Baktash A[1], Hornung BVH[1], Corver J[1], Harmanus C[1,2], Smits WK[1], Fawley W[3], Wilcox MH[4], Kumar N[5], Eyre DW[6], Indra A[7], Mellmann A[8], Kuijper EJ[1,2]

[1] Department of Medical Microbiology, Section Experimental Bacteriology, Leiden University Medical Center, Leiden, the Netherlands; [2] National Reference Laboratory Clostridium difficile, National Institute of Public Health and the Environment, Leiden University Medical Center, Leiden, the Netherlands; [3] National Infection Service, Public Health England, and University of Leeds, Leeds, United Kingdom; [4] Department of Microbiology, Leeds Teaching Hospitals & University of Leeds, Leeds, United Kingdom; [5] Microbiota Interactions Laboratory, Wellcome Sanger Institute, Hinxton, UK; [6] Big Data Institute, University of Oxford, Oxford, United Kingdom; [7] Institute of Medical Microbiology and Hygiene, Österreichische Agentur für Gesundheit und Ernährungssicherheit (AGES), Vienna, Austria; [8] Institute of Hygiene, University Hospital Münster, and National Reference Center for C. difficile, Münster Branch, Münster, Germany

Leiden University Medical Center

# PERFORMANCE OF CORE GENOME MULTILOCUS SEQUENCE TYPING COMPARED TO CAPILLARY-ELECTROPHORESIS PCR RIBOTYPING AND SNP ANALYSIS OF *Clostridioides difficile*

## Background

*Clostridioides difficile* is the most common cause of antibiotic-associated gastrointestinal infections. Capillary-electrophoresis (CE) PCR ribotyping is currently the gold standard for *C. difficile* typing, but lacks sufficient discriminatory power to fully resolve outbreaks.

## Aim of the study

This study compares the performance of core genome (cg) and whole genome (wg) multilocus sequence typing (MLST) with CE-PCR ribotyping and single nucleotide polymorphism (SNP) analysis

## Methods

**Ribotypes**
- 100 unique Ribotypes (RTs)

**CDI Outbreaks**
- RT078 outbreak in a Dutch general hospital (3 cases & 3 non-related strains)
- RT181 outbreak in a Greek rehabilitation clinic (15 cases)

**Sequence data**
- 541 *C. difficile* sequences (NCBI Sequence Read Archive)
- 74 sequenced RTs (Leeds-Leiden reference collection)
- 21 sequenced outbreak strains (RT078 & RT181)

**Molecular Typing & analysis**

| | |
|---|---|
| NGS | Illumina Novaseq 6000 (Genome Scan B.V.) |
| Typing | cgMLST (Ridom®SeqSphere v5.1 & Enterobase)<br>wgMLST (Enterobase) |
| SNP analysis | CSI Phylogeny 1.4 |

## Results

**Backward compatibility & performance of cgMLST**
- 82 RTs of 100 RTs had a unique profile, 18 clustered (≤ 6 targets/alleles) with multiple RTs (ranging 1-3 RTs) (Fig 1 & Table 1)
- Average intra-RT allele difference varied between RTs; RT056 has highest average allele difference and RT181 has the lowest (Fig. 2)

**Outbreak setting**
- The minimal allele difference between outbreak strains and non-outbreak strains varies per typing method
- The minimal allele difference between outbreak strains and non-outbreak in a RT078 outbreak ranged from 2-8, with the lowest for cgMLST (Seqsphere & Enterobase) and highest for SNP analysis (Fig 3a, Table 2)
- The outbreak and non-outbreak strains showed no minimal allele difference in the RT181 outbreak setting in all typing methods except for wgMLST with 2 allele differences between both groups (Fig 3b, Table 2)

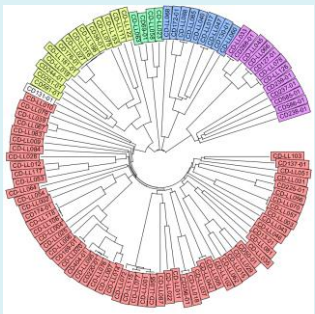

Figure 1: Neighbor joining tree based on cgMLST allele difference from 100 RTs. RTs from MLST Clade 1, 2, 3, 4 ,5 are colored red, yellow, green, blue and purple, respectively.

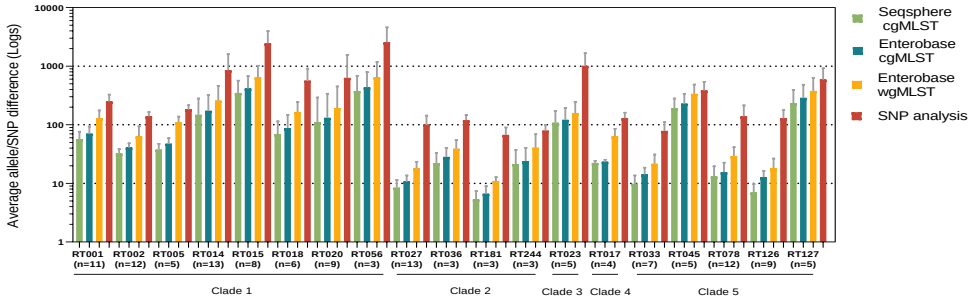


Figure 2: Average allele/SNP difference shown for RTs from MLST Clade 1 (RT001-RT056), Clade 2 (RT027-RT244), Clade 3 (RT023), Clade 4 (RT017) and Clade 5 (RT033-RT127).

