Journal of Medical Microbiology (2007), **56**, 1460–1466 **DOI** 10.1099/jmm.0.47322-0

Multilocus sequence typing analysis of *Shigella flexneri* isolates collected in Asian countries

Seon Young Choi,[1,2] Yoon-Seong Jeon,[1,3] Je Hee Lee,[1,2] Boram Choi,[1] Sun Hwa Moon,[1] Lorenz von Seidlein,[1] John D. Clemens,[1] Gordon Dougan,[4] John Wain,[4] Jun Yu,[4] Je Chul Lee,[5] Sung Yong Seol,[5] Bok Kwon Lee,[6] Jae-Hoon Song,[7] Manki Song,[1] Cecil Czerkinsky,[1] Jongsik Chun[1,2,3] and Dong Wook Kim[1]

Correspondence
Dong Wook Kim
dwkim@ivi.int

[1]International Vaccine Institute, San 4-8 Bongcheon 7 dong, Kwanak gu, Seoul 151-818, Republic of Korea

[2]School of Biological Sciences, Seoul National University, Seoul 151-742, Republic of Korea

[3]Interdisciplinary Program of Bioinformatics, Seoul National University, Seoul 151-742, Republic of Korea

[4]The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK

[5]School of Medicine, Kyungpook National University, Daegu 700-422, Republic of Korea

[6]Laboratory of Enteric Infections, Department of Microbiology, National Institute of Health Korea, Republic of Korea

[7]Division of Infectious Disease, Samsung Medical Center, Seoul, Republic of Korea

Received 30 March 2007
Accepted 6 July 2007

The multilocus sequence typing scheme used previously for phylogenetic analysis of *Escherichia coli* was applied to 107 clinical isolates of *Shigella flexneri*. DNA sequencing of 3423 bp throughout seven housekeeping genes identified eight new allele types and ten new sequence types among the isolates. *S. flexneri* serotypes 1–5, X and Y were clustered together in a group containing many allelic variants while serotype 6 formed a distinct group, as previously established.

INTRODUCTION

Shigellosis, a disease characterized by the destruction of the colonic epithelium in humans, is caused by *Shigella* spp. The estimated annual incidence of shigellosis exceeds 150 million cases worldwide and it is responsible for 1 million deaths per year mainly in the developing world (Kotloff *et al.*, 1999). However, a recent report based on the disease burden and epidemiology of shigellosis in six Asian countries suggests that the actual number of deaths due to this infection might be less than previously assumed in this area (von Seidlein *et al.*, 2006).

Abbreviations: EIEC, enteroinvasive *E. coli*; MLST, multilocus sequence typing; ST, sequence type.

The GenBank/EMBL/DDBJ accession numbers for the nucleotide sequences of the new allele types identified in this study are EF364101–EF364108 (*icd*123, EF364101; *adk*116, EF364102; *fum*C149, EF364103; *fum*C145, EF364104; *pur*A98, EF364105; *pur*A96, EF364106; *pur*A97, EF364107; *pur*A95, EF364108).

A full list of clinical *S. flexneri* isolates subjected to MLST analysis in this study and a phylogenetic tree generated by the neighbour-joining method for *S. flexneri* are available as supplementary material with the online version of this paper.

A total of 50 serotypes based on O-antigenic structure are recognized within the genus. There are four subtypes or species: *Shigella dysenteriae* (subtype A), *Shigella flexneri* (subtype B), *Shigella boydii* (subtype C) and *Shigella sonnei* (subtype D). Twenty and 15 serotypes have been defined for *S. boydii* and *S. dysenteriae*, respectively. All *S. sonnei* strains fall into a single serotype and *S. flexneri* is characterized by 14 serotypes. Many typing methods, including multilocus enzyme electrophoresis and multilocus sequence typing (MLST) using different sets of genes, have shown that *Shigella* spp. belong to the *Escherichia coli* superfamily, consisting of three main clusters and a number of outliers (Escobar-Paramo *et al.*, 2003; Pupo *et al.*, 2000; Wirth *et al.*, 2006; Yang *et al.*, 2007). *S. flexneri* serotypes 1–5, X and Y belong to the main cluster 3 (C3) along with *S. boydii* 12. *S. flexneri* 6 and *S. boydii* 2 and 4 belong to the subcluster 3 (SC3) of the main cluster 1, while the majority of *S. dysenteriae* and *S. boydii* (1, 3, 6, 8, 10 and 18) belong to the subclusters 1 and 2 (SC1, SC2) of the main cluster 1, respectively. *S. flexneri* variant Y contains the basic O-antigen of tetrasaccharide *N*-acetylglucosamine–rhamnose–rhamnose–rhamnose, serotypes 1–5 and X have glucosyl and/or *O*-acetyl modifications

