Clostridium difficile: investigating transmission patterns between infected and colonized patients using whole genome sequencing

L. Y. Kong MD[1], D. W. Eyre DPhil[2,3], J. Corbeil MD[4], F. Raymond PhD[4], A. S. Walker PhD[3], M.H. Wilcox MD[5], D. W. Crook MD[2, 6], S. Michaud MD[7], B. Toye MD[8], E. Frost PhD[7], N. Dendukuri PhD[9], I. Schiller MSc[10], A.M. Bourgault MD[1,11], A. Dascal MD[12], M. Oughton MD[12], Y. Longtin MD[12], L. Poirier MD[13], P. Brassard MD[10], N. Turgeon MD[14], R. Gilca MD[15], V. G. Loo MD[1]

[1]Division of Infectious Diseases and Department of Medical Microbiology, McGill University Health Centre, Montréal, Canada

[2]Nuffield Department of Medicine, John Radcliffe Hospital, Oxford, UK

[3]NIHR Oxford Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK

[4]Centre de recherche CHUQ, Université Laval, Québec, Canada

[5]Department of Microbiology, Leeds Teaching Hospitals and University of Leeds, Leeds, UK

[6]National Infection Service, Public Health England, London, UK

[7]Department of Microbiology and Infectiology, Université de Sherbrooke, Sherbrooke, Canada

[8]Division of Microbiology, Ottawa Hospital, University of Ottawa, Ottawa, Canada

[9]Technology Assessment Unit, McGill University Health Centre, Montréal, Canada

[10]Centre for Outcomes Research, McGill University Health Centre – Research Institute, Montréal, Canada

[11]Department of Microbiology, Centre Hospitalier de l'Université de Montréal, Montréal, Canada

[12]Division of Infectious Diseases, Jewish General Hospital, Montréal, Canada

[13]Department of Microbiology, Hôpital Maisonneuve-Rosemont, Montréal, Canada

[14]Department of Microbiology, Centre Hospitalier Universitaire de Québec – Hôtel-Dieu de Québec, Québec, Canada

[15] Quebec Institute of Public Health, Québec, Canada

Key words: *Clostridium difficile*; whole genome sequencing; colonization

Running title: *C. difficile* colonization & transmission

Corresponding author:

Ling Yuan Kong, MD

Division of Infectious Diseases and Department of Medical Microbiology

McGill University Health Centre

1001 Decarie Blvd

Montreal, QC H4A 3J1

Email : ling.kong@mail.mcgill.ca

Telephone : (514) 835-7093

Fax : (514) 412-4354

Alternate Corresponding Author:

Vivian Loo, MD

Division of Infectious Diseases and Department of Medical Microbiology

McGill University Health Centre

1001 Decarie Blvd

Montreal, QC H4A 3J1

Email: vivian.loo@mcgill.ca

Telephone: (514) 934-1934 ext. 42818

Fax : (514) 412-4354

Word count: Text 2855 Abstract 250

Figures: 1

Tables: 2

Summary of main point

Using whole genome sequencing of isolates from a cohort of patients with *Clostridium difficile* infection (CDI) and colonization, we found that incident CDI cases were more likely to be linked to an infected than colonized donor.

Abstract

Background

Whole genome sequencing (WGS) studies can enhance our understanding of the role of patients with asymptomatic *Clostridium difficile* colonization in transmission.

Methods

Isolates obtained from patients with *Clostridium difficile* infection (CDI) and colonization identified in a study conducted during 2006 - 2007 at six Canadian hospitals underwent typing by pulsed-field gel electrophoresis, multilocus sequence typing, and WGS. Isolates from incident CDI cases not in the initial study were also sequenced where possible. Ward movement and typing data were combined to identify plausible donors for each CDI case, as defined by shared time and space within predefined limits. Proportions of plausible donors for CDI cases that were colonized, infected, or both were examined.

Results

Five hundred and fifty-four isolates were sequenced successfully, 353 from colonized and 201 from CDI cases. The NAP1/027/ST1 strain was the most common strain, found in 124 (62%) of infected and 92 (26%) of colonized patients. A donor with a plausible ward link was found for 81 CDI cases (40%) using WGS with a threshold of ≤2 single nucleotide variants to determine relatedness. Sixty-five (32%) CDI cases could be linked to both infected and colonized donors. Exclusive linkages to infected and colonized donors were found for 28 (14%) and 12 (6%) CDI cases, respectively.

Conclusion

Colonized patients contribute to transmission, but CDI cases are more likely linked to other infected patients than colonized patients in this cohort with high rates of NAP1/027/ST1 strain, highlighting the importance of local prevalence of virulent strains in determining transmission dynamics.

Background

Clostridium difficile is a leading cause of healthcare-associated diarrhea and a major cause of morbidity and mortality for hospitalized patients[1]. Patients with symptomatic infection and asymptomatic colonization are both known to shed spores into the environment[2]. Currently recommended infection control measures focus on the detection and isolation of symptomatic patients, believed to be responsible for most healthcare-associated transmission events[3]. However, recent molecular studies using whole genome sequencing (WGS) have found that most new cases of *C. difficile* infection (CDI) in endemic settings could not be explained by transmission from symptomatic cases[4], raising interest in the role of colonized patients in transmission of *C. difficile*.

