Clin. Lab. 2016;62:XXX-XXX

ORIGINAL ARTICLE

Real-Time PCR Typing of *Escherichia coli* Based on Multiple Single Nucleotide Polymorphisms - a Convenient and Rapid Method

Malin Lager, Sara Mernelius, Sture Löfgren, Jan Söderman

Microbiology Laboratory, Division of Medical Diagnostics, Ryhov County Hospital, SE-55185 Jönköping, Sweden

SUMMARY

Background: **Healthcare-associated infections caused by *Escherichia coli* and antibiotic resistance due to extended-spectrum beta-lactamase (ESBL) production constitute a threat against patient safety. To identify, track, and control outbreaks and to detect emerging virulent clones, typing tools of sufficient discriminatory power that generate reproducible and unambiguous data are needed.**
Methods: **A probe based real-time PCR method targeting multiple single nucleotide polymorphisms (SNP) was developed. The method was based on the multi locus sequence typing scheme of Institute Pasteur and by adaptation of previously described typing assays.**
Results: **An 8 SNP-panel that reached a Simpson's diversity index of 0.95 was established, based on analysis of sporadic *E. coli* cases (ESBL n = 27 and non-ESBL n = 53). This multi-SNP assay was used to identify the sequence type 131 (ST131) complex according to the Achtman's multi locus sequence typing scheme. However, it did not fully discriminate within the complex but provided a diagnostic signature that outperformed a previously described detection assay. Pulsed-field gel electrophoresis typing of isolates from a presumed outbreak (n = 22) identified two outbreaks (ST127 and ST131) and three different non-outbreak-related isolates. Multi-SNP typing generated congruent data except for one non-outbreak-related ST131 isolate.**
Conclusions: **We consider multi-SNP real-time PCR typing an accessible primary generic *E. coli* typing tool for rapid and uniform type identification.**
(Clin. Lab. 2016;62:xx-xx. DOI: 10.7754/Clin.Lab.2015.150621)

Correspondence:
Jan Söderman
Microbiology Laboratory
Division of Medical Diagnostics
Ryhov County Hospital
Jönköping, Sweden
Phone: +46 36-322457
Fax: +46 36-180073
Email: jan.soderman@rjl.se

Manuscript accepted August 3, 2015

KEY WORDS

Escherichia coli, multi locus sequence typing, pulsed-field gel electrophoresis, real-time PCR, single nucleotide polymorphism, typing

INTRODUCTION

Healthcare-associated infections caused by bacteria constitute a threat against patient safety [1], and *Escherichia coli* accounts for the majority of these cases [2]. Furthermore, the global spread of extended-spectrum beta-lactamase (ESBL) producing strains has emerged as a significant problem [3,4], which in part has been explained by the appearance of the virulent sequence type 131 (ST131) [5], defined by Achtman's multi locus sequence typing (MLST) scheme [6].

Convenient typing tools of sufficient discriminatory power are needed to rapidly identify, track, and control outbreaks [7] and to disclose differences in virulence and transmission potential [8,9], as well as for monitoring of asymptomatic bacteriuria comparing consecutive samples [10,11] and for disclosing colonization of multiple types [12,13].

Pulsed-field gel electrophoresis (PFGE) demonstrates high discriminatory power and epidemiological concordance and is considered the 'gold standard' for typing of clinically important bacteria [7] including *E. coli* [14]. PFGE has revealed a large degree of genetic diversity within the ST131 complex of *E. coli* [15]. Nevertheless, PFGE has several limitations [7,16] and a recent study with the multiple-locus variable-number tandem-repeats analysis (MLVA) typing method [17] has shown promising results in the assessment of possible outbreaks of ESBL-producing *E. coli* [18]. Additionally, a MLST scheme based on ten genes has been suggested as a useful typing tool for ESBL-producing *E. coli* [19]. Even though the MLST method is labor intensive, it has potential to provide a more unambiguous type assignment compared to MLVA. However, both MLVA and MLST require a capillary electrophoresis instrument.

A typing method should be easy to perform and interpret, possess high typeability and reproducibility, and have adequate discriminatory ability and stability over time [7,16]. Preferably a typing method should be based on commonly available instrumentation. Real-time PCR is an easy to use, convenient method that is fast, reproducible, self-contained, and has been adopted by most clinical microbiology laboratories [20].

In this study, we present a rapid and accessible real-time PCR typing tool for *E. coli*, with a Simpson´s index of diversity of 0.95 based on detection of allele-specific probes of multiple single nucleotide polymorphisms (SNPs).