**Table 1: Clustering between ribotypes at different thresholds**

| Threshold (in alleles) | RT | amount of strains | RT | amount of strains | Clade | Threshold (in alleles) | RT | amount of strains | RT | amount of strains | Clade |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | 020 | 1/20 | 076 | 1/2 | 1 | 4 | 016 | 1/1 | 027 | 1/23 | |
| | 016 | 1/1 | 078 | 5/23 | 2 | | 027 | 6/23 | 036 | 1/4 | 2 |
| | | | 036 | 1/4 | | | | 9/23 | 176 | 5/16 | |
| | | | 176 | 4/16 | | | 033 | 2/46 | 288 | 2/2 | |
| | 027 | 3/23 | 036 | 2/4 | 2 | | 045 | 2/15 | 078 | 5/59 | |
| | | 10/23 | 176 | 13/16 | | | | 2/15 | 126 | 4/29 | |
| | | 2/23 | 198 | 1/2 | | | | 3/15 | 127 | 3/17 | |
| | 036 | 1/4 | 176 | 1/16 | 2 | | 066 | 1/3 | 078 | 1/59 | |
| | 033 | 2/46 | 288 | 2/2 | 5 | | 078 | 25/59 | 126 | 15/29 | |
| | 045 | 2/15 | 078 | 16/59 | 5 | 3 | 018 | 1/18 | 356 | 3/13 | 1 |
| | | 2/15 | 126 | 7/29 | | | 027 | 3/23 | 036 | 1/4 | 2 |
| | 066 | 1/3 | 078 | 3/59 | 5 | | | 6/23 | 176 | 3/16 | 2 |
| | | 1/3 | 126 | 1/29 | | | 045 | 1/15 | 078 | 1/59 | |
| | 078 | 40/59 | 126 | 24/29 | 5 | | | 1/15 | 126 | 1/29 | |
| 5 | 018 | 1/18 | 356 | 1/13 | 1 | | | 3/15 | 127 | 3/17 | |
| | 016 | 1/1 | 027 | 2/23 | 2 | | 078 | 17/59 | 126 | 12/29 | |
| | | | 176 | 1/16 | | 2 | 001 | 1/14 | 055 | 1/1 | 1 |
| | | | 198 | 1/2 | | | 018 | 1/18 | 356 | 3/13 | 1 |
| | 027 | 4/23 | 036 | 1/4 | 2 | | 016 | 1/1 | 027 | 1/23 | |
| | | 10/23 | 176 | 6/16 | | | 027 | 3/23 | 176 | 2/16 | |
| | | 2/23 | 198 | 1/2 | | | 045 | 2/15 | 126 | 2/29 | |
| | 036 | 1/4 | 176 | 2/16 | 2 | | | 1/15 | 127 | 2/17 | |
| | 033 | 1/46 | 288 | 1/2 | 5 | | 078 | 9/59 | 126 | 6/29 | |
| | 045 | 2/15 | 078 | 7/59 | 5 | 1 | 018 | 1/18 | 356 | 6/13 | 1 |
| | | 2/15 | 126 | 4/29 | | | 045 | 1/15 | 127 | 1/17 | |
| | | 3/15 | 127 | 2/17 | | 0 | 045 | 2/15 | 127 | 2/17 | |
| | 066 | 1/3 | 078 | 8/59 | 5 | | 066 | 2/3 | 078 | 1/59 | |
| | | 1/3 | 126 | 2/29 | | | | | | | |
| | 078 | 31/59 | 126 | 21/29 | 5 | | | | | | |

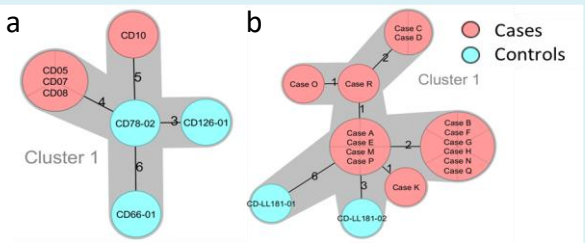


Figure 3: Minimum spanning tree of CDI due to RT078 and RT181 with added controls. (a) Cluster 1 from RT078 outbreak with cases (CD05-CD08), non-related strain (CD10) and added controls (CD66-01;RT066, CD78-02; RT078 and CD126-01;RT126). (b) Cluster 1 from RT181 outbreak with cases (CD181-01 to CD181-15) and added controls (CD-LL181-01 and CD-LL181-02; RT181).

**Table 2: comparison in range between outbreak and non-outbreak strains of RT078 and RT181**

| Typing method | Strains | Range | Minimal allele difference |
|---|---|---|---|
| Seqsphere cgMLST | 078 outbreak | 0 | 3 |
| | non-outbreak | 3-9 | |
| | 181 outbreak | 0-5 | overlap |
| | non-outbreak | 3-9 | |
| Enterobase cgMLST | 078 outbreak | 2-4 | 2 |
| | non-outbreak | 6-14 | |
| | 181 outbreak | 0-8 | overlap |
| | non-outbreak | 4-12 | |
| Enterobase wgMLST | 078 outbreak | 1 | 6 |
| | non-outbreak | 7-41 | |
| | 181 outbreak | 0-8 | 2 |
| | non-outbreak | 10-15 | |
| SNP analysis | 078 outbreak | 0 | 8 |
| | non-outbreak | 8-14 | |
| | 181 outbreak | 0-9 | overlap |
| | non-outbreak | 7-23 | |

## Conclusions

- cgMLST has the potential to be an alternative to CE-PCR ribotyping; it is reproducible, backward compatible (to certain extent), easy to standardiz and offers a higher discriminatory power.
- In outbreak settings of RT 078 and 181, the performance with the highest discriminatory power was obtained with Enterobase wgMLST and SNP analysis.
- Additional epidemiologic information is necessary to fully resolve outbreaks of monomorphic strains.

Correspondence: A. Baktash
(a.baktash.mm@lumc.nl)