Typing methods used to identify transmission leading to CDI include pulsed-field gel electrophoresis (PFGE), PCR ribotyping, and multilocus sequence typing (MLST), among others[2]. With the advent of high-throughput sequencing technologies, WGS is increasingly being adopted as a preferred typing/fingerprinting method with high discriminatory power, and so has been used in multiple molecular epidemiology studies on *C. difficile* transmission[4-7]. In this study, using WGS of isolates and epidemiological data from a prospective cohort study, we aimed to elucidate the role of patients colonized with *C. difficile* in onward transmission of infection.

Methods

Study population and definitions

A multicenter prospective study was conducted between March 6, 2006 and June 25, 2007 to determine host and pathogen factors for health care-associated *C. difficile* infection and colonization, with results previously published[8]. Briefly, data were collected in six Canadian, university-affiliated hospitals, on 15 study units (seven surgical units and eight medical units). The selected units were those with a historically high or low incidence of CDI. All patients 18 years or older admitted to these hospital units were eligible for participation. Exclusion criteria included hemodynamic instability, palliative status, neutropenia (absolute neutrophil count ≤1000 per cubic millimeter), or inability to participate in the informed-consent process.

Patients were followed daily until ward discharge, death, or withdrawal from the study. Rectal swabs or stool samples were obtained for culture on admission, weekly during hospitalization, and at onset of diarrhea (if applicable). Toxigenic *C. difficile* culture was performed on stool samples or rectal swabs using standard methods[9]. The cell cytotoxin neutralization assay was the diagnostic assay used in routine clinical care during the study period. Isolates were tested for presence of *tcd*A and *tcd*B using nucleic acid amplification methods[10, 11].

CDI was defined as the presence of diarrhea without an alternative explanation and a positive *C. difficile* cytotoxin assay or toxigenic culture, an endoscopic diagnosis of pseudomembranes, or a pathological diagnosis of CDI. Diarrhea was defined as at least

three loose stools within at least one 24-hour period. Asymptomatic *C. difficile* colonization was defined as a positive stool *C. difficile* culture in the absence of diarrhea. Non-toxigenic strains of *C. difficile* were defined as culture positive and *tcdB* negative.

In order to capture a more comprehensive picture of transmission, we also reviewed infection control data to determine the incidence of CDI cases in non-participants occurring on the study units during the study period. For one of the six participating hospitals, isolates were conserved for the purpose of infection control surveillance and were available for non-study incident CDI cases on study units; all incident CDI cases participated in the study for one other hospital. These isolates were included in the current analysis. Hospital and study unit admission and discharge dates were collected for every participant admitted to study units.

PFGE

Each isolate underwent PFGE using standard methods[12] at the time of the study. Strain relatedness was determined using the criteria of Tenover *et al* using BioNumerics (Applied Maths)[13]. The Dice coefficient was used to measure similarity between patterns.

DNA preparation, sequencing, mapping and single nucleotide polymorphism (SNP) detection

DNA was extracted using Purelink viral RNA/DNA minikit (Invitrogen, Burlington, ON, Canada) on a sub-cultured colony from frozen isolates. DNA was quantified using the

QuantiFluor dye (Promega). Sequencing libraries were prepared using the Nextera XT Sample Preparation Kit (Illumina, San Diego, CA, USA) with 1 ng of purified DNA per sample. Dual indices were added during library preparation. Library concentrations were normalized using bead normalization as described by the manufacturer. Ninety-six libraries were pooled per HiSeq lane. Sequencing was performed on the HiSeq 2500 sequencer (Illumina) using v3 chemistry, generating paired-end 101 bp reads. Reads and assemblies have been deposited in the European Nucleotide Archive database in project PRJEB11776.

Sequence reads were analyzed and assembled using a previously described pipeline developed specifically for bacterial genomes[4]. The set of reads from each isolate was mapped using Stampy v. 1.0.11 (without Burrows-Wheeler Aligner pre-mapping, using an expected substitution rate of 0.01)[14] to the *C. difficile* 630 reference genome (Genbank: AM180355.1)[15]. Base-pair calls were identified across all mapped non-repetitive core genome sites using SAMtools (version 0.1.19) mpileup with the extended base-alignment quality flag, using parameters based on bacterial sequences[4]. A consensus of ≥75% was required to support a nucleotide call, and calls were required to be homozygous under a diploid model. Only calls supported by ≥5 reads, including one in each direction were accepted.

Sequences were compared using single nucleotide polymorphisms (SNPs), obtaining differences between sequences from maximum likelihood phylogenies constructed using PhyML[16] with generalized time-reversible substitution model and “BEST” tree

topology search algorithm, corrected for the effect of recombination using ClonalFrameML[17] (with default settings). Sequence reads were also assembled *de novo* with Velvet[18] and MLSTs and toxigenic strains identified using BLAST searches of *de novo* assemblies (≥1000 nucleotide identities with *tcdA* or *tcdB* genes).

Transmission analysis

Isolates' PFGE, MLST and toxigenic status were first examined according to colonized or infected status. Ward movement and WGS data were then combined to identify plausible donors for each CDI case. Proportions of plausible donors that were colonized or infected were calculated. A donor was identified for an isolate when they were determined to be clonal (differed by ≤2 SNPs by WGS), and a plausible epidemiological link could be identified between the pair based on a previously described model[19], namely the pair shared a ward after the donor tested positive and before the recipient tested positive, shared a ward before either tested positive, or if the recipient occupied a ward after the donor tested positive and was discharged. Maximum infectious period of 8 weeks, incubation period of 12 weeks and ward contamination period of 26 weeks were allowed[20].