MATERIALS AND METHODS

Bacterial isolates

A total of 102 *E. coli* isolates were acquired from Microbiology Laboratory (Ryhov County Hospital, Jönköping, Sweden), representing two different study sets; (1) 80 sporadic *E. coli* isolates composed of 27 ESBL isolates from consecutive clinical patient samples and 53 non-ESBL isolates of different types established through the combined use of three different typing methods [12], and (2) 22 isolates collected from children, parents, and health care workers during an outbreak of ESBL producing *E. coli* at the neonatal ward at Ryhov County Hospital (Jönköping, Sweden) during September 2008 through February 2009.

Typing

To discriminate within the global ST131 complex, defined by Achtman's MLST scheme [6], polymorphic positions (Table 1; Supplementary Table 1) were chosen based on data from the MLST scheme of Institute Pasteur [21].

One additional polymorphic position was chosen according to Sheludchenko et al. [22]. ESBL-producing isolates belonging to ST131 were identified by *pabB* PCR adapted from Clermont et al. [23], or by sequencing the full set of genes defined by Achtman's MLST scheme [6].

Genomic DNA was real-time PCR amplified using primers according to the MLST scheme of Institute Pasteur [21], Clermont et al. [23] (*pabB*) and Sheludchenko et al. [22] (*lysP*), and analyzed using probes specific for each polymorphic variant (Table 1). Each 10 µL-reaction contained 1x Type-it Fast SNP Probe PCR master mix (Qiagen, Hilden, Germany), 1.5 pM of each forward and reverse primer (Sigma Aldrich, St. Louis, MO, USA), 0.25 pM of each probe (Life Technologies, Carlsbad, CA, USA) and 3 ng DNA. The amplification was done in either a 7500 Fast Real-time PCR (Life Technologies) or a 7900HT Fast Real-time PCR instrument (Life Technologies) with the following thermal cycling conditions: 50°C for 2 minutes and 95°C for 10 minutes, 35 cycles including 95°C for 60 seconds, 55°C for 60 seconds and 60°C for 90 seconds, and a final extension at 72°C for 30 seconds. Variants were assigned according to the probe with the lowest threshold cycle value. Individual SNPs and different SNP combinations were investigated for discriminatory power by Simpson's diversity index (Di) [24] using the Minimum SNP software [25]. For each isolate, SNP typing results were combined and every different pattern was assigned a separate multi-SNP type. Multi-SNP types were clustered by the unweighted pair group method with arithmetic mean method and the MEGA6 software [26]. PFGE was done according to Christiansson et al. [12] except that isolates that were degraded in a 0.5 x Tris-Borate-EDTA buffer were separated in 1 x HEPES buffer (16 mM HEPES, 16 mM sodium acetate, 0.8 mM EDTA, pH 7.5) for 28 hours at 4.0 V/cm.

RESULTS

Establishing a multi-SNP real-time PCR assay

Simpson´s index of diversity ranged from 0.20 to 0.66 for individual SNPs in the collection of sporadic *E. coli* (Supplementary Table 2) with low diversity figures for SNP 6 and SNP 7 which were selected to detect *E. coli* ST131.

Selection of a subset of SNPs for inclusion in a typing assay was based on high discrimination of sporadic ESBL *E. coli* (study set 1). A multi-SNP real-time PCR assay based on 8 SNPs reached a Simpson's index of diversity of 0.95, which was the maximum diversity obtainable by the complete set of 19 SNPs (Table 2). In total the multi-SNP real-time PCR assay identified 24 different multi-SNP types among the 80 isolates of study set 1 (Figure 1).

Table 1. **Genes and polymorphic positions used in multiple single nucleotide polymorphisms real-time PCR typing.**

