APMIS
ACTA PATHOLOGICA, MICROBIOLOGICA ET IMMUNOLOGICA SCANDINAVICA

APMIS 125: 1117–1124

DOI 10.1111/apm.12753

Epidemiological characterization of a nosocomial outbreak of extended spectrum β-lactamase *Escherichia coli* ST-131 confirms the clinical value of core genome multilocus sequence typing

HANNA WOKSEPP,[1,2] ANNA RYBERG,[3] LINDA BERGLIND,[4] THOMAS SCHÖN[1,2,5,6] and JAN SÖDERMAN[4]

[1]Department of Clinical Microbiology, Kalmar County Hospital; [2]Department of Medicine and Optometry, Linnaeus University, Kalmar; [3]Department of Clinical Microbiology, Växjö Central Hospital, Växjö; [4]Division of Medical Diagnostics, Region Jönköping County, Jönköping; [5]Department Infectious Diseases, Kalmar County Hospital, Kalmar; and [6]Department of Clinical and Experimental Medicine, Division of Medical Microbiology, Linköping University, Linköping, Sweden

Woksepp H, Ryberg A, Berglind L, Schön T, Söderman J. Epidemiological characterization of a nosocomial outbreak of extended spectrum β-lactamase *Escherichia coli* ST-131 confirms the clinical value of core genome multilocus sequence typing. APMIS 2017; 125: 1117–1124.

Enhanced precision of epidemiological typing in clinically suspected nosocomial outbreaks is crucial. Our aim was to investigate whether single nucleotide polymorphism (SNP) analysis and core genome (cg) multilocus sequence typing (MLST) of whole genome sequencing (WGS) data would more reliably identify a nosocomial outbreak, compared to earlier molecular typing methods. Sixteen isolates from a nosocomial outbreak of ESBL *E. coli* ST-131 in southeastern Sweden and three control strains were subjected to WGS. Sequences were explored by SNP analysis and cgMLST. cgMLST clearly differentiated between the outbreak isolates and the control isolates (>1400 differences). All clinically identified outbreak isolates showed close clustering (≥2 allele differences), except for two isolates (>50 allele differences). These data confirmed that the isolates with >50 differing genes did not belong to the nosocomial outbreak. The number of SNPs within the outbreak was ≤7, whereas the two discrepant isolates had >700 SNPs. Two of the ESBL *E. coli* ST-131 isolates did not belong to the clinically identified outbreak. Our results illustrate the power of WGS in terms of resolution, which may avoid overestimation of patients belonging to outbreaks as judged from epidemiological data and previously employed molecular methods with lower discriminatory ability.

Key words: Epidemiological typing; whole genome sequencing; *Escherichia coli* ST-131; single nucleotide polymorphism; cgMLST.

Hanna Woksepp, Department of Clinical Microbiology, Kalmar County Hospital, S-391 85 Kalmar, Sweden. e-mail: hanna.woksepp@ltkalmar.se

Severe bacterial infections are a major cause of morbidity and mortality (1) and thus a major threat to human health (2). Early identification of nosocomial outbreaks is crucial, especially the spread of high-risk clones that have accumulated mutations enabling multiresistance and increased pathogenicity (3). Current methods for analyzing epidemiological linkage between bacterial isolates include pulse field gel electrophoresis (PFGE) (4), multilocus sequence typing (MLST) (5, 6), different PCR-based methods such as ligation-mediated qPCR high-resolution melt analysis (LMqPCR HRMA) (7, 8) and more recently whole genome sequencing (WGS) (9, 10). WGS does not only provide information about epidemiological linkage but can also identify known antibiotic resistance genes (11) and virulence factors (12, 13). Furthermore, WGS also allows for the detection of plasmids (14).

Received 3 March 2017. Accepted 23 July 2017

Escherichia coli is a gram-negative rod and classified both as a commensal and as a human pathogen. Each *E. coli* genome encodes up to 5000 genes, of which approximately 2200 are part of the core genome (cg), and the remainder consists of a highly variable accessory genome (15–17), allowing for an extensive genomic diversity. *E. coli* genome size can vary up to 1 Mb, due to horizontal gene transfer with acquisition or loss of pathogenic islands or other genetic elements (18). A new widespread sequence type (ST) *E. coli* (ST-131), producing extended spectrum β-lactamases (ESBL), was identified in 2008 and has since then been reported to cause outbreaks over the entire globe (19, 20).

