AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology

Whole-Genome Multilocus Sequence Typing of Extended-Spectrum-Beta-Lactamase-Producing *Enterobacteriaceae*

Marjolein F. Q. Kluytmans-van den Bergh,[a,b] John W. A. Rossen,[c] Patricia C. J. Bruijning-Verhagen,[b] Marc J. M. Bonten,[b,d] Alexander W. Friedrich,[c] Christina M. J. E. Vandenbroucke-Grauls,[e] Rob J. L. Willems,[d] Jan A. J. W. Kluytmans,[a,b] on behalf of the SoM Study Group

Amphia Academy Infectious Disease Foundation, Amphia Hospital, Breda, the Netherlands[a]; Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands[b]; Department of Medical Microbiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands[c]; Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, the Netherlands[d]; Department of Medical Microbiology and Infection Control, VU University Medical Center, Amsterdam, the Netherlands[e]

Molecular typing has become indispensable in the detection of nosocomial transmission of bacterial pathogens and the identification of sources and routes of transmission in outbreak settings, but current methods are labor-intensive, are difficult to standardize, or have limited resolution. Whole-genome multilocus sequence typing (wgMLST) has emerged as a whole-genome sequencing (WGS)-based gene-by-gene typing method that may overcome these limitations and has been applied successfully for several species in outbreak settings. In this study, genus-, genetic-complex-, and species-specific wgMLST schemes were developed for *Citrobacter* spp., the *Enterobacter cloacae* complex, *Escherichia coli*, *Klebsiella oxytoca*, and *Klebsiella pneumoniae* and used to type a national collection of 1,798 extended-spectrum-beta-lactamase-producing *Enterobacteriaceae* (ESBL-E) isolates obtained from patients in Dutch hospitals. Genus-, genetic-complex-, and species-specific thresholds for genetic distance that accurately distinguish between epidemiologically related and unrelated isolates were defined for *Citrobacter* spp., the *E. cloacae* complex, *E. coli*, and *K. pneumoniae*. wgMLST was shown to have higher discriminatory power and typeability than *in silico* MLST. In conclusion, the wgMLST schemes developed in this study facilitate high-resolution WGS-based typing of the most prevalent ESBL-producing species in clinical practice and may contribute to further elucidation of the complex epidemiology of antimicrobial-resistant *Enterobacteriaceae*. wgMLST opens up possibilities for the creation of a Web-accessible database for the global surveillance of ESBL-producing bacterial clones.

The continuing global spread of extended-spectrum-beta-lactamase-producing *Enterobacteriaceae* (ESBL-E) constitutes a major public health threat in both health care and community settings (1–4). Extended-spectrum beta-lactamases (ESBL) confer resistance to the majority of beta-lactam antibiotics, including third-generation cephalosporins, which limits the options for antimicrobial therapy and results in increased morbidity and mortality and health care costs (5, 6). Infection control guidelines recommend prevention of the spread of ESBL-E in health care settings (7, 8). Molecular typing has become an important tool in infection control, as it enables the detection of nosocomial transmission of bacterial pathogens and the identification of sources and routes of transmission in outbreak settings. Different molecular typing methods have been used, ranging from fingerprint-based methods, like pulsed-field gel electrophoresis (PFGE) and amplified fragment length polymorphism (AFLP), to sequence-based methods, like multilocus sequence typing (MLST) (9–11). Although widely adopted, PFGE is labor-intensive and difficult to standardize and has limited interlaboratory reproducibility (12–14). AFLP is less time-consuming than PFGE but may have lower discriminatory power for typing of *Enterobacteriaceae* (13, 15). MLST targets 7 or 8 housekeeping genes, depending on the *Enterobacteriaceae* species involved. It is a relatively fast, accurate, and reproducible strain-typing method, which has enabled the creation of an unambiguous nomenclature for bacterial clones. However, sequence conservation in housekeeping genes limits its discriminatory power and thus its use in outbreak investigations (16–18). Whole-genome sequencing (WGS) has emerged as an ultimate typing tool that fits any bacterial species, study type, and laboratory (19–27). Knowledge of the sequence of the entire genome, rather than of only a few loci, has greatly increased the precision of molecular epidemiologic data and allows its use in infection control, where questions of similarity of isolates focus on recently diverged isolates. Although promising, the analysis, interpretation, and interlaboratory comparability of WGS-based typing results are still challenging. Whole-genome multilocus sequence typing (wgMLST) is a WGS-based typing method that extends MLST to the genome level (20). Today, such gene-by-gene comparison of WGS data has been applied successfully for several species in outbreak settings, but data for *Enterobacteriaceae* and nonoutbreak settings are still limited (22–27). The aim of this study was to develop representative wgMLST schemes for the

Received 2 August 2016 **Returned for modification** 26 August 2016
Accepted 8 September 2016

Accepted manuscript posted online 14 September 2016

Citation Kluytmans-van den Bergh MFQ, Rossen JWA, Bruijning-Verhagen PCJ, Bonten MJM, Friedrich AW, Vandenbroucke-Grauls CMJE, Willems RJL, Kluytmans JAJW, on behalf of the SoM Study Group. 2016. Whole-genome multilocus sequence typing of extended-spectrum-beta-lactamase-producing *Enterobacteriaceae*. J Clin Microbiol 54:2919–2927. doi:10.1128/JCM.01648-16.

Editor: N. A. Ledeboer, Medical College of Wisconsin

Address correspondence to Marjolein F. Q. Kluytmans-van den Bergh, marjoleinkluytmans@gmail.com.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/JCM.01648-16.

most prevalent ESBL-E species and to define thresholds for wgMLST-based genetic distance that reliably distinguish between epidemiologically related and unrelated ESBL-E isolates.

MATERIALS AND METHODS

ESBL-E isolates. A national collection of ESBL-E isolates was prospectively gathered in the SoM study, a multicenter cluster-randomized study comparing different isolation strategies for known ESBL-E carriers that was performed in 14 Dutch hospitals between 2011 and 2014 (28). ESBL-E isolates were obtained from routine clinical cultures and from perianal screening cultures that were taken during 690 repeated ward-based prevalence surveys performed in the participating hospitals. Descriptive characteristics, including species, date of culture, specimen type, patient, prevalence survey, ward, and hospital, were recorded for each isolate.

Microbiological methods. Detection of ESBL-E in clinical cultures was performed according to national and international guidelines (29, 30). Screening cultures consisted of perianal swabs that were preenriched in a selective tryptic soy broth and subsequently cultured on a selective ESBL screening agar (EbSA) (Cepheid Benelux, Apeldoorn, the Netherlands) as described previously (31). For all oxidase-negative Gram-negative bacteria that grew on the EbSA plate, species identification was performed with Vitek MS (bioMérieux, Marcy l'Etoile, France), using the Vitek MS Knowledge Base v1.1 (until January 2013) or v2.0 (from January 2013). The production of ESBL was phenotypically confirmed by the combination disk diffusion method with cefotaxime (30 μg), ceftazidime (30 μg), and cefepime (30 μg), both alone and combined with clavulanic acid (10 μg; Rosco, Taastrup, Denmark). Test results were considered positive if the inhibition zone around the disk was ≥5 mm larger for the combination with clavulanic acid (29, 30).