SNP number	Gene (position) [a]	Variant positions according to		Variant	Variant frequency [d]	Basis for variant selection [e]
		Gene sequence [b]	MLST sequence [c]			
1	*dinB* (250898 - 251953)	480	417	C/T	0.53/0.45	MLST-db
2		564	n.a.	A/G		Study set 1
3	*icdA* (1194346 - 1195596)	741	102	C/T	0.89/0.11	Study set 1
4		810..813	171..174	G..A/G..G/A..G	0.44/0.42/0.13	MLST-db
5	*lysP* (2245085 - 2246554)	330	n.a.	A/G	0.36/0.64	Scheludchenko et al. (22)
6	*pabB* (1892829 - 1894190)	267	n.a.	C/T		Clermont et al. (23)
7		573	n.a.	A/G		Clermont et al. (23)
8		720	102	C/T	0.44/0.56	MLST-db
9		872..873	254..255	AG/TA/AA	0.31/0.04/0.65	Study set 1
10	*polB* (65780 - 63429)	844..850	228..234	G..G/A..G/G..A	0.46/0.29/0.23	MLST-db
11		903..907	171..175	C..A/T..A/C..G	0.73/0.03/0.23	Study set 1
12	*putP* (1078528 - 1080036)	1095	105	C/A	0.37/0.62	MLST-db
13		1230	241	A/G	0.16/0.83	Study set 1
14	*trpA* (1315246 - 1314440)	249..250	99..100	TT/CC/TC	0.45/0.09/0.46	Study set 1
15		618	468	G/A	0.37/0.63	MLST-db
16	*trpB* (1316439 - 1315246)	196..199	429..432	A..G/C..T/C..G	0.35/0.48/0.16	Study set 1
17		256..262	366..372	A..C/G..C/G..T	0.71/0.08/0.17	MLST-db
18	*uidA* (1694095 - 1692284)	961	539	T/C	0.40/0.56	MLST-db
19		1456	45	A/C/T	0.16/0.76/0.07	Study set 1

[a] - Gene position according to accessions number NC_000913.2.
[b] - Within each gene nucleotides were numbered from the start of the gene, and the first nucleotide of the start codon was denoted + 1.
[c] - Position according to the multi locus sequencing typing (MLST) sequence of each gene (www.pasteur.fr/recherche/genopole/PF8/mlst/EColi.html; accessed 3 May 2011). n.a., not applicable, variant position outside reported MLST-sequence or gene not present in the Institute Pasteur's MLST scheme).
[d] - Variant frequencies based on *Escherichia coli* MLST database (MLST-db) maintained at Institute Pasteur, France (accessed 8 August 2008).
[e] - Variant positions were selected based on (1) the *E. coli* MLST-db maintained at Institute Pasteur, France (accessed 8 August 2008), (2) MLST sequencing of non-extended-spectrum beta-lactamase (non-ESBL) isolates of study set 1, and (3) selected from previous publications by Clermont et al. [23] and Scheludchenko et al. [22]. Assays for variants selected from Clermont et al. [23] and Scheludchenko et al. [22] were converted to the assay format of the present study (Supplementary Table 1).

Using the SNP method described by Clermont et al. (Table 1; [23]) nine ST131 were detected among the sporadic ESBL-producing *E. coli* (study set 1). According to the suggested multi-SNP typing assay, four additional isolate of study set 1 co-clustered with the ST131 isolates (Figure 1; one isolate each from individuals i27 and i57, and two isolates from i16), and for three of these isolates the ST131 status according to Achtman's MLST scheme [6] was confirmed by sequencing.

Simultaneous colonization of different *E. coli* types
In study set 1, five individuals who were living at a long term care facility were simultaneously colonised by different types of *E. coli* that are consistently identified with three different typing methods: PFGE, MLVA and Pheneplate [12]. Isolates from these individuals were al-

Table 2. **Accumulated Simpson's index of diversity of sporadic *E. coli.***

SNP number [a]	Accumulated Simpson's index of diversity
14	**0.66**
11	**0.83**
9	**0.89**
3	**0.92**
7	**0.93**
19	**0.94**
2	**0.94**
12	**0.95**
max Di [b]	**0.95**

SNP - single nucleotide polymorphism, [a] SNP numbers correspond to the numbers provided in Table 1, [b] Based on all 19 assays of Table 1.

so assigned to different multi-SNP types (Figure 1; individuals i4, i6, i12, i14, and i26 of set 1).

Recognition and tracking of outbreaks

Two overlapping outbreaks caused by different types of ESBL-producing *E. coli* and three non-outbreak-related isolates were initially detected by PFGE during an outbreak at a neonatal ward (study set 2; Figure 2). MLST sequencing [6] identified the isolates of the first outbreak as ST127 and the second outbreak as ST131 (Figure 1). One of the three non-outbreak related isolates was ST131. The two remaining PFGE types not involved in the outbreak were also classified as non-outbreak-related by MLST (Figure 1; ST129 and ST642). The multi-SNP real-time PCR typing discriminated the types in accordance with MLST (Figure 1) and provided a stable type assignment over time. ST127 corresponded to a new multi-SNP type, adding to the previously identified 24 types *E. coli*.