Comparison of clonal dissemination of *E. coli* using WGS has previously been performed by single nucleotide polymorphism (SNP) analysis by Mellmann et al. (10) and core genome (cg) MLST by Fischer et al. (21) and Kaas et al. (17, 22).

We have previously investigated a clinically defined nosocomial outbreak of ESBL-producing *E. coli,* but some of the isolates were difficult to evaluate by earlier generations of molecular methods such as high-resolution melt (HRM) analysis of PCR products and PFGE due to the limited resolution of the methods used (7). Thus, our aim was to investigate whether cgMLST and/or SNP analysis of WGS data could resolve whether or not the isolates belonged to the nosocomial outbreak.

MATERIALS AND METHODS

Selection of bacterial isolates

Sixteen isolates, from a clinically well-characterized nosocomial outbreak of ESBL-producing *E. coli* of type ST-131, and three, random, outbreak unrelated isolates of ESBL-producing *E. coli* of different ST-types (Table 1 and Fig. S1) were collected during 2008–2009, at the Department of Clinical Microbiology, Kalmar County Hospital, Sweden (7). The nosocomial outbreak isolates had an unusual susceptibility pattern for that geographical region (ESBL-producing *E. coli* with resistance to gentamicin and ciprofloxacin but susceptibility to trimethoprim-sulfamethoxazole), were related in time and place and thus hypothesized to belong to the same outbreak. The majority of the ST-131 isolates were collected from urine samples from elderly patients at a medical department and one nursing homes within Kalmar County. Eleven of sixteen were clinical cultures and five colonization cultures (isolate 3, 4, 6, 9 and 10). Two patients (carriers of isolates 1 and 2, respectively) suffered from septicemia caused by the isolate.

Sample preparation and WGS

DNA was extracted using either DNA tissue kit and the BioRobot EZ1 (Qiagen AB, Sollentuna, Sweden) (isolates 1, 6, 8 and 13) or MagNA Pure Compact system (Roche Diagnostics Scandinavia AB, Bromma, Sweden) (isolates 2–5, 7, 9–12, 14–18 and 23) according to the manufacturer's instructions. DNA concentrations were measured using Qubit dsDNA BR Assay system (Thermofisher Scientific, Waltham, MA, USA), and samples were diluted to

Table 1. The isolates included in the study with epidemiology, source of collection and sequence type

Isolate	Epidemiology	Date isolated	Community/Hospital acquired	Suspect outbreak isolate	Source	MLST
1	Medical ward, H1	2008 May	HA	Y	Blood	ST-131
2	Emergency, H1	2008 May	HA	Y	Blood	ST-131
3	Surgery ward, H1	2008 January	HA	Y	Feces	ST-131
4	Geriatric nursing home D	2008 October	CA	Y	Feces	ST-131
5	ICU, H1	2009 January	HA	Y	Urine	ST-131
6	ICU, H1	2008 October	HA	Y	Secretion	ST-131
7	HCC2	2008 May	CA	Y	Ulcer	ST-131
8	OB, H1	2008 August	HA	Y	Unspecified	ST-131
9	Geriatric nursing home D	2008 October	CA	Y	Feces	ST-131
10	Geriatric nursing home D	2008 October	CA	Y	Feces	ST-131
11	Geriatric nursing home D	2008 July	CA	Y	Urine	ST-131
12	Geriatric nursing home D	2008 September	CA	Y	Urine	ST-131
13	Medical ward, H1	2008 October	HA	Y	Urine	ST-131
14	Rehab/Geriatric ward, H1	2007 December	HA	Y	Urine	ST-131
15	HCC1	2008 December	CA	Y	Urine	ST-131
16	Emergency, H1	2009 May	HA	Y	Urine	ST-131
17	Rehab/Geriatric ward, H1	2008 January	HA	N	Secretion	ST-624
18	Medical ward, H2	2008 November	HA	N	Urine	ST-69
21	Surgery ward, H1	2009 January	HA	N	Urine	ST-410

H, Hospital; ICU, intensive care unit; HCC, Healthcare center; OB, Obstetrics and gynecology; Y, yes; N, no; HA, Hospital acquired; CA, Community acquired. Isolates 17 and 21 are unrelated control isolates from the same hospital and time period as isolates 1–16. Isolate 18 is an unrelated control isolate from the same time period but another hospital as isolates 1–16.

0.2 ng/μL using 10 mM Tris-HCl, pH 8.5. DNA concentrations of the diluted samples were confirmed by renewed measurements.

Sequencing libraries were prepared according to the manufacturer's instructions using the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA), and indexed to allow pooling of libraries.