WGS. EbSA plates were inoculated with phenotypically confirmed ESBL-E isolates and incubated for 18 to 24 h at 35 to 37°C. For each isolate, several colonies (about 5 μl) of the culture were suspended in 300 μl microbead solution, which was subjected to DNA extraction with the Ultraclean Microbial DNA isolation kit (Mo Bio Laboratories, Carlsbad, CA, USA). The DNA concentration and purity were measured using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and the Qubit double-stranded DNA (dsDNA) HS and BR assay kits (Life Technologies, Carlsbad, CA, USA). One nanogram of bacterial DNA was used for library preparation. The DNA library was prepared using the Nextera XT library preparation kit with the Nextera XT v2 index kit (Illumina, San Diego, CA, USA). The library fragment length was aimed at fragments with a median size of 575 bases and was assessed with the Genomic DNA ScreenTape assay with the 2200 TapeStation system (Agilent Technologies, Waldbronn, Germany). Subsequently, the library was sequenced on either a MiSeq sequencer, using the MiSeq reagent kit v2 generating 250-bp paired-end reads, or on a HiSeq 2500 sequencer, using the HiSeq Rapid SBS kit v2 generating 100-bp paired-end reads (Illumina, San Diego, CA, USA). Sequencing was aimed at a coverage of at least 60-fold. MiSeq data were processed with MiSeq control software v2.4.0.4 and MiSeq Reporter v2.4 and HiSeq data with bcl2fastq2 conversion software v1.8.4 (Illumina, San Diego, CA, USA). Quality trimming of reads was performed with CLC Genomics Workbench 7.0.4 (Qiagen, Hilden, Germany) using a minimum Phred (Q) score of 28. A random and blinded sample of eight isolates was sequenced in duplicate.

***De novo* assembly.** *De novo* assembly was performed using CLC Genomics Workbench 7.0.4 (Qiagen, Hilden, Germany) with optimal word sizes based on the maximum N50 (the largest scaffold length, N, such that 50% of the assembled genome size is contained in scaffolds with a length of at least N) value. For each assembled genome, the coverage (mean depth), the number of scaffolds, the N50, the maximum scaffold length, and the percentage of the expected genome size were determined. Assembled genomes were excluded from further analysis if they did not fulfill the following quality criteria: (i) coverage, at least 20; (ii) number of scaffolds, at most 1,000; (iii) N50, at least 15,000 bases; (iv) maximum scaffold length, at least 50,000 bases; and (v) assembled genome size, between 90% and 115% of the expected genome size.

WGS-based species identification, detection of ESBL-encoding genes, and MLST. Assembled genomes were uploaded to the online bioinformatics tools KmerFinder v2.0, ResFinder v2.1, and MLST v1.7 (Center for Genomic Epidemiology, Technical University of Denmark, Lingby, Denmark) (32–34). Species identification was based on the "winner takes it all" scoring method of KmerFinder. ESBL-encoding genes were reported when at least 60% of the length of the best matching gene in the ResFinder database was covered with a sequence identity of at least 90%. A nonstringent threshold for the percent alignment (60%) of ESBL-encoding genes was used to account for the fact that short-read sequence data are not optimal for the assembly of plasmids, which may occasionally result in incomplete assembly of ESBL-encoding genes. Conventional MLST schemes in MLST v1.7 were based on the publicly available PubMLST.org database (http://www.pubmlst.org) and were available for *Citrobacter freundii*, *Enterobacter cloacae*, *Escherichia coli*, *Klebsiella oxytoca*, and *Klebsiella pneumoniae* (see Table S1 in the supplemental material). For *E. coli*, the Achtman MLST scheme was used (35). A sequence type (ST) was assigned when all MLST loci in the assembled genome perfectly matched all MLST alleles of a known ST in the MLST database, i.e., 100% sequence identity and 100% alignment for all alleles (32). When all MLST loci perfectly matched known MLST alleles but the combination of alleles was unknown in the database, isolates were classified as "unknown ST, perfect match." These isolates were considered typeable with MLST based on their allele profiles. Isolates with a sequence identity below 100% for one or more of the MLST loci were classified as "unknown ST, minor mismatch," and isolates with an alignment below 100% for one or more of the MLST loci were classified as "unknown ST, major mismatch." Isolates with minor or major mismatches were considered nontypeable with MLST.

Definition of WGS-based typing schemes. The MLST+ target definer function of Ridom SeqSphere+ (Ridom, Münster, Germany) was used to define target gene sets for all species with at least 10 isolates available. Depending on the number of available complete reference genomes, typing schemes were developed at the genus level (*Citrobacter* spp.), genetic-complex level (*E. cloacae* complex), or species level (*E. coli*, *K. oxytoca*, and *K. pneumoniae*). For each scheme, one annotated and publicly available complete genome (chromosome) was used as a reference genome (see Table S2 in the supplemental material). Open reading frames were extracted from the reference genomes and filtered by applying a minimum-length filter that discarded all genes with a length below 50 bases; a start codon filter that discarded all genes that contained no start codon at the beginning of the gene; a stop codon filter that discarded all genes that contained no stop codon, more than one stop codon, or a premature stop codon; an excluded sequences filter that discarded plasmid sequences; and a homologous gene filter that discarded all genes with fragments that occurred in multiple copies within the genome (with a sequence identity of at least 90% and more than 100-base overlap). Subsequently, the remaining set of reference genome genes was compared to a varying number of publicly available query genomes using the Basic Local Alignment Search Tool (BLAST) (36). The type and number of query genomes that were used differed by species and were dependent on the availability of annotated genomes (see Table S2 in the supplemental material). A gene overlap filter was used to identify reference genome genes that overlapped by more than 4 bases and to filter out the shorter gene. In addition, a stop codon percentage filter was applied to filter out genes that contained no stop codon, more than one stop codon, or a premature stop codon in more than 80% of the query genomes.

Reference genome genes that passed the gene overlap filter and the stop codon percentage filter were designated whole-genome targets. Of those, genes that were present in all query genomes with a sequence identity of at least 90% and an alignment of 100% were classified as core genome targets; the remaining genes were classified as accessory genome

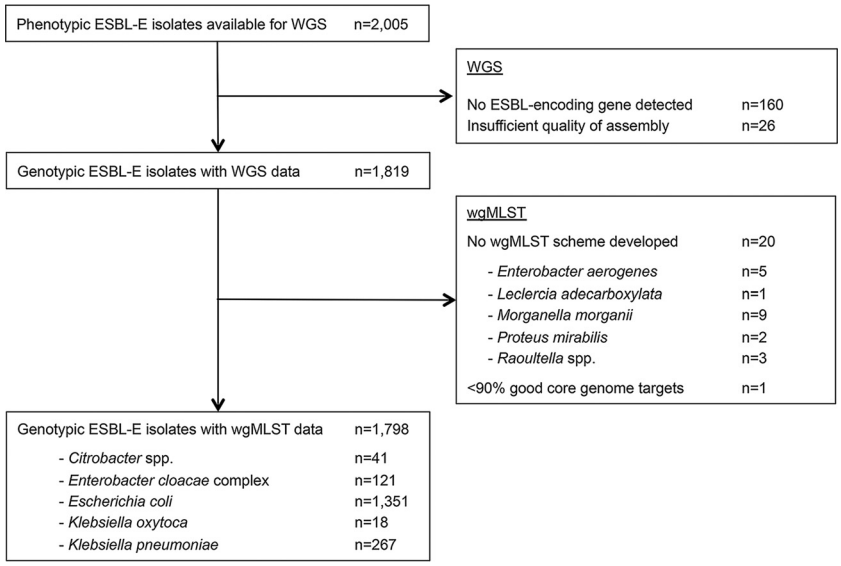

FIG 1 Availability of ESBL-E isolates for WGS and wgMLST.

targets. wgMLST schemes included all whole-genome targets; core genome MLST (cgMLST) schemes included only the core genome targets. An overview of the typing schemes, including the number of core, accessory, and whole-genome targets per typing scheme, is provided in Table S3 in the supplemental material.

