大韓獸醫學會誌 (2015) 第 55 卷 第 1 號
Korean J Vet Res(2015) 55(1) : 31~39
http://dx.doi.org/10.14405/kjvr.2015.55.1.31

<Original Article>

rpoB gene sequencing for phylogenetic analysis of avian pathogenic *Escherichia coli*

Hyuk-Joon Kwon[1,2,3,*], Won-Jin Seong[4], Tae-Eun Kim[5], Yong-Jin Won[6], Jae-Hong Kim[2,4,*]

[1]Laboratory of Poultry Clinical Science, [2]Research Institute for Veterinary Science, [3]Farm Animal Clinical Training & Research Center, and [4]Laboratory of Avian Diseases, College of Veterinary Medicine and BK21 for Veterinary Science, Seoul National University, Seoul 151-742 Korea
[5]BioPOA Co., Yongin 446-598, Korea
[6]Division of EcoScience, Ewha Womans University, Seoul 120-750, Korea

(Received: November 6, 2014; Accepted: January 8, 2015)

Abstract : The present study was conducted to determine the full *rpoB* and eight house-keeping gene sequences of 78 and 35, respectively, avian pathogenic *E. coli* (APEC) strains. Phylogenetic comparison with 66 *E. coli* and *Shigella* strains from GenBank and EMBL was also conducted. Based on the full *rpoB* sequence, 50 different *rpoB* sequence types (RSTs) were identified. RST 1 was assigned to a major RST that included 34.7% (50/144) of the analyzed strains. RST 2 to RST 50 were then assigned to other strains with higher nucleotide sequence similarity to RST 1 in order. RST 1, 11, and 23 were mixed with APEC along with human commensal and pathogenic strains while RST 2, 6, 9, 13-15, 22, 24, 25, 33, 34, 36, and 41 were unique to APEC strains. Only five APEC strains grouped into RST 32 and 47, which contained human pathogenic *E. coli* (HPEC). Thus, most of the APEC strains had genetic backgrounds different from HPEC strains. However, the minor APEC strains similar to HPEC should be considered potential zoonotic risks. The resolution power of multi-locus sequence typing (MLST) was better than RST testing. Nevertheless, phylogenetic analysis of *rpoB* was simpler and more economic than MLST.

Keywords : avian pathogenic *Escherichia coli*, *rpoB*, multi-locus sequence typing, phylogenetic analysis

Introduction

Escherichia (*E.*) *coli* strains have been divided into four major phylogenetic groups (A, B1, B2, and D) on the basis of phylogenetic analysis of multilocus enzyme electrophoresis (MLEE) alleles and a rapid and simple triplex polymerase chain reaction (PCR) technique [1, 30]. Human extraintestinal pathogenic *E. coli* (ExPEC) such as uropathogenic and neonatal meningitis *E. coli* strains was classified mainly into group B2 and, to a lesser extent, group D, while most commensal strains belonged to groups A and B1 [4, 14, 19, 29, 33]. The extraintestinal infection of avian pathogenic *E. coli* (APEC) induces colibacillosis in chickens, which is characterized by polyserositis, septicemic shock, and cellulitis [6, 11]. In contrast to human *E. coli*, the group A is most frequent in APEC strains, and phylogenetic group and clinical severity are also unrelated [7, 9, 10, 14, 19, 26, 33]. Recently pan- and core genomes were identified by comparing *E. coli* and *Shigella spp.* genomes and a hierarchical clustering of the variable genes resulted in clear separation of *E. coli* strains into pathotypes and clinically relevant serotypes [23]. But, full genome sequencing and analyses remain expensive and time-consuming. Multilocus sequence typing (MLST) has been developed for the understanding of global epidemiology and population structure of infectious agents [24]. For MLST of *E. coli*, 6-8 housekeeping genes are amplified and the 6-8 amplicons are purified and sequenced using both PCR primers (12-16 reactions). In addition, relatively high recombination rates of housekeeping genes can blur the interpretation of phylogenetic analysis results [13, 16, 32]. Therefore, a simple and economical alternative phylogenetic scheme with high discriminative power and free of recombination bias is needed. Among the minimal essential genes, partial or complete RNA polymerase beta subunit gene (*rpoB*) has been applied to phylogenetic analysis of various bacteria and differentiation of certain serotypes of *Salmonella enterica* [1, 31]. In this study, we determined the complete *rpoB* sequences of 78 and 8 house-keeping genes (*dinB*, *icdA*, *pabB*, *polB*, *putP*, *trpA*, *trpB* and *uidA*) sequences of 35 APEC strains which had been isolated from 1985-2005 from Korean chickens suffering from colibacillosis (Table 1), conducted a phylogenetic analysis together with 66 *E. coli*

*Corresponding authors
Tel: +82-2-880-1250, Fax: +82-2-880-1233
E-mails: kwonhj01@snu.ac.kr (Kwon HJ), kimhong@snu.ac.kr (Kim JH)

