Supporting Information

Sims and Kim 10.1073/pnas.1105168108

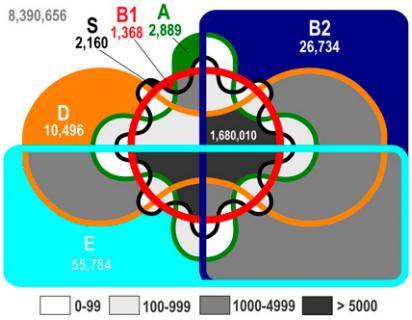

Fig. S1. Distinguishing *Escherichia coli*/*Shigella* phylogroup features. An Edwards' Venn diagram depicting the numbers of features present in all possible combinations of phylogroup intersections. The colored regions represent the characteristic features of each phylogroup. Distinguishing features are those features that are present in a single phylogroup but absent from all others. Other regions colored in grayscale represent groups of features that are shared by two or more phylogroups. Note the features used are universally present in all members of the phylogroups in each described region. The union of all sets contains 8,224,172 features, and the universal set contains 8,390,656.

Table S1. Feature counts

Phylogroup set	Feature count	Phylogroup set	Feature count	Phylogroup set	Feature count
V = A ∪ B1 ∪ B2 ∪ D ∪ E ∪ A	8,224,172	**V**′	166,484	**U** = Universal set	8,390,656
A	2,727,759	**B1**	2,848,252	**B2**	2,691,012
D	2,742,919	**E**	3,505,791	**S**	2,153,098
A ∩ (B1 ∪ B2 ∪ D ∪ E ∪ S)′	2,889	**A** ∩ **B1** ∩ **B2** ∩ (D ∪ E ∪ S)′	42	**A** ∩ **B1** ∩ **B2** ∩ **E** ∩ (D ∪ S)′	27
B1 ∩ (A ∪ B2 ∪ D ∪ E ∪ S)′	1,368	**A** ∩ **B1** ∩ **D** ∩ (B2 ∪ E ∪ S)′	107	**A** ∩ **B1** ∩ **B2** ∩ **S** ∩ (D ∪ E)′	360
B2 ∩ (A ∪ B1 ∪ D ∪ E ∪ S)′	26,734	**A** ∩ **B1** ∩ **E** ∩ (B2 ∪ D ∪ S)′	353	**A** ∩ **B1** ∩ **D** ∩ **E** ∩ (B2 ∪ S)′	237
D ∩ (A ∪ B1 ∪ B2 ∪ E ∪ S)′	10,496	**A** ∩ **B1** ∩ **S** ∩ (B2 ∪ D ∪ E)′	1,025	**A** ∩ **B1** ∩ **D** ∩ **S** ∩ (B2 ∪ E)′	840
E ∩ (A ∪ B1 ∪ B2 ∪ D ∪ S)′	55,784	**A** ∩ **B2** ∩ **D** ∩ (B1 ∪ E ∪ S)′	420	**A** ∩ **B1** ∩ **E** ∩ **S** ∩ (B2 ∪ D)′	5,819
S ∩ (A ∪ B1 ∪ B2 ∪ D ∪ E)′	2,160	**A** ∩ **B2** ∩ **E** ∩ (B1 ∪ D ∪ S)′	47	**A** ∩ **B2** ∩ **D** ∩ **E** ∩ (B1 ∪ S)′	191
A ∩ **B1** ∩ (B2 ∪ D ∪ E ∪ S)′	1,179	**A** ∩ **B2** ∩ **S** ∩ (B1 ∪ D ∪ E)′	0	**A** ∩ **B2** ∩ **D** ∩ **S** ∩ (B1 ∪ E)′	25
A ∩ **B2** ∩ (B1 ∪ D ∪ E ∪ S)′	255	**A** ∩ **D** ∩ **E** ∩ (B1 ∪ B2 ∪ S)′	67	**A** ∩ **B2** ∩ **E** ∩ **S** ∩ (B1 ∪ D)′	8
A ∩ **D** ∩ (B1 ∪ B2 ∪ E ∪ S)′	114	**A** ∩ **D** ∩ **S** ∩ (B1 ∪ B2 ∪ E)′	0	**A** ∩ **D** ∩ **E** ∩ **S** ∩ (B1 ∪ B2)′	19
A ∩ **E** ∩ (B1 ∪ B2 ∪ D ∪ S)′	348	**A** ∩ **E** ∩ **S** ∩ (B1 ∪ B2 ∪ D)′	80	**B1** ∩ **B2** ∩ **D** ∩ **E** ∩ (A ∪ S)′	98
A ∩ **S** ∩ (B1 ∪ B2 ∪ D ∪ E)′	20	**B1** ∩ **B2** ∩ **D** ∩ (A ∪ E ∪ S)′	120	**B1** ∩ **B2** ∩ **D** ∩ **S** ∩ (A ∪ E)′	234
B1 ∩ **B2** ∩ (A ∪ D ∪ E ∪ S)′	430	**B1** ∩ **B2** ∩ **E** ∩ (A ∪ D ∪ S)′	36	**B1** ∩ **B2** ∩ **E** ∩ **S** ∩ (A ∪ D)′	64
B1 ∩ **D** ∩ (A ∪ B2 ∪ E ∪ S)′	58	**B1** ∩ **B2** ∩ **S** ∩ (A ∪ D ∪ E)′	21	**B1** ∩ **D** ∩ **E** ∩ **S** ∩ (A ∪ B2)′	134
B1 ∩ **E** ∩ (A ∪ B2 ∪ D ∪ S)′	159	**B1** ∩ **D** ∩ **E** ∩ (A ∪ B2 ∪ S)′	13	**B2** ∩ **D** ∩ **E** ∩ **S** ∩ (A ∪ B1)′	143
B1 ∩ **S** ∩ (A ∪ B2 ∪ D ∪ E)′	149	**B1** ∩ **D** ∩ **S** ∩ (A ∪ B2 ∪ E)′	9	**A** ∩ **B1** ∩ **B2** ∩ **D** ∩ **E** ∩ (S)′	3,458
B2 ∩ **D** ∩ (A ∪ B1 ∪ E ∪ S)′	7,091	**B1** ∩ **E** ∩ **S** ∩ (A ∪ B2 ∪ D)′	377	**A** ∩ **B1** ∩ **B2** ∩ **D** ∩ **S** ∩ (E)′	15,391
B2 ∩ **E** ∩ (A ∪ B1 ∪ D ∪ S)′	1,414	**B2** ∩ **D** ∩ **E** ∩ (A ∪ B1 ∪ S)′	1,689	**A** ∩ **B1** ∩ **B2** ∩ **E** ∩ **S** ∩ (D)′	4,592
B2 ∩ **S** ∩ (A ∪ B1 ∪ D ∪ E)′	33	**B2** ∩ **D** ∩ **S** ∩ (A ∪ B1 ∪ E)′	102	**A** ∩ **B2** ∩ **D** ∩ **E** ∩ **S** ∩ (B1)′	529
D ∩ **E** ∩ (A ∪ B1 ∪ B2 ∪ S)′	3,088	**B2** ∩ **E** ∩ **S** ∩ (A ∪ B1 ∪ D)′	15	**A** ∩ **B1** ∩ **E** ∩ **D** ∩ **S** ∩ (B2)′	21,714
D ∩ **S** ∩ (A ∪ B1 ∪ B2 ∪ E)′	22	**D** ∩ **E** ∩ **S** ∩ (A ∪ B1 ∪ B2)′	28	**B1** ∩ **B2** ∩ **D** ∩ **E** ∩ **S** ∩ (A)′	2,630
E ∩ **S** ∩ (A ∪ B1 ∪ B2 ∪ D)′	150	**A** ∩ **B1** ∩ **B2** ∩ **D** ∩ (E ∪ S)′	259	**A** ∩ **B1** ∩ **B2** ∩ **D** ∩ **E** ∩ **S**	1,680,010

