

Insights into the phylogeny and evolution of cold shock proteins: from enteropathogenic *Yersinia* and *Escherichia coli* to eubacteria

Tao Yu ^{1,2,*}, Riikka Keto-Timonen ², Xiaojie Jiang ², Jussa-Pekka Virtanen ², Hannu Korkeala ²

¹ Department of Life Science and Technology, Xinxiang University, 453003 Xinxiang, China;
² Department of Food Hygiene and Environmental Health, University of Helsinki, FI-00014 Helsinki University, Finland; riikka.keto-timonen@helsinki.fi (R.K.-T.); xiaojie.jiang@helsinki.fi (X.J.); jussa-pekka.virtanen@helsinki.fi (J.-P.V.); hannu.korkeala@helsinki.fi (H. K.)
* Correspondence: yutao@xxu.edu.cn (T.Y.)

Table of Contents

1. Supplementary Figures
2. Supplementary Tables

1. Supplementary Figures

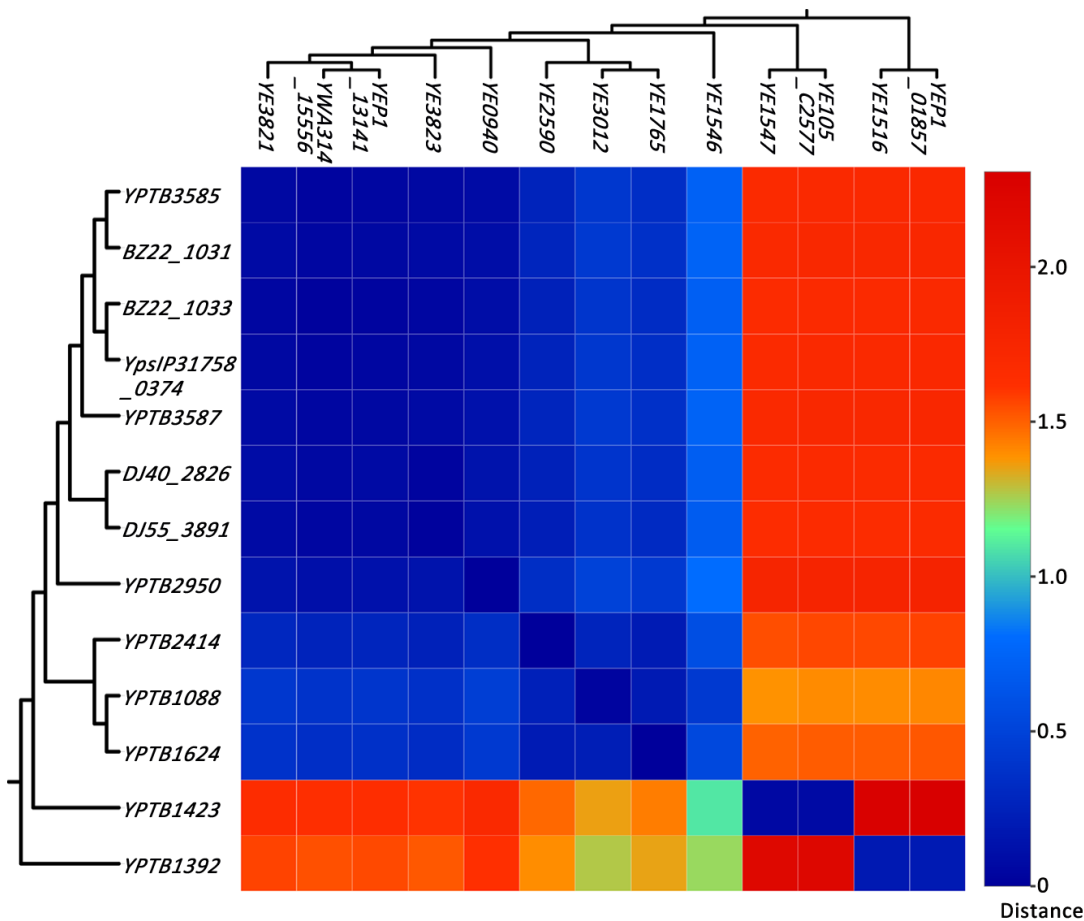


Figure S1. Heat map and hierarchical cluster showing the pairwise similarities between representative Csp proteins in enteropathogenic *Yersinia*. Cool colours (blue, green) indicate similar Csp pairs. Warm colours (red, yellow) indicate dissimilar Csp pairs.