Genetic distance. For each assembled genome, all whole-genome target genes that were present with a sequence identity of at least 90% and an alignment of 100% were extracted using Ridom SeqSphere+ software v 3.0 (Ridom, Münster, Germany). Genes that could not be identified in the assembled genome were classified as "missing targets." Genes with an alignment below 100% or a sequence identity below 90% were classified as "failed targets." The remaining genes were classified as "good targets." To evaluate the applicability and representativeness of the core genome target set, the percentage of good core genome targets was assessed for all assembled genomes. Assembled genomes for which the percentages of good core genome targets were below 90% were excluded from further analyses. All-to-all distance matrices, describing pairwise genetic distances, were constructed separately for core, accessory, and whole-genome target gene sets. The pairwise genetic distance was defined as the proportion of allele differences and was calculated by dividing the number of allele differences by the total number of good targets shared by both sequences, i.e., pairwise ignoring missing values.

Classification of pairwise comparisons. Pairwise comparisons of assembled genomes were classified according to the known epidemiological link between the isolates: (i) same patient; (ii) same survey but different patients; (iii) same ward but different surveys and patients; (iv) same hospital but different wards, surveys, and patients; and (v) no known epidemiological link, i.e., different hospitals, wards, surveys, and patients.

Definition of epidemiologically related and unrelated isolates. In the absence of a gold standard molecular typing method with perfect discriminatory power, epidemiological-link data were used to define related and unrelated isolates, where maximal contrast between related and unrelated isolates was sought. Epidemiologically related isolates were defined as being obtained from the same patient, belonging to the same ST, and cultured within a time window of 30 days. The last two restrictions were added to exclude potential bias from simultaneous carriage of multiple ESBL-E isolates from different clonal lineages and within-patient microevolution of isolates over time. Epidemiologically unrelated isolates were defined as being obtained from patients without any known epidemiological link.

Statistical analysis. The Kruskal-Wallis H test and sign test were used to test for differences in median genetic distance between species and between core and accessory genomes, respectively. All analyses were performed with IBM SPSS Statistics v22 (IBM Corp., Armonk, New York, USA).

Ethical considerations. This study was reviewed by the Medical Research and Ethics Committee (MREC) of the Elisabeth/TweeSteden Hospital (Tilburg, the Netherlands) in December 2010. The study was judged to be beyond the scope of the Medical Research Involving Human Subjects Act (in Dutch, Wet Medisch-wetenschappelijk Onderzoek met Mensen [WMO]), and a waiver of written informed consent was obtained (METC/jv/2010.034). Patients who participated in the study provided verbal consent for use of demographic, clinical, and culture data.

Accession number(s). All generated raw reads were submitted to the European Nucleotide Archive (ENA) of the European Bioinformatics Institute (EBI) under the study accession number PRJEB15226.

RESULTS

A total of 2,005 phenotypic ESBL-E isolates were collected and available for WGS (Fig. 1). No ESBL-encoding gene could be detected in the assembled genomes of 160 (8.0%) isolates. For 26 (1.3%) isolates, the assembly was of insufficient quality. A summary of assembly characteristics for 1,819 genotypic ESBL-E isolates with good quality WGS data is provided in Table S4 in the supplemental material. WGS-based species identification was in accordance with that of Vitek-MS for 1,808 isolates. For 11 isolates that were identified as *Citrobacter* spp. by Vitek-MS, KmerFinder did not provide a high-confidence result for any species, probably due to the limited number of *Citrobacter* sp. reference genomes in the database. No wgMLST scheme was developed for *Enterobacter aerogenes* ($n = 5$), *Leclercia adecarboxylata* ($n = 1$), *Morganella morganii* ($n = 9$), *Proteus mirabilis* ($n = 2$), or *Raoultella* spp. ($n = 3$). One *E. coli* isolate for which only 71.4% of core genome targets could be detected in the assembled genome was excluded from wgMLST analyses. For the remaining 1,798 ESBL-E isolates, the percentage of core genome targets present in the assembled genomes was at least 90.0% (see Table S5 in the supplemental material), which is indicative of the representativeness of the wgMLST schemes for the ESBL-E population studied. As expected, the percentage of accessory genome targets present in the

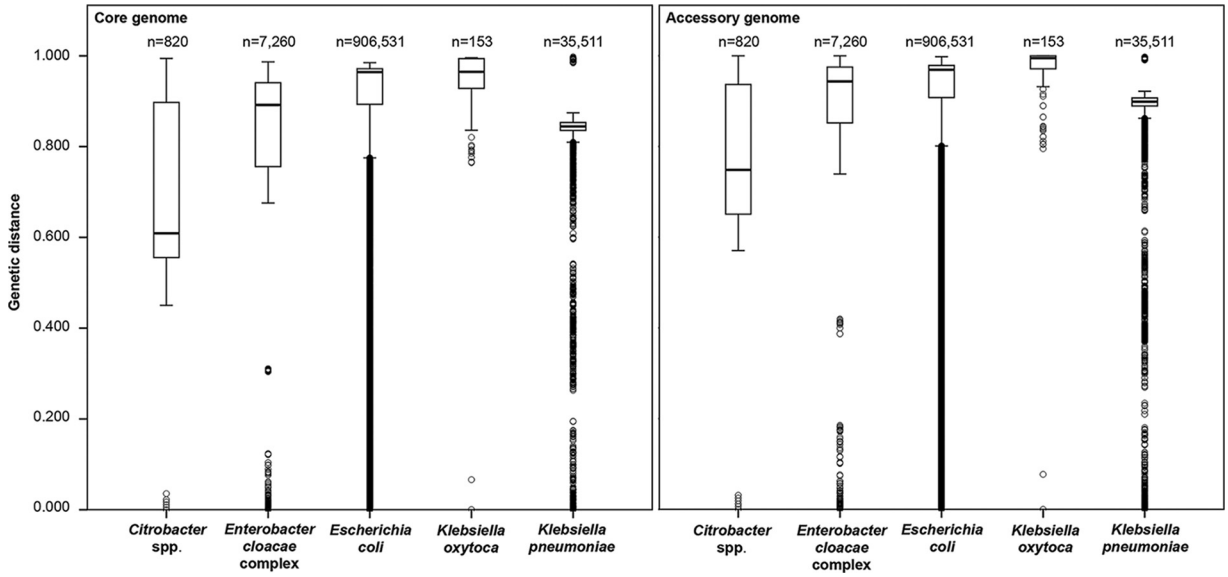

FIG 2 Distribution of genetic distances for 950,275 pairwise comparisons of extended-spectrum beta-lactamase-producing *Enterobacteriaceae*, shown separately for core genome and accessory genome. The boxes represent the interquartile (IQ) range. The whiskers extend to the highest and lowest values that are not greater than 1.5 times the IQ range. The lines across the boxes indicate the medians. The circles represent outliers, i.e., comparisons with values more than 1.5 times the IQ range. *n* defines the number of pairwise comparisons for each species. Kruskal-Wallis chi-square test results for difference between species were as follows: chi-square$_{\text{core genome}} = 22{,}685$, $P < 0.0005$; chi-square$_{\text{accessory genome}} = 15{,}980$, $P < 0.0005$. Sign test for the difference between core and accessory genomes, $P < 0.0005$.

assembled genomes varied considerably within and between species.