Table 1. RNA polymerase beta subunit gene (*rpoB*) sequences used in this study

Strain	Accession number	Strain	Accession number	Strain	Accession number	Strain	Accession number
E3	JN707606	E66	JN707643	E139	JN707680	FRIK966	ACXN01000121
E5	JN707607	E68	JN707644	E140	JN707681	FVEC1412	GG749233
E10	JN707608	E69	JN707645	E141	JN707682	H10407	FN649414
E11	JN707609	E70	JN707646	E147	JN707683	HS	CP000802
E13	JN707610	E80	JN707647	11368	AP010953	IAI1	CU928160
E15	JN707611	E84	JN707648	042	FN554766	IAI39	CU918264
E17	JN707612	E86	JN707649	101-1	AAMK02000037	IHE3034	CP001969
E18	JN707613	E89	JN707650	11128	A010960	KO11	CP002516
E20	JN707614	E90	JN707651	12009	AP010958	LF82	CU651637
E21	JN707615	E91	JN707652	536	CP000247	MS196-1	ADUD01000221
E22	JN707616	E95	JN707653	53638	AAKB02000001	NA114	CP002797
E23	JN707617	E101	JN707654	55989	CU928145	NRG 857C	CP001855
E24	JN707618	E102	JN707655	ABU83972	CP001671	OP50	ADBT01000723
E25	JN707619	E103	JN707656	APEC O1	CP000468	S88	CU928161
E26	JN707620	E105	JN707657	ATCC8739	CP000946	Sakai	BA000007
E27	JN707621	E107	JN707658	B088	GG749152	SB Sb227*	CP000036
E28	JN707622	E109	JN707659	B171	AAJX02000048	SB CDC3083-94	CP001063
E29	JN707623	E110	JN707660	B185	GG749189	SD Sd197†	CP000034
E31	JN707624	E111	JN707661	B7A	AAJT02000064	SD 1012	AAMJ01000024
E32	JN707625	E112	JN707662	BL21 (DE3)	AM946981	SE11	AP009240
E33	JN707626	E113	JN707663	BW2952	CP001396	SE15	AP009378
E35	JN707627	E117	JN707664	CB9615	CP001846	SF 2a 2457T‡	NC 004741
E36	JN707628	E119	JN707665	CFT073	AE014075	SF 2a 301	AE005674
E38	JN707629	E120	JN707666	DH1(K12)	CP001637	SF 5 8401	CP000266
E39	JN707630	E124	JN707667	E110019	AAJW02000004	SMS-3-5	CP000970
E42	JN707631	E125	JN707668	E22	AAJV02000051	SS Ss046	CP000038
E43	JN707632	E126	JN707669	E2348/69	FM180568	TB182A	EU891633
E48	JN707633	E127	JN707670	E24377A	CP000800	TW14359	C0001368
E49	JN707634	E129	JN707671	EC4042	ABHM0200001	UM146	CP002167
E51	JN707635	E130	JN707672	EC4045	ABHL02000001	UMN026	CU928163
E52	JN707636	E131	JN707673	EC4115	CP001164	UMNK88	CP002729
E53	JN707637	E132	JN707674	EC4206	ABHK02000001	UTI89	CP000243
E55	JN707638	E133	JN707675	ED1a	CP928162	W	CP002185
E59	JN707639	E135	JN707676	EDL933	AE005174	W3110(K12)	AP009048
E61	JN707640	E136	JN707677	ATCC 35469	CU928158		
E64	JN707641	E137	JN707678	F11	AAJU02000047		
E65	JN707642	E138	JN707679	FRIK2000	ACXO01000202		

SB: *Shigella boydii,* SD: *Shigella dysenteriae,* SF: *Shigella flexneri.*

and *Shigella* strains and a strain of *E. fergusonii* (ATCC 35469) in the GenBank (National Center for Biotechnology Information, USA) and EMBL (European Molecular Biology Laboratory, Germany) sequence databases to understand population structure of them, then evaluated whether it can be an alternative method of MLST.

Materials and Methods

Primers

rpoB genes of *E. coli* strains in the GenBank and EMBL were compared with Bioedit version 5.0.9.1 (Ibis Biosciences, USA) and a primer set to amplify the complete coding region of *rpoB* and seven primers for sequencing were

Table 2. Primers for *rpoB* amplification and sequencing

Primer	SequeEC4115nce (5' to 3')	Location*	Usage
P1	ATGGTTTACEC4206TCCTATACCGAGAA	1-23	PCR
P2	CAGATCCTYGACCTGTTCTT	655-674	Sequencing
P1R	CACTTTACCGTTAGCTTCGAT	783-763	Sequencing
P3	GAAGTTCAACCGTTCTCTGCT	1209-1229	Sequencing
P4	ATCGAAGAAGGCAACTACGTT	1825-1845	Sequencing
P5	CGAAGACTCCATCCTCGTAT	2436-2455	Sequencing
P6	AACAGCTGGCTGAGCAGTAT	3035-3054	Sequencing
P7	GAAGTTATGCGTCTGGCTGA	3502-3521	Sequencing
P8R	TTACTCGTCTTCCAGTTCGATGT	4029-4007	PCR

* The nucleotides of *rpoB* coding region were numbered from start codon to stop codon.

designed manually (Table 2).

PCR and sequencing

Total APEC DNA was extracted with a G-spin genomic DNA extraction kit (iNtRON Biotechnology, Korea) in accordance with the manufacturer's instructions. The PCR solution was composed of 10 × buffer (5 μL), dNTPs (2.5 mM, 1 μL), forward and reverse primers (10 pmol/μL, 1 μL each), Taq DNA polymerase (5 U/μL, 1 μL; Macrogen, Korea), distilled water (40 μL), and template DNA (50 ng/μL, 1 μL). Cycling conditions were 94°C, 3 min; 40 cycles at 94°C, 30 sec; 54°C, 20 sec; and 72°C, 4.5 min; and a final extension step at 72°C, 7 min. Amplicons were analyzed through electrophoresis on 1.0% agarose gels, and a 1 kb ladder was the molecular size marker (Macrogen). The PCR amplicons were purified using a PCRquick Spin Kit (iNtRON Biotechnology) in accordance with the manufacturer's instructions and were sequenced with an ABI3711 automatic sequencer (Macrogen).

Sequence analysis

The *in silico* phylogenetic grouping of *E. coli* and *Shigella* strains in GenBank was accomplished by querying sequences of forward and reverse primers of *chuA*, *yjaA*, and TspE4 in the BLASTN suite [2, 5]. Previously we revealed some mutations in yjaA.1 (the fifth C→T from the 3'-end), yjaA.2 (the fourth G→A from the 5'-end), and TspE4.C2 (the fifth C→T from the 5'end and the third A→G from the 3'-end) primers, and the mutated sequences were also queried [14]. According to the presence or absence of *chuA*, *yjaA*, and TspE4, the phylogenetic group was determined [5].

