E. coli* strain	ref|ZP_0346477.1, gb|EDV81558.1	Taxonomy ID: 340185
E. coli: **EC958**	Pathogenic O25b:H4-ST131 strain. Multi-drug resistant, phylogenetic group B2, (England)	WGS, contig00117: 9100-9924(bp)	(3)
E. coli: **M605**	Native Australian bat strain (*Nyctophilus geoffroyi*, NSW, Australia)	gb|EGI13263.1, ZP_08351234.1	David Gordon, Taxonomy ID: 656417
E. coli **MS16-3**	Reference genome for human microbiome project	gb|EFU59366.1	Taxonomy ID: 749542
E. coli: **PCN033**	Porcine ExPEC strain. Multi-antibiotic resistant (China)	gb|EGP21814.1	(4)

Supplementary Table 1 (continued).

E. coli: **SE15**	Commensal human strain (Japan)	ref|YP_003352001.1 dbj|BAI57551.1	(5)
E. coli: **SMS-3-5**	Environmental strain, multi-antibiotic resistant, (Shipyard Creek, USA)	ref|YP_001745894.1, gb|ACB18411.1	(6)
E. coli: **TA206**	Long-nosed bandicoot strain (*Perameles nasuta*, NSW, Australia)	gb|EGI25163.1, ref|ZP_0836853.1	David Gordon, Taxonomy ID: 656440
E. coli: **W26**	Cow faecal strain, for comparative genomic analysis. (South Korea)	gb| EID69191.1	Taxonomy ID: 1090926
Escherichia **sp.: H605**	Clade IV strain Human faecal strain, (Canberra)	WGS, supercontig 8: 791700-792530	David Gordon, Taxonomy ID: 754321
Escherichia **sp.: TW09276**	Environmental *Escherichia* clade III strain, (freshwater beach sample, Michigan)	WGS contig_154: 807-1	(7)
Escherichia **sp.: TW09308**	Environmental *Escherichia* clade V strain, (freshwater beach sample, Michigan)	ref|ZP_09457944.1	(7)
Escherichia **sp.: TW11588**	Environmental *Escherichia* clade IV strain, (freshwater beach sample, Michigan)	WGS, contig_001: 2720933-2720109	(7)
Escherichia **sp.: TW14182**	Environmental *Escherichia* clade IV strain, (freshwater beach sample, Michigan)	WGS, contig_073: 10242-9418	(7)

Supplementary Table 2. Identity and similarity scores of RegA homologues[a].