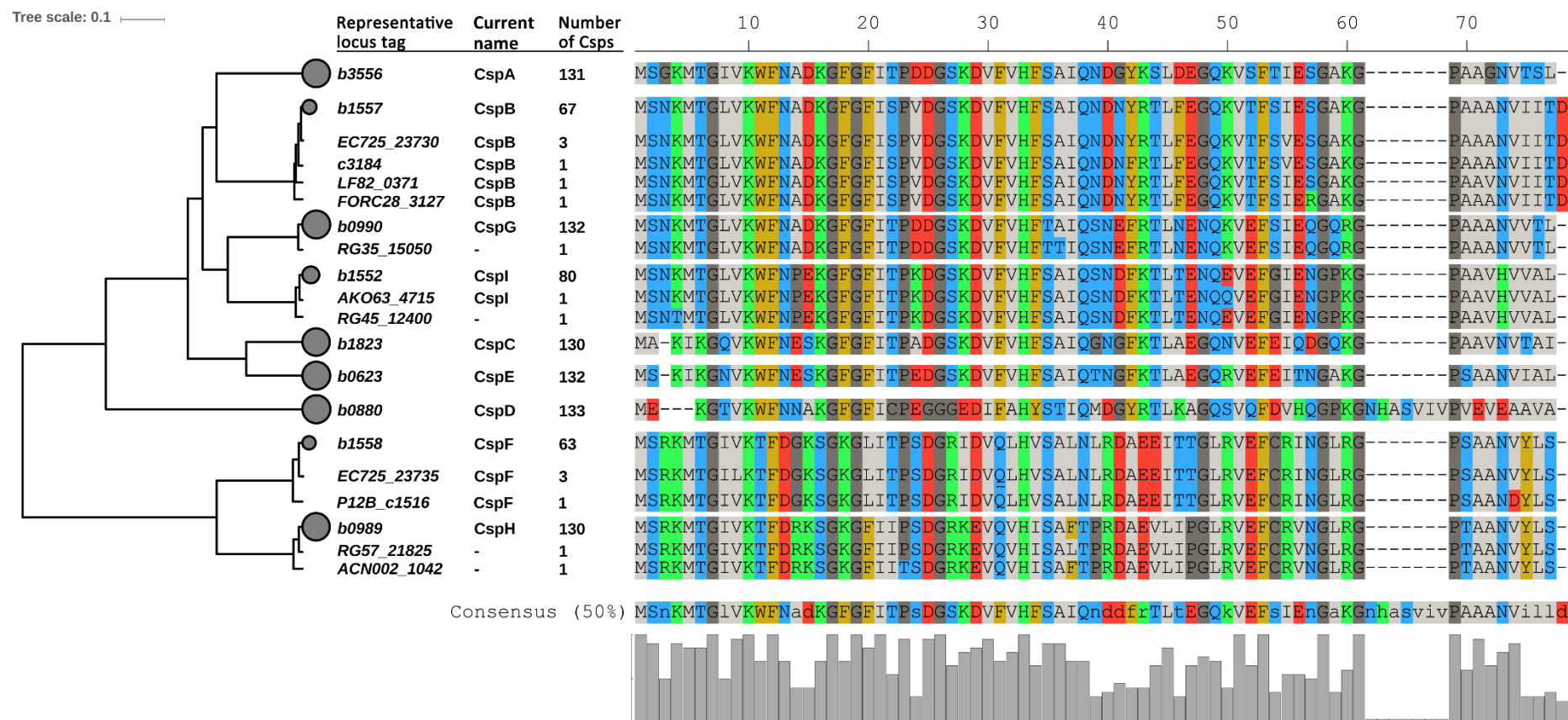


Figure S2. Phylogenetic tree for 20 Csp sequence patterns in *Escherichia coli*. A phylogenetic tree composed on 20 Csp sequence patterns in 135 *E. coli* strains. Monophyletic nodes have been collapsed and are represented by circles. The number of Csp in a collapsed node is indicated by the circle radius. Hyphens indicate *csp* genes with no current name. The histogram shows Csp sequence conservation.