The determined overlapping sequences were assembled into a single complete sequence using ChromasPro version 1.5 (Technelysium, Australia). All complete *rpoB* sequences available in the GenBank and EMBL sequence databases were collected by a BLAST search with full coding region *rpoB* sequence of K12 W3110 strain (AP009048) (Table 1). Nucleotide similarity, variable nucleotide comparison, and translation of *rpoB* nucleotides were performed with Bioedit version 5.0.9.1 (Ibis Biosciences). Although the primer regions (1-23 and 4007-4029) of APEC strains were not determined the nucleotide sequences of 67 reference strains in the GenBank and EMBL databases were conserved. Therefore, the primer sequences were included to generate complete *rpoB* sequences of APEC strains for phylogenetic analysis. Phylogenetic analysis was performed with MEGA version 5 software (neighbor-joining method with Tamura-Nei distance and 1,000 repeats of bootstrapping). Shimodaira-Hasegawa test was performed with PAUP (ver. 4.0; Sinauer Associates, USA). The accession numbers of *E. coli* strains are summarized in Table 1.

To compare MLST with *rpoB* phylogenetic analysis, the nucleotide sequences of eight housekeeping genes of 35 APEC strains were determined as previous [13] and those of 38 references *E. coli* strains were extracted from the GenBank and EMBL sequence databases [13]. The genes were *dinB* (DNA polymerase), *icdA* (isocitrate dehydrogenase), *pabB* (p-aminobenzoate synthase), *polB* (polymerase PolII), *putP* (proline permease), *trpA* (tryptophan synthase subunit A), *trpB* (tryptophan synthase subunit B), and *uidA* (beta-glucuronidase). The nucleotide sequences were queried to the database of the MLST Pasteur website (Genotyping of Pathogens and Public Health of Institut Pasteur, France), and their allele profiles and STs were determined. The phylogenetic tree was constructed based upon eight concatenated housekeeping genes as *rpoB* [13].

Results

Complete *rpoB* sequences of APEC strains

The complete coding region of *rpoB* of *E. coli* was successfully amplified by the PCR primer set and the sequencing primer set was specific enough to clearly read the whole region. For nucleotide analysis, the primer sequences of the forward and reverse PCR primers were included. The complete coding region of *rpoB* of *E. coli* was 4,029 nucleotides including the stop codon. The nucleotide similarities between *E. coli* strains were 98.4%~100%. The numbers of variable nucleotide loci were 180, and synonymous and nonsynonymous nucleotide changes were 175 and 5, respectively.

Among the tested APEC strains, E132 has an amino acid change, R994C, and GenBank strains ED1a, OP50, 536, and LF82 possessed single amino acid changes (P288S, D711N, A1043S, and A1263E, respectively).

In silico phylogenetic grouping

Phylogenetic groups of the reference strains were determined by the presence of nucleotide sequences of primers for phylogenetic grouping of *E. coli*. Forty-seven reference strains were grouped into phylogenetic group A (14/47), B1 (9/47), B2 (15/47), and D (9/47). The phylogenetic group of each reference strain is depicted in Fig. 1. All strains of *Shigella* (*S.*) *boydii*, *S. flexneri*, and *S. sonnei* were group A, but *S. dysenteriae* Sd197 was group B2.

Phylogenetic analysis

The 78 APEC and 66 *E. coli* and *Shigella* strains in the GenBank and EMBL sequence databases were grouped into 50 RSTs (RST 1-RST 50) on the basis of phylogenetic analysis of the full *rpoB* nucleotide sequences (Fig. 1). RST 1 was assigned to a major RST covering 34.5% (50/145) of analyzed strains and RST 2 to RST 50 were assigned to other strains in order with higher nucleotide sequence similarity to RST 1. The RST numbers are preliminary except RST 1, but RST code is sequence specific. The RST code of RST 1 is RST-0-1. The middle and end numbers of the RST code denote the number of nucleotide difference from RST 1 and identifier, respectively. The RST code, prototype strain, and the number of member were summarized in Table 3, and for better understanding RST numbers were used instead of RST codes in the manuscript. RSTs 11 (13 strains), 47 (8), 35 (6), 23 (5), 38 (5), 32 (4), 25 (3), 2 (2), 4 (2), 5 (2), 15 (2), 22 (2), 33 (2), 40 (2), and 42 (2) contained more than one strain, and other RSTs contained only one strain. RST 1, 11, and 23 were mixed with APEC, and human commensal and virulent strains, but RSTs 2, 6, 9, 13, 14, 15, 22, 24, 25, 33, 34, 36, and 41 were unique to APEC strains (Table 3). Only five APEC strains grouped into RSTs 32 (1) and 47 (4), which contained virulent human *E. coli* strains (Fig. 1).

The separation of phylogenetic groups, A, B1, B2, and D was unclear in the *rpoB* tree. The phylogenetic groups of APEC strains in the present study were defined in the previous work [14]. RST 1 was composed of diverse phylogenetic groups, and A, B1, B2, and D comprised 17, 15, 8, and 2 strains, respectively. RST 11 and 23 were composed mainly of group A, but E35 and E90 were group D. RST 12, 13, 17-21, 24, and 31 were composed of group A, and RST 2-5 and 8-10 were composed of B1. RST 25, 27, 40, 42, 43, and 45-50 were composed of group B2, and RST 22, 32, 33, and 35-39 were composed of group D.

