Supplementary information

Defining and Evaluating a Core Genome Multilocus Sequence Typing Scheme for Genome-Wide Typing of *Clostridium difficile*

Stefan Bletz,[a,b] Sandra Janezic,[c,d] Dag Harmsen,[e] Maja Rupnik,[c,d,] Alexander Mellmann[a,b#]

Table S1. List of BAPS grouping results per MLST sequence type (generated on March 31, 2016; based on all available sequence types). If known, the clade assignment is also given. NA, not assigned.

Table S2. List of core genome genes used for the *C. difficile* cgMLST scheme. Locus tags designations were taken from the reference strain 630 (GenBank acc. no. NC_009089).

Table S3. Metadata and allelic profiles of all *C. difficile* reference strains and isolates used for cgMLST scheme definition and evaluation. NA, not assigned. For alleles, not assigned is given either due to gene absence or to failed quality criteria.

Table S4. List of *Clostridium difficile* genomes downloaded from NCBI (available data until Oct 21, 2015) used for evaluation of cgMLST scheme. NA, not assigned alleles either due to gene absence or to failed quality criteria.

Figure S1. Minimum-spanning tree of all isolates downloaded from NCBI used for evaluation of the cgMLST scheme and comparison with classical MLST. All isolates that comprised a valid MLST ST were included (n=2,929 isolates of the in total 2,954 isolates [see Table S4]).

Each node represents a unique allelic cgMLST profile and the size of the nodes represents the number of isolates. The 20 most common MLST STs were given as colors of the circles, which are additionally named with the ST. The number on connecting lines displays the number of differing alleles between the genotypes in a pairwise comparison. Grey shades display isolates that are highly similar, i. e. differing in ≤ 6 cgMLST target genes. The analysis parameter “pairwise ignore missing data” was used during analysis of the up to 2,270 cgMLST target genes.