Tree scale: 0.1

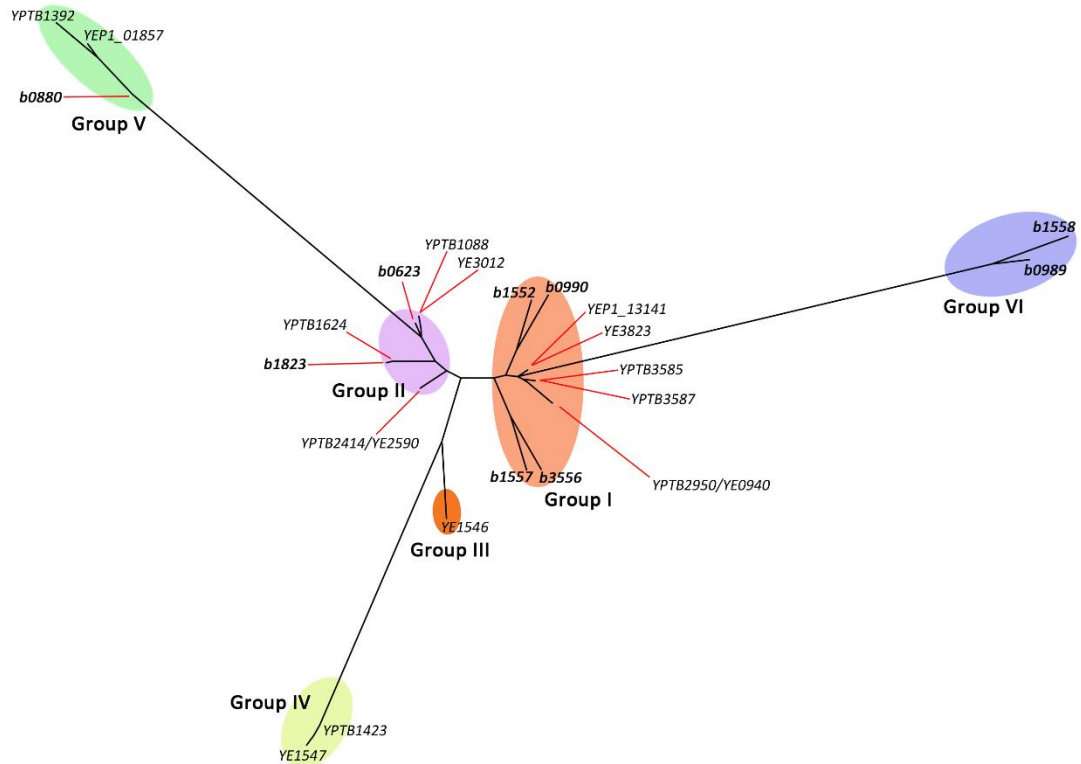
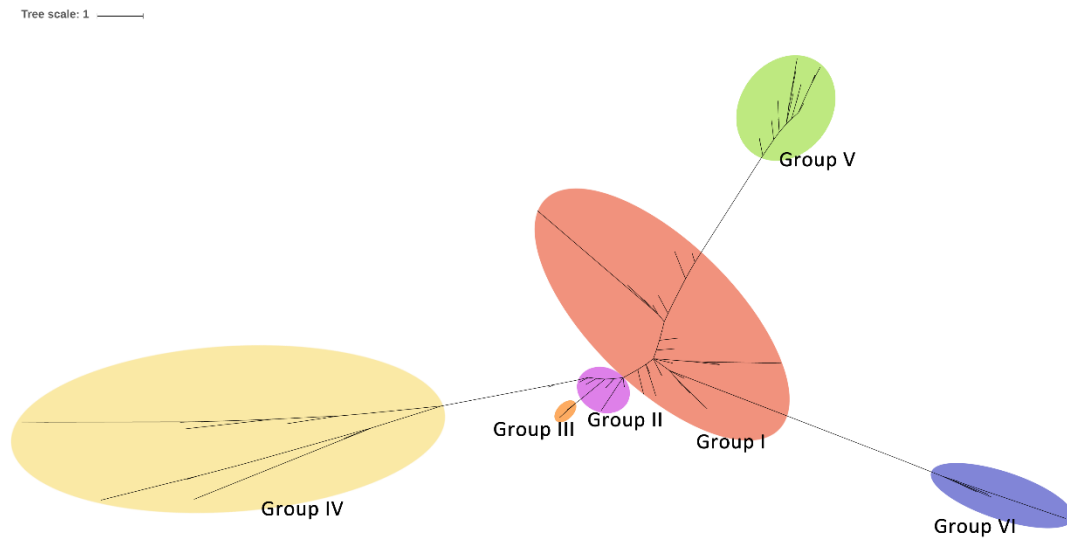