The analyses were first done for all available isolates, then restricted to two hospitals where 80% or more of all incident CDI cases occurring on study units during the study period were sequenced, whether or not part of the prospective study.

Results

Five hundred and thirteen of 568 isolates from the cohort study were available for sequencing. An additional 52 isolates from 77 incident CDI cases from one of the participating hospitals were included for a total of 565 isolates. The participation rate in the initial prospective cohort study was 57.1% of eligible patients admitted to the study units. For one hospital contributing 9.6% of isolates, all incident CDI cases on study units were captured in the study. Figure 1 provides a breakdown of sample sources and patient statuses.

Overall, 554 (98%) samples were sequenced successfully, from 550 patients (4 patients contributed 2 samples). There were 353 samples from colonized patients and 201 from infected patients. Two isolates did not have a PFGE pattern available, and 17 isolates could not be assigned to a known MLST.

The epidemic NAP1/ST1(ribotype 027) strain was the most commonly occurring strain among both infected and colonized patients, found in 124 (62%) and 92 (26%) patients, respectively. However, the majority of colonized patients carried strains from a variety of different sequence types (Figure 2). Strains from 27 different sequence types were found among infected patients, whereas a greater variety with 41 sequence types was found among colonized patients. The majority (74%) of colonized patients carried toxigenic strains.

By comparing all samples from infected patients with prior samples from within the cohort, using a threshold of ≤2 SNPs to determine relatedness, overall 105 (52%) cases

could be linked genetically to a prior sample (Table 1); 65 patients (32%) could be linked to both infected and colonized donors. More cases were found to be related to isolates only from infected patients than isolates only from colonized patients, 28 cases (14%) and 12 cases (6%) respectively. Within all 105 cases related to a previous infected or colonized donor using WGS, a donor with a plausible ward link could be found for 81 patients (77%; 40% of all 201 cases). Nearly all the identified donors were of the epidemic NAP1/ST1 strain. Only 7 patients with genetic and ward links were found to have non-NAP1/ST1 donors, including 3 linked to colonized donors only, 3 linked to infected donors only and one to both infected and colonized donors.

Restricting analyses to the 2 hospitals with most complete data (Table 2), overall similar patterns were observed, including for those cases substantiated with ward links. Thirty out of 117 cases (26%) could be linked to isolates from both infected and colonized patients and 26 (22%) to isolates from only infected patients, whereas only 4 (3%) were linked to samples from only colonized patients. Of 46 cases with a ward link, 30 (26% of all 117 cases) had an exclusive link to an infected donor, and only 2 (2% of all 117 cases) had an exclusive link to a colonized donor.

Discussion

The role of colonized patients in transmission of CDI has been subject of several previous molecular epidemiology studies [7, 19, 21]. Curry et al. used multilocus variable number tandem repeats analysis genotyping and concluded that 29% of 56 incident CDI cases

could be linked to colonized patients [21]. Using WGS, Eyre *et al*. did not find evidence of any onward transmission from 18 asymptomatic colonized patients to CDI cases [19].

Using WGS, we investigated the contribution of colonized and infected patients in onward transmission toward incident CDI cases. In our larger cohort, 52% of cases could be linked to a previous patient. This is higher than previously reported rates [4], in part because our study includes both infected and colonized patients as sources, although higher linkage rates to symptomatic patients, 93/201 (46%) of cases, were also found. This difference may be explained in part by the diagnostic laboratory methods used. In the study by Eyre et al, the laboratory method used was immunoassay whereas in our study, the laboratory method was toxigenic culture which has a higher sensitivity than enzyme immunoassay for detecting *C. difficile*. Therefore, more patients would have been classified as CDI and a higher linkage would be made with CDI patients. However, patients met the case definition for CDI and did not have an alternative explanation for diarrhea. In addition, the high incidence of CDI of 28.1 cases per 10,000 patient-days in our cohort reflected the epidemic setting of the study, with a large pool of symptomatic patients, and a higher infection-to-colonization ratio compared to other cohorts[22]. The high proportion of infected patients is likely explained by the predominance of the NAP1/ST1 strain, which is more virulent and likely to cause infection[8].

Examining data from all units, an incident CDI case was 2.3 times more likely to be linked to an infected patient only than to a colonized patient only, whereas in the subset of hospitals with most complete data, this was 6 times more likely. Within the hospitals

where data were most complete, exclusive linkage to colonized donors was less common; however, in these hospitals the proportion of infected cases sequenced (77-86%) was substantially higher than on the other units (26-27%) due to availability of additional isolates. In both analyses, many cases could be linked to both infected and colonized patients, reflecting the outbreak setting in which the cohort study took place and the relatively slow rate of *C. difficile* evolution relative to the time between transmitted cases, enabling additional potential transmission links to be identified.

Our analyses suggest that colonized patients may be a source of onward transmission to incident CDI cases, but that spread from infected donors is likely more frequent. This could plausibly be explained by lower levels of shedding seen in colonized patients (without diarrhea) as compared with infected patients [23]. Onward transmission events from colonized individuals to infected patients in our cohort frequently carried the epidemic NAP1/ST1 strain, possibly reflecting strain-specific characteristics, such as higher transmissibility [24] (increasing the chance of acquisition) and higher propensity to cause symptomatic infection and thereby increasing detection. For example, NAP1/ST1 may be shed more profusely and persist more effectively in the environment. A study using WGS to track transmission similar to ours, but examining only ribotype-027 (NAP1/ST1) strains within one UK hospital, found that 60% of their genetically-related strains were circulated by ward-based contamination [7]. However, another possibility for the greater degree of linkage is the relatively recent emergence of this fluoroquinolone-resistant NAP1/ST1, resulting in less population-wide genetic diversity, and thus increasing the chance of observing genetic linkage without direct transmission.