∩ is the intersection, ∪ is the union, and ′ represents the complement. V′ therefore indicates all features that are not present in any phylogroup, and **A** ∩ (B1 ∪ B2 ∪ D ∪ E ∪ S)′ indicates all features that are present in all phylogroup A species but absent in all other phylogroups.

Table S2. Phylogroup distinguishing features (l = 24) and associated genes

Group	Gene	Hits
S	*ipaH*, invasion plasmid antigen	15,636
	Phage tail fiber protein/phage integrase/misc phage component	5,846+
	Insertion sequence/transposase	3,674+
	SitB	231
	SitD	175
	tRNA-Arg phage associated (SF4456)	165
	SitA	112
	SitC	63
B1	*mvpA*, plasmid maintenance protein	590
	speF, ornithine decarboxylase	301
	Transcriptional regulation protein (ECE24377A_2756)	259
	YehB, putative outer membrane usher protein precursor (ECE24377A_1244)	175
	asnB, asparagine synthetase	161
	hopD, leader peptidase (ECE24377A_3806)	154
	mvpT, postsegregational killing toxin (ECE24377A_2756)	132
	eprK, type III secretion system lipoprotein precursor	114
	yidJ, sulfatase (ECE24377A_4186)	105
	ygiZ, conserved inner membrane protein	80
D	Invasion/intimin protein (EcoSMS35_4024)	2,532
	ydbA, autotransported OMP involved in cell adhesion (EcoSMS35_1767)	802
	IpfC, long polar fimbrial usher operon protein	789
	IpfD, long polar fimbrial usher operon protein	660
	Invasion protein regulator (EcoSMS35_4023)	906
	sipB, type III cell invasion protein	743
	sipD, type III effector protein	752
	Chromosome segregation protein (EcoSMS35_4086)	686
	Unknown function (EcoSMS35_4035)	613
	Exported protein (EcoSMS35_4687)	552

Other Supporting Information Files

Dataset S1 (TXT)
Dataset S2 (TXT)