on those sugar residues and serotypes 6 and 6a have a different O-antigen structure from the others (Cheah *et al.*, 1991). The genome sequence analyses of two *S. flexneri* serotype 2a strains, 2457T and Sf301, one 5b serotype strain, Sf8401, *S. sonnei*, *S. dysenteriae* serotype 1 and *S. boydii* serotype 4 have been completed (Jin *et al.*, 2002; Nie *et al.*, 2006; Wei *et al.*, 2003; Yang *et al.*, 2005). The complete DNA sequences of the large virulence plasmid of *Shigella* species have also been published (Buchrieser *et al.*, 2000; Jiang *et al.*, 2005).

MLST of bacterial species was first reported in 1998, and today there are several schemes and public websites for more than 33 species in operation (Maiden, 2006). The evolution of *E. coli*, particularly the development of virulence and the origin of pathogenic strains, has been studied using MLST analysis (Wirth *et al.*, 2006). The free-access MLST database of *E. coli* is based on analysis of 462 globally distributed strains including the 72 ECOR collection, and data are continuously being added. By MLST analysis, pathogenic enterohaemorrhagic *E. coli*, enteropathogenic *E. coli*, enteroinvasive *E. coli* (EIEC) and *Shigella* appear to have originated independently and have been continually evolving from several lineages (Wirth *et al.*, 2006). *S. flexneri* can be categorized into two sequence type (ST) complexes, each containing a number of STs (Wirth *et al.*, 2006). According to this classification, *S. flexneri* serotypes 1–5 and variants X and Y make up the ST245 complex, and serotypes 6 and 6a belong to the ST243 complex. However, this study focused on various strains of *E. coli* in general and only one strain of each serotype of *Shigella* spp. was included. Detailed analysis of more clinical isolates is therefore warranted for phylogenetic analysis of *Shigella* spp. and hence we applied the MLST scheme to a sample of *S. flexneri* isolates collected mainly in Korea and Asian countries since the 1980s. We found that ST245 is common throughout *S. flexneri* serotypes 1–4 and X and confirmed that *S. flexneri* serotypes 6 and 6a formed a distinct group from other serotypes. Approximately 10 % of the *S. flexneri* isolates were represented by new STs and this suggests that there are many variable STs yet to be identified in this species.

METHODS

Strains. A total of 107 clinical *S. flexneri* isolates collected in Korea, China, the Philippines, Singapore, Sri Lanka and Taiwan were transported to the International Vaccine Institute, Seoul, Korea (Supplementary Table S1 in JMM Online). Sixty-seven isolates were collected throughout Korea from 1981 to 2000 (Jeong *et al.*, 2003; Kim *et al.*, 2004; Seol *et al.*, 2006), 17 were collected from Hebei Province, China, in 2002 (von Seidlein *et al.*, 2006), and 23 isolates were collected from the Philippines, Singapore, Sri Lanka and Taiwan in 2002. The isolates were grown at 37 °C overnight in trypticase soy broth or agar as required and serotyped by slide agglutination (Denka Seiken). The genome sequences of *S. flexneri* serotype 2a strain 2457T (Wei *et al.*, 2003), Sf301 (Jin *et al.*, 2002) and serotype 5b strain Sf8401 (Nie *et al.*, 2006) were entered into MLST analysis after generating allele types *in silico*. Two laboratory strains, *S. flexneri* serotype 2a YSH6000 (Ohya *et al.*, 2005) and serotype 5a M90T (Allaoui *et al.*, 1992), were also included in this analysis.