The limitations in our study include the incomplete sampling in the participating hospitals. Overall, we only obtained fecal samples from 57% of eligible participants, and did not capture all CDI cases on all study units. Incomplete sampling leads to the proportion of linked cases being under-estimated as some potential transmission donors are missed. Patients who were ineligible in the initial cohort study represent another pool of potential missed linkages, since previously determined eligibility criteria (e.g. neutropenia) for the prospective study do not necessarily translate to a ward-based transmission analysis study. Ideally, studies focused on ward-based transmission would be less restrictive, given the very low risk posed to patients of undergoing rectal swabs. Increased participation could have been achieved by waiving written informed consent and obtaining verbal consent and implementation or ward-based communication tools explaining the option to opt-out.

When limiting the analyses to two hospitals with more than 80% incident cases contributing isolates for sequencing, rates of linkage to infected patients increased, but this could represent sampling bias given more infected donors were available. Finally, although all transmission events were inferred from the genetic data, other sources, such as patients not included in analyses, including ineligible patients, and the environment were not sampled and may be other reservoirs of *C. difficile* leading to CDI.

Our study provides new insight into the epidemiology of transmission between colonized and infected patients, by deriving data from the largest cohort to date of colonized and

infected patients along with geographic ward information. We also confirm the utility of WGS in conjunction with epidemiological data to track transmission, which is increasingly studied including in healthcare epidemiological models.

Conclusion

Patients colonized with *C. difficile* without diarrhea contribute to the transmission of infection, but more transmission events appear to originate from infected patients with diarrhea. Certain strains, such as the epidemic NAP1/ST1 strain, may be more transmissible and virulent, and hence more likely to cause more symptomatic infection following contact with infected and asymptomatically colonized patients. Thus, the relative contribution of colonized and infected patients toward onward transmission is likely dependent on the local prevalence of virulent strains.

Funding

This work was supported by the funding source of the prospective cohort study, the Consortium de Recherche sur le *Clostridium difficile*. The partners of the Consortium include: Fonds de la recherche en santé du Québec; Canadian Institutes of Health Research; Ministère de la Santé et des Services sociaux du Québec; Institut national de santé publique du Québec; Health Canada; Centre hospitalier de l'Université de Montréal; McGill University Health Centre; Centre hospitalier universitaire de Québec; and Centre hospitalier universitaire de Sherbrooke. As the current study uses existing information collected during the study, the funding sources did not have any role in the design of the current study, its analysis, interpretation of data, and writing of the report.

Potential conflicts of interest

L.Y.K. No conflict. D.W.E. No conflict. J.C. No conflict. F.R. No conflict. A.S.W. No conflict. M.H.W. reports grants and personal fees from Actelion, Cubist, Astellas, Merck, Sanofi-Pasteur, Summit, Seres, Biomerieux; and personal fees from Valneva, Alere, Ferring, and Synthetic Biologics. He reports personal fees from Astra Zeneca, Nabriva, Pfizer, Roche, The Medicine Company, Basilea, Bayer, Allergan, Menarini, Motif Biosciences, AiCuris, Antabio, and Spero; and grants and personal fees from Abbott, European Tissue Symposium, Paratek, Tetraphase, and Surface Skins outside of the submitted work. D.W.C. reports grants from Wellcome Trust and MRC (UK) and NIHR (UK). S.M. No conflict. B.T. No conflict. E.F. No conflict. N.D. No conflict. I.S. No conflict. A.M.B. No conflict. A.D. has received grants from Fonds de la recherche en santé du Québec and Canadian Institutes of Health Research. M.O. reports grants and personal fees from Merck, and grants from Sanofi and Summit outside of the submitted work. Y.L. reports grants from Merck and Becton Dickinson outside of the submitted work. L.P. reports grants from Consortium de recherché sur le *Clostridium difficile*. P.B. No conflict. N.T. reports grants from Consortium de recherche sur le *Clostridium difficile.* She reports payment for lectures from FMOQ and Pendopharm outside of the submitted work. R.G. reports grants from Sanofi Pasteur and Pfizer. V.G.L. reports grants from Consortium de recherche sur le *Clostridium difficile*, and consultancy fees and payment for lectures from Merck outside of the submitted work.