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1		100	100	100	97.1	98.1	96.4	96.7	96.7	96.7	70.3	70.7	71	63	62.7	65.6	67.2	66.8	64.1	64.5	48.9
2	100		100	100	97.1	98.1	96.4	96.7	96.7	96.7	70.3	70.7	71	63	62.7	65.6	67.2	66.8	64.1	64.5	48.9
3	100	100		100	97.1	98.1	96.4	96.7	96.7	96.7	70.3	70.7	71	63	62.7	65.6	67.2	66.8	64.1	64.5	48.9
4	100	100	100		97.1	98.1	96.4	96.7	96.7	96.7	70.3	70.7	71	63	62.7	65.6	67.2	66.8	64.1	64.5	48.9
5	99.6	99.6	99.6	99.6		98.1	94.8	95.2	95.2	95.2	70.3	72.5	71.7	62.5	62.1	65.8	67.1	67.1	64.3	64.3	48
6	99.6	99.6	99.6	99.6	100		93.5	93.8	93.8	93.8	68.1	69.2	69.2	60.9	60.5	63.4	63.3	63.3	62.7	62.7	46.8
7	98.2	98.2	98.2	98.2	97.8	95.3		99.6	99.6	99.6	70.7	70.7	71	63	62.7	65.6	67.2	67.2	64.1	64.1	49.6
8	98.6	98.6	98.6	98.6	98.1	95.7	99.6		100	100	70.7	70.7	71	63	62.7	65.6	67.6	67.2	64.1	64.5	49.6
9	98.6	98.6	98.6	98.6	98.1	95.7	99.6	100		100	70.7	70.7	71	63	62.7	65.6	67.6	67.2	64.1	64.5	49.6
10	98.6	98.6	98.6	98.6	98.1	95.7	99.6	100	100		70.7	70.7	71	63	62.7	65.6	67.6	67.2	64.1	64.5	49.6
11	87	87	87	87	87	84.8	87	87	87	87		70.3	73.9	59.8	59.8	62.7	67.2	66.4	63.4	64.1	45.3
12	88	88	88	88	88.1	85.9	88.4	88	88	88	85.5		83.3	62.7	62.3	63.8	69.3	69.3	66.3	66.3	46.4
13	86.6	86.6	86.6	86.6	86.2	84.1	87.7	87.7	87.7	87.7	85.9	93.5		63.8	63.4	64.9	70.1	70.5	67.4	67	48.6
14	78.6	78.6	78.6	78.6	78.4	76.4	78.3	78.3	78.3	78.3	79	80.4	82.2		99.6	89.5	84.6	84.6	84.1	84.1	46.4
15	78.6	78.6	78.6	78.6	78.4	76.4	78.3	78.3	78.3	78.3	79	80.4	82.2	100		89.1	84.6	84.6	83.7	83.7	46.4
16	81.2	81.2	81.2	81.2	80.7	79	80.8	80.8	80.8	80.8	79.3	81.9	82.6	95.3	95.3		85.9	86.3	86.6	86.2	46
17	84.2	84.2	84.2	84.2	84.2	81.7	83.4	83.8	83.8	83.8	83.8	85.9	85.9	94.2	94.2	94.2		97.1	97.1	100	48.1
18	83.8	83.8	83.8	83.8	83.8	81.3	83.4	83.4	83.4	83.4	83.4	86.3	86.3	94.6	94.6	94.6	99.2		100	97.1	47.3
19	81.5	81.5	81.5	81.5	81.4	79.3	80.8	80.8	80.8	80.8	80.4	82.6	82.6	93.8	93.8	94.2	99.2	100		97.5	44.6
20	81.9	81.9	81.9	81.9	81.8	79.7	80.8	81.2	81.2	81.2	80.8	82.2	82.2	93.5	93.5	93.8	100	99.2	99.3		45.3
21	70.9	70.9	70.9	70.9	70.8	69.1	71.9	71.6	71.6	71.6	67.6	69.8	72.7	68	68	68.7	69.3	69.3	66.9	66.9	

[a] Protein sequence identity (top right triangle) and similarity (bottom left triangle) scores. Data shown are percentage scores, generated with the MATGAT program (8); shaded to indicate level of similarity (with green indicating 100% through to red indicating <50%). Strains are as follows: 1. *E. coli* SE15, 2. *E. coli* AA86, 3. *E. coli* M605, 4. *E. coli* EC958, 5. *Escherichia* clade IV H605, 6. *Escherichia* clade III TW09276[b], 7. *E. coli* W26, 8. *E. coli* E22, 9. *E. coli* PCN033, 10. *E. coli* 97.0259, 11. *E. coli* 4_1_47FAA, 12. *C. rodentium* ICC168, 13. *E. coli* B088, 14. *Escherichia* clade IV TW14182, 15. *Escherichia* clade IV TW11588, 16. *Escherichia* clade V TW09308, 17. *E. coli* TA206[b], 18. *E. coli* B093[b], 19. *E. coli* SMS-3-5, 20. *E. coli* MS16-3, 21. *E. coli* NC101[c]

[b] Strains TA206 and B093 have truncated RegA protein sequences due to frameshift mutations. TW09276 is truncated (7 aa) at the C-terminal due to the end of the contig sequence. The figures given in this table are comparisons of the remaining protein sequence compared to the equivalent sequences in the other strains.

[c] The *E. coli* NC101 strain possesses a RegA-like protein (accession gb|EFM54601.1), and has been shown here as an outlier.