Genomic DNA preparation, DNA sequencing and MLST analysis. Genomic DNA was prepared from agar-grown cultures using the Prepman Ultra kit (Applied Biosystems). Seven housekeeping genes were amplified using the primers and PCR conditions described previously (http://web.mpiib-berlin.mpg.de/mlst/dbs/Ecoli/documents/primersColi_html) (Wirth *et al.*, 2006) for MLST analysis: *adk* (adenylate kinase), *fumC* (fumarate hydratase), *gyrB* (DNA gyrase), *icd* (isocitrate/isopropylmalate dehydrogenase), *mdh* (malate dehydrogenase), *purA* (adenylosuccinate dehydrogenase) and *recA* (ATP/GTP binding motif). The purified PCR products were sequenced in both directions using a Big Dye cycle sequencing kit (Applied Biosystems) and sequencing was performed on an ABI 3770 automatic sequencer according to the manufacturer's instructions. Sequence data for each isolate were added to a group of known sequences to be aligned simultaneously and edited through jPHYDIT (Jeon *et al.*, 2005). Allele type numbers were assigned by comparing the sequences to those in the *E. coli* MLST database. New ST numbers were assigned to novel STs that included new allele type(s) and/or new combination of allele types. Phylogenetic trees were constructed with combined sequences using BioNumerics software (Applied Maths) by the neighbour-joining method, based on distances estimated using the Jukes and Cantor coefficient (Holmquist *et al.*, 1972). Minimal spanning trees were generated in BioNumerics based on the degree of allele sharing.

RESULTS AND DISCUSSION

Eight new allele types (*adk*, 1; *fumC*, 2; *icd*, 1; *purA*, 4) were found among the 107 clinical isolates and 2 laboratory strains of *S. flexneri*. Moreover, 10 new STs were identified which contained the new allele type(s) and/or new combinations of known allele types. Two ST complexes were reported in *S. flexneri* previously as shown in Table 1 (Wirth *et al.*, 2006). ST145 (allele types 1,10,1,1,1,1,1 in the order of *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* and *recA*,) and ST262 (1,6,1,1,1,1,1) that fell in the ST243 complex were recognized in *S. flexneri* serotypes 6 and 6a. Other serotypes of *S. flexneri* were grouped in the ST245 complex (ST245 : 6,61,6,11,13,3,50). Nine new STs differing in one to three allele loci from ST245 were identified in this study (Tables 2 and 3). In the study of Wirth *et al.* (2006), ST245 was found in serotypes 1a, 1b, 3a, 3c, 4b and X of *S. flexneri*; this work confirms that a number of isolates of serotypes 2a, 2b, 3b, 4a and 4c (antigenic formula IV:7,8) also fell in this ST (Tables 2 and 3).

Novel STs associated with serotypes

The new STs found in this study and the variation sites of each allele are listed in Tables 3 and 4. One of the serotype 1b isolates (IB012) which originated in Korea in 1991 was typed as ST627 on the basis of variation in the *purA* locus (*purA*97) at nt 477 (T to C). Among serotype 2a isolates ST626 was found in a representative (IB047) from Taiwan, characterized by variation at nt 449 in the *adk* locus. A further new allele type, 149, was identified in the *fumC* locus of one of the serotype 2a isolates (IB037) from China. This isolate contained a repeat of six nucleotides

Table 1. Previously reported results of MLST analysis of *S. flexneri* strains (Wirth *et al.*, 2006)

Allele numbers for each locus are shown.

Strain	ST	ST complex	*adk*	*fumC*	*gyrB*	*icd*	*mdh*	*purA*	*recA*	Serotype*
M1362	245	ST245	6	61	6	11	13	3	50	1a
M1349	245	ST245	6	61	6	11	13	3	50	1b
M1383	268	ST245	6	61	1	11	13	3	50	2a
M1342	240	ST245	6	61	4	11	13	3	50	2b
M1363	245	ST245	6	61	6	11	13	3	50	3a
M1366	255	ST245	6	61	6	11	48	3	50	3b
M1377	245	ST245	6	61	6	11	13	3	50	3c
M1379	264	ST245	6	61	68	11	13	3	55	4a
M1361	245	ST245	6	61	6	11	13	3	50	4b
M1356	248	ST245	6	74	6	66	13	3	50	5
M1376	245	ST245	6	61	6	11	13	3	50	X
M1370	259	ST245	6	78	6	11	13	3	50	Y
15546/98	245	ST245	6	61	6	11	13	3	50	ND
M1382	145	ST243	1	10	1	1	1	1	1	6
16200/97	145	ST243	1	10	1	1	1	1	1	ND
M1375	262	ST243	1	6	1	1	1	1	1	6a

*ND, Serotype not determined.

(ATGAAC, nt 316–321) between nt 321 and 322 which gave rise to a new ST, ST651. A laboratory strain, YSH6000, of serotype 2a contained a new allele type, *purA*98, which differed in one nucleotide (position 633) from the common allele type *purA*3, resulting in ST633.

We analysed 17 serotype 3a isolates: one from the Philippines (in 2002) and 16 Korean isolates (1986 and 1987). The Philippines isolate was assigned ST628; this ST did not contain new allele types, but was composed of a new combination of previously recognized allele types.