Figure S3. Phylogenetic tree for representative Csp in enteropathogenic *Yersinia* and *Escherichia coli*. A phylogenetic tree composed of the representative Csp in foodborne pathogenic *Yersinia* and *E. coli* (Abundance ratio > 0.3). The phylogenetic groups of representative Csp are represented by colour. The complete tree with full bootstrap values is available in Newick format as Supplementary Data 3.



27

28 **Figure S4.** Phylogenetic tree for all Csps in Enterobacteriales. A phylogenetic tree composed on 322 non-
 29 redundant Csps in Enterobacteriales, using cold shock domain sequences. The phylogenetic groups of
 30 Csps are represented by colour. The complete tree with full bootstrap values is available in Newick format
 31 as Supplementary Data 4.

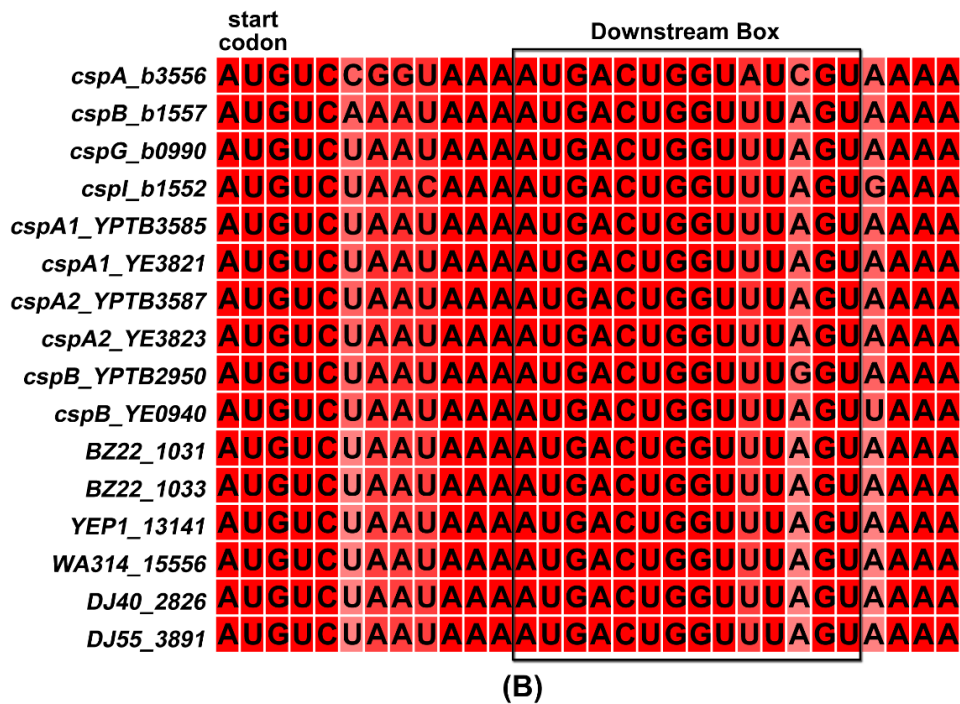
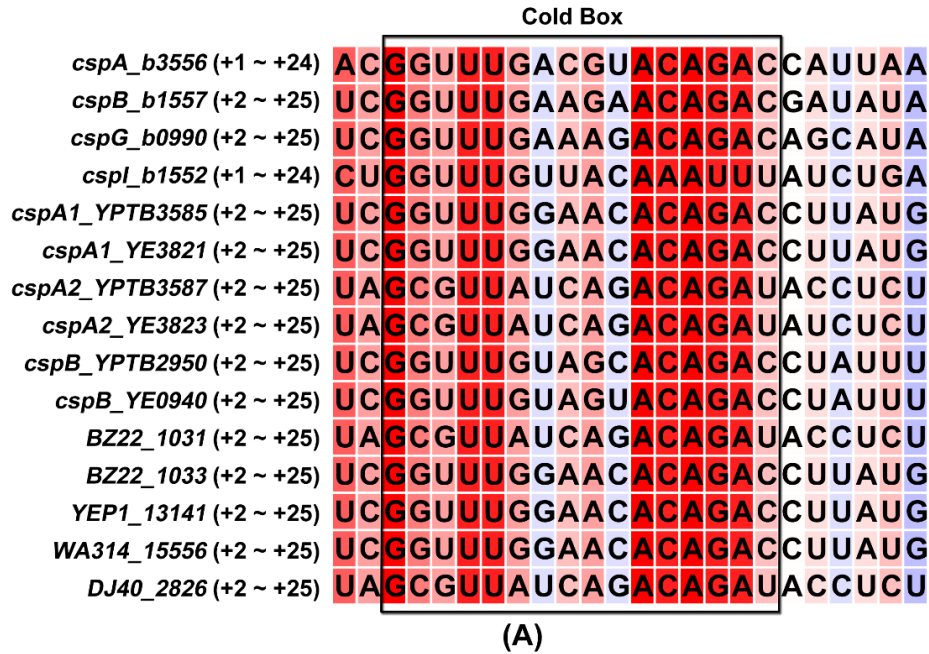


Figure S5. Messenger RNA Sequence comparison of four cold-inducible Csp's in *E. coli* and putative cold-inducible *csp* genes in enteropathogenic *Yersinia*. Messenger RNA Sequence alignment of the 5'-UTR (A) and