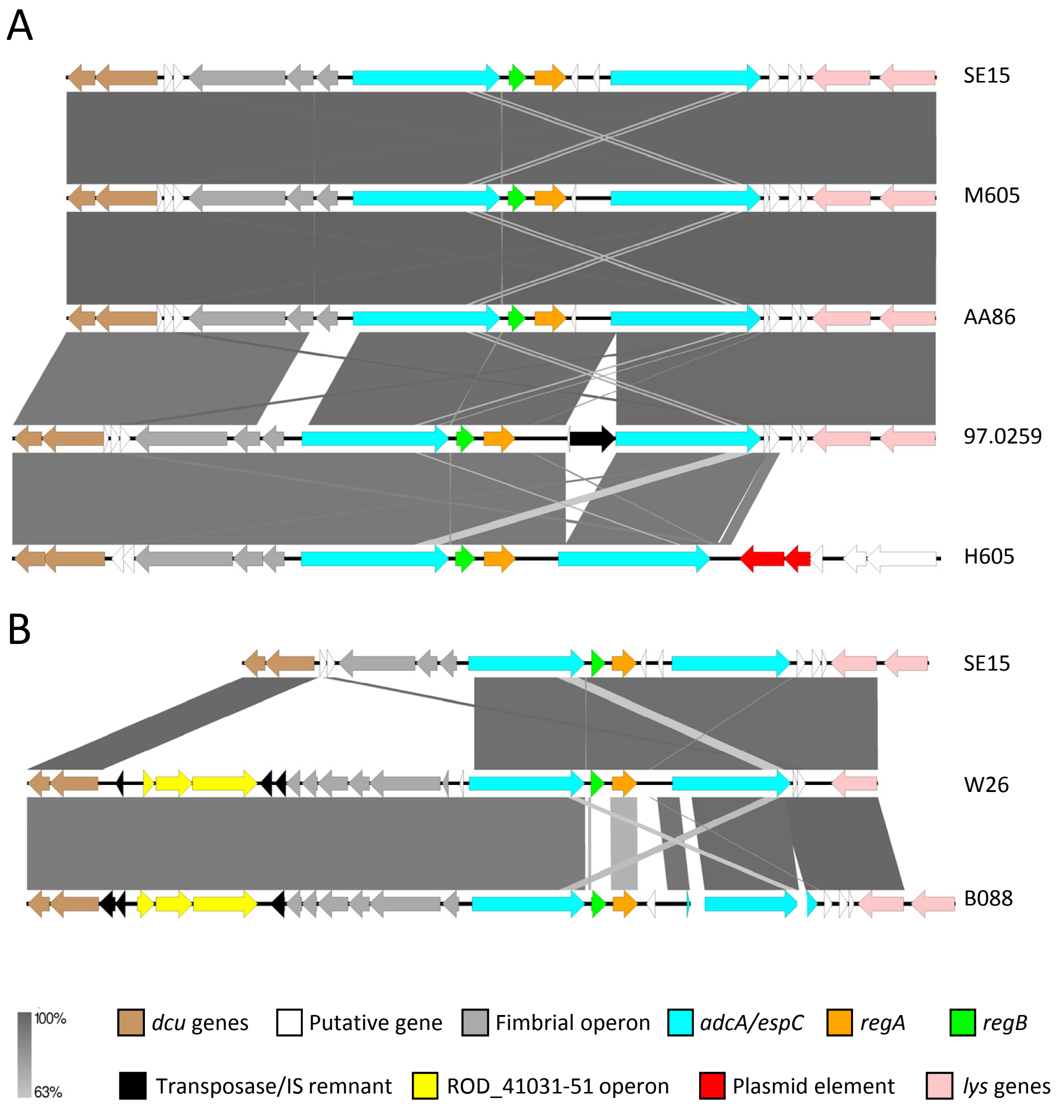

Figure S1. Comparison of the *regA* loci in Group 1 strains. The *regA* genetic loci from Group 1 strains (names shown on the right) have been compared using Easyfig 2.0 software (9). Coding regions are represented by colored arrows, with predicted function shown in the color key shown below the figure. The level of nucleotide identity (Blastn) between adjacent strains is indicated by the grey shading between the loci, and shown in the key (bottom left). **A.** Group 1 strains that share *E. coli* SE15 *regA* locus organization. **B.** SE15 locus organization (top) compared to *E.* coli strains, W26 and B088. These strains have the *kfc* operon instead of the fimbrial cluster found in SE15 and other Group 1 strains. The *regA* locus in the draft genome of B088 is located over three scaffolded contigs, with gaps indicated here by breaks in the DNA sequence (black line).