Table 2. MLST results of 107 clinical isolates and 3 laboratory strains (YSH6000, M90T and Sf8401) of *S. flexneri* analysed in this study

ST complex	ST	Serotype	No. of isolates	*adk*	*fumC*	*gyrB*	*icd*	*mdh*	*purA*	*recA*
ST245	ST245	1a	2	6	61	6	11	13	3	50
		1b	10							
		2a	61*							
		2b	2							
		3b	1							
		4a	1							
		4c	2							
		X	3							
	ST626	2a	1	116	61	6	11	13	3	50
	ST651		1	6	149	6	11	13	3	50
	ST633		1	6	61	6	11	13	98	50
	ST627	1b	1	6	61	6	11	13	97	50
	ST628	3a	1	6	61	6	11	13	3	55
	ST629		16	6	145	6	11	13	3	55
	ST631	5a	1	6	74	6	123	13	3	50
	ST634	5b	1†	6	61	6	123	13	3	50
	ST630	Y	1	6	61	6	11	6	95	7
ST243	ST145	6	3	1	10	1	1	1	1	1
	ST632		1	1	10	1	1	1	96	1

*Serotype 2a 2457T and Sf301 genome sequenced strains belong to ST245 (Jin *et al.*, 2002; Wei *et al.*, 2003).
†Serotype 5b strain Sf8401 from the genome sequence data (Nie *et al.*, 2006).

Table 3. MLST STs found in *S. flexneri*

New STs identified in this study and ST245 and ST145 found in this study are shown in bold. STs previously reported but not found in this study are shown in plain text.

Serotype	ST	*adk*	*fumC*	*gyrB*	*icd*	*mdh*	*purA*	*recA*
1a	**245**	**6**	**61**	**6**	**11**	**13**	**3**	**50**
1b	**245**	**6**	**61**	**6**	**11**	**13**	**3**	**50**
	627	**6**	**61**	**6**	**11**	**13**	**97**	**50**
2a	**245**	**6**	**61**	**6**	**11**	**13**	**3**	**50**
	626	**116**	**61**	**6**	**11**	**13**	**3**	**50**
	651	**6**	**149**	**6**	**11**	**13**	**3**	**50**
	633	**6**	**61**	**6**	**11**	**13**	**98**	**50**
	268	6	61	1	11	13	3	50
2b	**245**	**6**	**61**	**6**	**11**	**13**	**3**	**50**
	240	6	61	4	11	13	3	50
3a	245	6	61	6	11	13	3	50
	628	**6**	**61**	**6**	**11**	**13**	**3**	**55**
	629	**6**	**145**	**6**	**11**	**13**	**3**	**55**
3b	**245**	**6**	**61**	**6**	**11**	**13**	**3**	**50**
	255	6	61	6	11	48	3	50
4a	**245**	**6**	**61**	**6**	**11**	**13**	**3**	**50**
	264	6	61	68	11	13	3	55
4c	**245**	**6**	**61**	**6**	**11**	**13**	**3**	**50**
5	248	6	74	6	66	13	3	50
5a	**631**	**6**	**74**	**6**	**123**	**13**	**3**	**50**
5b	**634**	**6**	**61**	**6**	**123**	**13**	**3**	**50**
6	**145**	**1**	**10**	**1**	**1**	**1**	**1**	**1**
	632	**1**	**10**	**1**	**1**	**1**	**96**	**1**
	262	1	6	1	1	1	1	1
X	**245**	**6**	**61**	**6**	**11**	**13**	**3**	**50**
Y	**630**	**6**	**61**	**6**	**11**	**6**	**95**	**7**
	259	6	78	6	11	13	3	50

Allele *recA*55, which was found in serotype 4a in a previous study (Wirth *et al.*, 2006), was identified in this isolate. The Korean serotype 3a also contained *recA*55 but had an additional new allele, *fumC*145, and defined ST629.