The amplified and indexed libraries were purified using Agencourt AMPure XP beads (Beckman Coulter, Indianapolis, IN, USA), according to instructions for the Nextera XT DNA Library Preparation Kit (Illumina), and the size distributions of the purified libraries were assessed using Agilent High Sensitivity DNA chip (Agilent Technologies Inc., Santa Clara, CA, USA).

Normalized libraries were pooled in equal volumes, and sequenced as 12-plex using the MiSeq v3 chemistry and the MiSeq desktop sequencer (Illumina). All libraries were sequenced using a 2 × 300 bp, paired-end workflow. Base calling, demultiplexing and adapter trimming were conducted using MiSeq Reporter version 2.5.1 (Illumina). Isolates 1, 6, 8 and 13 were subjected to paired-end WGS using the Illumina HiSeq system. Sequencing libraries were prepared from 1.5 μg of DNA using the TruSeq DNA sample prep kit v2 (Illumina) according to the manufacturer's instructions (guide # 15026486 revA). Paired-end sequencing with 100 bp read length was performed on a HiSeq2000 instrument (Illumina) with v3 sequencing chemistry according to the manufacturer's instructions (HiSeq control software 1.5.15.1/RTA 1.13.48). Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala, Sweden www.sequencing.se.

Quality control, trimming and *de novo* assembly were performed using CLCbio Genomics workbench 7.5 (23) with default settings. The sequences reported in this study have been deposited in European Nucleotide Archive Sequence Read Archive under study number PRJEB15588.

In silico MLST, SNP analysis and detection of plasmids and resistance genes using the Center for Genomic Epidemiology server

Contigs were uploaded to Center for Genomic Epidemiology (CGE) server (www.genomicepidemiology.org, April 16, 2016) and analyzed using MLST E. coli #1 (24, 25), PlasmidFinder (26) and ResFinder (11) using default settings. Phylogeny was inferred using the default settings (except that the minimum depth at SNP position was set to 20x) of the CSI Phylogeny service (22) and the original fastq-files, by calling SNPs with respect to either a well-characterized ST-131 reference sequence (JJ1886, accession number NC_022648.1) (27) or the contigs from a *de novo* assembly of the index isolate (isolate 14). Also JJ1886 was included in the analysis when the index isolate was used as reference.

MLST and cgMLST analysis using SeqSphere

SeqSphere+ (3.2.1 64-bit, Ridom GmbH, Münster, Germany) (28) was used in pipeline mode to implement a workflow consisting of read trimming, *de novo* assembly using Velvet (29) and conventional MLST (5) and cgMLST. The cgMLST scheme was developed from 181 genomes (Table S1), rendering 1602 targets for cgMLST. Minimum spanning trees were created after the exclusion of targets not found or failed. The reference sequence of the outbreak unrelated ST-131 isolate JJ1886 (NC_022648.1) was included in the analysis.

Visualization of isolate differences using BRIG

All isolates were aligned to the ST-131 reference sequence (NC_022648.1) of *E. coli* JJ1886 using NCBI BLASTn+ and BLAST Ring Image Generator (BRIG) (30).

RESULTS

Analysis of resistance genes, plasmid typing and in silico MLST analysis

All isolates carried the *aac(6′)Ib-cr* gene-mediating resistance to aminoglycosides and fluoroquinolones, and gene-encoding resistance to β-lactam antibiotics (*blaOXA-1* and *blaCTX-M-15* as well as *blaTEM-1B*), expect for 8 and 13, which lacked the *blaTEM-1B* gene. All isolates carried the Inc FII plasmid, and all but one of the non-ST131 isolates carried plasmids belonging to the Inc FIA type. Isolates 8, 13, 17, 18 and 21 also carried Inc FIB plasmids. All of the suspected outbreak isolates were of the ST-131 type (Table 1; isolates 1 – 16), and the control isolates were of type ST-624, ST-69 and ST-410, respectively (Table 1; isolates 17, 18 and 21).

Comparison between clinically defined nosocomial outbreak isolates and epidemiological clustering with cgMLST

Forty-one of 1602 cgMLST targets were excluded, due to the presence of failed comparisons between the WGS data and the target genes, from the cluster analysis (Table S2). Among the suspected outbreak isolates, except isolate 8 and 13, there were 0–2 allele differences per isolate in the cgMLST analysis (Fig. 1). Using cgMLST, isolates 8 and 13 differed by 79 and 53 alleles, respectively, from the outbreak isolates. The ST-131 reference isolate JJ1886 had 30 allele differences to the nosocomial outbreak. The three non-ST-131 control isolates differed at >1400 genes from the suspected outbreak isolates of type ST-131 (Fig. 1).