References

1. Evans CT, Safdar N. Current Trends in the Epidemiology and Outcomes of Clostridium difficile Infection. Clin Infect Dis **2015**; 60 Suppl 2: S66-71.
2. Martin JS, Monaghan TM, Wilcox MH. Clostridium difficile infection: epidemiology, diagnosis and understanding transmission. Nat Rev Gastroenterol Hepatol **2016**; 13(4): 206-16.
3. Cohen SH, Gerding DN, Johnson S, et al. Clinical practice guidelines for Clostridium difficile infection in adults: 2010 update by the society for healthcare epidemiology of America (SHEA) and the infectious diseases society of America (IDSA). Infect Control Hosp Epidemiol **2010**; 31(5): 431-55.
4. Eyre DW, Cule ML, Wilson DJ, et al. Diverse sources of C. difficile infection identified on whole-genome sequencing. N Engl J Med **2013**; 369(13): 1195-205.
5. Didelot X, Eyre DW, Cule M, et al. Microevolutionary analysis of Clostridium difficile genomes to investigate transmission. Genome Biol **2012**; 13(12): R118.
6. Eyre DW, Golubchik T, Gordon NC, et al. A pilot study of rapid benchtop sequencing of Staphylococcus aureus and Clostridium difficile for outbreak detection and surveillance. BMJ Open **2012**; 2(3).
7. Kumar N, Miyajima F, He M, et al. Genome-Based Infection Tracking Reveals Dynamics of Clostridium difficile Transmission and Disease Recurrence. Clin Infect Dis **2016**; 62(6): 746-52.
8. Loo VG, Bourgault AM, Poirier L, et al. Host and pathogen factors for Clostridium difficile infection and colonization. N Engl J Med **2011**; 365(18): 1693-703.

9. Clabots CR, Gerding SJ, Olson MM, Peterson LR, Gerding DN. Detection of asymptomatic Clostridium difficile carriage by an alcohol shock procedure. J Clin Microbiol **1989**; 27(10): 2386-7.
10. Spigaglia P, Mastrantonio P. Molecular analysis of the pathogenicity locus and polymorphism in the putative negative regulator of toxin production (TcdC) among Clostridium difficile clinical isolates. J Clin Microbiol **2002**; 40(9): 3470-5.
11. Goncalves C, Decre D, Barbut F, Burghoffer B, Petit JC. Prevalence and characterization of a binary toxin (actin-specific ADP-ribosyltransferase) from Clostridium difficile. J Clin Microbiol **2004**; 42(5): 1933-9.
12. Fawley WN, Wilcox MH. Pulsed-field gel electrophoresis can yield DNA fingerprints of degradation-susceptible Clostridium difficile strains. J Clin Microbiol **2002**; 40(9): 3546-7; author reply 7.
13. Tenover FC, Arbeit RD, Goering RV, et al. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J Clin Microbiol **1995**; 33(9): 2233-9.
14. Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res **2011**; 21(6): 936-9.
15. Sebaihia M, Wren BW, Mullany P, et al. The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. Nat Genet **2006**; 38(7): 779-86.
16. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol **2003**; 52(5): 696-704.

17. Didelot X, Wilson DJ. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLoS Comput Biol **2015**; 11(2): e1004041.
18. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **2008**; 18(5): 821-9.
19. Eyre DW, Griffiths D, Vaughan A, et al. Asymptomatic Clostridium difficile colonisation and onward transmission. PLoS One **2013**; 8(11): e78445.
20. Walker AS, Eyre DW, Wyllie DH, et al. Characterisation of Clostridium difficile hospital ward-based transmission using extensive epidemiological data and molecular typing. PLoS Med **2012**; 9(2): e1001172.
21. Curry SR, Muto CA, Schlackman JL, et al. Use of multilocus variable number of tandem repeats analysis genotyping to determine the role of asymptomatic carriers in Clostridium difficile transmission. Clin Infect Dis **2013**; 57(8): 1094-102.
22. Longtin Y, Paquet-Bolduc B, Gilca R, et al. Effect of Detecting and Isolating Clostridium difficile Carriers at Hospital Admission on the Incidence of C difficile Infections: A Quasi-Experimental Controlled Study. JAMA Intern Med **2016**; 176(6): 796-804.
23. Donskey CJ, Kundrapu S, Deshpande A. Colonization versus carriage of Clostridium difficile. Infect Dis Clin North Am **2015**; 29(1): 13-28.
24. Eyre DW, Fawley WN, Rajgopal A, et al. Comparison of Control of Clostridium difficile Infection in Six English Hospitals Using Whole-Genome Sequencing. Clin Infect Dis **2017**.

