International Journal of Infectious Diseases 120 (2022) 68–76

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

Multiplex polymerase chain reaction typing scheme based on *Escherichia coli* O157:H7 Sakai prophage (Sp)-associated genes

István Tóth[a], Eva Bagyinszky[a,1], Domonkos Sváb[a,*]

[a] *Veterinary Medical Research Institute, Hungária krt. 21., H-1143, Budapest, Hungary*

ARTICLE INFO

Article history:
Received 11 March 2022
Revised 6 April 2022
Accepted 7 April 2022

Keywords:
pathogenic *Escherichia coli*
O157
EHEC
prophage
typing
multiplex PCR

ABSTRACT

Objectives: *Escherichia coli* strains of the O157 serogroup include significant foodborne pathogens: enterohemorrhagic *E. coli* (EHEC) and enteropathogenic *E. coli*, which are responsible for a considerable number of hospitalizations and deaths worldwide each year. There is a constant need for rapid, reliable, and easy-to-use methods for their identification, typing, and phylogenetic classification. In this study, we proposed a new multiplex polymerase chain reaction (PCR)–based typing system for pathogenic *E. coli*, focusing on the O157 serogroup.

Methods: We designed primers targeting 12 lambdoid prophage regions carried by the prototypic polylysogenic strain of EHEC, the O157:H7 Sakai strain. The reactions were tested *in vitro* as well as *in silico* with the PubMLST database.

Results: The PCR assays can be grouped into four multiplex reactions, and their results can be given as a four-digit code. *In vitro* and *in silico* testing showed that these Sakai prophage regions are prevalent not only in *E. coli* O157 strains but also in Shiga toxigenic *E. coli* non-O157 strains and the method provides appropriate resolution.

Conclusions: The proposed method could be a valuable tool in epidemiologic tracing and preliminary phylogenetic grouping of this diverse group of pathogens.

1. Introduction

Intestinal pathogenic *Escherichia coli* strains of the O157 serogroup include the most notorious pathogenic strains of the enterohemorrhagic (EHEC) pathotype (Pennington, 2010) as well as strains of enteropathogenic *E. coli* (EPEC) and strains with atypical virulence arrays (Tóth et al., 2009). EHEC strains and especially those of the O157:H7 serotype are capable of causing hemorrhagic colitis and the life-threatening complication, hemolytic uremic syndrome (Bielaszewska and Karch, 2005). Shiga toxin–producing *E. coli* (STEC, of which EHEC is a subset producing intimin as well as Shiga toxin) are responsible for an estimated 2.8 million hospitalizations and 230 deaths worldwide annually (Majowicz et al., 2014).

In recent decades, *E. coli* O157 has been a subject of intense whole genome sequencing (WGS)-based studies, with valuable insights gained about their genomic structure and phylogenetic relations. Regarding the former, one of the most important notions is the significant role of prophages in the virulence and genomic variability of the strains (Asadulghani et al., 2009; Ogura et al., 2009; Shaaban et al., 2016). The key virulence factors of STEC and EHEC, the genes encoding Shiga toxin (Stx), are also carried by lambdoid prophages, which are inducible and transducible in several cases (Muniesa and Schmidt, 2014; Rodríguez-Rubio and Muniesa, 2021; Tóth et al., 2016). Within the genome of the prototypic EHEC O157:H7 Sakai strain, 18 prophages were identified and characterized (Asadulghani et al., 2009; Ogura et al., 2009).

Throughout adaptation, prophages become degraded, and their transfer potential is often lost (Bobay et al., 2013; Liu et al., 2020), but their recombination capabilities have an evolutionary potential (Asadulghani et al., 2009; Sváb et al., 2015) to the extent that the O157 strains were called ‘phage factories’ with the capability to release recombinant phages when induced (Ohnishi et al., 2001) .

The WGS-based investigations created opportunities for the design of more precise and broad-ranged, rapid nucleotide sequence–based identification methods of pathogens, and this was also true for *E. coli* O157, especially because the genomic variability and

* Corresponding author. Phone and fax no.: +36 1 252 2455
E-mail address: svab.domonkos@vmri.hu (D. Sváb).
[1] Present address: Graduate School of Environment Department of Industrial and Environmental Engineering, Gachon University, Seongnam, 13120, Republic of Korea

https://doi.org/10.1016/j.ijid.2022.04.015

Table 1
List of strains used for testing of the PCR scheme with their assigned types. The type is the result of the typing PCR reactions transformed into a four-digit numeric code.