Comparison between clinically defined nosocomial outbreak isolates and epidemiological clustering with SNP analysis

The SNP analysis using CSI Phylogeny and the earliest isolated outbreak isolate (isolate 14) as a reference showed that the three control non-ST-131

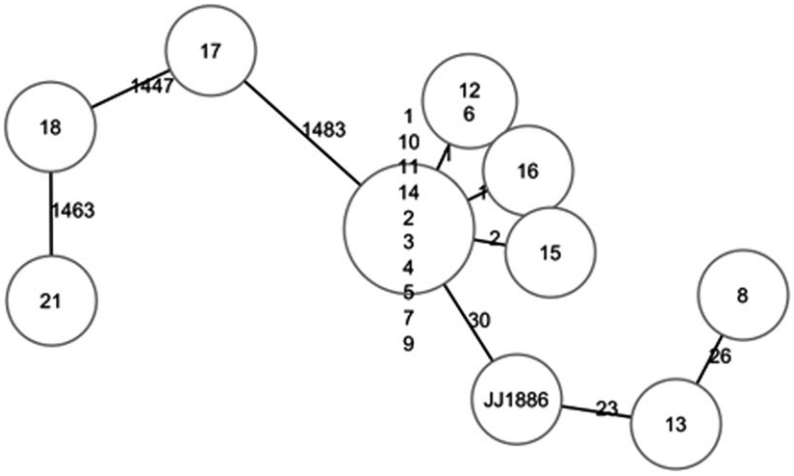

Fig. 1. Minimum spanning tree of all sequenced isolates and JJ1886 as analyzed by cgMLST. The numbers on the connecting lines give the number of differing genes.