DISCUSSION

We have developed a rapid, reliable, and convenient tool for typing of *E. coli*. The assay is based on real-time PCR and variant-specific probes. Assaying 8 SNPs reached a Simpson´s index of diversity of 0.95 recommended by van Belkum et al. [27] as a lower limit of discrimination for a typing assay. The discriminatory ability of the multi-SNP typing assay was sufficient to accurately differentiate between multiple types of E. coli that were simultaneously colonized in individuals living at a long term care facility, and the assay delivered stable and reliable typing results over time.

Therefore, multi-SNP typing might provide a convenient and efficient tool for detecting the simultaneous presence of multiple types of *E. coli* [12] in colonization and disease. The concurrent presence of susceptible and resistant isolates could have an impact on antibiotic treatment, and multiclonal colonization might also impair epidemiological conclusions. Also, it is important to resolve issues regarding persistence, relapse with the same type, or reinfection with a new type [11,28,29]. During the investigated outbreak, multi-SNP typing demonstrated concordance with PFGE by identifying two overlapping outbreak of ESBL-producing *E. coli* at a neonatal ward. Isolates of one of these overlapping outbreaks corresponded to ST131. Similar to MLST, multi-SNP typing failed to identify one ST131 isolate as non-outbreak-related, which, however, was detected by PFGE. Complementary typing is warranted to resolve the ST131 complex of isolates.

All ST131 isolates co-clustered according to multi-SNP typing, even though the assay was developed based on other genes than the genes of the MLST scheme that defined ST131. In our hands, detection of ST131 according to Clermont et al. [23] failed to recognize three of the ST131 isolates.

Other assays targeting multiple SNP markers by means of PCR rely on a capillary electrophoresis instrument for sequencing [6,21,30] or genotyping [31] or a real-time PCR based assessment of 8 SNPs with variant-specific priming using 17 separate reactions in duplicates [22]. Our method is based on the assessment of 8 SNPs using hydrolysis probes. A probe-based typing method requires fewer reactions, compared to variant-specific priming. Typing based on hydrolysis probes in conjunction with our mode of interpretation can handle at least three different variants per polymorphic position in a single reaction. Thus, even though four of the included polymorphic positions required three probes, our method required only 8 reactions and the assay cost per isolate may be calculated to €20. Apart from being accessible, probe-based real-time PCR generates reliable typing results in a timely manner and should allow for uniform type identification between laboratories.

We suggest that our multi-SNP real-time PCR assay is suitable as a primary generic *E. coli* typing tool. How-