Out of 35 APEC strains all 8 amplicons were available from 33 strains but 2 strains (E39 and E59) were lack of *icdA* and *dinB* amplicons, respectively. The fourteen APEC and five reference strains showed one or two nucleotide mismatches in one or two house-keeping genes, and the closest

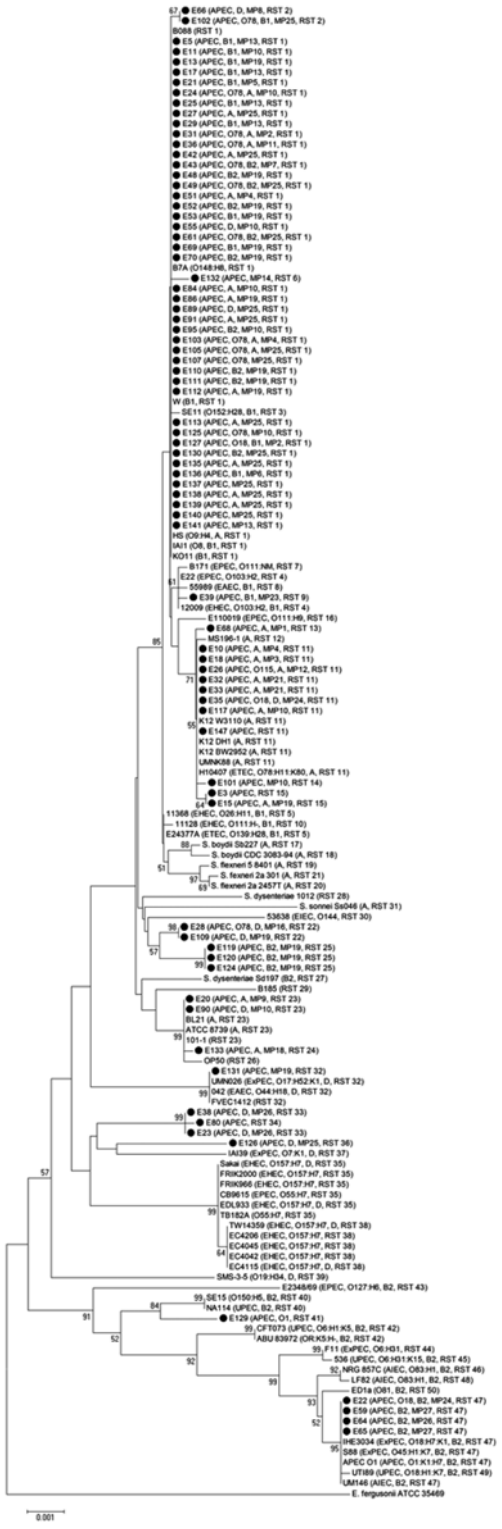

Fig. 1. Phylogenetic analysis of complete *rpoB* of *E. coli* strains. Phylogenetic analysis was performed with MEGA version 5 software (neighbor-joining method with Tamura-Nei distance and 1,000 repeats of bootstrapping). Phylogenetic groups (A, B1, B2 and D), serotypes, molecular pathotypes (MP1-MP27) and *rpoB* sequence types (RST1-RST50) were represented in parenthesis.

Table 1. *rpoB* sequence type (RST) system

RST number	RST code	Number of nucleotide difference	Prototype	Member (ST/phylogenetic group);[Number of member]
1	RST-0-1	0	E89	E24 (7/A), E31 (7/A), E43(7/B2), E49 (7/B2), E61 (7/B2), E70 (7/B2), E103 (7/A), E105 (7/A), E48 (66/B2), E36 (66/A), E55 (88/D), E89 (88/D), KO11 (360/B1), W (360/B1), E69 (NEW1/B1), E127 (NEW2/B1), IAI1 (NEW3/B1), HS (NEW4/A), E5 (nd*/B1), E11 (nd/B1), E13 (nd/B1), E21 (nd/B1), E17 (nd/B1), E27 (nd/A), E29 (nd/B1), E25 (nd/B1), E42 (nd/A), E51 (nd/A), E52 (nd/B2), E53 (nd/B1), E86 (nd/A), E84 (nd/A), E91 (nd/A), E95 (nd/B2), E107 (nd/nd), E110 (nd/B2), E111 (nd/B2), E112 (nd/A), E113 (nd/A), E125 (nd/nd), E130 (nd/B2), E135 (nd/A), E136 (nd/B1), E137 (nd/nd), E138 (nd/A), E139 (nd/A), E140 (nd/nd), E141 (nd/nd), B088 (nd/nd), B7A (nd/nd) [50]
2	RST-1-1	1	E66	E102 (21/B1), E66 (NEW/5/D) [2]
3	RST-1-2	1	SE11	SE11 (19/B1) [1]
4	RST-1-3	1	E22(EPEC)	E22(nd/nd), 12009 (135/B1) [2]
5	RST-1-4	1	E24377A	E24377A (468/B1), 11368 (481/B1) [2])
6	RST-2-1	2	E132	E132 (nd/nd) [1]
7	RST-2-2	2	B171	B171 (nd/nd) [1]
8	RST-2-3	2	55989	55989 (290/B1) [1]
9	RST-2-4	2	E39	E39 (nd/B1) [1]
10	RST-2-5	2	11128	11128 (480/B1) [1]
11	RST-3-1	3	E26	E26 (446/A), E10 (nd/A), E18 (nd/A), E32 (nd/A), E33 (nd/A), E35 (nd/D), E117 (nd/A), E147 (155/nd), K12 BW2952 (2/A), K12 DH1 (nd/A), K12 W3110 (nd/A), H10407 (132/A), UMNK88 (NEW6/A) [13]
12	RST-4-1	4	MS196-1	MS196-1 (nd/nd) [1]
13	RST-4-2	4	E68	E68 (2/A) [1]
14	RST-4-3	4	E101	E101 (NEW7/nd) [1]
15	RST-4-4	4	E15	E15 (NEW8/A), E3 (nd/nd) [2]
16	RST-4-5	4	E110019	E110019 (nd/nd) [1]
17	RST-5-1	5	*S. boydii* Sb227	*S. boydii* Sb227 (nd/A) [1]
18	RST-6-1	6	*S. boydii* CDC3083-94	*S. boydii* CDC3083-94 (nd/A) [1]
19	RST-6-2	6	*S. flexneri* 5 8401	*S. flexneri* 5 8401 (nd/A) [1]
20	RST-6-3	6	*S. flexneri* 2a 2457T	*S. flexneri* 2a 2457T (nd/A) [1]
21	RST-7-1	7	*S. flexneri* 2a 301	*S. flexneri* 2a 301 (nd/A) [1]
22	RST-8-1	8	E28	E28 (305/D), E109 (533/D) [2]
23	RST-9-1	9	E90	E90 (83/D), BL21 (83/A), E20 (398/A), ATCC8739 (NEW9/A), 101-1 (nd/nd) [5]
24	RST-10-1	10	E133	E133 (nd/A) [1]
25	RST-11-1	11	E119	E119 (NEW10/B2), E120 (NEW10/B2), E124 (NEW10/B2) [3]
26	RST-11-2	11	OP50	OP50 (nd/nd) [1]
27	RST-11-3	11	*S. dysenteriae* Sd197	*S. dysenteriae* Sd197 (nd/B2) [1]
28	RST-15-1	15	*S. dysenteriae* 1012	*S. dysenteriae* 1012 (nd/nd)[1]
29	RST-17-1	17	B185	B185 (nd/nd) [1]
30	RST-18-1	18	53638	53638 (nd/nd) [1]
31	RST-20-1	20	*S. sonnei* Ss046	*S. sonnei* Ss046 (nd/A) [1]
32	RST-21-1	21	E131	E131 (nd/nd), UMN026 (3/D), 042 (NEW11), FVEC1412 (nd/nd) [4]
33	RST-24-1	24	E38	E38 (48/D), E23 (NEW12/D) [2]
34	RST-25-1	25	E80	E80 (nd/nd) [1]