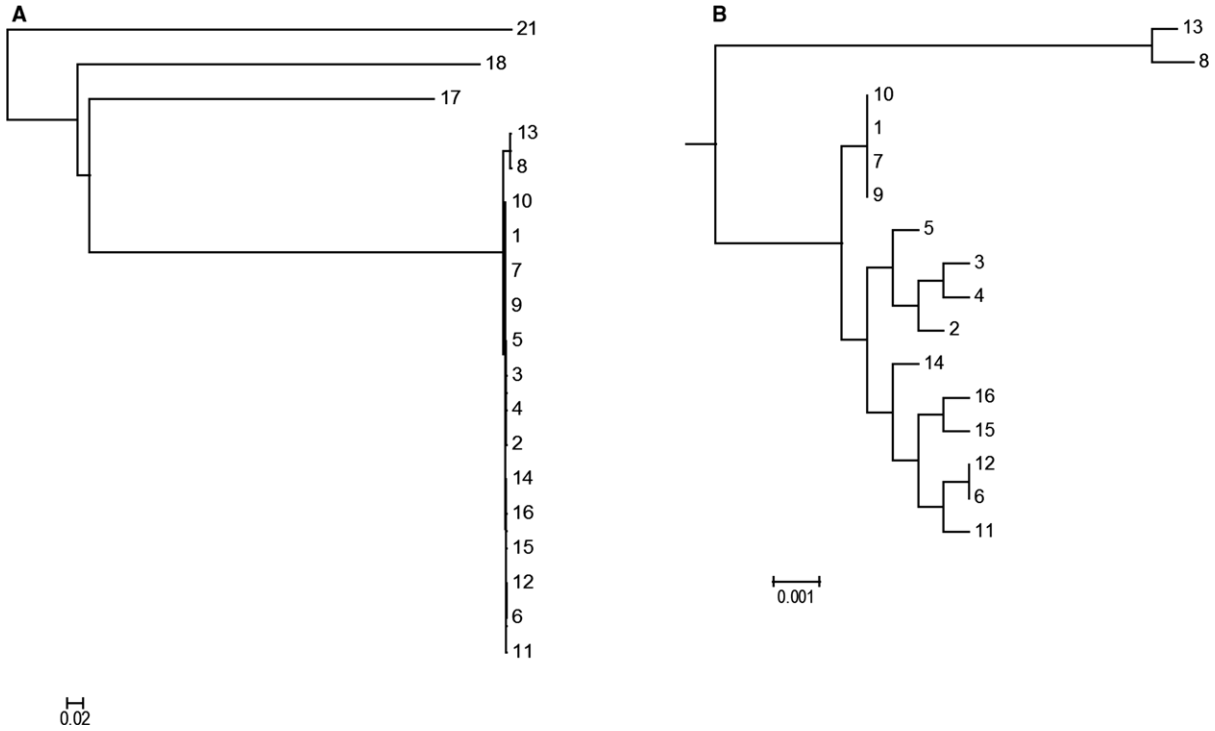

Fig. 2. Rectangular tree over SNP analysis performed on all sequenced isolates (A) and only the ST-131 isolates (B). The SNP analysis identified ≥28 374 SNPs between the three control non-ST-131 isolates and the index isolate 14 (A). Within the ST-131 group (B), isolates 8 and 13 deviated with a high number of SNPs compared to the index isolate 14 (>700), compared to <7 SNPs between the other ST-131 outbreak isolates (isolates 1–7, 9–12 and 15–16) and isolate 14.

isolates (isolates 17, 18 and 21) differed by at least 28374 SNPs from the ST-131 group of isolates (Fig. 2A). Within the group of suspected outbreak isolates (ST-131), isolates 8 and 13 clearly deviated with a high number of SNPs compared to isolate 14 (743 and 719, respectively) (Fig. 2B). A maximum of seven SNPs were identified between the other ST-131 outbreak isolates (isolates 1–7, 9–12 and 15–16) and isolate 14 (Fig. 2B). The ST-131 reference genome JJ1886 not belonging to the

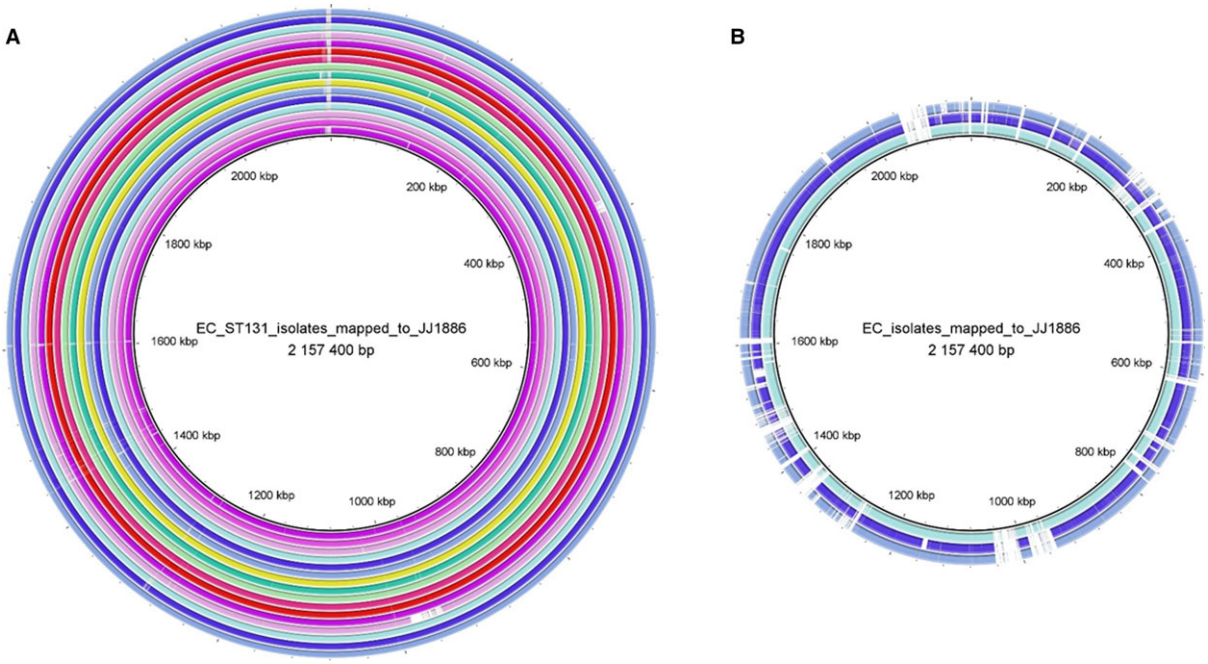

Fig. 3. BLAST comparisons of isolates 1–16 (A) and isolates 17, 18 and 21 (B) against JJ1886, respectively. Gene absence/presence and divergence are visualized as the color represents sequence identity on a sliding scale, the brighter the lower percentage identity.

nosocomial isolates differed at 14 positions from isolate 14. Similar results were obtained using the outbreak unrelated ST-131 isolate JJ1886 (NC_022648.1) as reference, as the number of SNPs between isolates 1–7;9–12;14–16 and isolates 8 and 13 were >20 and within isolates 1–7;9–12;14–16 were <5 (data not shown). Furthermore, using JJ1886 as reference the number of SNPs between the ST-131 isolates and the non-ST-131 isolates was >11 700 (data not shown). By visualization using BRIG, it was clear that isolates 8 and 13 were more different to JJ1886 than the other ST-131 isolates as they display more point-wise differences (Fig. 3A). The control isolates 17, 18 and 21 contained several regions with less similar sequences spread throughout the genome (Fig. 3B).