Strain	Pathotype	Serotype	SP type	Phage type[a]	Reference
Sakai	EHEC	O157:H7	1111	14	Hayashi et al., 2001
EDL933	EHEC	O157:H7	1111	21	Perna et al., 2001
34	EHEC	O157:H7	1114	21	Tóth et al., 2009
52	EHEC	O157:H7	1131	33	Tóth et al., 2009
254	EHEC	O157:H7	1111	21	Tóth et al., 2009
R4	EHEC	O157:H7	1114	21	Tóth et al., 2009
R67	EHEC	O157:H7	1131	21	Tóth et al., 2009
F67	EHEC	O157:H7	1114	21	Tóth et al., 2009
318	EHEC	O157:NM	1117	8	Tóth et al., 2009
319	EHEC	O157:NM	1113	8	Tóth et al., 2009
320	EHEC	O157:NM	1113	8	Tóth et al., 2009
321	EHEC	O157:NM/H7	7334	8	Tóth et al., 2009
4979	EHEC	O157:H7	1133	8	Tóth et al., 2009
64	EPEC	O157:H7	4423	8	Tóth et al., 2009
65	EPEC	O157:H7	4413	33	Tóth et al., 2009
67	EPEC	O157:H7	4433	33	Tóth et al., 2009
68	EPEC	O157:H7	4455	8	Tóth et al., 2009
103	EPEC	O157:H7	4455	50d	Tóth et al., 2009
121	EPEC	O157:H7	4455	50d	Tóth et al., 2009
122	EPEC	O157:H7	4455	50d	Tóth et al., 2009
B20	atypical	O157:H12	5888	NT-R	Tóth et al., 2009
B47	atypical	O157:NM	2763	NT-R	Tóth et al., 2009
B54	atypical	O157:H12	5768	NT-R	Tóth et al., 2009
T16	atypical	O157:H43	5788	NC	Tóth et al., 2009
T34	atypical	O157:H43	5788	21	Tóth et al., 2009
T22	atypical	O157:H43	5717	NC	Tóth et al., 2009
T4	atypical	O157:H12	5837	NT-R	Tóth et al., 2009
T49	atypical	O157:H37	5737	NC	Tóth et al., 2009
T50	atypical	O157:H43	5738	NC	Tóth et al., 2009
E2348/69	EPEC	O127:H6	6817	N/A	Iguchi et al., 2009
28C	ExPEC	O75	8878	N/A	Dozois et al., 1997
493/89	EHEC	O157:NM	4413	N/A	Karch et al., 1993
536	UPEC	O6:K15:H31	8838	N/A	Schneider et al., 2004
CFT073	UPEC	O6:H1:K	8837	N/A	Mobley et al., 1990
TB156A	EPEC	O55:H7	8887	N/A	Schmidt et al., 1999
O42	EAEC	O44:H18	8757	N/A	Chaudhuri et al., 2010
IHE3034	ExPEC	O18:K1:H7	8734	N/A	Korhonen et al., 1985
C600	nonpathogenic	K-12	8865	N/A	Appleyard, 1954
HNCMB20045		*Shigella sonnei*	2514	N/A	Hungarian National Collection of Medical Bacteria
HNCMB20081		*Shigella dysenteriae*	6717	N/A	Hungarian National Collection of Medical Bacteria

[a] Phage type is given according to the typing scheme of Ahmed et al., 1987. The abbreviations stand for the following: NT, nontypeable; R, phage resistant; NC, noncharacteristic PT; d, derivative; N/A, not applicable.

abundance of its isolates warrants the need for reliable identification and classification methods.

From a practical perspective, the rapid identification of a pathogen is a key issue, both for choosing the optimal treatment as well as for epidemiological tracing. For a precise identification, polymerase chain reaction (PCR)-based methods are widely used, with multilocus variable number of tandem repeat analysis (MLVA) viewed as a precise tool for epidemiological tracing (Van Belkum, 2007). As for the intestinal pathotypes of *E. coli,* there have been MLVA systems developed specifically for EHEC strains of the O157 serogroup (Lee et al., 2019) as well as for members of the O111, and O26 serogroups (Izumiya et al., 2010), which were later refined by Wakabayashi et al (2021). Schemes of detection for the non-O157, the so-called 'big 6' serogroups of STEC, have also been published (Izumiya et al., 2020; Timmons et al., 2016). The general limitation of the method, however, is its narrow target spectrum (Timmons et al., 2016). On the other hand, prophages are widespread in *E. coli*, especially in the O157 serogroup, and play a role in its genomic variability and evolution (Davies et al., 2016; Fortier and Sekulovic, 2013).

In this work, we mapped and monitored the presence of prophages from one of the most well-known prototypic EHEC O157:H7 strain, the Sakai (Hayashi et al., 2001).

We proposed a PCR-based identification scheme on the basis of the sequences of the prophages. Demonstrating the prevalence of these prophages among pathogenic *E. coli* strains, we developed a practical tool for quick genotyping and epidemiological tracing of STEC strains and those of the O157 serogroup.

2. Materials and methods

2.1. Bacterial strains

EHEC, EPEC, and atypical (*stx-, eae-*) *E.coli* O157, non-O157 EPEC, enteroaggregative *E. coli* (EAEC) as well as uropathogenic (UPEC), extraintestinal pathogenic (ExPEC), nonpathogenic laboratory strain *E. coli* K-12 C600, one *Shigella sonnei*, and one *S. dysenteriae* strain were included among the strains on which the typing scheme was tested. All strains used in the study are listed in Table 1.

2.2. Primer design

Primers were designed manually on the basis of the nucleotide sequences of the Sakai prophages (Sp) within the genome of EHEC O157:H7 prototype strain Sakai (GenBank BA000007; Asadulghani et al., 2009; Hayashi et al., 2001). Care was taken that the targets be present in only one copy in the Sakai genome and that their size falls between 100–400 bp but of different length each to allow the multiplexing of the reactions. For the same reason, care was taken that none of the primers of different reactions would be able hybridize with each other. The complete list and nucleotide sequences of primers are shown in Table 2.

Table 2
List of primers used in the PCR typing scheme. The reaction group column indicates the reactions which can be grouped as multiplexes. The reaction number indicates the designation of the reaction within the multiplex, which is necessary for the transformation of the results into a numeric code.