Table 3. continued

RST number	RST code	Number of nucleotide difference	Prototype	Member (ST/phylogenetic group);[Number of member]
35	RST-28-1	28	Sakai	Sakai (296/D), EDL933 (296/D), CB9615 (553/nd), FRIK966 (nd/nd), FRIK2000 (nd/nd), TB182A (nd/nd)[6]
36	RST-28-2	28	E126	E126 (45/D) [1]
37	RST-28-3	28	IAI39	IAI39 (254/D) [1]
38	RST-29-1	29	TW14359	TW14359 (NEW13/D), EC4115 (NEW13/D), EC4042 (nd/nd), EC4045 (nd/nd), EC4206 (nd/nd) [5]
39	RST-31-1	31	SMS-3-5	SMS-3-5 (39/D) [1]
40	RST-35-1	35	SE15	SE15 (506/B2), NA114 (43/B2) [2]
41	RST-36-1	36	E129	E129 (nd/nd) [1]
42	RST-40-1	40	CFT073	CFT073 (4/B2), ABU 83972 (4/B2) [2]
43	RST-43-1	43	E2348/69	E2348/69 (491/B2) [1]
44	RST-47-1	47	F11	F11 (nd/nd) [1]
45	RST-48-1	48	536	536 (NEW14/B2) [1]
46	RST-49-1	49	NRG857C	NRG857C (64/B2) [1]
47	RST-49-2	49	E22	E22 (1/B2), E59 (nd/B2), E64 (1/B2), E65 (1/B2), IHE3034 (1/B2), S88 (1/B2), UM146 (1/B2), APEC O1 (418/B2) [8]
48	RST-50-1	50	LF82	LF82 (64/B2) [1]
49	RST-50-2	50	UTI89	UTI89 (1/B2) [1]
50	RST-50-3	50	ED1a	ED1a (149/B2) [1]

nd: not determined, *S.*: *Shigella*.

allele was used to determine the STs. The sixteen APEC (8 strains) and reference strains (8 strains) were classified into 14 new STs. The RST 1 (18 strains), 2 (2 strains), 5 (2 strains), 11 (5 strains), 22 (2 strains), 23 (4 strains), 32 (2 strains), 33 (2 strains), 35 (3 strains), 40 (2 strains), 42 (2 strains), and 47 (7 strains) were further divided into several STs and the tested 30 RSTs were divided into 49 STs (Table 3).

Although the global tree topologies of *rpoB* and MLST (MLS tree) were turned out to be significantly different (SH test, $p < 0.001$), the local clustering of the APEC and reference *E. coli* strains in the *rpoB* tree was, however, notably similar to the MLS tree (Fig. 2). APEC strains of RST 1 were indistinguishable each other in *rpoB* tree but they formed several different clusters in MLS tree. RST 1, 2, 3, 4, 5, 8, 10, 11, 13, 14 and 15 formed a cluster in both trees. RST 23 strains formed a cluster and RST 22 and 25 formed a cluster in both trees. RST 40, 41, 42, 43, 45, 46, 47, 49 and 50 formed a cluster in both trees. K12 strains and B strains formed different clusters in both phylogenetic trees. Pathogenic O157:H7 and O55:H7 strains formed a cluster in both trees, but TW14359, EC4206, EC4045, EC4042, and EC4115 were a single nucleotide different from Sakai, FRIK966, FRIK2000, CB9615, EDL933 and TB182A, respectively in *rpoB* sequence. The *rpoB* sequences of *Shigella spp.* were closely related to *E. coli*, *S. boydii and S. flexneri* strains formed different cluster with each other, but *S. dysenteriae* strains 1012 and Sd197 did not form a cluster.