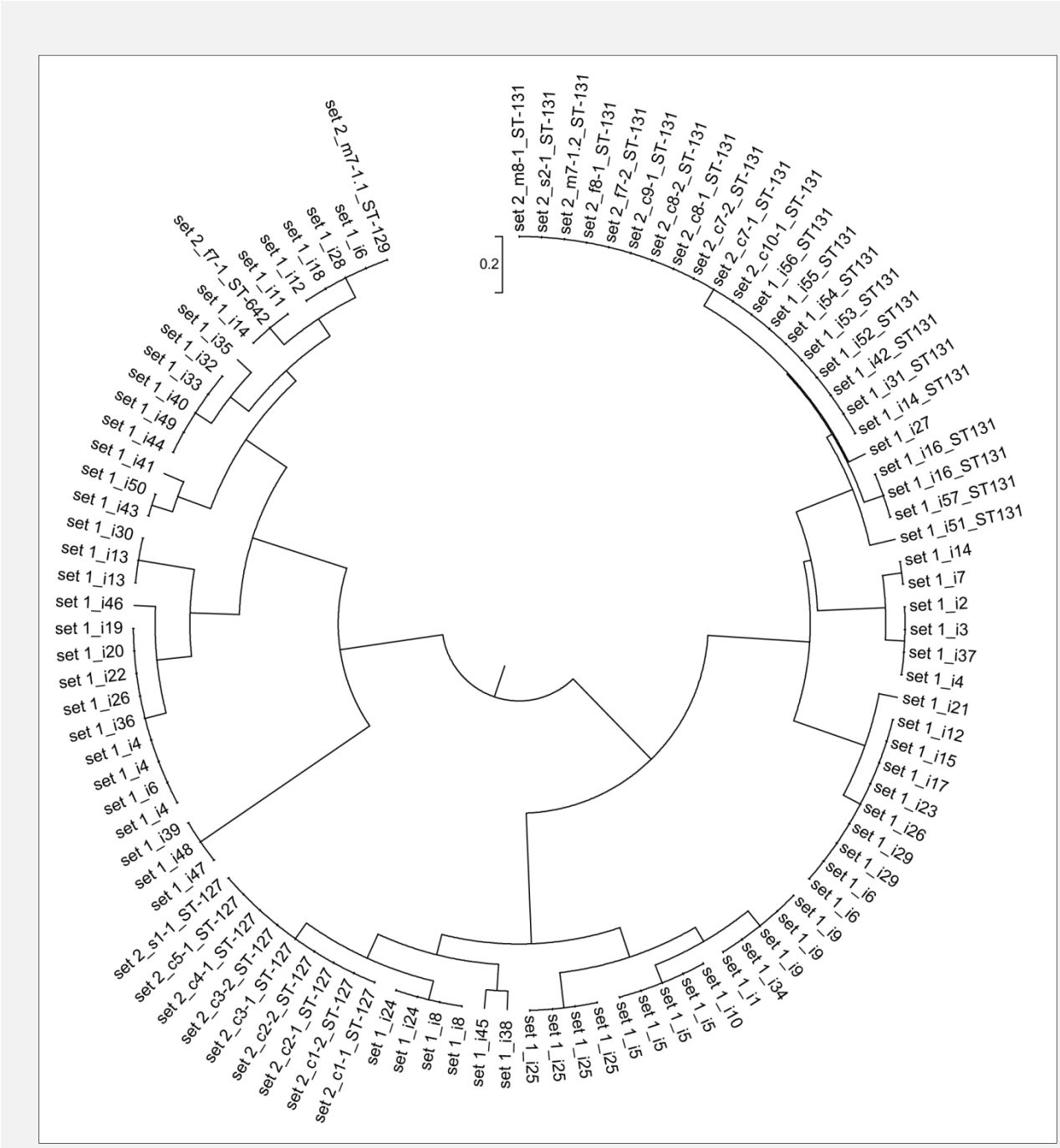

Figure 1. Cluster analysis of *E. coli* isolates from study sets 1 and 2.

Individual isolates were identified by "study set_sampled individual_MLST type (if available)". For study set 2, each individual was assigned a prefix indicating child (c), father (f), mother (m), or staff (s), and a suffix indicating if the sample is the first (-1) or last sample (-2). The suffix also indicated if several isolates were analysed at the sampling occasion (e.g., -1.1 and -1.2).

ever, taking next generation sequencing data into consideration it is likely that even more informative polymorphic positions may improve the assay. We consider probe-based typing an attractive assay format and propose it as a convenient format for outbreak-specific assays based on next generation sequencing data [32].

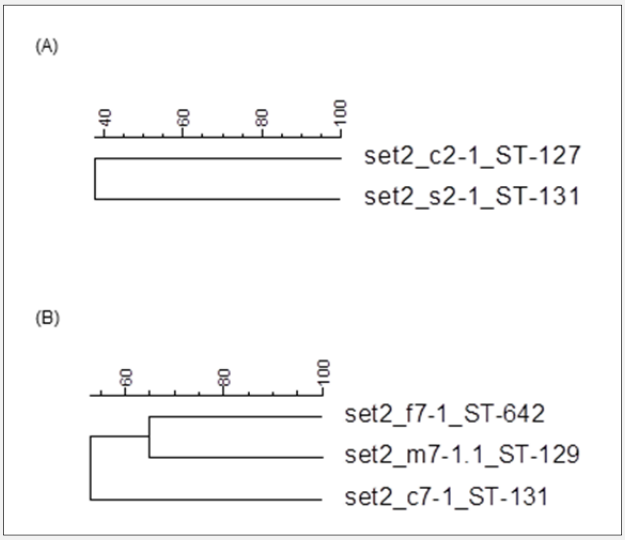

Figure 2. Dendrograms based on one isolate per PFGE type identified among isolates sampled during two overlapping outbreaks at a neonatal ward at Ryhov County Hospital (Jönköping, Sweden).

Isolates were typeable using either 1 x HEPES buffer (A; isolate set2_c2-1_ST-127 belonged to the first outbreak) or 0.5 x Tris-Borate-EDTA buffer (B; isolate set2_c7-1_ST-131 belonged to the second outbreak). Individual isolates were named as follow "study set_sampled individual_MLST type", and each individual was assigned a prefix indicating child (c), father (f), mother (m), or staff (s), and a suffix indicating if the sample is the first (-1) or last sample (-2). The suffix also indicated if several isolates were analysed at the sampling occasion (e.g., -1.1 and -1.2).

Acknowledgement:
We appreciate the support of Futurum, the Academy for Healthcare, Jönköping County Council, for funding part of this work.

Financial disclosures:
Part of this work has received financial support from Futurum - The Academy for Healthcare, Jönköping County Council.