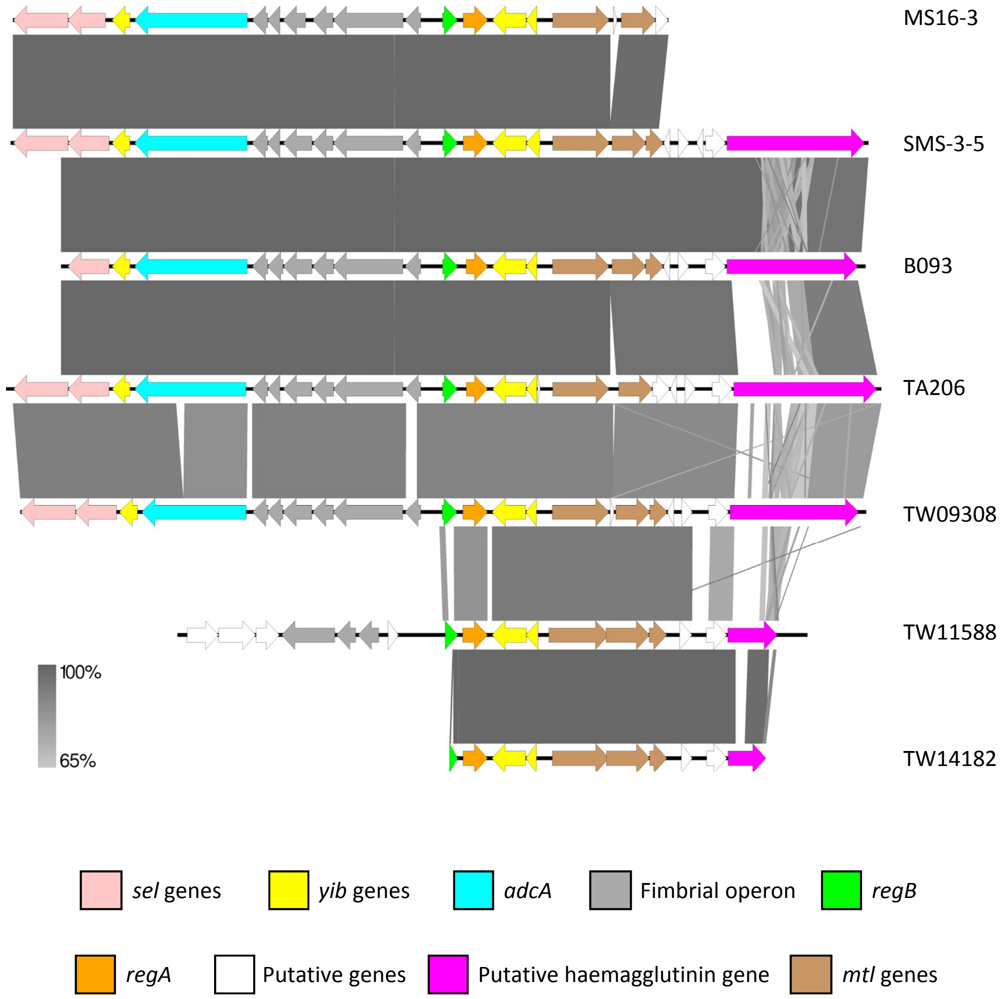

Figure S2. Comparison of the *regA* loci in Group 2 strains. The *regA* genetic loci from Group 2 strains (listed right) were compared using Easyfig 2.0 software (9). Coding regions are represented by colored arrows, with predicted functions shown in the color key below the figure. The level of nucleotide homology between adjacent strains is indicated by the grey shading between the loci, and shown in the key (bottom left). Most loci contain the *kfc* fimbrial operon, with the exception of *Escherichia* clade IV strain TW11588. The contig containing the *regA* locus of *Escherichia* clade IV strain TW14182 is shown with TW11588 for comparative purposes, but is of insufficient length to confirm that the locus belongs with Group 2.

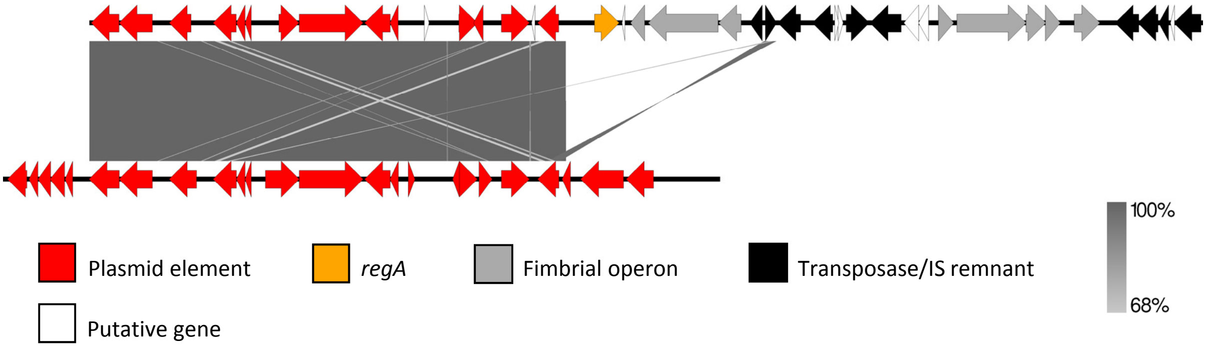

Figure S3. The *regA* locus in *E. coli* 4_1_47FAA. Organization of the RegA locus of *E. coli* 4_1_47FAA (top), compared to the F-plasmid (bottom) by using the Easyfig 2.0 software (9). Coding regions are represented by colored arrows, with identity or functional groups corresponding to the color key found below. The level of nucleotide homology between adjacent strains is indicated by the grey shading between the loci, and corresponds to the scale bar (bottom right).