Multiplex group	Reaction within group	Sakai prophage	Putative gene function	Primer name	Primer sequence (5′->3′)	Product length (bp)	Position in Sakai genome (Genbank BA000007)
1	**1**	**Sp1**	hypothetical protein	31.1-f	CGCCAGCTAAATCGAACCGCAT	333	307795-308128
				31.1-r	CGGCTGATGATGACGACTTACTG		
	2	**Sp3**	hypothetical protein	84.3-f	CAGCAGATTGAAGCAGCACTCG	483	928750-929233
				84.3-r	GAATAAGAGCTGAGTCGTGCGG		
	3	**Sp4**	host specificity protein	106.5-f	ACGATTGAGCTGACACCGGGC	190	1205485 - 1205675
				106.5-r	CCGGGCTTAATGTGCGGGCC		
2	**1**	**Sp5**	DNA-binding protein	110.2-f	CGAAGGGGCAACCGCGAAAATA	421	1264526 - 1264947
				110.2-r	CCCTTGTTACTTTCAGCATTCCG		
	2	**Sp6**	phage repressor	122.1-f	GGTGATGGTTTGTGGGAGAGGT	340	1544591-1544931
				122.1-r	TTGGGGGCTTAACGAATACCCC		
	3	**Sp8**	tail protein	124.3-f	GCGTGCAGGTAATGGTAATCCG	483	1651512- 1651995
				124.3-r	TTTAATGCCGTCCTGTTCCTGAGA		
3	**1**	**Sp10**	host specificity protein	145.3-f	TCCGGCATTTTCCCTGACACCA	353	1967113- 1967466
				145.3-r	TATCGCGTGCCTCCTGGGTTAT		
	2	**Sp9**	integrase	133.1-f	GGCATCTAACGGTCTGGTGCC	235	1758290 - 1758525
				133.1-r	CAGCAGAAGCGAACAGCCGTCT		
	3	**Sp11**	tail protein, small	164.1-f	TGACATCCACCACATCCGCAGAA	135	2171798- 2171933
				164.1-r	TGTGAGGAAGAGCAGACGGAGA		
4	**1**	**Sp15**	superinfection inhibition	220.4-f	TACAGCGAATGCCAAATACGCTC	388	2935386-2935774
				220.4-r	TCACCCCTACAGAGAGCAAAAGAG		
	2	**Sp14**	hypothetical protein	204.4-f	CCAAAATACATCCACCCACCGCA	407	2702464- 2702871
				204.4-r	AACGCATAGAAGAGCTGGAGGC		
	3	**Sp17**	antiterminator protein	276.1-f	CAGGTGGGTTGGGTAAGGTTTG	425	3487859-3488284
				276.1-r	GATGGCTGCTATGGGGATGGC		

Table 3
The typing scheme transforming the results of multiplex PCR reactions numbered according to Table 2 into a numeric code, analogously to the phage typing scheme outlined by Farmer (1970).

Type	Reaction 1	Reaction 2	Reaction 3
1	+	+	+
2	+	+	-
3	+	-	+
4	-	+	+
5	+	-	-
6	-	+	-
7	-	-	+
8	-	-	-

2.3. In vitro *testing of the PCR system*

The reactions were performed using deoxyribonucleic acid (DNA) extracted from overnight cultures of bacterial strains grown on Luria-Bertani agar listed in Table 1 by boiling a 10 µl loopful of cells in sterile, distilled water. From every DNA sample, 2 µl was added to a final reaction mixture of 25 µl.

All reactions were performed using Taq DNA polymerase (Fermentas/Thermofisher, Vilnius, Lithuania) according to the manufacturer's instructions. The primers were suspended in distilled water for a stock solution of 100 µM and used in a working concentration of 0.8 µM, with 0.2 µl of stock solution of each primer added to a 25 µl of total reaction volume. The heat profile of all reactions was 3 minutes of initial denaturation at 94°C, then 30 cycles consisting of a denaturation at 94°C for 30 seconds, annealing at 68°C for 30 seconds, and extension at 72°C for 30 seconds. The reactions were ended with a final extension step of 72°C for 5 minutes. Results of the reactions were visualized and evaluated by gel electrophoresis.

The reaction profile of each strain was given as a four-digit code similarly to the phage typing scheme of Farmer (1970), as shown in Table 3.

2.4. *In silico testing of the PCR system*

Using the PubMLST database, in silico PCR was performed (Jolley et al., 2018) separately for each of the primer pairs on all *Escherichia* genomes present in the database in December 2021 and

Table 4
Results of *in silico* PCR against EHEC, STEC, as well as atypical and commensal O157 strains in the PubMLST database. The types of the strains are presented as a four-digit code according to the scheme outlined in Table 3.

Pathotype	PubMLST id	Isolate designation	PCR type
Nonpathogenic	872	Ec_str._K-12_MG1655star	8888
EHEC	159	E_coli_234765	8858
	466	Sakai	1111
	801	Ec_O157:H7_str._EC4206	2464
	802	Ec_O157:H7_str._EC4045	2464
	803	Ec_O157:H7_str._EC4042	2464
	804	Ec_O157:H7_str._EC4196	2464
	805	Ec_O157:H7_str._EC4113	2464
	806	Ec_O157:H7_str._EC4076	2464
	807	Ec_O157:H7_str._EC4401	2464
	808	Ec_O157:H7_str._EC4486	2464
	809	Ec_O157:H7_str._EC4501	1111
	810	Ec_O157:H7_str._EC869	4444
	811	Ec_O157:H7_str._EC508	2744
	812	Ec_O157:H7_str._EC4024	2464
	814	Ec_O157:H7_str._TW14588	1111
	825	Ec_O157:H7_str._FRIK966	4444
	826	Ec_O157:H7_str._FRIK2000	4444
	836	Ec_O157:H7_str._EC4009	2464
	863	Ec_O157:H7_str._EC4191	2464
	879	Ec_TW10509	8868
	889	Ec_TW10598	8888
	890	Ec_TW10722	8825
	891	Ec_TW10828	8858
	892	Ec_TW11681	8868
	893	Ec_TW14425	8868
	896	Ec_O157:H7_str._1044	1111
	897	Ec_O157:H7_str._EC1212	4414
	898	Ec_O157:H7_str._1125	2444
	899	Ec_WV_060327	6687
	900	Ec_EC4100B	7835
	901	Ec_O157:H7_str._G5101	4348
	902	Ec_O157:H-_str._493-89	4728
	903	Ec_O157:H-_str._H_2687	4728
	904	Ec_O55:H7_str._3256-97	4768
	905	Ec_O55:H7_str._USDA_5905	6868
	906	Ec_O157:H7_str._LSU-61	6748
	922	Ec_96.0497	8888
	928	Ec_96.154	8788
	938	Ec_TW07793	5487
	953	Ec_TX1999	8888
	985	Ec_O103:H25_str._NIPH-11060424	8618
	987	Ec_O157:H-_str._493-89	6624
	1012	Ec_O157:H43_str._T22	5788
	1146	Ec_NCCP15658	1444
	1150	Ec_O103:H25_str._CVM9340	8626
	1151	Ec_O103:H2_str._CVM9450	7168
	1153	Ec_O26:H11_str._CVM10026	7147
	1163	Ec_O26:H11_str._CVM10021	7347
	1164	Ec_O26:H11_str._CVM10030	8344
	1165	Ec_O26:H11_str._CVM10224	7344
	1166	Ec_O26:H11_str._CVM9952	7544
	1169	Ec_PA9	2444
	1170	Ec_PA22	1444
	1171	Ec_PA25	2464
	1172	Ec_PA28	2444
	1173	Ec_PA40	6444
	1174	Ec_PA42	1111
	1175	Ec_TW06591	4466
	1176	Ec_TW07945	2444
	1177	Ec_TW10246	2464
	1178	Ec_TW11039	1621
	1179	Ec_TW09098	2444
	1180	Ec_TW09109	4424
	1181	Ec_TW10119	1464
	1182	Ec_TW14301	4466
	1183	Ec_EC4421	1441
	1184	Ec_EC4422	1441
	1185	Ec_EC4013	2464
	1186	Ec_EC4436	2444
	1187	Ec_EC4437	2444
	1188	Ec_EC1738	4424
	1189	Ec_EC1734	2464