Declaration of Interest:
No declaration of interest exists.

References:

1. Peleg AY, Hooper DC. Hospital-acquired infections due to gram-negative bacteria. N Engl J Med 2010 May 13;362(19):1804-13. PubMed PMID: 20463340. Pubmed Central PMCID: 3107499. Epub 2010/05/14. eng.

2. Sievert DM, Ricks P, Edwards JR, et al. Antimicrobial-resistant pathogens associated with healthcare-associated infections: summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2009-2010. Infection Control and Hospital Epidemiology 2013 Jan;34(1):1-14. PubMed PMID: 23221186. Epub 2012/12/12. eng.

3. Kohlenberg A, Schwab F, Ruden H. Wide dissemination of extended-spectrum beta-lactamase (ESBL)-producing Escherichia coli and Klebsiella spp. in acute care and rehabilitation hospitals. Epidemiology and Infection 2012 Mar;140(3):528-34. PubMed PMID: 21676355. Epub 2011/06/17. eng.

4. Rodriguez-Bano J, Navarro MD. Extended-spectrum beta-lactamases in ambulatory care: a clinical perspective. Clin Microbiol Infect 2008 Jan;14 Suppl 1:104-10. PubMed PMID: 18154533. Epub 2007/12/25. eng.

5. Peirano G, Pitout JD. Molecular epidemiology of Escherichia coli producing CTX-M beta-lactamases: the worldwide emergence of clone ST131 O25:H4. Int J Antimicrob Agents 2010 Apr;35(4): 316-21. PubMed PMID: 20060273. Epub 2010/01/12. eng.

6. Wirth T, Falush D, Lan R, et al. Sex and virulence in Escherichia coli: an evolutionary perspective. Mol Microbiol 2006 Jun;60(5): 1136-51. PubMed PMID: 16689791. Pubmed Central PMCID: 1557465. Epub 2006/05/13. eng.

7. Sabat AJ, Budimir A, Nashev D, et al. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill 2013;18(4):20380. PubMed PMID: 23369389. Epub 2013/02/02. eng.

8. Moissenet D, Salauze B, Clermont O, et al. Meningitis caused by Escherichia coli producing TEM-52 extended-spectrum beta-lactamase within an extensive outbreak in a neonatal ward: epidemiological investigation and characterization of the strain. J Clin Microbiol 2010 Jul;48(7):2459-63. PubMed PMID: 20519482. Pubmed Central PMCID: 2897521. Epub 2010/06/04. eng.

9. Adler A, Gniadkowski M, Baraniak A, et al. Transmission dynamics of ESBL-producing Escherichia coli clones in rehabilitation wards at a tertiary care centre. Clin Microbiol Infect 2012 Dec;18(12):E497-505. PubMed PMID: 22963432. Epub 2012/09/12. eng.

10. Nicolle LE, Bradley S, Colgan R, et al. Infectious Diseases Society of America guidelines for the diagnosis and treatment of asymptomatic bacteriuria in adults. Clin Infect Dis 2005 Mar 1;40 (5):643-54. PubMed PMID: 15714408. Epub 2005/02/17. eng.

11. Rahav G, Pinco E, Bachrach G, Bercovier H. Molecular epidemiology of asymptomatic bacteriuria in the elderly. Age and Ageing. 2003 Nov;32(6):670-3. PubMed PMID: 14600011. Epub 2003/11/06. eng.

12. Christiansson M, Melin S, Matussek A, Lofgren S, Soderman J. MLVA is a valuable tool in epidemiological investigations of Escherichia coli and for disclosing multiple carriage. Scand J Infect Dis 2011 Apr 1. PubMed PMID: 21453204. Epub 2011/04/02. eng.

13. Lautenbach E, Bilker WB, Tolomeo P, Maslow JN. Impact of diversity of colonizing strains on strategies for sampling Escherichia coli from fecal specimens. J Clin Microbiol 2008 Sep;46(9): 3094-6. PubMed PMID: 18650357. Pubmed Central PMCID: 2546773. Epub 2008/07/25. eng.

14. Ribot EM, Fair MA, Gautom R, et al. Standardization of pulsed-field gel electrophoresis protocols for the subtyping of Escherichia coli O157:H7, Salmonella, and Shigella for PulseNet. Foodborne Pathog Dis 2006 Spring;3(1):59-67. PubMed PMID: 16602980. Epub 2006/04/11. eng.