Discussion

The nucleotide sequence variability of *rpoB* among the compared *E. coli* and *Shigella* strains was 98.4-100%. The *rpoB* sequence similarity for speciation is 98%, and *Shigella* strains could be classified into *E. coli* as previous [1]. Mutations in *rpoB* can cause rifampicin-resistance and can bias phylogenetic analysis with partial sequences including rifampicin-resistance determining region (RRDR) [12]. But, since the *rpoB* sequences of *E. coli* and *Shigella* strains presently showed enough variability and no mutations in the RRDR, they may be useful for phylogenetic analysis of *E. coli*.

E. coli strains have been divided into four phylogenetic groups. The B2 group is regarded as pathogenic, in contrast to A and B1, which are mainly commensals. Previous studies have reported frequencies of A, B1, B2, and D groups in APEC strains as 34.5%~71.0%, 4.1%~23.3%, 7.9%~44.5%, and 12.0%~29.9%, respectively, and, on average, group A is the most frequent in APEC strains [7, 9, 10, 14, 19, 33]. The relatively high frequencies of A group in APEC and human commensals, and of B2 and D in human ExPEC, may reflect different pathogenicity evolution of APEC from human *E. coli* strains. But, recently restricted distribution of ExPEC in certain phylogenetic groups was explained by sampling bias because most of studies used selected ExPEC strains [3, 15, 17, 21, 25]. A MLST study that employed unselected human bacteremic strains revealed unrestricted distribution into all

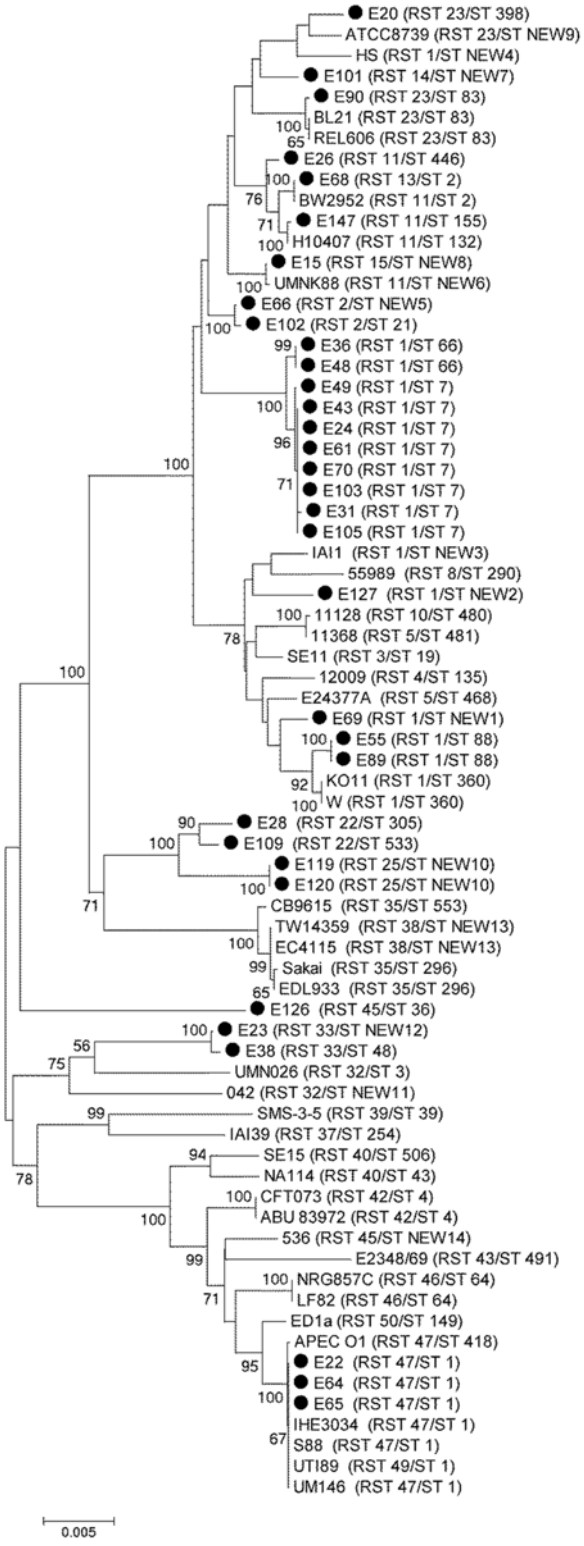

Fig. 2. Phylogenetic analysis of the APEC and reference strains of *E. coli* on the basis of 8 concatenated house-keeping genes of MLST scheme (Jaureguy *et al.* [13]). Phylogenetic analysis was performed with MEGA version 5 software (neighbor-joining method with Tamura-Nei distance and 500 repeats of bootstrapping).

phylogenetic groups, therefore, further phylogenetic studies with unselected ExPEC may be valuable [13].

Clonal division as well as distribution of virulence factors in human and poultry septicemic O2 and O78 ExPEC strains is host-independent [28], but the presence of major RST 1, which covered 56.4% (44/78) of the APEC strains, and RSTs unique to APEC strains are reminiscent of host specificity of APEC. APEC from septicemia are more virulent to chicks than human *E. coli* strains from newborn meningitis [28], therefore, extensive sequencing and analysis of full *rpoB* of poultry and human commensal and pathogenic *E. coli* may make clear the host-dependent distribution of *E. coli*.

The derivation of four phylogenetic groups from the RST 1 reflect the divergent evolution of phylogenetic groups, but phylogenetic groups B2 and D evolved convergently from RSTs 1, 11, 22, 25, 27, 32, 33, 35~40, 42, 43, and 45~50 via horizontal gene transfers and recombinations. Similar evolution patterns were observed in serotypes and molecular pathotypes (MP) (Fig. 1) [14]. For examples RST1 evolved divergently to O8, O9, O18 and O78 serotypes but serotypes O18 and O78 evolved convergently from RSTs 1, 11, 47 and 49, and 1, 2, 11 and 22, respectively. RST 1 evolved divergently to MPs 2, 4, 5, 6, 7, 10, 11, 13, 14, 19, 25 and 26 but MP19 and MP25 evolved convergently from RSTs 1, 15, 25 and 32, and RSTs 1, 2 and 36, respectively. The evolution of *rpoB* depends on point mutations and it changes more slowly than serotypes, phylogenetic groups, and pathotypes, which depend on horizontal gene transfers and recombinations. Therefore, *rpoB* may be valuable to chronicle and may be suitable as a landmark molecule to understand the rapid and complicated evolution of *E. coli* strains.