The sequence identity of the ESBL-encoding genes that were identified was 100% for 1,751 (97.6%) ESBL-E isolates. For the remaining 47 (2.4%) isolates, the percent sequence identity with known ESBL-encoding genes ranged from 99.1% to 99.9%. The percent alignment of the ESBL-encoding genes was 100% for 1,789 isolates (99.5%) and ranged from 66.6% to 97.9% for the other 9 (0.5%) isolates.

The all-to-all genetic distance matrices comprised 950,275 unique pairwise comparisons. Figure 2 shows the species-specific distributions of genetic distance for core and accessory genome targets. Pairwise genetic distances ranged from 0.000 to (approximately) 1.000 for both core and accessory genomes and for all species. However, the distribution of genetic distances differed significantly between species ($P < 0.0005$) and between core and accessory genomes ($P < 0.0005$). For the core genome, median genetic distances ranged from 0.609 for *Citrobacter* spp. to 0.965 for *K. oxytoca* and, for the accessory genome, from 0.748 for *Citrobacter* spp. to 0.995 for *K. oxytoca*. For all species, the median genetic distance was higher for the accessory genome than for the core genome. Summary statistics of the distribution of numbers of alleles compared; numbers of different alleles; and genetic distances for core genome, accessory genome, and whole genome are presented by species in Table S6 in the supplemental material. To assess the added value of the accessory genome in terms of genetic distance, the difference between whole-genome and core genome genetic distances was calculated for each pairwise comparison (Fig. 3). For the majority of pairwise comparisons (82.0% [779,316/950,275]), the whole-genome genetic distance was larger than the core genome distance, although the absolute difference differed between species. For the remaining comparisons, the inclusion of accessory genome targets reduced the genetic distance. Five *E. coli* and three *K. pneumoniae* isolates were sequenced in duplicate. The results were found to be reproducible, with genetic distances ranging from 0.000 to 0.001 for both core and whole-genome targets.

Table 1 provides the distributions of genetic distances for isolates with the same (conventional) ST and isolates with different

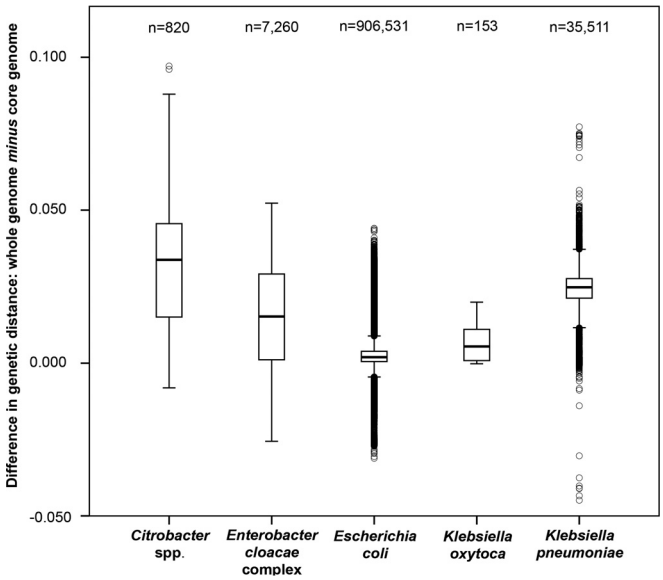

FIG 3 Distribution of the differences in genetic distance between whole genome and core genome for 950,275 pairwise comparisons of extended-spectrum beta-lactamase-producing *Enterobacteriaceae*. The boxes represent the IQ range. The whiskers extend to the highest and lowest values that are not greater than 1.5 times the IQ range. The lines across the boxes indicate the medians. The circles represent outliers, i.e., pairwise comparisons with values more than 1.5 times the IQ range. *n* defines the number of pairwise comparisons for each species. Kruskal-Wallis chi-square test results for difference between species were as follows: chi-square = 69,760, $P < 0.0005$.

TABLE 1 Genetic distances for pairwise comparisons of ESBL-E isolates within and between conventional MLST STs[a]

MLST and species	Within ST (n = 68,482)			Between ST (n = 834,045)		
		Genetic distance			Genetic distance	
	Pairwise comparison (n)	Median	Range	Pairwise comparison (n)	Median	Range
cgMLST						
Citrobacter spp.	14	0.000	[0.000–0.036]	31	0.897	[0.537–0.917]
E. cloacae complex	430	0.006	[0.000–0.104]	5,135	0.881	[0.088–0.987]
E. coli	67,108	0.022	[0.000–0.507]	796,847	0.966	[0.019–0.985]
K. oxytoca	1	0.066	[0.066–0.066]	65	0.992	[0.789–0.996]
K. pneumoniae	929	0.020	[0.000–0.542]	31,967	0.844	[0.090–0.998]
wgMLST						
Citrobacter spp.	14	0.000	[0.000–0.027]	31	0.913	[0.558–0.933]
E. cloacae complex	430	0.007	[0.000–0.129]	5,135	0.902	[0.102–0.988]
E. coli	67,108	0.027	[0.000–0.520]	796,847	0.968	[0.021–0.986]
K. oxytoca	1	0.070	[0.070–0.070]	65	0.994	[0.795–0.996]
K. pneumoniae	929	0.022	[0.000–0.566]	31,967	0.870	[0.104–0.997]

[a] Excluding 47,748 (5.0%) pairwise comparisons for which no ST could be assigned to either of the isolates.

STs. As expected, wgMLST and cgMLST enabled further differentiation of isolates that belong to the same ST. The broadest within-ST ranges in genetic distance were observed for *E. coli* and *K. pneumoniae*, with ranges from 0.000 up to 0.520 and from 0.000 up to 0.566 (wgMLST), respectively. Although median genetic distances clearly differed for within-ST and between-ST comparisons, the ranges of genetic distances overlapped for *E. cloacae*, *E. coli*, and *K. pneumoniae*, indicating that some isolates with different STs can be genetically more closely related than isolates with the same ST. For 47,748 (5.0%) pairwise comparisons, no ST could be assigned to either of the isolates because of minor or major mismatches for at least one of the seven alleles involved.

The distributions of genetic distances for epidemiologically related and unrelated isolates were clearly different. Figure 4 shows the distributions of wgMLST-based genetic distances for epidemiologically related and unrelated ESBL-producing *E. coli*. Comparable distributions were observed for epidemiologically related and unrelated isolates of the other ESBL-E species (see Fig. S1 in the supplemental material) and for cgMLST-based genetic distances (data not shown). Based on these distributions for epidemiologically related and unrelated isolates, thresholds for genetic distance were set at the minimal value that classified 100% of epidemiologically related isolate pairs as clonally related (Table 2). The absence of a pair of epidemiologically related *K. oxytoca* isolates in the ESBL-E strain collection precluded the definition of thresholds for *K. oxytoca*. Thresholds differed between species, with two to three times higher values for *E. coli* than for the other three species. Thresholds for cgMLST- and wgMLST-based genetic distances were similar for *Citrobacter* spp. and the *E. cloacae* complex but differed slightly for *E. coli* (a lower threshold for wgMLST) and *K. pneumoniae* (a lower threshold for cgMLST). For all genetic distance thresholds, the corresponding percentage of presumed epidemiologically unrelated isolate pairs that were classified as clonally related was low (0.1% to 0.5%), which indicates high discriminatory power, i.e., the ability to assign different types to different isolates (37).