(*continued on next page*)

Table 4 (*continued*)

Pathotype	PubMLST id	Isolate designation	PCR type
	1190	Ec_EC1863	2444
	1195	Ec_EC302/04	8887
	1208	Ec_FRIK920	6444
	1209	Ec_PA7	2411
	1210	Ec_PA34	2444
	1214	Ec_FRIK1999	4444
	1216	Ec_FRIK2001	4444
	1217	Ec_PA4	1111
	1218	Ec_PA23	1111
	1219	Ec_PA45	1111
	1220	Ec_TT12B	1714
	1224	Ec_TW15901	6888
	1226	Ec_TW00353	6888
	1230	Ec_EC1735	2444
	1231	Ec_EC1736	2444
	1232	Ec_EC1737	2444
	1233	Ec_EC1848	2464
	1234	Ec_EC1849	2464
	1235	Ec_EC1850	2464
	1236	Ec_EC1856	2464
	1237	Ec_EC1864	2444
	1238	Ec_EC1866	2444
	1239	Ec_EC1868	2444
	1240	Ec_EC1869	2444
	1253	Ec_O26:H11_str._CFSAN001629	7344
	1453	Ec_ATCC_700728	6766
	1454	Ec_PA11	2146
	1455	Ec_PA19	2466
	1456	Ec_PA13	2466
	1457	Ec_PA2	2766
	1458	Ec_PA47	2466
	1459	Ec_PA48	6566
	1460	Ec_PA8	2466
	1464	Ec_PA35	2464
	1716	Ec_O157_str._NCCP15739	2141
	1717	Ec_O157_str._NCCP15738	8888
	1801	Ec_O157:H7_str._F8092B	4444
	1804	Ec_B102	2462
	1805	Ec_B107	1411
	1806	Ec_B26-1	1411
	1807	Ec_B26-2	1411
	1808	Ec_B28-1	2466
	1809	Ec_B28-2	2466
	1810	Ec_B29-1	2466
	1811	Ec_B29-2	2466
	1812	Ec_B36-1	2466
	1813	Ec_B36-2	2466
	1814	Ec_B7-1	2444
	1815	Ec_B7-2	2444
	1816	Ec_B93	2446
	1817	Ec_B94	2446
	1818	Ec_B95	2464
	1819	Ec_TW07509	8887
	1828	Ec_14A	6464
	1830	Ec_B103	2462
	1831	Ec_B104	2462
	1832	Ec_B105	2462
	1833	Ec_B106	2462
	1834	Ec_B108	2166
	1835	Ec_B109	2166
	1836	Ec_B112	2464
	1837	Ec_B113	2466
	1838	Ec_B114	2466
	1839	Ec_B15	2162
	1840	Ec_B17	2162
	1841	Ec_B40-1	2462
	1842	Ec_B40-2	2462
	1843	Ec_B49-2	2462
	1844	Ec_B5-2	2162
	1845	Ec_B83	2462
	1846	Ec_B84	2462
	1847	Ec_B85	2462
	1848	Ec_B86	2461
	1849	Ec_B89	2466
	1850	Ec_B90	2466

(*continued on next page*)

Table 4 (*continued*)