the first 10 codon nucleotides (B) of cold-inducible *cspA* (*b3556*), *cspB* (*b1557*), *cspG* (*b0990*) and *cspI* (*b1552*) in *E. coli* and putative cold-inducible *csp* genes in enteropathogenic *Yersinia* (DJ55_3891 was not included in (A) due to the absence of 5'-UTR sequence information). The highly homologous sequences (cold box and downstream box) are boxed and labeled above the boxes. Red indicates conservative sites and blue indicates unconservative sites.

40 **2. Supplementary Tables**

41 **Table S1.** Information of Csps in *Yersinia enterocolitica* and *Yersinia pseudotuberculosis*

Species	No. of named Csp	No. of Strain	Representative strain
<i>Y. enterocolitica</i> subsp. <i>palearctica</i>	87	11	<i>Y. enterocolitica</i> subsp. <i>palearctica</i> YE-P1
<i>Y. enterocolitica</i> subsp. <i>enterocolitica</i>	17	2	<i>Y. enterocolitica</i> subsp. <i>enterocolitica</i> 8081
<i>Y. pseudotuberculosis</i>	343	45	<i>Y. pseudotuberculosis</i> IP 32953
Total	447	58	

42

Table S2. Information for each Csp pattern in enteropathogenic *Yersinia*.