DISCUSSION

Using cgMLST analysis from WGS data when investigating a clinically identified nosocomial outbreak of *E. coli* ST-131, we were able to clearly identify two isolates (8 and 13) not belonging to the outbreak. In a previous analysis of the clinically defined epidemiological outbreak (7), PFGE and LMqPCR HRMA were able to differentiate between ST-131 isolates and non-ST-131 isolates but had difficulties in judging on differences between isolates 8 and 13 and the rest of the ST-131 suspected outbreak isolates. This prompted us to proceed to WGS followed by cgMLST and SNP analysis, where isolates 8 and 13 were clearly differentiated from the other ST-131 isolates. Isolate 8 and 13 were collected at the same hospital as the other ST-131 isolates, which had an epidemiological connection and all showed the same resistance pattern, which at the time, was unusual for that area. The isolates concluded to belong to the outbreak are all isolated from patients with possible links to the rehabilitation and geriatric clinic at hospital 1 or a geriatric care facility (Fig. S1). Isolate 8 was collected from an obstetric clinic at hospital 1, where, to our knowledge, no other patients with ST-131 had been identified. Isolate 13 was collected at the medical clinic, but the cgMLST and SNP strongly indicated that this isolate was not connected to the other isolates and there was no strong epidemiological link such as in the geriatric health facility.

Our results contribute to the knowledge of molecular typing methods as it applies modern WGS-based methods on a nosocomial outbreak of clinical relevance where data on earlier generation

molecular typing methods were available (7). Additionally, there is still a need to establish cgMLST as a tool in clinical practice and standardization of cgMLST is urgently needed. To our knowledge, there are no systematic studies investigating the utility of cgMLST in well-established nosocomial outbreaks for the successful clone, *E. coli* ST-131.

As suggested by others, residents in nursing homes comprise a reservoir of *E. coli* ST-131 with high carriage and infection level (20). As transmission between humans is readily reported (31, 32), standardized, fast, simple and preferably automated methods for the evaluation of possible outbreak scenarios are warranted. In addition, as WGS by massive parallel sequencing is becoming the new standard for investigating bacterial outbreaks, standardized analysis pipelines are required. Furthermore, ways of communicating typing results between laboratories using an international standard such as MLST are desirable. This automation is to some part possible with programs such as SeqSphere (28) or BacTyper (33) using a pipeline mode where all sequences designated to a pre-set folder can be subjected to assembly and subsequently outbreak analysis. In fact, Mellmann et al. (34) performed real-time microbial WGS on multi-drug-resistant (MDR) *E. coli* from a tertiary care hospital. In that study, real-time epidemiological typing was performed by WGS and subsequent cgMLST analysis, which could avoid unnecessary isolation of MDR *E. coli* from the clinical judgement only. They did not observe any increase in MDR *E. coli* transmission, and the costs were reduced due to avoided costs of isolation (34).

By including the genome sequence of the non-related reference isolate JJ1886 with the same MLST type (ST-131) in the cgMLST analysis, it was possible to suggest a threshold of differing alleles within a cluster. A previous study has shown a threshold of ≤10 differing alleles in the core genome for *E. coli*. (33) As isolate 8 and 13 had >40 differing alleles from the outbreak isolates, they were judged as non-related both to each other and to the outbreak.

Both cgMLST and SNP analysis resulted in the same clustering, with the exception of higher differentiation within the cluster of outbreak isolates using SNP analysis (Figs 1 and 2). Nevertheless, the reference-based SNP analysis of *E. coli* might not be the most comprehensible approach as *E. coli* is a very diverse organism with a high level of recombination (35), which may result in false-negative clustering because short non-allelic sequences may be incorrectly incorporated during assembly and then misinterpreted as polymorphisms (17, 36). In contrast to using the index isolate (14) for comparison in the SNP analysis, using JJ1886 resulted in fewer differences between isolate 8 and 13 and the rest of the ST-131 isolates. This is probably due to the fact that a large part of the sequences from the outbreak suspected ST-131 isolates is not present in JJ1886 (data not shown). Hence, the SNP analysis using JJ1886 as reference would be based on a smaller amount of data. Thus, using a universal reference could decrease the resolution, while using only local isolates would allow for only local comparisons and not between different outbreaks and laboratories. For bacteria such as *M. tuberculosis* with a more clonal evolution, SNP analysis might be favorable. However, as some of the SNPs will depend on the methodology and assembly processes (22), results may not be comparable between laboratories or over time. As both approaches of analysis (SNP vs cgMLST) resulted in the same clustering, our results indicate that they are not biased by the methodology selected.