Thresholds for genetic distance were subsequently used to type all 1,798 ESBL-E isolates, and the results were compared with those of *in silico* MLST (Table 3). Epidemiological concordance was observed for all three typing methods, i.e., the percentage of isolates identified as clonally related was highest for within-patient comparisons and decreased with waning degrees of epidemiological linkage. The typeability, i.e., the percentage of isolates to which a genotype could be assigned, was significantly higher for cgMLST and wgMLST (99.98%) than for *in silico* MLST (95.0%; $P < 0.0005$) (37). In addition, cgMLST and wgMLST had higher discriminatory power than *in silico* MLST, i.e., classification of presumed epidemiologically unrelated isolates as clonally related was significantly less frequent for cgMLST (0.5%) and wgMLST (0.3%) than for *in silico* MLST (7.1%, $P < 0.0005$). Extending cgMLST to wgMLST resulted in a 0.2% increase in the overall percentage of correct classification of presumed epidemiologically

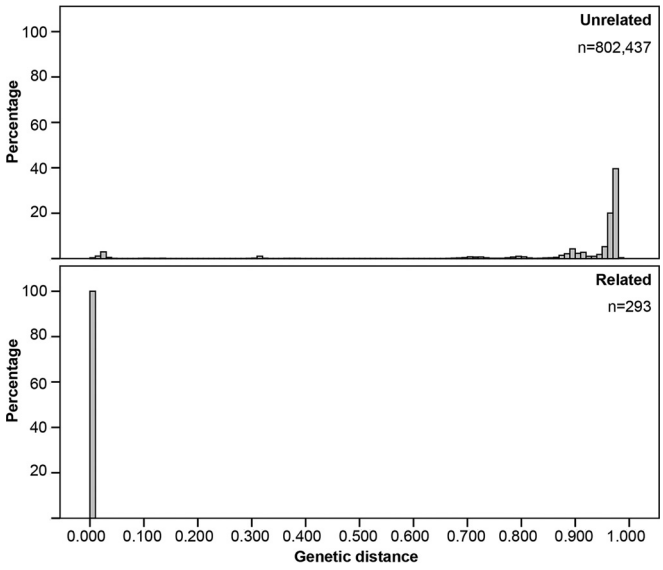

FIG 4 Distributions of wgMLST-based genetic distances for pairwise comparisons of epidemiologically related and unrelated extended-spectrum beta-lactamase-producing *Escherichia coli* isolates. Epidemiologically related isolates were defined as being obtained from the same patient, belonging to the same ST, and cultured within a time window of 30 days. Epidemiologically unrelated isolates were defined as being obtained from patients without a known epidemiological link.

TABLE 2 Thresholds for genetic distance with corresponding classification of epidemiologically unrelated ESBL-E isolates[a]

MLST and species	Threshold for genetic distance[b]	Threshold-based classification of epidemiologically unrelated isolates[c]		
		Pairwise comparison (*n*)	Clonally related	
			n	%
cgMLST				
Citrobacter spp.	0.0030	693	1	0.1
E. cloacae complex	0.0035	6,256	31	0.5
E. coli	0.0105	802,437	3,999	0.5
K. pneumoniae	0.0035	30,808	75	0.2
wgMLST				
Citrobacter spp.	0.0030	693	1	0.1
E. cloacae complex	0.0035	6,256	17	0.3
E. coli	0.0095	802,437	2,485	0.3
K. pneumoniae	0.0045	30,808	96	0.3

[a] The absence of a pair of epidemiologically related *K. oxytoca* isolates precluded the definition of thresholds for *K. oxytoca*.
[b] Thresholds were set at the minimal value for genetic distance that classified 100% of epidemiologically related ESBL-E isolates as clonally related.
[c] Epidemiologically related isolates were those obtained from the same patient, belonging to the same ST, and cultured within a time window of 30 days. Epidemiologically unrelated isolates were those obtained from patients without a known epidemiological link. Clonally related isolates were those at a genetic distance below or equal to the threshold value.

TABLE 3 Typing results for 950,275 pairwise comparisons of ESBL-E isolates with different degrees of epidemiological linkage by MLST, cgMLST, and wgMLST

Epidemiological link between isolates and typing method	Pairwise comparison					
	Clonally related		Clonally unrelated		Nontypeable[a]	
	n	%	*n*	%	*n*	%
Same patient (*n* = 726)						
MLST	550	75.8	136	18.7	40	5.5
cgMLST	564	77.7	161	22.2	1	0.1
wgMLST	565	77.8	160	22.0	1	0.1
Same survey (*n* = 1,180)						
MLST	168	14.2	930	78.8	82	6.9
cgMLST	129	10.9	1,051	89.1	0	0.0
wgMLST	131	11.1	1,049	88.9	0	0.0
Same ward (*n* = 14,291)						
MLST	1,165	8.2	12,465	87.2	661	4.6
cgMLST	329	2.3	13,960	97.7	2	0.0
wgMLST	301	2.1	13,988	97.9	2	0.0
Same hospital (*n* = 93,740)						
MLST	6,908	7.4	82,414	87.9	4,418	4.7
cgMLST	921	1.0	92,813	99.0	6	0.0
wgMLST	778	0.8	92,956	99.2	6	0.0
Other hospital (*n* = 840,338)						
MLST	59,691	7.1	738,100	87.8	42,547	5.1
cgMLST	4,106	0.5	836,088	99.5	144	0.0
wgMLST	2,599	0.3	837,595	99.7	144	0.0

[a] MLST, unknown ST, minor or major mismatch (*n* = 47,748); cgMLST/wgMLST, no cutoff available for *K. oxytoca* (*n* = 153).

unrelated isolates as clonally unrelated ($P < 0.0005$). In addition, reclassification analysis of individual comparisons showed that extension to wgMLST improved classification for 1,551 of 840,338 (0.2%) epidemiologically unrelated isolate pairs, i.e., presumed epidemiologically unrelated isolate pairs were reclassified from clonally related (cgMLST) to clonally unrelated (wgMLST) (Table 4). On the other hand, 44 of 840,338 (0.005%) presumed epidemiologically unrelated isolate pairs were classified as clonally related by wgMLST, while they were classified as clonally unrelated by cgMLST.

DISCUSSION

In this study, representative genus-, genetic-complex-, and species-specific wgMLST schemes were developed for the five most prevalent *Enterobacteriaceae* species in clinical practice and used to type a large national collection of ESBL-E isolates obtained from patients in Dutch hospitals. Genus-, genetic-complex-, and species-specific thresholds for genetic distance that accurately distinguished between epidemiologically related and unrelated isolates were defined for *Citrobacter* spp., the *E. cloacae* complex, *E. coli*, and *K. pneumoniae*. wgMLST was shown to have higher discriminatory power and typeability than *in silico* MLST.

After the introduction of WGS as a typing tool with high potential, gene-by-gene comparison of WGS data has been applied successfully for several species in outbreak settings (22–27). To our knowledge, this large multicenter study is the first to apply wgMLST for typing of *Enterobacteriaceae* in nonoutbreak settings and to include a comprehensive evaluation of its performance in a large collection of isolates without a clear epidemiological link. The availability of detailed epidemiological data enabled the definition of thresholds for wgMLST-based genetic distance that can be used to distinguish between epidemiologically related and unrelated isolates in hospital settings and provided estimates of the probability of classifying presumed epidemiologically unrelated isolates as clonally related in settings with low levels of transmission.