Pattern ¹	Cluster	Species	Representative locus tag	Currently used name ^{2,3}	Abundance ratio ⁴
1	A	<i>Y. pseudotuberculosis</i>	YPTB3585	<i>cspA1</i>	0.84
2	A	<i>Y. pseudotuberculosis</i>	BZ22_1031	<i>cspG</i>	0.07
3	A	<i>Y. pseudotuberculosis</i>	BZ22_1033	<i>cspG</i>	0.07
4	A	<i>Y. pseudotuberculosis</i>	YpsIP31758_0374	-	0.09
5	A	<i>Y. pseudotuberculosis</i>	YPTB3587	<i>cspA2</i>	0.31
6	A	<i>Y. enterocolitica</i>	YE3821	<i>cspA1</i>	0.15
7	A	<i>Y. enterocolitica</i>	YEP1_13141	-	0.77
8	A	<i>Y. enterocolitica</i>	YWA314_15556	-	0.15
9	A	<i>Y. enterocolitica</i>	YE3823	<i>cspA2</i>	0.77
10	A	<i>Y. pseudotuberculosis</i>	DJ40_2826	<i>cspG</i>	0.16
11	A	<i>Y. pseudotuberculosis</i>	DJ55_3891	<i>cspG</i>	0.13
12	A	<i>Y. pseudotuberculosis</i>	YPTB2950	<i>cspB</i>	1.00
12	A	<i>Y. enterocolitica</i>	YE0940	<i>cspB</i>	1.00
13	B	<i>Y. pseudotuberculosis</i>	YPTB2414	<i>cspC</i>	0.96
13	B	<i>Y. enterocolitica</i>	YE2590	<i>cspC2</i>	1.00
14	B	<i>Y. pseudotuberculosis</i>	YPTB1624	<i>cspC</i>	1.02
14	B	<i>Y. enterocolitica</i>	YE1765	<i>cspC1</i>	0.15
15	B	<i>Y. pseudotuberculosis</i>	YPTB1088	<i>cspE</i>	1.00
16	B	<i>Y. enterocolitica</i>	YE3012	<i>cspE</i>	1.00
17	C	<i>Y. enterocolitica</i>	YE1546	<i>cspE2</i>	1.00
18	D	<i>Y. pseudotuberculosis</i>	YPTB1423	<i>cspB</i>	0.98
19	D	<i>Y. enterocolitica</i>	YE1547	<i>cspB</i>	0.92
20	D	<i>Y. enterocolitica</i>	YE105_C2577	-	0.08
21	E	<i>Y. pseudotuberculosis</i>	YPTB1392	<i>cspD</i>	1.00
22	E	<i>Y. enterocolitica</i>	YE1516	<i>cspD</i>	0.15

44 **Table S2.** Continued.

Pattern ¹	Cluster	Species	Representative locus tag	Currently used name ^{2,3}	Abundance ratio ⁴
23	E	<i>Y. enterocolitica</i>	<i>YEP1_01857</i>	-	0.85

45 ¹ Csps with same amino acid sequence.

46 ² Naming convention as in *E. coli*, with the exception of Csps in cluster C and D.

47 ³ Hyphen indicates a *csp* gene with no current name.

48 ⁴ Abundance ratio represents Csp number per strain for each Csp pattern.

49 **Table S3.** Consistency analysis of 104 Csp sequences in *Yersinia enterocolitica*

Pattern	Pattern size	Representative locus tag	Currently used name ^{1,2}	Length (aa)
6	2	YE3821	<i>cspA1</i>	70
7	10	YEP1_13141	-	70
8	2	YWA314_15556	-	70
9	10	YE3823	<i>cspA2</i>	70
12	13	YE0940	<i>cspB</i>	70
13	13	YE2590	<i>cspC2</i>	70
14	2	YE1765	<i>cspC1</i>	69
16	13	YE3012	<i>cspE</i>	69
17	13	YE1546	<i>cspE2</i> ³	69
19	12	YE1547	<i>cspB</i> ³	69
20	1	YE105_C2577	-	69
22	2	YE1516	<i>cspD</i>	85
23	11	YEP1_01857	-	85

50 ¹ Naming convention as in *E. coli* except YE1546 and YE1547.