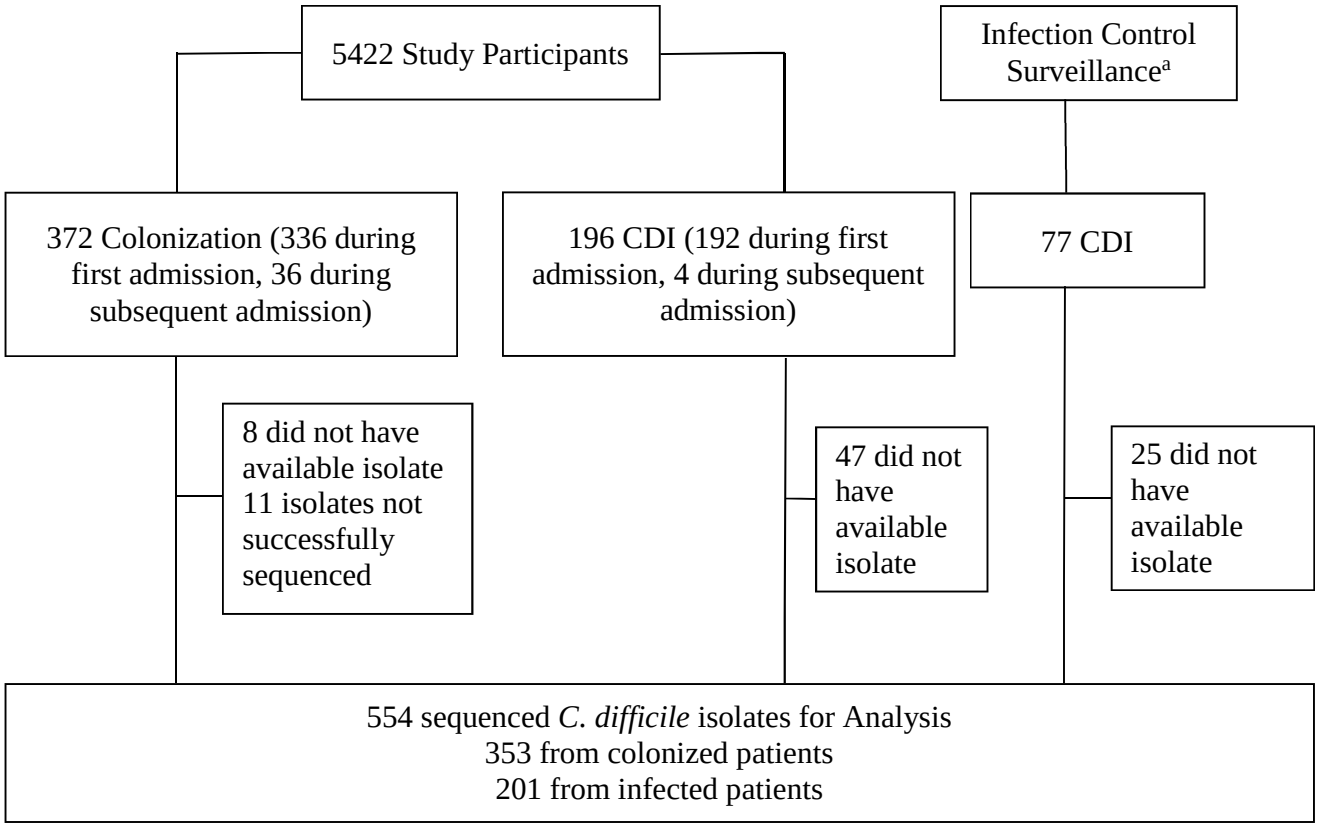
Table 1. Proportions of CDI cases genetically and epidemiologically linked to prior infected and colonized donors using WGS – all hospitals (201 cases)

	Genetically linked, n (%)	NAP1/027/ST1 among genetically linked donors, n (%)	Genetic and ward link, n (%)	NAP1/027/ST1 among genetically and ward linked donors, n (%)
Linked to prior case	105 (52)	95 (91)	81 (40)	74 (91)
Linked to infected patients only	28 (14)	23 (82)	34 (17)	31 (91)
Linked to colonized patients only	12 (6)	8 (67)	19 (10)	16 (84)
Linked to both infected and colonized patients	65 (32)	64 (99)	28 (14)	27 (96)

Table 2. Proportions of CDI cases genetically and epidemiologically linked to prior infected and colonized donors using WGS – 2 hospitals (117 cases)

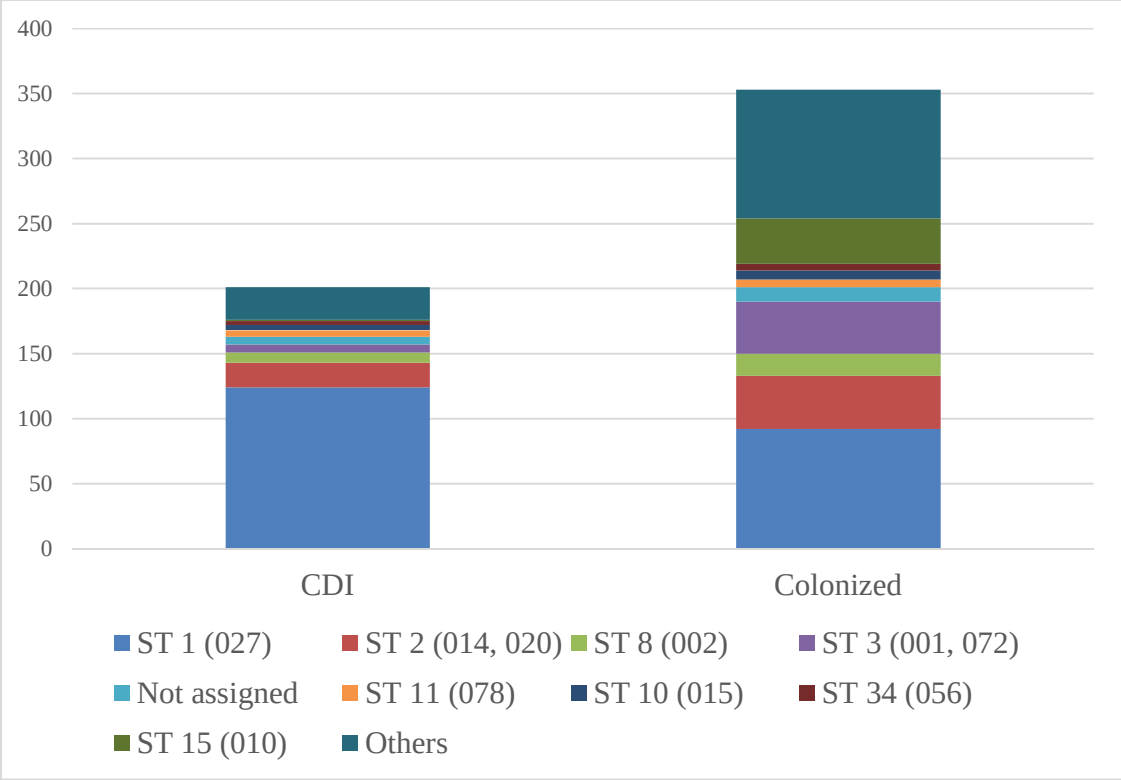
Possible source	Genetically linked, n (%)	NAP1/027/ST1 among genetically linked donors, n (%)	Genetic and ward link, n (%)	NAP1/027/ST1 among genetically and ward linked donors, n (%)
Linked to prior case	60 (51)	53 (88)	46 (39)	42 (91)
Linked to infected patients only	26 (22)	21 (81)	30 (26)	27 (90)
Linked to colonized patients only	4 (3)	3 (75)	2 (2)	2 (100)
Linked to both infected and colonized patients	30 (26)	29 (97)	14 (12)	13 (93)