Available studies on WGS-based typing have shown that reliable phylogenies with good resolution can be inferred from the core genome (16). It has been suggested, though, that resolution at the isolate level may be further improved by the additional targeting of more variable genes from the accessory genome (16). In the present study, wgMLST resulted in a larger genetic distance between isolates than cgMLST for the majority of isolate comparisons and in less frequent classification of presumed epidemiologically unrelated isolates as clonally related and thus had slightly higher discriminatory power. Based on these findings, the use of

TABLE 4 Reclassification of clonal (un)relatedness of epidemiologically unrelated ESBL-E isolates after extension of cgMLST to wgMLST

Pairwise comparison by cgMLST	Pairwise comparison by wgMLST (*n*)		
	Clonally related	Clonally unrelated	Nontypeable
Clonally related	2,555	1,551	0
Clonally unrelated	44	836,044	0
Nontypeable	0	0	144[a]

[a] All nontypeable isolates were *K. oxytoca* isolates.

wgMLST is preferred over cgMLST for the *Enterobacteriaceae* species studied here. However, extension with the accessory genome was shown to result in a decrease in genetic distance for a minority of isolates and in rare cases even in classification of presumed epidemiologically unrelated isolates as clonally related.

Thresholds for genetic distance that distinguished epidemiologically related from unrelated isolates differed between species and were highest for *E. coli*. These high thresholds for *E. coli* may reflect higher mutation and recombination rates or may be related to within-patient microevolution of coexisting isolates with common ancestry. Estimates of the mutation and recombination rates of *Enterobacteriaceae* are limited and range from 0 to 10 mutations per genome per year (38, 39). These rates reflect fixed substitution rates; short-term mutation (polymorphism) rates may be higher and are known to increase under antibiotic pressure or other environmental stress (40).

Due to the lack of *K. oxytoca* isolate pairs that fulfilled the definition of epidemiologically related isolates, thresholds for genetic distance for the species are still to be defined. The distribution of genetic distances for epidemiologically unrelated *K. oxytoca* isolates was comparable to that of the other *Enterobacteriaceae* species and might be used to evaluate proposed thresholds in future studies.

wgMLST was shown to have high discriminatory power. With genetic distance thresholds set at the minimal value that classified 100% of epidemiologically related isolates as clonally related, only 0.3% of presumed epidemiologically unrelated isolates were classified as clonally related. Thus, wgMLST may enable high-resolution detection of nosocomial transmission of antimicrobial-resistant *Enterobacteriaceae*. Classification of presumed epidemiologically unrelated isolates as clonally related may be due to a residual lack of resolution of wgMLST but may also reflect the presence of an unidentified epidemiological link in the hospital or community setting. Other possible explanations may be related to clone-specific characteristics, such as a low mutation rate, converging evolution, or recent introduction into the human population.

Interestingly, about 20% of within-patient comparisons were found to pertain to clonally unrelated isolates, indicating that patients are cocolonized with multiple ESBL-producing isolates. This finding may reflect multiple acquisitions of ESBL-E but may also be the result of within-patient horizontal transfer of ESBL-encoding mobile genetic elements.

An important strength of wgMLST is the use of alleles and not nucleotide polymorphisms as units of comparison. Irrespective of the number of nucleotide polymorphisms involved, each allelic change is numbered as a single event, i.e., an allelic change is related to at least one point mutation, but can also contain several nucleotide changes. Thus, allele-based methods largely reduce the conflicting signals of recombination in determining genetic relatedness of bacteria (41–45). Another major advantage of this gene-by-gene allelic profiling method is the immediate and automated assignment of novel allele variants, which enables a standardized and universal nomenclature and is a prerequisite for successful interlaboratory exchange of data (46). Most studies on wgMLST have used the absolute number of allele differences to compare isolates, which is highly dependent on the number of targets that are compared (22, 23, 26, 27). In the present study, differences in the numbers of targets compared were taken into account by defining genetic distance as the proportion of allele differences. This enabled the unbiased comparison of genetic distances for different pairwise comparisons and the definition of thresholds for genetic distance.

A limitation of allele-based typing is that the analysis is reduced to coding regions only, as second-generation sequencing platforms currently in use produce only relatively short reads that do not facilitate high-quality assembly of highly repetitive intergenic regions. This might change when third-generation sequencing instruments that produce much longer reads from a single molecule become widely available. Nevertheless, the current wgMLST approach will be sustainable, as it will maintain backward compatibility with the expansion of typing schemes, as we see today with the *in silico* extraction of conventional MLST sequence types from WGS data.

In conclusion, the wgMLST schemes developed in this study facilitate high-resolution WGS-based typing of the most prevalent ESBL-producing species in clinical practice and may contribute to further elucidation of the complex epidemiology of antimicrobial-resistant *Enterobacteriaceae*. wgMLST opens up possibilities for the creation of a Web-accessible database for the global surveillance of ESBL-producing bacterial clones.

ACKNOWLEDGMENTS

We are grateful to the infection control practitioners, research nurses, and microbiology technicians in the participating hospitals for their contributions to the collection of epidemiological and microbiological data.

SoM Study Group: Peter van Keulen, Jan Kluytmans, Marjolein Kluytmans-van den Bergh, Kees Verduin, Carlo Verhulst, and Ina Willemsen (Amphia Hospital, Breda, the Netherlands); Els de Brauwer (Atrium MC, Heerlen, the Netherlands); Margreet Vos (Erasmus MC, Rotterdam, the Netherlands); Erika van Elzakker (Haga Hospital, The Hague, the Netherlands); Gijs Ruijs (Isala Clinics, Zwolle, the Netherlands); Guy Oudhuis and Paul Savelkoul (Maastricht UMC, Maastricht, the Netherlands); Bram Diederen (Red Cross Hospital, Beverwijk, the Netherlands); Anton Buiting (St. Elisabeth/TweeSteden Hospital, Tilburg, the Netherlands); Alexander Friedrich, Greetje Kampinga, Erwin Raangs, Sigrid Rosema, and John Rossen (UMC Groningen, Groningen, the Netherlands); Joost Hopman and Andreas Voss (UMC St Radboud, Nijmegen, the Netherlands); Marc Bonten, Martin Bootsma, Patricia Bruijning-Verhagen, Jan Kluytmans, Marjolein Kluytmans-van den Bergh, Annet Troelstra, and Rob Willems (UMC Utrecht, Utrecht, the Netherlands); Christina Vandenbroucke-Grauls (VU University Medical Center, Amsterdam, the Netherlands); Nashwan al Naiemi (Ziekenhuisgroep Twente, Almelo/Hengelo, the Netherlands).

Author contributions were as follows: conception and design of the study, M.F.Q.K.-V.D.B., J.W.A.R., A.W.F., and J.K.; acquisition of data, M.F.Q.K.-V.D.B., J.W.A.R., M.B., A.W.F., C.M.J.E.V.-G., and J.A.J.W.K.; analysis and/or interpretation of data, M.F.Q.K.-V.D.B., J.W.A.R., P.C.J.B.-V., M.J.M.B., R.J.L.W., and J.A.J.W.K.; drafting of the manuscript, M.F.Q.K.-V.D.B.; revising the manuscript, J.W.A.R., P.C.J.B.-V., M.J.M.B., A.W.F., C.M.J.E.V.-G., R.J.L.W., and J.A.J.W.K.; approval of the manuscript, M.F.Q.K.-V.D.B., J.W.A.R., P.C.J.B.-V., M.J.M.B., A.W.F., C.M.J.E.V.-G., R.J.L.W., and J.A.J.W.K.

We have no financial or other relationships relevant to the study to declare.

FUNDING INFORMATION

This work, including the efforts of Marjolein F.Q. Kluytmans-van den Bergh, Marc J.M. Bonten, Alexander W. Friedrich, Christina M. J. E. Vandenbroucke-Grauls, and Jan A. J. W. Kluytmans, was funded by ZonMw (Netherlands Organisation for Health Research and Development) (205100010).