Pathotype	PubMLST id	Isolate designation	PCR type
	1851	Ec_Tx1686	6466
	1852	Ec_Tx3800	2464
	2157	Ec_ATCC_35150	7718
	2265	Ec_O157:H7_str._H093800014	1466
STEC	909	Ec_STEC_7v	8777
	914	Ec_1.2741	8878
	915	Ec_97.0246	4417
	916	Ec_5.0588	5885
	917	Ec_97.0259	8888
	919	Ec_95.0941	8888
	920	Ec_1.2264	8767
	921	Ec_97.0264	8758
	924	Ec_3.2608	7168
	925	Ec_93.0624	7768
	926	Ec_4.0522	7817
	927	Ec_JB1-95	7618
	929	Ec_5.0959	7868
	930	Ec_9.1649	8888
	931	Ec_9.0111	8788
	932	Ec_4.0967	7757
	933	Ec_2.3916	8867
	934	Ec_3.3884	8788
	940	Ec_900105_(10e)	7344
	942	Ec_STEC_B2F1	8788
	943	Ec_STEC_C165-02	8888
	946	Ec_STEC_94C	8788
	947	Ec_STEC_DG131-3	8868
	948	Ec_STEC_EH250	8868
	950	Ec_STEC_H.1.8	7388
	951	Ec_STEC_MHI813	8887
	952	Ec_STEC_S1191	8828
	955	Ec_STEC_O31	8888
	956	Ec_DEC2B	6867
	988	Ec_O113:H21_str._CL-3	8788
	989	Ec_O91:H21_str._B2F1	8788
	990	Ec_O121:H19_str._MT#2	7848
	991	Ec_O45:H2_str._03-EN-705	8448
	992	Ec_O145:H28_str._4865/96	8818
	1065	Ec_DEC1A	6867
	1066	Ec_DEC1B	6867
	1067	Ec_DEC1C	6867
	1068	Ec_DEC1D	6867
	1069	Ec_DEC1E	8887
	1070	Ec_DEC2A	6867
	1071	Ec_DEC2C	6867
	1072	Ec_DEC2D	8887
	1073	Ec_DEC2E	6867
	1074	Ec_DEC3A	1711
	1075	Ec_DEC3B	1411
	1076	Ec_DEC3C	1411
	1077	Ec_DEC3D	1421
	1078	Ec_DEC3E	2444
	1079	Ec_DEC3F	4424
	1080	Ec_DEC4A	1476
	1081	Ec_DEC4B	2464
	1082	Ec_DEC4C	4444
	1083	Ec_DEC4D	4444
	1084	Ec_DEC4E	1464
	1085	Ec_DEC4F	1411
	1086	Ec_DEC5A	4768
	1087	Ec_DEC5B	4768
	1088	Ec_DEC5C	6728
	1089	Ec_DEC5D	6768
	1090	Ec_DEC5E	4768
	1091	Ec_DEC6A	8868
	1092	Ec_DEC6B	8768
	1093	Ec_DEC6C	8788
	1094	Ec_DEC6D	8448
	1095	Ec_DEC6E	8748
	1096	Ec_DEC7A	8888
	1097	Ec_DEC7B	8888
	1098	Ec_DEC7C	8888
	1099	Ec_DEC7D	8888

(*continued on next page*)

Table 4 (*continued*)

Pathotype	PubMLST id	Isolate designation	PCR type
	1100	Ec_DEC7E	8888
	1101	Ec_DEC8A	7817
	1102	Ec_DEC8B	7617
	1103	Ec_DEC8C	8147
	1104	Ec_DEC8D	8747
	1105	Ec_DEC8E	7847
	1106	Ec_DEC9A	7787
	1107	Ec_DEC9B	7467
	1108	Ec_DEC9C	8777
	1109	Ec_DEC9D	7767
	1110	Ec_DEC9E	7747
	1111	Ec_DEC10A	7374
	1112	Ec_DEC10B	8467
	1113	Ec_DEC10C	7344
	1114	Ec_DEC10D	7487
	1115	Ec_DEC10E	8888
	1116	Ec_DEC10F	3787
	1117	Ec_DEC11A	7787
	1118	Ec_DEC11B	7787
	1119	Ec_DEC11C	7467
	1120	Ec_DEC11D	8853
	1121	Ec_DEC11E	8428
	1122	Ec_DEC12A	8788
	1123	Ec_DEC12B	8488
	1124	Ec_DEC12C	8788
	1125	Ec_DEC12D	8787
	1126	Ec_DEC12E	8788
	1127	Ec_DEC13A	8868
	1128	Ec_DEC13B	8868
	1129	Ec_DEC13C	8765
	1130	Ec_DEC13D	8765
	1131	Ec_DEC13E	8765
	1132	Ec_DEC14A	8888
	1133	Ec_DEC14B	8888
	1134	Ec_DEC14C	8788
	1135	Ec_DEC14D	8788
	1136	Ec_DEC15A	8825
	1137	Ec_DEC15B	8855
	1138	Ec_DEC15C	8855
	1139	Ec_DEC15D	8855
	1140	Ec_DEC15E	8855
	1145	Ec_NCCP15657	8284
	1152	Ec_O111:H8_str._CVM9574	7817
	1161	Ec_O111:H8_str._CVM9602	7817
	1162	Ec_O111:H8_str._CVM9634	7827
	1254	Ec_O111:H11_str._CFSAN001630	8147
	1255	Ec_O111:H8_str._CFSAN001632	7847
	1271	Ec_97.0007	2365
	1495	Ec_O91	8888
	1802	Ec_95NR1	7628
	2156	Ec_95JB1	7628
O157 commensal	935	Ec_2.4168	8888
	936	Ec_3.2303	6888
	937	Ec_3003	8887

January 2022, allowing for a 1-nucleotide mismatch and a maximum product length of 1000 nucleotides in each case. Tests were run separately for all isolates (4556 genomes), isolates labeled as EHEC (151 genomes), environmental strains of the O157 serogroup (3 genomes), and STEC (120 genomes).

3. Results

We designed and tested a PCR system targeting 12 prophage genes of the prototypic EHEC strain O157:H7 Sakai. The system could be applied as individual reactions as well as grouped into four triplex reactions, specific for groups of 3 genetic regions indicated in Table 2. Sample results of the multiplexed reactions are shown on Fig. 1.

The schematic grouping used in interpreting the results is shown in Table 3. The results of the four multiplex reactions can lead to one of the numbered patterns listed. The patterns obtained by running all four reactions can be given as a numeric code, which is specific for a narrow group of strains.