All other serotypes of *S. flexneri* contained *icd*11 while serotype 5, 5a, and 5b isolates contained *icd*66 or *icd*123, which have common variation sites at positions 207 and 258 but differ at position 625 (Table 4); M90T, a laboratory strain of serotype 5a, contained allele type *icd*123, which differed by two nucleotides (207 and 258) from the common allele type *icd*11 of the ST245 complex. Allele *fumC*74, with a variation site at nt 578, was also identified in M90T, resulting in a new ST, ST631. *S. flexneri* 5b strain Sf8401, for which the genome sequence is available (Nie *et al.*, 2006), contained the same allele type *icd*123 as M90T but differed from M90T in the *fumC* locus, and was assigned to ST634. The *icd* locus might be useful for differentiating serotype 5 from other serotypes or for further analyses such as single nucleotide polymorphism analysis for *S. flexneri*. Although ST245 is common in other serotypes of *S. flexneri* as indicated above, it was not found in serotype 5 and subtype isolates (M1356, M90T and Sf8401) investigated here (Allaoui *et al.*, 1992; Nie *et al.*, 2006; Wirth *et al.*, 2006). Nevertheless, since serotype 5 strains are rarely isolated (3 of 1976 clinical isolates in the report of von Seidlein *et al.*, 2006), further clinical isolates should be examined before the absence of ST245 can be confirmed.

The single isolate of serotype Y tested was assigned to ST630 characterized by allele *mdh*6, which was found by Wirth *et al.* (2006) in five strains of *S. boydii* (ST149 complex) and one EIEC strain (ST250 complex), but not in *S. flexneri*. A new allele type of the *purA* locus was found in ST630 (*purA*95) and *recA*7 that was not previously found in *S. flexneri* but was present in all *S. sonnei* strains and some *S. dysenteriae* strains (Wirth *et al.*, 2006). Although ST630 cannot be included in the ST245 complex since ST245 and ST630 differed by more than three loci, it remains clear that the two STs are closely related.

We identified a new allele type, *purA*96, in a serotype 6 *S. flexneri* isolate from Taiwan (C to T at nt 453), resulting in an ST632. The STs of four serotype 6 isolates, including ST632, fell in the ST243 complex, which is distinct from the ST245 complex. Some previously reported STs were not found in this study (Table 3), perhaps due to the geographical difference and limited duration of the strain collection.

Phylogenetic analysis

The distribution of each of the *S. flexneri* STs in the minimal spanning tree is shown in Fig. 1, and the phylogenetic tree generated by the neighbour-joining method is shown in supplementary Fig. S1 in JMM Online. Both analyses showed that the ST245 complex and the ST243 complex are distinct groups. Since serotype Y is considered a parental form of other serotypes, analysis of more Y isolates may provide helpful information on the origin of *S. flexneri*. Given that ST245 is found in various serotypes and several STs could be recognized within a serotype, categorizing each serotype according to the STs was not possible. However, serotype 5 and subtype strains constitute a separate lineage within the ST245 complex by both analysis methods.

Along with the chromosomal loci analysis, the evolution of the virulence plasmid of *Shigella* and EIEC has been extensively studied (Lan *et al.*, 2001, 2003; Yang *et al.*, 2007). Here, we focused on the chromosomal loci according to the MLST analysis scheme of *E. coli*. It is important to combine the analysis of the chromosomal loci as well as the analysis of the virulence plasmid in the future for better understanding of the origin and pathogenicity of *Shigella* spp. Two hypotheses are suggested for the origin of *Shigella* spp. from the ancestral *E. coli*. Escobar-Paramo *et al.* (2003) concluded from the sequence analysis results of comparison of four chromosomal genes and three virulence plasmid genes that a single transfer of the virulence plasmid occurred into an ancestral *E. coli*, which diversified into *Shigella* and EIEC (Escobar-Paramo *et al.*,

Table 4. Variable site alignment of the MLST alleles of *S. flexneri*

STs found in this study are shown in bold. Nucleotide position numbers above the sequences are counted from the translation start codon of each gene.

ST complex	ST	*adk*	*fumC*	*gyrB*	*icd*	*mdh*	*purA*	*recA*	Serotype
		44	2344555	44466777777	223333355566	233	446678	3345	
		49	4656017	05626011356	050134505802	302	573661	3613	
		92	3055408	26079814566	783862722265	413	373333	9315	
245	**245**	GG	GTCAAGC	TTCTACCACTC	CTCGTCTTCCTC	CCA	CTGCCA	TTCC	
	240	..	.....	CC.CG..GTCT	.....	...	.....	....	2b
	248	..	.....T	.....	TC.....T	...	.....	....	5
	255	..	.....	.....	.....	.T.	.....	....	3b
	259	..	A.....	.....	.....	...	.....	....	Y
	264	..	.....	CATCGT.GT..	.....	...	.....	.C.T	4a
	268	..	.....	C...G.TG.CT	.....	...	.....	....	2a
	626	A.	.....	.....	.....	...	.....	....	2a
	627	..	.....	.....	.....	...	.C....	....	1b
	628	..	.....	.....	.....	...	.....	.C.T	3a
	629	..	..T....	.....	.....	...	.....	.C.T	3a
	630	..	.....	.....	.....	..C	....T.	.C..	Y
	631	..	.....T	.....	TC.....	...	.....	....	5a
	633	..	.....	.....	.....	...	..A...	....	2a
	634	..	.....	.....	TC.....	...	.....	....	5b
243	145	.A	.C.GGA.	C...G.TG.CT	..TTCGCAGTG.	T.C	...T.C	CCT.	6
	262	.A	.C.G...	C...G.TG.CT	..TTCGCAGTG.	T.C	...T.C	CCT.	6a
	632	.A	.C.GGA.	C...G.TG.CT	..TTCGCAGTG.	T.C	T..T.C	CCT.	6
Codon position		23	3323332	33133333331	333333313331	312	333313	3331	