REFERENCES

1. **European Centre for Disease Prevention and Control.** 2015. Antimicrobial resistance surveillance in Europe 2014: annual report of the European Antimicrobial Resistance Surveillance Network (EARS-Net). Euro-

pean Centre for Disease Prevention and Control, Stockholm, Sweden. http://ecdc.europa.eu/en/publications/_layouts/forms/Publication_DispForm.aspx?List=4f55ad51-4aed-4d32-b960-af70113dbb90&ID=1400.
2. **Paterson DL, Bonomo RA.** 2005. Extended-spectrum beta-lactamases: a clinical update. Clin Microbiol Rev **18:**657–686. http://dx.doi.org/10.1128/CMR.18.4.657-686.2005.
3. **Dutch Foundation of the Working Party on Antibiotic Policy.** 2015. NethMap 2015: consumption of antimicrobial agents and antimicrobial resistance among medically important bacteria in the Netherlands in 2014. Dutch Foundation of the Working Party on Antibiotic Policy, Rotterdam, Netherlands. http://www.swab.nl/swab/cms3.nsf/uploads/4F5A0D8E6F0DD139C1257E6E0051833A/$FILE/NethmapMaran2015%20_webversie.pdf.
4. **Woerther PL, Burdet C, Chachaty E, Andremont A.** 2013. Trends in human fecal carriage of extended-spectrum beta-lactamases in the community: toward the globalization of CTX-M. Clin Microbiol Rev **26:**744–758. http://dx.doi.org/10.1128/CMR.00023-13.
5. **Ammerlaan HSM, Troelstra A, Kruitwagen CLJJ, Kluytmans JAJW, Bonten MJM.** 2009. Quantifying changes in incidences of nosocomial bacteraemia caused by antibiotic-susceptible and antibiotic-resistant pathogens. J Antimicrob Chemother **63:**1064–1070. http://dx.doi.org/10.1093/jac/dkp036.
6. **Rottier WC, Ammerlaan HSM, Bonten MJM.** 2012. Effects of confounders and intermediates on the association of bacteraemia caused by extended-spectrum beta-lactamase-producing *Enterobacteriaceae* and patient outcome: a meta-analysis. J Antimicrob Chemother **67:**1311–1320. http://dx.doi.org/10.1093/jac/dks065.
7. **Kluytmans-Vandenberg MFQ, Kluytmans JAJW, Voss A.** 2005. Dutch guideline for preventing nosocomial transmission of highly resistant microorganisms (HRMO). Infection **33:**309–313. http://dx.doi.org/10.1007/s15010-005-5079-z.
8. **Tacconelli E, Cataldo MA, Dancer SJ, De Angelis G, Falcone M, Frank U, Kahlmeter G, Pan A, Petrosillo N, Rodríguez-Baño J, Singh N, Venditti M, Yokoe DS, Cookson B.** 2014. ESCMID guidelines for the management of the infection control measures to reduce transmission of multidrug-resistant Gram-negative bacteria in hospitalized patients. Clin Microbiol Infect **20**(Suppl 1)**:**S1–S55.
9. **Goering RV.** 1993. Molecular epidemiology of nosocomial infection: analysis of chromosomal restriction fragment patterns by pulsed-field gel electrophoresis. Infect Control Hosp Epidemiol **14:**595–600.
10. **Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russel JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A **95:**3140–3145. http://dx.doi.org/10.1073/pnas.95.6.3140.
11. **Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M.** 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res **23:**4407–4414. http://dx.doi.org/10.1093/nar/23.21.4407.
12. **van Belkum A, van Leeuwen W, Kaufmann ME, Cookson B, Forey F, Etienne J, Goering R, Tenover F, Steward C, O'Brien F, Grubb W, Tassios P, Legakis N, Morvan A, El Solh N, de Ryck R, Struelens M, Salmenlinna A, Vuopio-Varkila J, Kooistra M, Talens A, Witte W, Verbrugh H.** 1998. Assessment of resolution and intercenter reproducibility of results of genotyping *Staphylococcus aureus* by pulsed-field gel electrophoresis of SmaI macrorestriction fragments: a multicenter study. J Clin Microbiol **36:**1653–1659.
13. **Hahm BK, Maldonado Y, Schreiber E, Bhunia AK, Nakatsu CH.** 2003. Subtyping of foodborne and environmental isolates of *Escherichia coli* by multiplex-PCR, rep-PCR, PFGE, ribotyping and AFLP. J Microbiol Methods **53:**387–389. http://dx.doi.org/10.1016/S0167-7012(02)00259-2.
14. **Tenover FC, Arbeit R, Archer G, Biddle J, Byrne S, Goering R, Hancock G, Hébert A, Hill B, Hollis R, Jarvis WR, Kreiswirth B, Eisner W, Maslow J, McDougal LK, Miller M, Mullligan M, Pfaller MA.** 1994. Comparison of traditional and molecular methods of typing isolates of *Staphylococcus aureus*. J Clin Microbiol **32:**407–415.
15. **Kluytmans JAJW, Overdevest ITMA, Willemsen I, Kluytmans-van den Bergh MFQ, van der Zwaluw K, Heck M, Rijnsburger M, Vandenbroucke-Grauls CMJE, Savelkoul PHM, Johnston BD, Gordon D, Johnson JR.** 2013. Extended-spectrum beta-lactamase-producing *Escherichia coli* from retail chicken meat and humans: comparison of strains, plasmids, resistance genes, and virulence factors. Clin Infect Dis **56:**478–487. http://dx.doi.org/10.1093/cid/cis929.
16. **Kaas RS, Friis C, Ussery DW, Aarestrup FM.** 2012. Estimating variation within genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. BMC Genomics **13:**577. http://dx.doi.org/10.1186/1471-2164-13-577.
17. **Sabat AJ, Budimir A, Nashev D, Sá-Leão R, van Dijl JM, Laurent F, Grundmann H, Friedrich AW on behalf of the ESCMID Study Group of Epidemiological Markers (ESGEM).** 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill **18:**20380.
18. **Urwin R, Maiden MCJ.** 2003. Multi-locus sequence typing: a tool for global epidemiology. Trends Microbiol **11:**479–487. http://dx.doi.org/10.1016/j.tim.2003.08.006.
19. **Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, Ip CLC, Wilson DJ, Didelot X, O'Connor L, Lay R, Buck D, Kearns AM, Shaw A, Paul J, Wilcox MH, Donnelly PJ, Peto TEA, Walker AS, Crook DW.** 2012. A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. BMJ Open **2:**e001124. http://dx.doi.org/10.1136/bmjopen-2012-001124.
20. **Maiden MCJ, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND.** 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Microbiol **11:**728–736. http://dx.doi.org/10.1038/nrmicro3093.
21. **Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Bielaszewska Prager R, Brzoska PM, Moore RL, Guenther S, Rothberg JM, Karch H.** 2011. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. PLoS One **6:**e22751. http://dx.doi.org/10.1371/journal.pone.0022751.
22. **Ruppitsch W, Pietzka A, Prior K, Bletz S, Lasa Fernandez H, Allerberger F, Harmsen D, Mellman A.** 2015. Defining and evaluating a core genome multilocus sequence typing scheme for whole-genome sequence-based typing of *L. monocytogenes*. J Clin Microbiol **53:**2869–2876. http://dx.doi.org/10.1128/JCM.01193-15.
23. **de Been M, Pinholt M, Top J, Bletz S, Mellmann A, van Schaik W, Brouwer E, Rogers M, Kraat Y, Bonten M, Corander J, Westh H, Harmsen D, Willems RJ.** 2015. Core genome multilocus sequence typing scheme for high-resolution typing of *Enterococcus faecium*. J Clin Microbiol **53:**3788–3797. http://dx.doi.org/10.1128/JCM.01946-15.
24. **Ferdous M, Friedrich AW, Grundmann H, de Boer RF, Croughs PD, Islam MA, Kluytmans van den Bergh MFQ, Kooistra-Smid AMD, Rossen JWA.** 2016. Molecular characterization and phylogeny of Shiga toxin-producing *Escherichia coli* isolates obtained from two Dutch regions using whole genome sequencing. Clin Microbiol Infect **22:**642.e1-9. http://dx.doi.org/10.1016/j.cmi.2016.03.028.
25. **Kingry LC, Rowe LA, Respicio-Kingry LB, Beard CB, Schriefer ME, Petersen JM.** 2016. Whole genome multilocus sequence typing as an epidemiologic tool for *Yersinia pestis*. Diagn Microbiol Infect Dis **84:**275–280. http://dx.doi.org/10.1016/j.diagmicrobio.2015.12.003.
26. **Kohl TA, Diel R, Harmsen D, Rothgänger J, Meywald Walter K, Merker M, Weniger T, Niemann S.** 2014. Whole-genome-based *Mycobacterium tuberculosis* surveillance: a standardized, portable, and expandable approach. J Clin Microbiol **52:**2479–2486. http://dx.doi.org/10.1128/JCM.00567-14.
27. **Leopold SR, Goering RV, Witten A, Harmsen D, Mellmann A.** 2014. Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. J Clin Microbiol **52:**2365–2370. http://dx.doi.org/10.1128/JCM.00262-14.
28. **Nederlands Trial Register.** 2016. Single- or multiple-occupancy room isolation of patients colonised with ESBL-producing enterobacteriaceae. Trial ID NTR2799. Nederlands Trial Register, Amsterdam, the Netherlands. http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=2799. Accessed 31 July 2016.
29. **Netherlands Society for Medical Microbiology.** 2012. NVMM guideline laboratory detection of highly resistant microorganisms, version 2.0. http://www.nvmm.nl/richtlijnen/hrmo-laboratory-detection-highly-resistant-microorganisms.
30. **European Committee on Antimicrobial Susceptibility Testing.** 2013. EUCAST guidelines for detection of resistance mechanisms and specific resistances of clinical and/or epidemiological importance, version