To date, B2 groups of APEC were suspected to be potential zoonotic pathogens, but there are various B2 groups derived from RSTs 1, 25, and 47. RSTs 1 and 25 are major and unique genotypes of APEC strains, respectively, and only RST 47 is composed of APEC strains and human pathogenic strains. In our previous study, E64 and E22 of RST 47, and E43 of RST 1 were classified into lethality classes 1, 2 and 3, of which LD_{50} ranged $\leq 5 \times 10^6$ cfu, 5×10^6 to 10^8 cfu, and $\geq 5 \times 10^8$, respectively, and they were different in pathogenicity [8, 14]. In the same context, studies on the pathogenicity of D group strains that show extended distribution in different RSTs, such as 22, 32, 33, and 35~39 may be interesting, and the presence of minor but highly-pathogenic APEC strains closely related to human ExPEC should be considered as a potential zoonotic risk in the poultry industry [18].

The further differentiation of 30 RSTs into 49 STs by MLST represented higher resolution power of MLST than RST but the local clustering of the APEC and reference *E. coli* strains in the *rpoB* tree was, however, notably similar to the MLS tree. The B and K12 strains formed different clusters, and pathogenic B2 group strains formed a unique cluster, as reported previously [23]. The clustering of the O157 and O55 strains could be explained by the report that O157 emerged from O55 by only the acquisition of a single large

segment spanning the *rfb-gnd* gene cluster [22]. *Shigellas* are closely related to enteroinvasive *E. coli* (EIEC) and cannot be separated from other *E. coli* strains by MLST-based phylogenetic analyses [20, 27]. According to the phylogenetic analysis with *rpoB*, *S. boydii, S. sonnei* Ss046, and *S. flexneri* strains formed different clusters with one another, but they could not be separated from other *E. coli* strains. On the basis of *rpoB* sequence, *S. dysenteriae* 1012 and Sd197 strains were clearly different each other, reflecting the probable situation where ancestral *E. coli* strains with different genetic backgrounds experienced convergent evolution to become *S. dysenteriae* [20, 27]. Although pangenome clustering of *E. coli* and *Shigella* strains on their variable gene content has revealed clear separation of *E. coli* and *Shigella* and clustering of *S. dysenteriae* 1012 and Sd197, it requires whole genome sequencing and analysis [23].

In conclusion, the phylogenetic analysis of complete *rpoB* sequence can be more cost-effective method than MLST to understand population structure and clonal evolution of *E. coli* strains in human and animals.

Acknowledgments

This research was supported by Technology Development Program for Agriculture and Forestry, Ministry for Food, Agriculture, Forestry and Fisheries, and Technique innovation research program of Small and Medium Business Administration, the Republic of Korea.