Figure 1. Flowchart of patients and isolates included in analysis



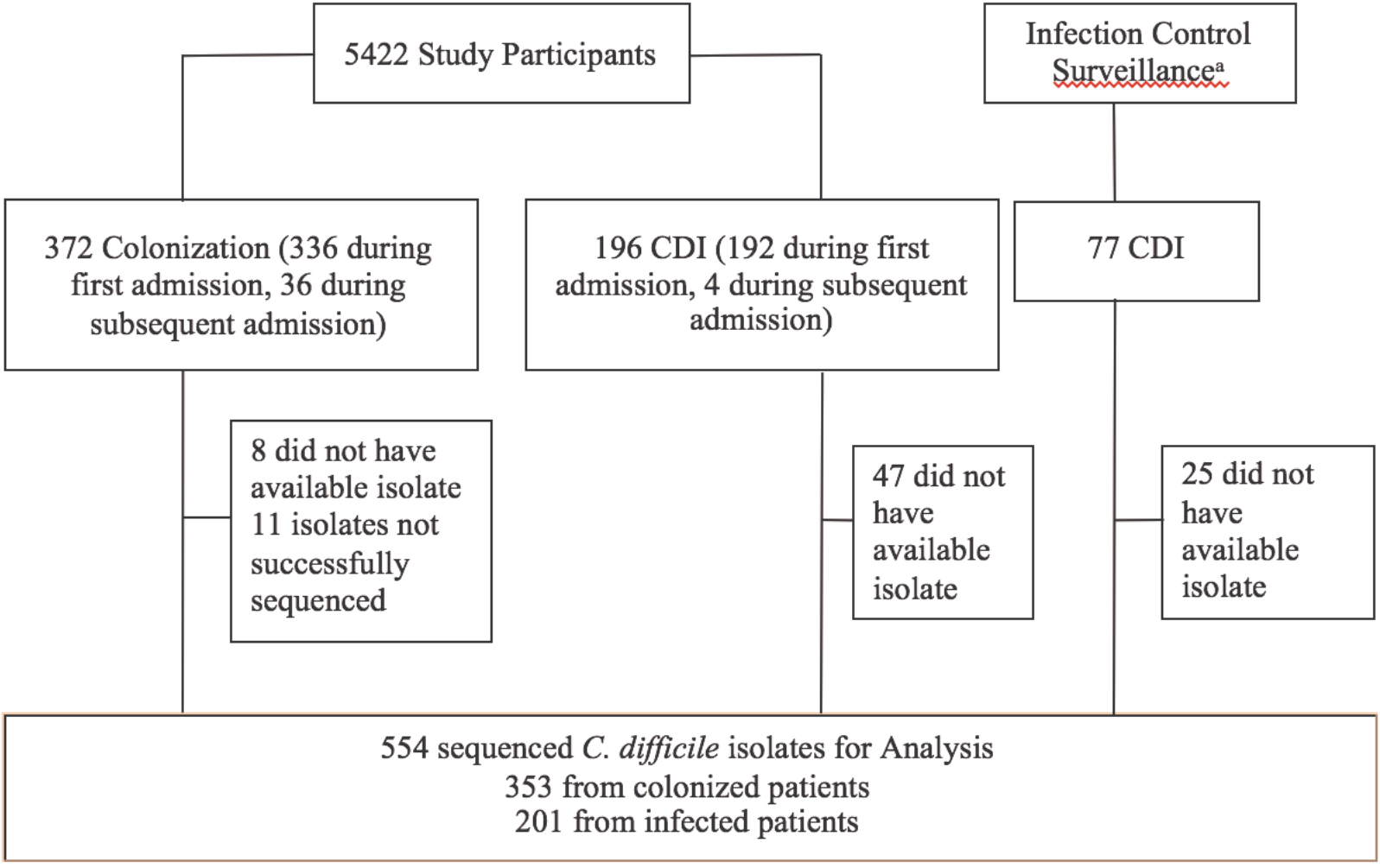
[a]Infection control surveillance isolates were available for one site