Our study have several limitations. The sample investigated is small and is restricted to the isolates in the nosocomial outbreak. Furthermore, sequencing was not performed for all isolates at the same time and two different Illumina platforms were used, which may cause differences in sequence errors affecting downstream analysis.

Most studies today (9, 13, 28) analyze a single colony when performing epidemiological analyses, which may lead to a biased interpretation as heterogeneities in the strains giving rise to the infection may not be represented (37). Furthermore, for epidemiological purposes and due to the lateral gene transfer within *Enterobacteriacae*, it should be considered that the resistance gene-containing plasmids should be primarily targeted in addition to the bacterial core genome (38).

In this study, we introduced WGS-based methods such as cgMLST analysis to confirm epidemiologic links (Fig. S1) within a clinically identified nosocomial outbreak of *E. coli* ST-131 previously investigated using old generation typing methods such as PFGE. Even if WGS-based methods lack standardized analysis protocols, cgMLST and SNP analyses did finally resolve the epidemiological links within the nosocomial outbreak due to improved discriminatory power.

Part of the sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

FUNDING

This work was supported by the research council of South-east Sweden (FORSS) and Marianne and Marcus Wallenberg foundation.

CONFLICT OF INTERESTS

The authors report no conflict of interests.

ETHICAL APPROVAL

All procedures were performed in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable standards.

REFERENCES

1. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012. Intensive Care Med 2013;39:165–228.
2. Spellberg B, Bartlett JG, Gilbert DN. The future of antibiotics and resistance. N Engl J Med 2013;368:299–302.
3. Canton R, Ruiz-Garbajosa P. Co-resistance: an opportunity for the bacteria and resistance genes. Curr Opin Pharmacol 2011;11:477–85.
4. van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, et al. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin Microbiol Infect 2007;13:1–46.
5. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci USA 1998;95:3140–5.
6. Perez-Losada M, Cabezas P, Castro-Nallar E, Crandall KA. Pathogen typing in the genomics era: MLST and the future of molecular epidemiology. Infect Genet Evol 2013;16:38–53.
7. Woksepp H, Jernberg C, Tarnberg M, Ryberg A, Brolund A, Nordvall M, et al. High-resolution melting-curve analysis of ligation-mediated real-time PCR for rapid evaluation of an epidemiological outbreak of extended-spectrum-beta-lactamase-producing Escherichia coli. J Clin Microbiol 2011;49:4032–9.
8. Woksepp H, Ryberg A, Billstrom H, Hallgren A, Nilsson LE, Marklund BI, et al. Evaluation of high-resolution melting curve analysis of ligation-mediated real-time PCR, a rapid method for epidemiological typing of ESKAPE (Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa, and Enterobacter Species) pathogens. J Clin Microbiol 2014;52:4339–42.
9. de Been M, Pinholt M, Top J, Bletz S, Mellmann A, van Schaik W, et al. Core genome multilocus sequence typing scheme for high- resolution typing of enterococcus faecium. J Clin Microbiol 2015;53:3788–97.
10. Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, et al. Prospective genomic characterization of the German enterohemorrhagic Escherichia coli O104:H4 outbreak by rapid next generation sequencing technology. PLoS One 2011;6:e22751.
11. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 2012;67:2640–4.
12. Pak TR, Kasarskis A. How next-generation sequencing and multiscale data analysis will transform infectious disease management. Clin Infect Dis 2015;61:1695–702.
13. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. J Clin Microbiol 2014;52:1501–10.
14. Brolund A, Franzen O, Melefors O, Tegmark-Wisell K, Sandegren L. Plasmidome-analysis of ESBL-producing escherichia coli using conventional typing and high-throughput sequencing. PLoS One 2013;8:e65793.
15. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, et al. Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genet 2009;5:e1000344.
16. Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, Gajer P, et al. The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. J Bacteriol 2008;190:6881–93.
17. Kaas RS, Friis C, Ussery DW, Aarestrup FM. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse Escherichia coli genomes. BMC Genom 2012;13:577.
18. Croxen MA, Finlay BB. Molecular mechanisms of Escherichia coli pathogenicity. Nat Rev Microbiol 2010;8:26–38.
19. Nicolas-Chanoine MH, Bertrand X, Madec JY. Escherichia coli ST131, an intriguing clonal group. Clin Microbiol Rev 2014;27:543–74.
20. Rogers BA, Sidjabat HE, Paterson DL. Escherichia coli O25b-ST131: a pandemic, multiresistant, community-associated strain. J Antimicrob Chemother 2011;66:1–14.
21. Fischer J, Hille K, Ruddat I, Mellmann A, Köck R, Kreienbrock L. Simultaneous occurrence of MRSA and ESBL-producing Enterobacteriaceae on pig farms and in nasal and stool samples from farmers. Vet Microbiol 2017;200:107–13.
22. Kaas RS, Leekitcharoenphon P, Aarestrup FM, Lund O. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. PLoS One 2014;9:e104984.
23. Leopold SR, Goering RV, Witten A, Harmsen D, Mellmann A. Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis

for typing and detection of virulence and antibiotic resistance genes. J Clin Microbiol 2014;52:2365–70.
24. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, et al. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 2012;50:1355–61.
25. Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, et al. Sex and virulence in Escherichia coli: an evolutionary perspective. Mol Microbiol 2006;60:1136–51.
26. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother 2014;58:3895–903.
27. Andersen PS, Stegger M, Aziz M, Contente-Cuomo T, Gibbons HS, Keim P, et al. Complete Genome sequence of the Epidemic and highly virulent CTX-M-15-producing H30-Rx subclone of Escherichia coli ST131. Genome Announc 2013;1:pii: e00988–13.
28. Kohl TA, Diel R, Harmsen D, Rothganger J, Walter KM, Merker M, et al. Whole-genome-based Mycobacterium tuberculosis surveillance: a standardized, portable, and expandable approach. J Clin Microbiol 2014;52:2479–86.
29. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 2008;18:821–9.
30. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST ring image generator (BRIG): simple prokaryote genome comparisons. BMC Genom 2011;12:402.
31. Ender PT, Gajanana D, Johnston B, Clabots C, Tamarkin FJ, Johnson JR. Transmission of an extended-spectrum-beta-lactamase-producing Escherichia coli (sequence type ST131) strain between a father and daughter resulting in septic shock and Emphysematous pyelonephritis. J Clin Microbiol 2009;47:3780–2.
32. Johnson JR, Anderson JT, Clabots C, Johnston B, Cooperstock M. Within-household sharing of a fluoroquinolone-resistant Escherichia coli sequence type ST131 strain causing pediatric osteoarticular infection. Pediatr Infect Dis J 2010;29:473–5.
33. Alm E. NGS surveillance and diagnostics, from development to routine. NEXT Generation Sequencing – NGS. Solna, Sweden: The Public Health Agency, 2016.
34. Mellmann A, Bletz S, Boking T, Kipp F, Becker K, Schultes A, et al. Real-time Genome sequencing of resistant bacteria provides precision infection control in an institutional setting. J Clin Microbiol 2016;54:2874–81. https://doi.org/10.1128/JCM.00790-16.
35. Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, Khanh Nhu NT, Roberts LW, Stanton-Cook M, et al. Sequential acquisition of virulence and fluoroquinolone resistance has shaped the evolution of Escherichia coli ST131. MBio 2016;7:2.
36. Chen TW, Gan RC, Chang YF, Liao WC, Wu TH, Lee CC, et al. Is the whole greater than the sum of its parts? De novo assembly strategies for bacterial genomes based on paired-end sequencing. BMC Genom 2015;16:648.
37. Paterson GK, Harrison EM, Murray GG, Welch JJ, Warland JH, Holden MT, et al. Capturing the cloud of diversity reveals complexity and heterogeneity of MRSA carriage, infection and transmission. Nat Commun 2015;6:6560.
38. Nyberg LK, Quaderi S, Emilsson G, Karami N, Lagerstedt E, Muller V, et al. Rapid identification of intact bacterial resistance plasmids via optical mapping of single DNA molecules. Sci Rep 2016;6:30410.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article:

Figure S1. Time line and location of retrieved *E. coli* isolates.
Table S1. The cgMLST target definer results from SeqSphere + .
Table S2. Results from cluster analysis using cgMLST scheme defined by SeqSphere + .