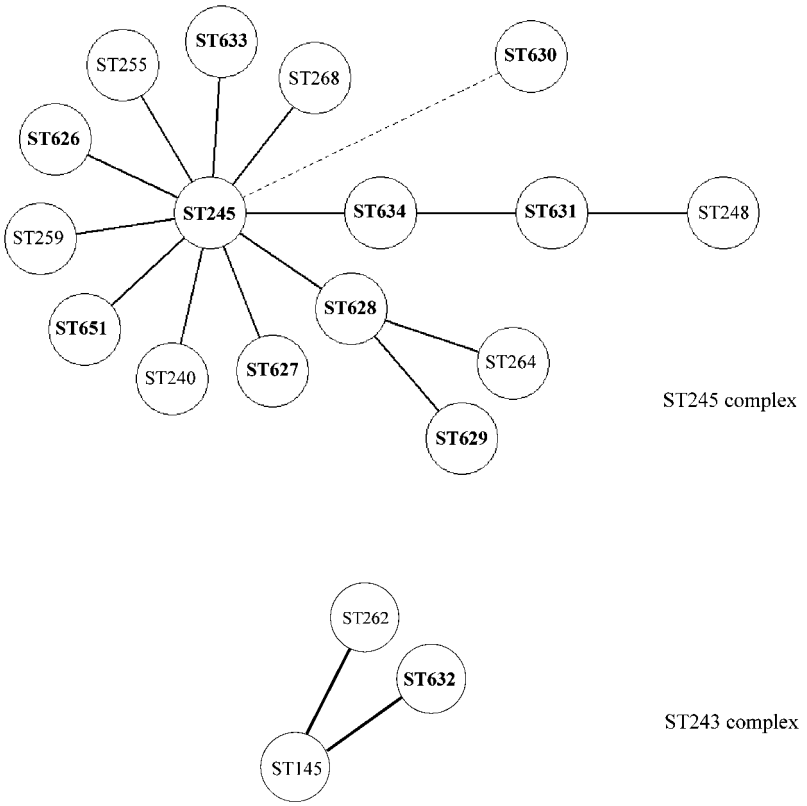

Fig. 1. Distribution of STs within two ST complexes (ST245 and ST243) of *S. flexneri* in a minimal spanning tree. Thick black lines indicate the connection between STs sharing six alleles. The dotted line between ST245 and ST630 indicates that they share four alleles. STs identified in this study are shown in bold. ST243, which includes *S. boydii* 1, 3, 6, 8, 10 and 18 and *S. dysenteriae* 5 and 7, is not shown in the ST243 complex tree.

2003). In contrast, Reeves and colleagues proposed multiple origins of *Shigella* from the MLST results of eight housekeeping genes (Lan & Reeves, 2002; Pupo *et al.*, 2000), and, most recently, Yang *et al.* (2007) suggested a compromise hypothesis of lateral transfer of the virulence plasmid among *E. coli*, EIEC and *Shigella*. However, those studies included only single or two representatives of each serotype and used different analysis methods. To understand the origin of *Shigella* spp., a unified and more expanded analysis method is essential as well as the inclusion of more strains. Since the diversity of ST complexes of *Shigella* spp. is less complicated compared to other pathogenic *E. coli* (enterohaemorrhagic *E. coli* and enteropathogenic *E. coli*) (Wirth *et al.*, 2006), it would be feasible to analyse the evolutionary history of *Shigella* spp. starting with closely related groups or species such as *S. flexneri* and *S. sonnei* and expanding to other groups or species. A good model would be the report of the evolutionary history of *Salmonella enterica* serovar Typhi (Roumagnac *et al.*, 2006), in which the sequence variations of 200 gene fragments from 105 representative globally collected strains of *Salmonella* Typhi were analysed. A similar approach was applied to an *E. coli* collection (Hommais *et al.*, 2005). For the application of such analysis methods, it is important to have a set of globally and chronologically collected representative *Shigella* strains. We expect that some isolates from this study will provide useful information for further analysis. We have collected around 3000 isolates of *Shigella* spp. anticipating identification of the more informative strains among them.