51 ²Hyphen indicates a *csp* gene with no current name.

52 ³ Nomenclature is inappropriate.

53 **Table S4.** Consistency analysis of 343 Csp sequences in *Yersinia pseudotuberculosis*

Pattern	Pattern size	Representative locus tag	Currently used name ^{1,2}	Length (aa)
1	38	YPTB3585	<i>cspA1</i>	70
2	3	BZ22_1031	-	70
3	3	BZ22_1033	-	70
4	4	YpsIP31758_0374	-	70
5	14	YPTB3587	<i>cspA2</i>	70
10	7	DJ40_2826	-	70
11	6	DJ55_3891	-	70
12	45	YPTB2950	<i>cspB</i>	70
13	43	YPTB2414	<i>cspC</i>	70
14	46	YPTB1624	<i>cspC</i>	69
15	45	YPTB1088	<i>cspE</i>	69
18	44	YPTB1423	<i>cspB</i> ³	69
21	45	YPTB1392	<i>cspD</i>	87

54 ¹ Naming convention as in *E. coli* except YPTB1423.

55 ² Hyphen indicates a *csp* gene with no current name.

56 ³ Nomenclature is inappropriate.

57

Table S5. Consistency analysis of Csp sequences in 135 *Escherichia coli* strains

Pattern	Pattern size	Representative <i>csp</i> gene	Currently used name ¹	Length (aa)
1	131	<i>b3556</i>	<i>cspA</i>	70
2	67	<i>b1557</i>	<i>cspB</i>	71
3	3	<i>EC725_23730</i>	<i>cspB</i>	71
4	1	<i>c3184</i>	<i>cspB</i>	71
5	1	<i>LF82_0371</i>	<i>cspB</i>	71
6	1	<i>FORC28_3127</i>	<i>cspB</i>	71
7	130	<i>b1823</i>	<i>cspC</i>	69
8	133	<i>b0880</i>	<i>cspD</i>	74
9	132	<i>b0623</i>	<i>cspE</i>	69
10	63	<i>b1558</i>	<i>cspF</i>	70
11	3	<i>EC725_23735</i>	<i>cspF</i>	70
12	1	<i>P12B_c1516</i>	<i>cspF</i>	70
13	132	<i>b0990</i>	<i>cspG</i>	70
14	1	<i>RG35_15050</i>	-	70
15	130	<i>b0989</i>	<i>cspH</i>	70
16	1	<i>RG57_21825</i>	-	70
17	1	<i>ACN002_1042</i>	-	70
18	80	<i>b1552</i>	<i>cspI</i>	70
19	1	<i>AKO63_4715</i>	<i>cspI</i>	70
20	1	<i>RG45_12400</i>	-	70

58

¹Hyphen indicates a *csp* gene with no current name.

59

60

Table S6. Evolutionary distance between nine representative Csps in *Escherichia coli*

	CspA	CspB	CspC	CspD	CspE	CspF	CspG	CspH	CspI
CspA	0	0.28	0.61	1.38	0.50	1.49	0.39	1.38	0.43
CspB	0.28	0	0.59	1.36	0.48	1.47	0.37	1.36	0.41
CspC	0.61	0.59	0	0.89	0.23	1.62	0.52	1.51	0.56
CspD	1.38	1.36	0.89	0	1.00	2.39	1.29	2.28	1.33
CspE	0.50	0.48	0.23	1.00	0	1.51	0.41	1.40	0.45
CspF	1.49	1.47	1.62	2.39	1.51	0	1.40	0.25	1.44
CspG	0.39	0.37	0.52	1.29	0.41	1.40	0	1.29	0.26
CspH	1.38	1.36	1.51	2.28	1.40	0.25	1.29	0	1.33
CspI	0.43	0.41	0.56	1.33	0.45	1.44	0.26	1.33	0