References

1. **Adékambi T, Drancourt M, Raoult D.** The *rpoB* gene as a tool for clinical microbiologists. Trends Microbiol 2009, **17**, 37-45.
2. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** Basic local alignment search tool. J Mol Biol 1990, **215**, 403-410.
3. **Bidet P, Mahjoub-Messai F, Blanco J, Blanco J, Dehem M, Aujard Y, Bingen E, Bonacorsi S.** Combined multilocus sequence typing and O serogrouping distinguishes *Escherichia coli* subtypes associated with infant urosepsis and/or meningitis. J Infect Dis 2007, **196**, 297-303.
4. **Bingen E, Picard B, Brahimi N, Mathy S, Desjardins P, Elion J, Denamur E.** Phylogenetic analysis of *Escherichia coli* strains causing neonatal meningitis suggests horizontal gene transfer from a predominant pool of highly virulent B2 group strains. J Infect Dis 1998, **177**, 642-650.
5. **Clermont O, Bonacorsi S, Bingen E.** Rapid and simple determination of the *Escherichia coli* phylogenetic group. Appl Environ Microbiol 2000, **66**, 4555-4558.
6. **Dho-Moulin M, Fairbrother JM.** Avian pathogenic *Escherichia coli* (APEC). Vet Res 1999, **30**, 299-316.
7. **Dissanayake DR, Wijewardana TG, Gunawardena GA, Poxton IR.** Distribution of lipopolysaccharide core types among avian pathogenic *Escherichia coli* in relation to the major phylogenetic groups. Vet Microbiol 2008, **132**, 355-363.
8. **Dozois CM, Dho-Moulin M, Brée A, Fairbrother JM, Desautels C, Curtiss R 3rd.** Relationship between the Tsh autotransporter and pathogenicity of avian *Escherichia coli* and localization and analysis of the Tsh genetic region. Infect Immun 2000, **68**, 4145-4154.
9. **Ewers C, Antão EM, Diehl I, Philipp HC, Wieler LH.** Intestine and environment of the chicken as reservoirs for extraintestinal pathogenic *Escherichia coli* strains with zoonotic potential. Appl Environ Microbiol 2009, **75**, 184-192.
10. **Ghanbarpour R, Sami M, Salehi M, Ouromiei M.** Phylogenetic background and virulence genes of *Escherichia coli* isolates from colisepticemic and healthy broiler chickens in Iran. Trop Anim Health Prod 2011, **43**, 153-157.
11. **Gross WG, Gyles CL.** Diseases due to *Escherichia coli* in Poultry. In: Gyles CL (ed.). *Escherichia coli* in Domestic Animal and Humans. pp. 237-259, CAB International, Wallingford, 1994.
12. **Herrera L, Jiménez S, Valverde A, García-Aranda MA, Sáez-Nieto JA.** Molecular analysis of rifampicin-resistant *Mycobacterium tuberculosis* isolated in Spain (1996-2001). Description of new mutations in the *rpoB* gene and review of the literature. Int J Antimicrob Agents 2003, **21**, 403-408.
13. **Jaureguy F, Landraud L, Passet V, Diancourt L, Frapy E, Guigon G, Carbonnelle E, Lortholary O, Clermont O, Denamur E, Picard B, Nassif X, Brisse S.** Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. BMC Genomics 2008, **9**, 560.
14. **Jeong YW, Kim TE, Kim JH, Kwon HJ.** Pathotyping avian pathogenic *Escherichia coli* strains in Korea. J Vet Sci 2012, **13**, 145-152.
15. **Johnson JR, Delavari P, Stell AL, Whittam TS, Carlino U, Russo TA.** Molecular comparison of extraintestinal *Escherichia coli* isolates of the same electrophoretic lineages from humans and domestic animals. J Infect Dis 2001, **183**, 154-159.
16. **Johnson JR, Owens KL, Clabots CR, Weissman SJ, Cannon SB.** Phylogenetic relationships among clonal groups of extraintestinal pathogenic *Escherichia coli* as assessed by multi-locus sequence analysis. Microbes Infect 2006, **8**, 1702-1713.
17. **Johnson JR, Stell AL, O'Bryan TT, Kuskowski M, Nowicki B, Johnson C, Maslow JN, Kaul A, Kavle J, Prats G.** Global molecular epidemiology of the O15:K52:H1 extraintestinal pathogenic *Escherichia coli* clonal group: evidence of distribution beyond Europe. J Clin Microbiol 2002, **40**, 1913-1923.
18. **Johnson TJ, Kariyawasam S, Wannemuehler Y, Mangiamele P, Johnson SJ, Doetkott C, Skyberg JA, Lynne AM, Johnson JR, Nolan LK.** The genome sequence of avian pathogenic *Escherichia coli* strain O1:K1:H7 shares strong similarities with human extraintestinal pathogenic E. coli genomes. J Bacteriol 2007, **189**, 3228-3236.
19. **Johnson TJ, Wannemuehler Y, Johnson SJ, Stell AL, Doetkott C, Johnson JR, Kim KS, Spanjaard L, Nolan LK.** Comparison of extraintestinal pathogenic *Escherichia coli* strains from human and avian sources reveals a mixed subset representing potential zoonotic pathogens. Appl Environ Microbiol 2008, **74**, 7043-7050.
20. **Lan R, Alles MC, Donohoe K, Martinez MB, Reeves**

PR. Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and *Shigella* spp. Infect Immun 2004, **72**, 5080-5088.

21. **Le Gall T, Clermont O, Gouriou S, Picard B, Nassif X, Denamur E, Tenaillon O.** Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. Mol Biol Evol 2007, **24**, 2373-2384.
22. **Leopold SR, Magrini V, Holt NJ, Shaikh N, Mardis ER, Cagno J, Ogura Y, Iguchi A, Hayashi T, Mellmann A, Karch H, Besser TE, Sawyer SA, Whittam TS, Tarr PI.** A precise reconstruction of the emergence and constrained radiations of *Escherichia coli* O157 portrayed by backbone concatenomic analysis. Proc Natl Acad Sci U S A 2009, **106**, 8713-8718.
23. **Lukjancenko O, Wassenaar TM, Ussery DW.** Comparison of 61 sequenced *Escherichia coli* genomes. Microb Ecol 2010, **60**, 708-720.
24. **Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG.** Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 1998, **95**, 3140-3145.
25. **Manges AR, Johnson JR, Foxman B, O'Bryan TT, Fullerton KE, Riley LW.** Widespread distribution of urinary tract infections caused by a multidrug-resistant *Escherichia coli* clonal group. N Engl J Med 2001, **345**, 1007-1013.
26. **Martínez JA, Soto S, Fabrega A, Almela M, Mensa J, Soriano A, Marco F, Jimenez de Anta MT, Vila J.** Relationship of phylogenetic background, biofilm production, and time to detection of growth in blood culture vials with clinical variables and prognosis associated with *Escherichia coli* bacteremia. J Clin Microbiol 2006, **44**, 1468-1474.
27. **Pupo GM, Lan R, Reeves PR.** Multiple independent origins of Shigella clones of *Escherichia coli* and convergent evolution of many of their characteristics. Proc Natl Acad Sci U S A 2000, **97**, 10567-10572.
28. **Ron EZ.** Host specificity of septicemic *Escherichia coli*: human and avian pathogens. Curr Opin Microbiol 2006, **9**, 28-32.
29. **Russo TA, Johnson JR.** Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli*: ExPEC. J Infect Dis 2000, **181**, 1753-1754.
30. **Selander RK, Caugant DA, Whittam TS.** Genetic structure and variation in natural populations of *Escherichia coli*. In: Neidhardt FC, Ingraham JL (eds.). *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology. Vol. 1. pp. 1625-1648, American Society for Microbiology, Washington, D.C., 1987.
31. **Seong WJ, Kwon HJ, Kim TE, Lee DY, Park MS, Kim JH.** Molecular serotyping of *Salmonella enterica* by complete *rpoB* gene sequencing. J Microbiol 2012, **50**, 962-969.
32. **Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MCJ, Ochman H, Achtman M.** Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol 2006, **60**, 1136-1151.
33. **Zhao L, Gao S, Huan H, Xu X, Zhu X, Yang W, Gao Q, Liu X.** Comparison of virulence factors and expression of specific genes between uropathogenic *Escherichia coli* and avian pathogenic *E. coli* in a murine urinary tract infection model and a chicken challenge model. Microbiology 2009, **155**, 1634-1644.