1.0. http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Resistance_mechanisms/EUCAST_detection_of_resistance_mechanisms_v1.0_20131211.pdf.
31. **Kluytmans-van den Bergh MFQ, Verhulst C, Willemsen LE, Verkade E, Bonten MJM, Kluytmans JAJW.** 2015. Rectal carriage of extended-spectrum beta-lactamase-producing *Enterobacteriaceae* in hospitalized patients: selective preenrichment increases yield of screening. J Clin Microbiol **53:**2709–2712. http://dx.doi.org/10.1128/JCM.01251-15.
32. **Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lunda O.** 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol **50:**1355–1361. http://dx.doi.org/10.1128/JCM.06094-11.
33. **Larsen MV, Cosentino S, Lukjancenko O, Saputra D, Rasmussen S, Hasman H, Sicheritz-Pontén T, Aarestrup FM, Ussery DW, Lund O.** 2014. Benchmarking of methods for genomic taxonomy. J Clin Microbiol **52:**1529–1539. http://dx.doi.org/10.1128/JCM.02981-13.
34. **Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV.** 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother **67:**2640–2644. http://dx.doi.org/10.1093/jac/dks261.
35. **Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MCJ, Ochman H, Achtman M.** 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol **60:**1136–1151. http://dx.doi.org/10.1111/j.1365-2958.2006.05172.x.
36. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
37. **van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, Fussing V, Green J, Feil E, Gerner-Smidt P, Brisse S, Struelens M for the European Society of Clinical Microbiology and Infectious Diseases (ESCMID) Study Group on Epidemiological Markers (ESGEM).** 2007. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin Microbiol Infect **13**(Suppl 3)**:**S1–S46. http://dx.doi.org/10.1111/j.1469-0691.2007.01732.x.
38. **Mathers AJ, Stoesser N, Sheppard AE, Pankhurst L, Gless A, Yeh AJ, Didelot X, Turner SD, Sebra R, Kasarskis A, Peto T, Crook D, Sifri CD.** 2015. *Klebsiella pneumoniae* carbapenemase (KPC)-producing *K. pneumoniae* at a single institution: insights into endemicity from whole-genome sequencing. Antimicrob Agents Chemother **59:**1656–1663. http://dx.doi.org/10.1128/AAC.04292-14.
39. **Reeves PR, Liu B, Zhou Z, Li D, Guo D, Ren Y, Clabots C, Lan R, Johnson JR, Wang L.** 2011. Rates of mutation and host transmission for an *Escherichia coli* clone over 3 years. PLoS One **6:**e26907. http://dx.doi.org/10.1371/journal.pone.0026907.
40. **Long H, Miller SF, Strauss C, Zhao C, Cheng L, Ye Z, Griffin K, Te R, Lee H, Chen CC, Lynch M.** 2016. Antibiotic treatment enhances the genome wide mutation rate of target cells. Proc Natl Acad Sci U S A **113:**E2498–E2505. http://dx.doi.org/10.1073/pnas.1601208113.
41. **Feil EJ, Spratt BG.** 2001. Recombination and the population structures of bacterial pathogens. Annu Rev Microbiol **55:**561–590. http://dx.doi.org/10.1146/annurev.micro.55.1.561.
42. **Guttman DS, Dykhuizen DE.** 1994. Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. Science **266:**1380–1383. http://dx.doi.org/10.1126/science.7973728.
43. **Hao W, Allen VG, Jamieson FB, Low DE, Alexander DC.** 2012. Phylogenetic incongruence in *E. coli* O104: understanding the evolutionary relationships of emerging pathogens in the face of homologous recombination. PLoS One **7:**e33971. http://dx.doi.org/10.1371/journal.pone.0033971.
44. **Spratt BG, Hanage WP, Feil EJ.** 2001. The relative contributions of recombination and point mutation to the diversification of bacterial clones. Curr Opin Microbiol **4:**602–606. http://dx.doi.org/10.1016/S1369-5274(00)00257-5.
45. **Touchon M, Hoede C, Tenaillon E, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O, Calteau A, Chiapello H, Clermont O, Cruveiller S, Danchin A, Diard M, Dossat C, El Karoui M, Frapy E, Garry L, Ghigo JM, Gilles AM, Johnson J, Le Bouguénec Lescat M, Mangenot S, Martinez-Jéhanne V, Matic I, Nassif X, Oztas S, Petit MA, Pichon C, Rouy Z, Saint Ruf C, Schneider D, Tourret J, Vacherie B, Vallenet D, Médigue C, Rocha EPC, Denamur E.** 2009. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. PLoS Genet **5:**e1000344. http://dx.doi.org/10.1371/journal.pgen.1000344.
46. **Aarestrup FM, Brown EW, Detter C, Gerner-Smidt P, Gilmour MW, Harmsen D, Hendriksen RS, Hewson R, Heyman DL, Johansson K, Ijaz K, Keim PS, Koopmans M, Kroneman A, Lo Fo Wong D, Lund O, Palm D, Sawanpanyalert P, Sobel J, Schlundt J.** 2012. Integrating genome-based informatics to modernize global disease monitoring information sharing, and response. Emerg Infect Dis **18:**e1. http://dx.doi.org/10.3201/eid1811.120453.