Altogether, 40 strains were tested in vitro. The selected strains represented the enteric *E. coli* pathotypes as well as two ExPEC, two UPEC, and two *Shigella* strains. A total of 30 strains belonged to the O157 serogroup. The pathogenic strains carried at least one Sp marker gene. Most of the prophage markers genes were detected in EHEC and EPEC O157 strains, and different Sp gene patterns were observed in O157 strains, representing different patho- and serotpyes. Less genes were present in the other non-O157 enteric strains and a maximum of five genes were detected in the ExPEC isolates. One *S. sonnei* strain carried Stx2 phage (Sp5) marker gene, but *stx2* gene was not detected. Only two (Sp9 and Sp15) of the Sp marker genes were detected in nonpathogenic laboratory strain *E. coli* K-12 C600.

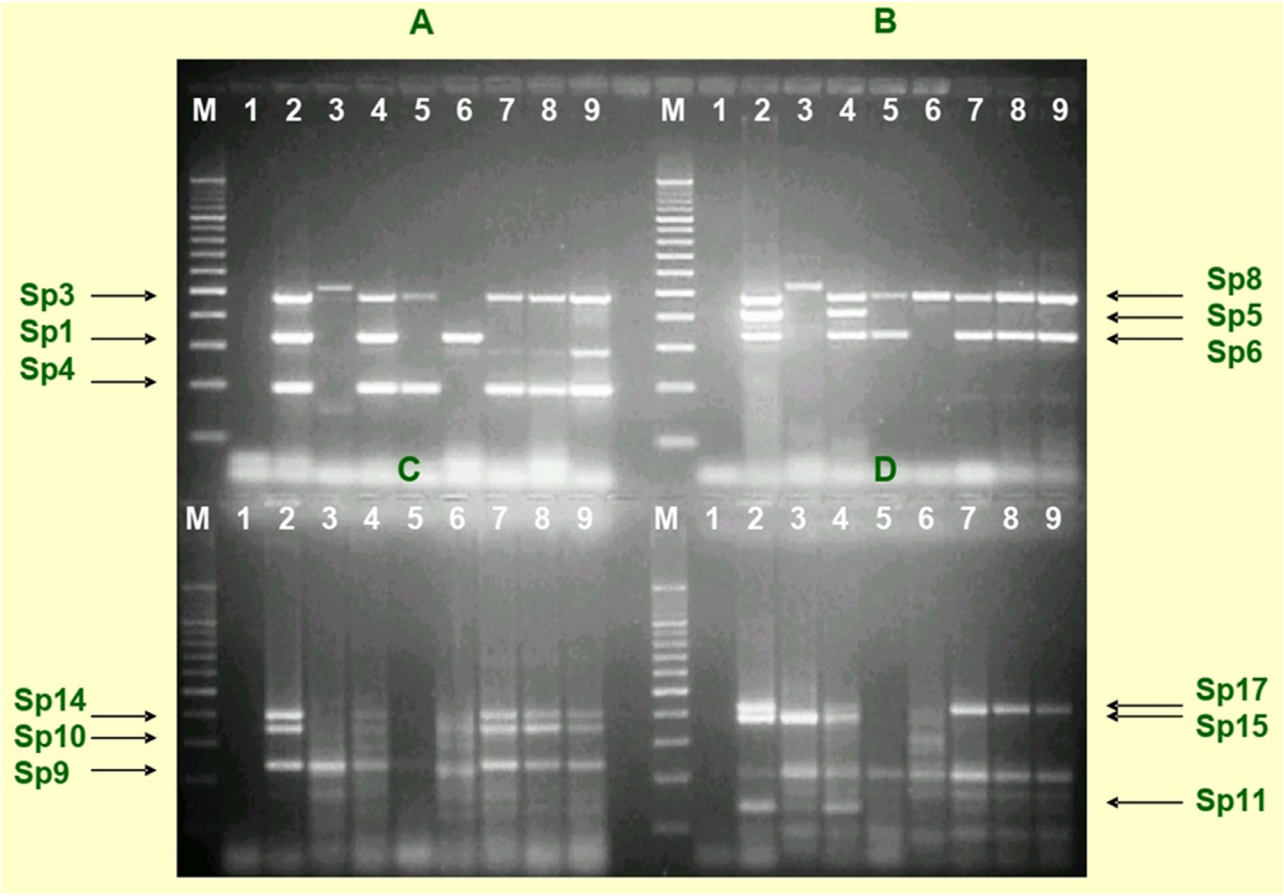

Figure 1. Sample gel electrophoresis of multiplex reactions of the typing scheme. A) reaction 1, B) reaction 2, C) reaction 3, D) reaction 4. Samples: M, marker; 1, no DNA; 2, O157:H7 Sakai (EHEC); 3, *E. coli* C600 (K-12, nonpathogenic); 4, O157:H7 strain 254 (EHEC); 5, O157:H7 strain 68 (EPEC); 6, O157:H43 strain T22 (atypical); 7-9, O157:NM strains (EHEC)

The code for each of the test strains is given in Table 1.

The virtual PCR scanning of the *Escherichia* genomes contained in the PubMLST database has shown that 1775 (39%) of the 4556 isolates fell into the 8888 type, being negative for all reactions. The rest of the strains could be typed by the system in a meaningful way, carrying at least one of the target lambdoid prophage regions.

Narrowing down the scope of strains for the EHEC and STEC strains, of the 151 and 120 genomes, only four and 14 fell into the 'untypeable' 8888 category. The three O157 strains of environmental origin were of types 8888, 6888, and 8887. The 274 investigated strains represented 96 types altogether, the most frequent ('2464') being represented by 24 strains, and 52 strains representing a single type each.

Some noteworthy differences could be identified when comparing the pattern types of the EHEC and STEC strains: the 'prototypic' 1111 pattern were only present among the EHEC strains. Type 2462 was represented by 11 EHEC strains, whereas none of the STEC showed this type. Type 2444 and 2464 were only represented by one STEC strain each, whereas among the EHEC strains, 18 and 22 strains showed these types, respectively.