IncHI1 plasmid

Wei *et al.* (2003) reported that pSf-R27, a *Salmonella enterica* serovar Typhi R27-like plasmid, was found in *S. flexneri* 2a 2457T during genome sequencing of this strain, but not in the other 142 isolates (Wei *et al.*, 2003). PCR primers (incH-F, CGAAATCGGTCCAACCCATTG; incH-R, CGACAACTCATCAGAAGCGTCAAC) specific to repHI1A of R27 (Wain *et al.*, 2003) were used to screen for the presence of pSf-R27 among our *S. flexneri* isolates; however, no isolates contained this plasmid.

ACKNOWLEDGEMENTS

We thank Philippe J. Sansonetti and Claude Parsot (Institut Pasteur, Paris, France) for providing strain M90T and C. Sasakawa (Tokyo University, Japan) for strain YSH6000. This work was supported by the International Vaccine Institute, the Swedish International Development Cooperation Agency (SIDA) and the Korean Ministry of Education. S.-Y. C., J. H. L. and D. W. K. were supported by grant R01-2006-000-10255-0 from the Basic Research Program of the Korea Science and Engineering Foundation. G. D., J. W. and J. Y. are funded by the Wellcome Trust of Great Britain.

REFERENCES

Allaoui, A., Mounier, J., Prevost, M. C., Sansonetti, P. J. & Parsot, C. (1992). *icsB*: a *Shigella flexneri* virulence gene necessary for the lysis of protrusions during intercellular spread. *Mol Microbiol* **6**, 1605–1616.

Buchrieser, C., Glaser, P., Rusniok, C., Nedjari, H., D'Hauteville, H., Kunst, F., Sansonetti, P. & Parsot, C. (2000). The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. *Mol Microbiol* **38**, 760–771.

Cheah, K. C., Beger, D. W. & Manning, P. A. (1991). Molecular cloning and genetic analysis of the rfb region from *Shigella flexneri* type 6 in *Escherichia coli* K-12. *FEMS Microbiol Lett* **67**, 213–218.

Escobar-Paramo, P., Giudicelli, C., Parsot, C. & Denamur, E. (2003). The evolutionary history of *Shigella* and enteroinvasive *Escherichia coli* revised. *J Mol Evol* **57**, 140–148.

Holmquist, R., Cantor, C. & Jukes, T. (1972). Improved procedures for comparing homologous sequences in molecules of proteins and nucleic acids. *J Mol Biol* **64**, 145–161.

Hommais, F., Pereira, S., Acquaviva, C., Escobar-Paramo, P. & Denamur, E. (2005). Single-nucleotide polymorphism phylotyping of *Escherichia coli*. *Appl Environ Microbiol* **71**, 4784–4792.

Jeon, Y. S., Chung, H., Park, S., Hur, I., Lee, J. H. & Chun, J. (2005). jPHYDIT: a JAVA-based integrated environment for molecular phylogeny of ribosomal RNA sequences. *Bioinformatics* **21**, 3171–3173.

Jeong, Y. S., Lee, J. C., Kang, H. Y., Yu, H. S., Lee, E. Y., Choi, C. H., Tae, S. H., Lee, Y. C., Cho, D. T. & Seol, S. Y. (2003). Epidemiology of nalidixic acid resistance and TEM-1- and TEM-52-mediated ampicillin resistance of *Shigella sonnei* isolates obtained in Korea between 1980 and 2000. *Antimicrob Agents Chemother* **47**, 3719–3723.

Jiang, Y., Yang, F., Zhang, X., Yang, J., Chen, L., Yan, Y., Nie, H., Xiong, Z., Wang, J. & other authors (2005). The complete sequence and analysis of the large virulence plasmid pSS of *Shigella sonnei*. *Plasmid* **54**, 149–159.

Jin, Q., Yuan, Z., Xu, J., Wang, Y., Shen, Y., Lu, W., Wang, J., Liu, H., Yang, J. & other authors (2002). Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. *Nucleic Acids Res* **30**, 4432–4441.