15. Johnson JR, Nicolas-Chanoine MH, DebRoy C, et al. Comparison of Escherichia coli ST131 pulsotypes, by epidemiologic traits, 1967-2009. Emerging Infectious Diseases 2012 Apr;18(4): 598-607. PubMed PMID: 22469129. Pubmed Central PMCID: 3309662. Epub 2012/04/04. eng.

16. van Belkum A, Tassios PT, Dijkshoorn L, et al. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin Microbiol Infect 2007 Oct;13 Suppl 3:1-46. PubMed PMID: 17716294. Epub 2007/11/06. eng.

17. Lobersli I, Haugum K, Lindstedt BA. Rapid and high resolution genotyping of all Escherichia coli serotypes using 10 genomic repeat-containing loci. J Microbiol Methods 2012 Jan;88(1):134-9. PubMed PMID: 22088357. Epub 2011/11/18. eng.

18. Karami N, Helldal L, Welinder-Olsson C, Ahren C, Moore ER. Sub-typing of extended-spectrum-beta-lactamase-producing isolates from a nosocomial outbreak: application of a 10-loci generic Escherichia coli multi-locus variable number tandem repeat analysis. PLoS One 2013;8(12):e83030. PubMed PMID: 24391735. Pubmed Central PMCID: 3877011. Epub 2014/01/07. eng.

19. Nemoy LL, Kotetishvili M, Tigno J, et al. Multilocus sequence typing versus pulsed-field gel electrophoresis for characterization of extended-spectrum beta-lactamase-producing Escherichia coli isolates. J Clin Microbiol 2005 Apr;43(4):1776-81. PubMed PMID: 15814998. Pubmed Central PMCID: 1081380. Epub 2005/04/09. eng.

20. Maurin M. Real-time PCR as a diagnostic tool for bacterial diseases. Expert Review of Molecular Diagnostics 2012 Sep;12(7): 731-54. PubMed PMID: 23153240. Epub 2012/11/17. eng.

21. Jaureguy F, Landraud L, Passet V, et al. Phylogenetic and genomic diversity of human bacteremic Escherichia coli strains. BMC Genomics 2008;9:560. PubMed PMID: 19036134. Pubmed Central PMCID: 2639426. Epub 2008/11/28. eng.

22. Sheludchenko MS, Huygens F, Hargreaves MH. Highly discriminatory single-nucleotide polymorphism interrogation of Escherichia coli by use of allele-specific real-time PCR and eBURST analysis. Appl Environ Microbiol 2010 Jul;76(13):4337-45. PubMed PMID: 20453128. Pubmed Central PMCID: 2897452. Epub 2010/05/11. eng.

23. Clermont O, Dhanji H, Upton M, et al. Rapid detection of the O25b-ST131 clone of Escherichia coli encompassing the CTX-M-15-producing strains. J Antimicrob Chemother 2009 Aug;64 (2):274-7. PubMed PMID: 19474064. Epub 2009/05/29. eng.

24. Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol 1988 Nov;26(11):2465-6. PubMed PMID: 3069867. Pubmed Central PMCID: 266921. Epub 1988/11/01. eng.

25. Robertson GA, Thiruvenkataswamy V, Shilling H, et al. Identification and interrogation of highly informative single nucleotide polymorphism sets defined by bacterial multilocus sequence typing databases. J Med Microbiol 2004 Jan;53(Pt 1):35-45. PubMed PMID: 14663103. Epub 2003/12/10. eng.

26. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution 2013 Dec;30(12):2725-9. PubMed PMID: 24132122. Pubmed Central PMCID: 3840312. Epub 2013/10/18. eng.

27. van Belkum A, Tassios PT, Dijkshoorn L, et al. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin Microbiol Infect 2007 Oct;13 Suppl 3:1-46. PubMed PMID: 17716294. Epub 2007/11/06. eng.

28. Geerlings SE, Brouwer EC, Gaastra W, Hoepelman AI. Is a second urine specimen necessary for the diagnosis of asymptomatic bacteriuria? Clin Infect Dis 2000 Sep;31(3):E3-4. PubMed PMID: 11017857. Epub 2000/10/06. eng.

29. Salvador E, Wagenlehner F, Kohler CD, et al. Comparison of asymptomatic bacteriuria Escherichia coli isolates from healthy individuals versus those from hospital patients shows that long-term bladder colonization selects for attenuated virulence phenotypes. Infection and Immunity 2012 Feb;80(2):668-78. PubMed PMID: 22104113. Pubmed Central PMCID: 3264318. Epub 2011/11/23. eng.