The full results of the *in silico* PCR, including the type codes for each isolate, are shown in Table 4.

4. Discussion

In the current study, we created a PCR-based typing system for *E. coli* strains of the O157 serogroup, targeting 12 prophage genes present in the prototypic EHEC strain O157:H7 Sakai, grouped analogously to the typing scheme of Farmer (1970), which was originally used in phage typing for *E. coli* O157:H7 (Ahmed et al., 1987). All the reactions target lambdoid phages, and in the overwhelming majority of cases, the targets are only present in one copy in a given genome (six of the reactions gave multiple products with the in silico PCR in a total of 57 instances; but in practice, it may not influence the results of the test, as the predicted products are of the same size in all cases).

The size of the reaction products may enable the reactions to be multiplexed, and all of them can be run with the same reaction conditions, including the annealing temperature without the loss of specificity. These features make the method rapid and easy to use.

By choosing a high number of genes, which could be grouped into multiplex reactions and its results converted into a code, we obtained a system which provides an adequate level of resolution within the target bacterial group, as the 274 in silico investigated strains were sorted into 96 types.

With the 'democratization' of WGS, there has been an upsurge of genomic data regarding pathogenic *E. coli* (Denamur et al., 2021), but there is still a need for rapid classification and identification of strains, especially within the abundantly isolated and sequenced (>1200 whole genomes in GenBank as of January 2022) O157 serogroup, containing several significant pathogenic strains. Our hypothesis was that sequences of prophage origin harbored by a high number of pathogenic strains in the O157 as well as other serogroups could be reliably used for identification, as was shown earlier in the case of *Salmonella* serovar Typhimurium (Fang et al., 2012). The conducted in silico test showed that the multiplex PCR system is indeed reliable. The prototypic EHEC strains showed the

same patterns as in vitro by testing with this method and that the scheme provides adequate resolution for simple and rapid genotyping of STEC strains as well as those of the O157 serogroup. The method could be a valuable help in epidemiologic tracing and preliminary phylogenetic grouping of this highly diverse group of pathogens.

Funding source

This work was supported by the National Research, Development, and Innovation Office (grant number K 124335). Domonkos Sváb was supported by the János Bolyai Research scholarship of the Hungarian Academy of Sciences.

Ethical approval statement

No ethical approval was needed for the study.

Conflict of interest

The authors declare no conflict of interest.

References

Ahmed R, Bopp C, Borczyk A, Kasatiya S. Phage-typing scheme for *Escherichia coli* O157:H7. J Infect Dis 1987;155:806–9.

Appleyard RK. Segregation of New Lysogenic Types during Growth of a Doubly Lysogenic Strain Derived from *Escherichia coli* K12. Genetics 1954;39:440–52.

Asadulghani M, Ogura Y, Ooka T, Itoh T, Sawaguchi A, Iguchi A, et al. The defective prophage pool of *Escherichia coli* O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. PLoS Pathog 2009;5. doi:10.1371/journal.ppat.1000408.

Bielaszewska M, Karch H. Consequences of enterohaemorrhagic *Escherichia coli* infection for the vascular endothelium. Thromb Haemost 2005;94:312–18. doi:10.1160/TH05-04-0265.

Bobay L-M, Rocha EPC, Touchon M. The adaptation of temperate bacteriophages to their host genomes. Mol Biol Evol 2013;30:737–51. doi:10.1093/molbev/mss279.

Chaudhuri RR, Sebaihia M, Hobman JL, Webber MA, Leyton DL, Goldberg MD, et al. Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative *Escherichia coli* strain 042. PLoS ONE 2010;5:e8801. doi:10.1371/journal.pone.0008801.

Denamur E, Clermont O, Bonacorsi S, Gordon D. The population genetics of pathogenic *Escherichia coli*. Nat Rev Microbiol 2021;19:37–54. doi:10.1038/s41579-020-0416-x.

Dozois CM, Clément S, Desautels C, Oswald E, Fairbrother JM. Expression of P, S, and F1C adhesins by cytotoxic necrotizing factor 1-producing *Escherichia coli* from septicemic and diarrheic pigs. FEMS Microbiol Lett 1997;152:307–12. doi:10.1111/j.1574-6968.1997.tb10444.x.

Fang N-X, Huang B, Hiley L, Bates J, Savill J. A rapid multiplex DNA suspension array method for *Salmonella* Typhimurium subtyping using prophage-related markers. J Microbiol Meth 2012;88:19–27. doi:10.1016/j.mimet.2011.10.002.

Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, et al. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res 2001;8:11–22. doi:10.1093/dnares/8.1.11.

Iguchi A, Thomson NR, Ogura Y, Saunders D, Ooka T, Henderson IR, et al. Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. J Bacteriol 2009;191:347–54. doi:10.1128/JB.01238-08.

Izumiya H, Lee K, Ishijima N, Iyoda S, Ohnishi M. Multiple-locus variable-number tandem repeat analysis for non-O157 Shiga toxin-producing *Escherichia coli:* focus on serogroups O103, O121, O145, O165, and O91. Jpn J Infect Dis 2020;73:481–90. doi:10.7883/yoken.JJID.2020.095.

Izumiya H, Pei Y, Terajima J, Ohnishi M, Hayashi T, Iyoda S, et al. New system for multilocus variable-number tandem-repeat analysis of the enterohemorrhagic *Escherichia coli* strains belonging to three major serogroups: O157, O26, and O111. Microbiol Immunol 2010;54:569–77. doi:10.1111/j.1348-0421.2010.00252.x.

Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res 2018;3:124. doi:10.12688/wellcomeopenres.14826.1.

Karch H, Böhm H, Schmidt H, Gunzer F, Aleksic S, Heesemann J. Clonal structure and pathogenicity of Shiga-like toxin-producing, sorbitol-fermenting *Escherichia coli* O157:H-. J Clin Microbiol 1993. doi:10.1128/jcm.31.5.1200-1205.1993.