Kim, S., Kim, J., Kang, Y., Park, Y. & Lee, B. (2004). Occurrence of extended-spectrum β-lactamases in members of the genus *Shigella* in the Republic of Korea. *J Clin Microbiol* **42**, 5264–5269.

Kotloff, K. L., Winickoff, J. P., Ivanoff, B., Clemens, J. D., Swerdlow, D. L., Sansonetti, P. J., Adak, G. K. & Levine, M. M. (1999). Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. *Bull World Health Organ* **77**, 651–666.

Lan, R. & Reeves, P. R. (2002). *Escherichia coli* in disguise: molecular origins of *Shigella*. *Microbes Infect* **4**, 1125–1132.

Lan, R., Lumb, B., Ryan, D. & Reeves, P. R. (2001). Molecular evolution of large virulence plasmid in *Shigella* clones and enteroinvasive *Escherichia coli*. *Infect Immun* **69**, 6303–6309.

Lan, R., Stevenson, G. & Reeves, P. R. (2003). Comparison of two major forms of the *Shigella* virulence plasmid pINV: positive selection is a major force driving the divergence. *Infect Immun* **71**, 6298–6306.

Maiden, M. C. (2006). Multilocus sequence typing of bacteria. *Annu Rev Microbiol* **60**, 561–588.

Nie, H., Yang, F., Zhang, X., Yang, J., Chen, L., Wang, J., Xiong, Z., Peng, J., Sun, L. & other authors (2006). Complete genome sequence of *Shigella flexneri* 5b and comparison with *Shigella flexneri* 2a. *BMC Genomics* **7**, 173.

Ohya, K., Handa, Y., Ogawa, M., Suzuki, M. & Sasakawa, C. (2005). IpgB1 is a novel *Shigella* effector protein involved in bacterial invasion of host cells: its activity to promote membrane ruffling via RAC1 and CDC42 activation. *J Biol Chem* **280**, 24022–24034.

Pupo, G. M., Lan, R. & Reeves, P. R. (2000). Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution

of many of their characteristics. *Proc Natl Acad Sci U S A* **97**, 10567–10572.

Roumagnac, P., Weill, F. X., Dolecek, C., Baker, S., Brisse, S., Chinh, N. T., Le, T. A., Acosta, C. J., Farrar, J. & other authors (2006). Evolutionary history of *Salmonella* Typhi. *Science* **314**, 1301–1304.

Seol, S. Y., Kim, Y. T., Jeong, Y. S., Oh, J. Y., Kang, H. Y., Moon, D. C., Kim, J., Lee, Y. C., Cho, D. T. & Lee, J. C. (2006). Molecular characterization of antimicrobial resistance in *Shigella sonnei* isolates in Korea. *J Med Microbiol* **55**, 871–877.

von Seidlein, L., Kim, D. R., Ali, M., Lee, H., Wang, X., Thiem, V. D., Canh do, G., Chaicumpa, W., Agtini, M. D. & other authors (2006). A multicentre study of *Shigella* diarrhoea in six Asian countries: disease burden, clinical manifestations, and microbiology. *PLoS Med* **3**, e353.doi:10.1371/journal.pmed.0030353

Wain, J., Diem Nga, L. T., Kidgell, C., James, K., Fortune, S., Song Diep, T., Ali, T., Ó Gaora, P., Parry, C. & other authors (2003). Molecular analysis of incHI1 antimicrobial resistance plasmids from *Salmonella* serovar Typhi strains associated with typhoid fever. *Antimicrob Agents Chemother* **47**, 2732–2739.

Wei, J., Goldberg, M. B., Burland, V., Venkatesan, M. M., Deng, W., Fournier, G., Mayhew, G. F., Plunkett, G., III, Rose, D. J. & other authors (2003). Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. *Infect Immun* **71**, 2775–2786.

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L. H., Karch, H., Reeves, P. R., Maiden, M. C. & other authors (2006). Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol* **60**, 1136–1151.

Yang, F., Yang, J., Zhang, X., Chen, L., Jiang, Y., Yan, Y., Tang, X., Wang, J., Xiong, Z. & other authors (2005). Genome dynamics and diversity of *Shigella* species, the etiologic agents of bacillary dysentery. *Nucleic Acids Res* **33**, 6445–6458.

Yang, J., Nie, H., Chen, L., Zhang, X., Yang, F., Xu, X., Zhu, Y., Yu, J. & Jin, Q. (2007). Revisiting the molecular evolutionary history of *Shigella* spp. *J Mol Evol* **64**, 71–79.