O558
1-hour Oral Session
Harnessing whole-genome sequencing for diagnostics and typing

Whole-genome multilocus sequence typing of ESBL-producing Escherichia coli: discriminatory power and optimal cut-off

Marjolein Kluytmans - van den Bergh*[1], John W.A. Rossen[2], Rob Willems[3], Alexander W. Friedrich[2], Christina Vandenbroucke-Grauls[4], Marc J. Bonten[5], Jan A.J.W. Kluytmans[6], On Behalf of the Som Study Group ,[7]

[1]*Julius Center for Health Sciences and Primary Care, Umc Utrecht, Utrecht, Breda, Netherlands*

[2]*Department of Medical Microbiology, University of Groningen, University Medical Center Groningen, Groningen, Netherlands*

[3]*Umc Utrecht, Utrecht, Netherlands*

[4]*Dept Medical Microbiology & Infection Control, Vu University Medical Center, Amsterdam, Netherlands*

[5]*University Medical Center Utrecht, Medical Microbiology, Utrecht, Netherlands*

[6]*Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Breda, Netherlands*

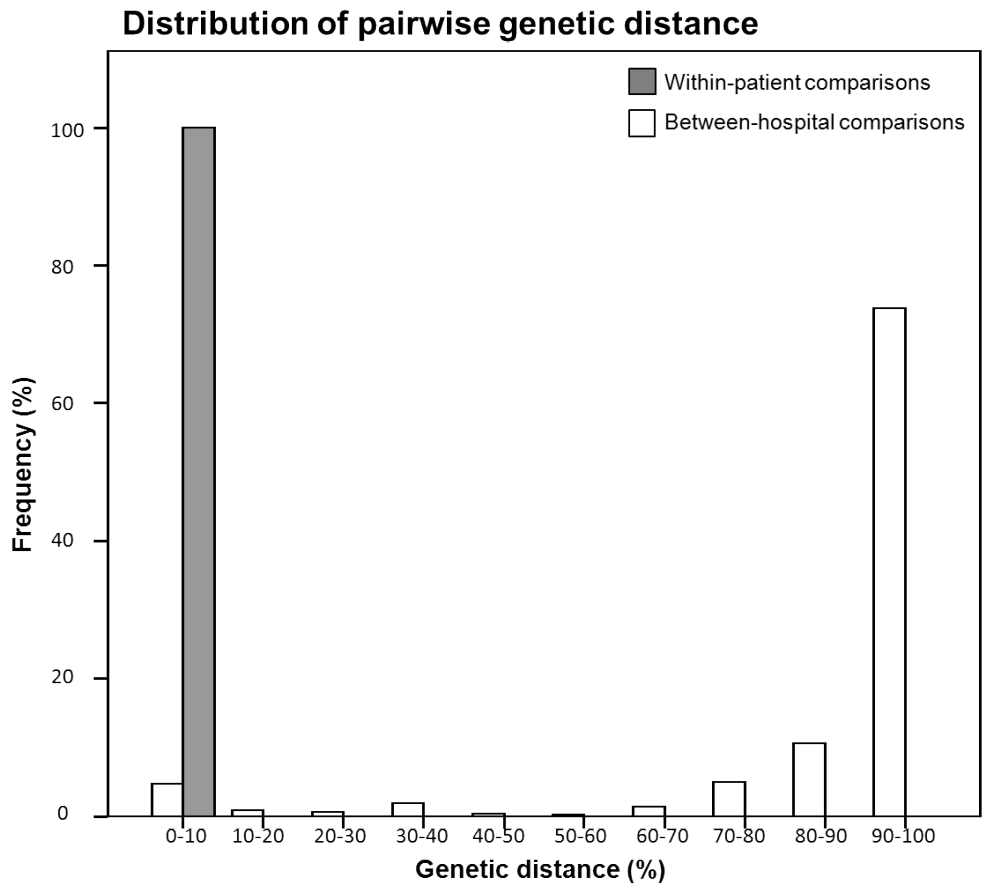
[7]*,, ,, Netherlands*

Background: Whole genome multilocus sequence typing (wgMLST) is a gene-by-gene typing method that extends MLST to the genome level. The discriminatory power of wgMLST is expected to be high, but studies on proposed cut-off values for wgMLST-based genetic distance that can be applied to identify transmission are limited. The aim of this study was to evaluate the discriminatory power of wgMLST typing for ESBL-producing *Escherichia coli* (ESBL-Ec) and to define a cut-off value for wgMLST-based genetic distance to distinguish between related and unrelated ESBL-Ec isolates.

Material/methods: A national collection of ESBL-Ec isolates was prospectively gathered in the SoM study, a multi-centre cluster-randomised study comparing different isolation strategies for known ESBL-E carriers. ESBL-Ec isolates were obtained from routine clinical cultures and perianal screening cultures. Phenotypic confirmation of ESBL production was performed according to the EUCAST guideline. Whole genome shotgun sequencing (WGS) was performed for all phenotypic confirmed ESBL-E isolates on either a MiSeq or a HiSeq 2500 sequencer (Illumina). *De novo* assembly was performed using CLC genomics Workbench 7.0.4 (Qiagen). The MLST+ target definer function of SeqSphere+ (Ridom) was used to define an *ad hoc* wgMLST scheme, comprising 2,764 core genome targets and 1,785 accessory genome targets. This typing scheme was used to calculate pairwise genetic distances. Pairwise comparisons were classified according to their known epidemiologic link, i.e. within-patient comparisons were classified as ‘related’, between-hospital comparisons as ‘unrelated’. All other pairwise comparisons were disregarded, in order to prevent bias by nosocomial transmission of ESBL-Ec isolates. The discriminatory power of wgMLST-based genetic distance to identify related and unrelated isolates was assessed by estimating the area under the curve (AUC) in

ROC curve analysis. The Youden index (YI), was used to estimate the optimal cut-off value for genetic distance.

Results: A total of 1,477 ESBL-Ec isolates were collected, comprising 1,090,026 pairwise comparisons. Of these, 438 were within-patient comparisons and 966,071 were between-hospital comparisons. The time interval between cultures ranged from 0 to 460 days (median 10 days) for within-patient comparisons, and from 0 to 1,044 days (median 278 days) for between-hospital comparisons. Genetic distance varied from 0.0% to 1.7% for within-patient comparisons, and from 0.0% to 99.9% for between-hospital comparisons (Figure). The ROC AUC was 1.000 [95% CI 1.000 – 1.000], indicating a high discriminatory power to identify related and unrelated isolates. The YI-based optimal cut-off value for genetic distance was 1.0% (sensitivity 99.5%; specificity 99.7%; YI 0.992), whereas a sensitivity of 100% was reached at a cut-off value of 1.7% (specificity 99.1%; YI 0.991).



Conclusions: The wgMLST scheme described in this study has a high discriminatory power to distinguish between related and unrelated ESBL-Ec isolates. The optimal cut-off value for genetic distance to discriminate between related and unrelated ESBL-Ec isolates is 1.0%.