Korhonen TK, Valtonen MV, Parkkinen J, Väisänen-Rhen V, Finne J, Orskov F, et al. Serotypes, hemolysin production, and receptor recognition of *Escherichia coli* strains associated with neonatal sepsis and meningitis. Infect Immun 1985;48:486–91. doi:10.1128/iai.48.2.486-491.1985.

Lee K, Izumiya H, Iyoda S, Ohnishi M, Working Group EHEC. Effective Surveillance Using Multilocus Variable-Number Tandem-Repeat Analysis and Whole-Genome Sequencing for Enterohemorrhagic *Escherichia coli O157*. Appl Environ Microbiol 2019;85 e00728-19. doi:10.1128/AEM.00728-19.

Liu Z, Deng Y, Ji M, Sun W, Fan X. Prophages domesticated by bacteria promote the adaptability of bacterial cells. BIOCELL 2020;44:157–66. doi:10.32604/biocell.2020.09917.

Majowicz SE, Scallan E, Jones-Bitton A, Sargeant JM, Stapleton J, Angulo FJ, et al. Global Incidence of Human Shiga Toxin–Producing Escherichia coli Infections and Deaths: A Systematic Review and Knowledge Synthesis. Foodborne Pathog Dis 2014;11:447–55. doi:10.1089/fpd.2013.1704.

Mobley HL, Green DM, Trifillis AL, Johnson DE, Chippendale GR, Lockatell CV, et al. Pyelonephritogenic *Escherichia coli* and killing of cultured human renal proximal tubular epithelial cells: role of hemolysin in some strains. Infect Immun 1990;58:1281–9. doi:10.1128/iai.58.5.1281-1289.1990.

Muniesa M, Schmidt H. Shiga toxin-encoding phages: multifunctional gene ferries. In: Morabito S, editor. Pathogenic *Escherichia coli:* Molecular and Cellular Microbiology. Norfolk: Caister Academic Press; 2014. p. 117–38.

Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, et al. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. PNAS 2009;106:17939–44. doi:10.1073/pnas.0903585106.

Ohnishi M, Kurokawa K, Hayashi T. Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? Trends Microbiol 2001;9:481–5. doi:10.1016/S0966-842X(01)02173-4.

Pennington H. *Escherichia coli* O157. The Lancet 2010;376:1428–35. doi:10.1016/S0140-6736(10)60963-4.

Perna NT, Plunkett G, Burland V, Mau B, Glasner JD, Rose DJ, et al. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature 2001;409:529–33. doi:10.1038/35054089.

Rodríguez-Rubio L, Muniesa M, Schüller S, Bielaszewska M. Isolation and Characterization of Shiga Toxin Bacteriophages. Shiga Toxin-Producing *E. coli:* Methods and Protocols. New York, NY: Springer US; 2021. p. 119–44.

Schmidt H, Geitz C, Tarr PI, Frosch M, Karch H. Non-O157:H7 Pathogenic Shiga Toxin-Producing *Escherichia coli:* Phenotypic and Genetic Profiling of Virulence Traits and Evidence for Clonality. J Infect Dis 1999;179:115–23. doi:10.1086/314537.

Schneider G, Dobrindt U, Brüggemann H, Nagy G, Janke B, Blum-Oehler G, et al. The Pathogenicity Island-Associated K15 Capsule Determinant Exhibits a Novel Genetic Structure and Correlates with Virulence in Uropathogenic *Escherichia coli* Strain 536. Infect Immun 2004. doi:10.1128/IAI.72.10.5993-6001.2004.

Shaaban S, Cowley LA, McAteer SP, Jenkins C, Dallman TJ, Bono JL, et al. Evolution of a zoonotic pathogen: investigating prophage diversity in enterohaemorrhagic *Escherichia coli* O157 by long-read sequencing. Microb Genom 2016;2. doi:10.1099/mgen.0.000096.

Sváb D, Bálint B, Maróti G, Tóth I. A novel transducible chimeric phage from *Escherichia coli* O157:H7 Sakai strain encoding Stx1 production. Infect Genet Evol 2015;29:42–7. doi:10.1016/j.meegid.2014.10.019.

Timmons C, Trees E, Ribot EM, Gerner-Smidt P, LaFon P, Im S, et al. Multiple-locus variable-number tandem repeat analysis for strain discrimination of non-O157 Shiga toxin-producing *Escherichia coli*. J Microbiol Meth 2016;125:70–80. doi:10.1016/j.mimet.2016.04.005.

Tóth I, Schmidt H, Kardos G, Lancz Z, Creuzburg K, Damjanova I, et al. Virulence genes and molecular typing of different groups of *Escherichia coli* O157 strains in cattle. Appl Environ Microbiol 2009;75:6282–91. doi:10.1128/AEM.00873-09.

Tóth I, Sváb D, Bálint B, Brown-Jaque M, Maróti G. Comparative analysis of the Shiga toxin converting bacteriophage first detected in *Shigella sonnei*. Infect Genet Evol 2016;37:150–7. doi:10.1016/j.meegid.2015.11.022.

Van Belkum A. Tracing isolates of bacterial species by multilocus variable number of tandem repeat analysis (MLVA). FEMS Immunol Med Microbiol 2007;49:22–7. doi:10.1111/j.1574-695X.2006.00173.x.

Wakabayashi Y, Harada T, Kawai T, Takahashi Y, Umekawa N, Izumiya H, et al. Multilocus Variable-Number Tandem-Repeat Analysis of Enterohemorrhagic *Escherichia coli* Serogroups O157, O26, and O111 Based on a De Novo Look-Up Table Constructed by Regression Analysis. Foodborne Pathog Dis 2021;18:647–54. doi:10.1089/fpd.2020.2921.