Figure 2. Multilocus sequence types by infected or colonized status

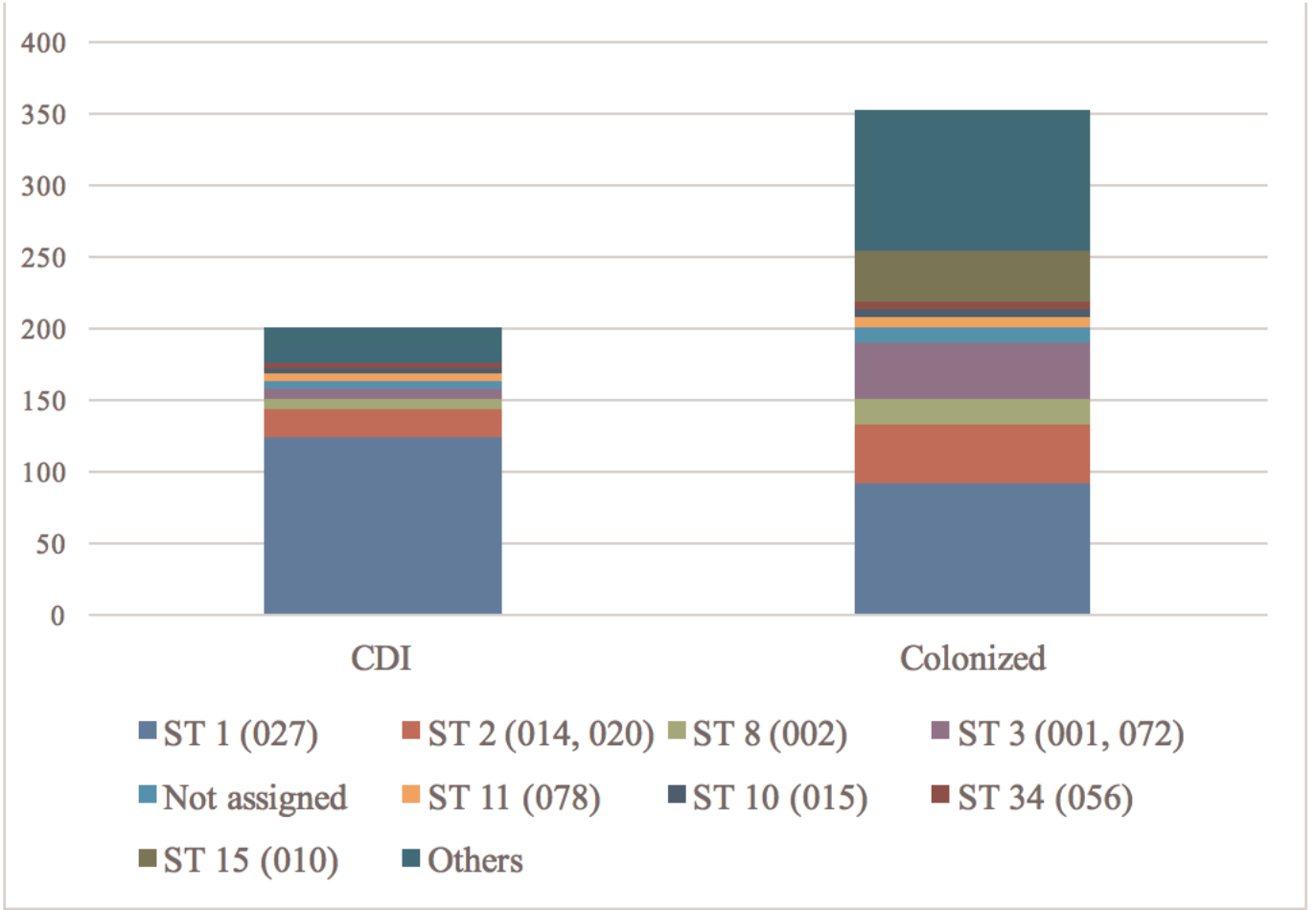


ST: Sequence type

PCR ribotype in parentheses



[a]Infection control surveillance isolates were available for one site


400
350
300
250
200
150
100
50
0
CDI
Colonized
ST 1 (027)
ST 2 (014, 020)
ST 8 (002)
ST 3 (001, 072)
Not assigned
ST 11 (078)
ST 10 (015)
ST 34 (056)
ST 15 (010)
Others

Supplementary Table. Detailed participation rates among hospitals in initial cohort.

	Hospital 1	Hospital 2	Hospital 3	Hospital 4	Hospital 5	Hospital 6	Total
Admissions	2320	1912	1617	2623	1851	1981	12304
Eligible patients	1823	1259	1326	2167	1688	1326	9502
Participants (% eligible)	1078 (59)	932 (74)	861 (65)	1159 (53)	850 (50)	542 (41)	5422 (57)
Patients testing positive in original cohort (colonized, infected)	118 (63, 55)	82 (54, 28)	64 (42, 22)	171 (112, 59)	80 (64, 16)	53 (37, 16)	568 (372, 196)
Isolates successfully sequenced from original cohort (colonized, infected)	111 (62, 49)	62 (50, 12)	53 (34, 19)	155 (110, 45)	75 (62, 13)	46 (35, 11)	502 (353, 149)
Number of CDI cases not enrolled in study	77	Unknown	0	Unknown	Unknown	Unknown	Unknown
Infection control CDI isolates sequenced	52	0	0	0	0	0	52
Total isolates successfully sequenced (colonized, infected)	163 (62, 101)	62 (50, 12)	53 (34, 19)	155 (110, 45)	75 (62, 13)	46 (35, 11)	554 (353, 201)
Proportion of all CDI cases included in analyses	77% (101/132)	Unknown	100% (19/19)	Unknown	Unknown	Unknown	Unknown