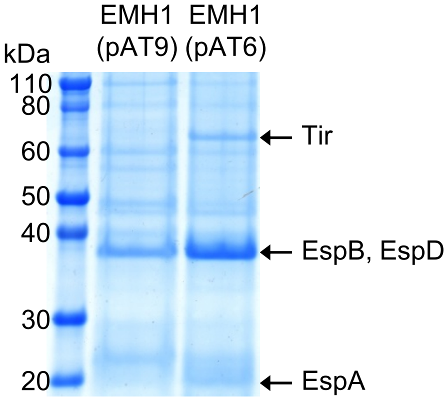

Figure S4. Effect of RegA$_{CR}$ and RegA$_{SMS}$ on the production of LEE proteins in *C. rodentium*. *C. rodentium* strains EMH1(pAT6[RegA$_{CR}$]) and EMH1(pAT9[RegA$_{SMS}$]) were grown in DMEM to OD$_{600}$ = 0.8 and the supernatants were collected by centrifugation. Supernatants were filtered through a 0.22-µm pore-size filter (Sartorius), and the proteins in the supernatants were precipitated with 10% (v/v) trichloroacetic acid and washed with 25% (v/v) acetone. Proteins in the supernatants were separated by SDS-PAGE using 12% Bis Tris gel and stained with Colloidal Coomassie G250. Predicted LEE proteins are indicated at the right of the gel.

References

1. **Petty NK, Bulgin R, Crepin VF, Cerdeno-Tarraga AM, Schroeder GN, Quail MA, Lennard N, Corton C, Barron A, Clark L, Toribio AL, Parkhill J, Dougan G, Frankel G, Thomson NR.** 2010. The *Citrobacter rodentium* genome sequence reveals convergent evolution with human pathogenic *Escherichia coli*. J. Bacteriol. **192:**525-538.

2. **Yi H, Cho Y-J, Hur H-G, Chun J.** 2011. Genome sequence of *Escherichia coli* AA86, isolated from cow feces. J. Bacteriol. **193:**3681.

3. **Totsika M, Beatson SA, Sarkar S, Phan M-D, Petty NK, Bachmann N, Szubert M, Sidjabat HE, Paterson DL, Upton M, Schembri MA.** 2011. Insights into a multidrug resistant *Escherichia coli* pathogen of the globally disseminated ST131 lineage: Genome analysis and virulence mechanisms. PLoS ONE **6:**e26578.

4. **Tan C, Xu Z, Zheng H, Liu W, Tang X, Shou J, Wu B, Wang S, Zhao GP, Chen H.** 2011. Genome sequence of a porcine extraintestinal pathogenic *Escherichia coli* strain. J. Bacteriol. **193:**5038.

5. **Toh H, Oshima K, Toyoda A, Ogura Y, Ooka T, Sasamoto H, Park S-H, Iyoda S, Kurokawa K, Morita H, Itoh K, Taylor TD, Hayashi T, Hattori M.** 2010. Complete genome sequence of the wild-type commensal *Escherichia coli* strain SE15, belonging to phylogenetic group B2. J. Bacteriol. **192:**1165-1166.

6. **Fricke WF, Wright MS, Lindell AH, Harkins DM, Baker-Austin C, Ravel J, Stepanauskas R.** 2008. Insights into the environmental resistance gene pool from the genome sequence of the multidrug-resistant environmental isolate *Escherichia coli* SMS-3-5. J. Bacteriol. **190:**6779-6794.

7. **Luo C, Walk ST, Gordon DM, Feldgarden M, Tiedje JM, Konstantinidis KT.** 2011. Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. Proc. Natl. Acad. Sci. U.S.A. **108:**7200-7205.

8. **Campanella J, Bitincka L, Smalley J.** 2003. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinformatics **4:**29.

9. **Sullivan MJ, Petty NK, Beatson SA.** 2011. Easyfig: a genome comparison visualizer. Bioinformatics **27:**1009-1010.