30. Tartof SY, Solberg OD, Riley LW. Genotypic analyses of uropathogenic Escherichia coli based on fimH single nucleotide polymorphisms (SNPs). J Med Microbiol 2007 Oct;56(Pt 10): 1363-9. PubMed PMID: 17893175. Epub 2007/09/26. eng.

31. Hommais F, Pereira S, Acquaviva C, Escobar-Paramo P, Denamur E. Single-nucleotide polymorphism phylotyping of Escherichia coli. Appl Environ Microbiol 2005 Aug;71(8):4784-92. PubMed PMID: 16085876. Pubmed Central PMCID: 1183324. Epub 2005/08/09. eng.

32. Pritchard L, Holden NJ, Bielaszewska M, Karch H, Toth IK. Alignment-free design of highly discriminatory diagnostic primer sets for Escherichia coli O104:H4 outbreak strains. PLoS One 2012;7(4):e34498. PubMed PMID: 22496820. Pubmed Central PMCID: 3320637.

Supplementary Table 1. Fluorescently labelled probes for allelic discrimination.

SNP number [a]	Dye	Probe sequence [b]
1	FAM	CAACGG[T]CAGTTTGT
	VIC	CAACGG[C]CAGTTTG
2	FAM	CAAAGTCTC[A]GCGGC
	VIC	CAAAGTCTC[G]GCGGC
3	FAM	CCAG[C]TGGTAGCC
	VIC	CGCCAG[T]TGGTAGC
4	FAM	CTCTTTGCC[G]GT[A]TT
	VIC	CTCTTTGCC[G]GT[G]TT
	NED	TCTCTTTGCC[A]GT[G]TT
5	FAM	ACGGC[A]ATAGTCAC
	VIC	ACGGC[G]ATAGTCA
6	FAM	CTGGATCG[T]GCAGACA
	VIC	TGGATCG[C]GCAGACA
7	FAM	CGCGA[A]CAGTACGG
	VIC	CCGCGA[G]CAGTAC
8	FAM	CTAAA[T]GGCGCGCGGT
	VIC	CTAAA[C]GGCGCGCG
9	FAM	CAAGCAG[AG]AAACT
	VIC	CAAGCAG[TA]AAACT
	NED	AGCAG[AA]AAACTG
10	FAM	CC[G]AGCAC[G]CCAT
	VIC	CC[G]AGCAC[A]CCATA
	NED	CC[A]AGCAC[G]CCATA
11	FAM	TTAC[C]CTG[A]CGTTTGG
	VIC	TTAC[T]CTG[A]CGTTTGGC
	NED	TTAC[C]CTG[G]CGTTTG
12	FAM	CTGGC[C]TGTTTACG
	VIC	CTGGC[A]TGTTTACGC
13	FAM	TTGGCGC[A]GCGTT
	VIC	CTTTGGCGC[G]GCGT
14	FAM	CCAGCA[CC]TCAAA
	VIC	TGCCAGCA[TT]TCA
	NED	TGCCAGCA[TC]TCA
15	FAM	CTGC[G]CCTCCATT
	VIC	CTGC[A]CCTCCATTG
16	FAM	ATGAAGC[A]CT[G]GAAG
	VIC	ATGAAGC[C]CT[T]GAAG
	NED	ATGAAGC[C]CT[G]GAAG
17	FAM	CC[A]CAACA[C]GCGTAT
	VIC	CC[G]CAACA[C]GCGTA
	NED	CC[G]CAACA[T]GCGTA
18	FAM	TTCCTGAT[C]AACCACAAA
	VIC	TCCTGAT[T]AACCACAAAC
19	FAM	ATACGG[A]GTGACATCG
	VIC	ATACGG[C]GTGACATC
	NED	ATACGG[T]GTGACATCG

SNP - single nucleotide polymorphism, [a] - Further information on each SNP can be found in Table 1 in the main text, [b] - Differently labeled probes for polymorphic discrimination were designed using Primer Express® 2.0 (Life Technologies, Carlsbad, CA,USA).

Supplementary Table 2. Simpson's index of diversity per SNP for sporadic *E. coli* (study set 1).

SNP number [a]	Simpson's index of diversity
1	0.50
2	0.50
3	0.45
4	0.55
5	0.50
6	0.20
7	0.29
8	0.51
9	0.59
10	0.59
11	0.56
12	0.47
13	0.48
14	0.66
15	0.51
16	0.56
17	0.51
18	0.47
19	0.64

SNP - single nucleotide polymorphism, [a] SNP - numbers correspond to the numbers provided in Table 1 in the main text.