61

62 **Table S7.** Groups of representative *csp* genes in enteropathogenic *Yersinia* and *E. coli*

Group	Representative <i>csp</i> gene(s) in <i>Y. enterocolitica</i> ¹	Representative <i>csp</i> gene(s) in <i>Y. pseudotuberculosis</i> ¹	Representative <i>csp</i> gene(s) in <i>E. coli</i> ¹	Currently used name(s) ²	RNP1 ³	RNP2 ³
I	YE3821, YE3823, YE0940, YEP1_13141, YWA314_15556, DJ40_2826, DJ55_3891	YPTB3585, YPTB3587, YPTB2950, BZ22_1031, BZ22_1033, YpsIP31758_0374	b3556, b1557, b0990, b1552	<i>cspA</i> , <i>cspB</i> , <i>cspG</i> , <i>cspI</i>	KGFGFI ^T / _S ^P	VFVHF
II	YE2590, YE1765, YE3012	YPTB2414, YPTB1624, YPTB1088	b1823, b0623	<i>cspC</i> , <i>cspE</i>	KGFGFITP	VFVHF
III	YE1546	-	-	<i>cspE2</i>	KGFGFIEQ	VFVHF
IV	YE1547, YE105_C2577	YPTB1423	-	<i>cspB</i>	EGYGFISP	VYV ^S / _N K
V	YE1516, YEP1_01857	YPTB1392	b0880	<i>cspD</i>	KGFGFICP	IFAHY
VI	-	-	b1558, b0989	<i>cspF</i> , <i>cspH</i>	SGKG ^F / _L I ^I / _T ^P	VQ ^V / _L H ^I / _V

63 ¹Hyphen indicates no representative *csp* gene in this species.

64 ² Naming convention as in *E. coli*, with the exception of CspS in group III and IV.

65 ³ RNA-binding motif.

Table S8. Distribution of strains included in Csp phylogenic analysis in Enterobacterales

Family	Genus	Strain number	Csp number	Csps per strain
Enterobacteriaceae	<i>Candidatus Moranella</i>	1	2	2.0
	<i>Candidatus Riesia</i>	1	1	1.0
	<i>Cedecea</i>	1	4	4.0
	<i>Citrobacter</i>	4	16	4.0
	<i>Cronobacter</i>	6	31	5.2
	<i>Enterobacter</i>	6	26	4.3
	<i>Escherichia</i>	5	34	6.8
	<i>Klebsiella</i>	2	12	6.0
	<i>Kluyvera</i>	2	10	5.0
	<i>Leclercia</i>	1	4	4.0
	<i>Raoultella</i>	2	13	6.5
	<i>Salmonella</i>	2	11	5.5
	<i>Shigella</i>	4	22	5.5
	<i>Shimwellia</i>	1	5	5.0
	<i>Yokenella</i>	1	5	5.0
Erwiniaceae	<i>Buchnera</i>	1	2	2.0
	<i>Erwinia</i>	7	27	3.9
	<i>Pantoea</i>	5	17	3.4
	<i>Tatumella</i>	2	6	3.0
	<i>Wigglesworthia</i>	1	1	1.0
Hafniaceae	<i>Edwardsiella</i>	2	14	7.0
	<i>Hafnia</i>	1	7	7.0
	<i>Obesumbacterium</i>	1	7	7.0
Morganellaceae	<i>Moellerella</i>	2	11	5.5
	<i>Photorhabdus</i>	3	17	5.7
	<i>Proteus</i>	2	8	4.0
	<i>Providencia</i>	6	37	6.2
	<i>Xenorhabdus</i>	5	24	4.8
Pectobacteriaceae	<i>Brenneria</i>	1	3	3.0
	<i>Dickeya</i>	2	8	4.0
	<i>Pectobacterium</i>	3	12	4.0
	<i>Sodalis</i>	1	3	3.0
Yersiniaceae	<i>Rahnella</i>	1	6	6.0
	<i>Serratia</i>	7	44	6.3
	<i>Yersinia</i>	12	84	7.0
Total		104	534	5.1