

1 Title: Genetic determinants of trehalose utilization are not associated with severe *Clostridium difficile*  
2 infection

3

4 Working title: Trehalose in *C. difficile* infection

5

6 Authors: Katie Saund<sup>1</sup>, Krishna Rao<sup>2</sup>, Vincent B. Young<sup>1,2</sup>, Evan S. Snitkin<sup>1,2</sup>

7 <sup>1</sup>Department of Microbiology and Immunology

8 <sup>2</sup>Department of Internal Medicine/Division of Infectious Diseases

9 University of Michigan, Ann Arbor, Michigan

10

11 Contact info for corresponding author: Evan S. Snitkin. Email: [esnitkin@med.umich.edu](mailto:esnitkin@med.umich.edu)

12 Keywords:

- 13 • trehalose
- 14 • comparative genomics
- 15 • *Clostridium difficile*
- 16 • case-control study

17

18

19

20

21

22

23

24

25

26

27

28

29 **ABSTRACT**

30 In a case-control study of patients with *C. difficile* infection we found no statistically significant association  
31 between the presence of trehalose utilization variants in infecting *C. difficile* strains and development of  
32 severe infection. These results do not support trehalose utilization conferring enhanced virulence in the  
33 context of human *C. difficile* infections.

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

## 57 INTRODUCTION

58 *Clostridium difficile* infection (CDI) is a health care-associated infection that can result in a range  
59 of patient outcomes. Of greatest concern is the development of severe disease, defined as intensive care  
60 unit admission, intraabdominal surgery (such as colectomy), and/or death attributable to the infection[1].  
61 Specific patient factors such as age, antibiotic use, and female gender have been associated with severe  
62 infection. The genetic background of the infecting *C. difficile* isolate may also influence clinical outcome.  
63 Prior studies reported an increased risk of severe infection for ribotypes RT027 and RT078[2,3]. One  
64 recently proposed mechanism for increased virulence of specific *C. difficile* lineages is the differential  
65 capacity to utilize the dietary disaccharide trehalose[4]. These authors observed enhanced trehalose  
66 utilization in strains of both RT027 and RT078, and demonstrated increased virulence of an RT027 strain  
67 in a mouse model of infection when physiologically relevant quantities of trehalose were given. This report  
68 also noted the coincidence between the introduction of trehalose as a food additive in 2000 and the global  
69 emergence of RT027 and RT078 shortly thereafter.

70 Subsequently, Eyre et al. examined clinical *C. difficile* isolates for the ability to use trehalose,  
71 noting that variants conferring improved trehalose utilization were not confined to successful epidemic  
72 lineages. Moreover, Eyre et al. found no evidence of association between a trehalose utilization variant  
73 and 30-day all-cause mortality in RT015, a ribotype in which enhanced trehalose utilization is variably  
74 present[5]. Here, we set out to more comprehensively evaluate the potential contribution of trehalose  
75 utilization to clinical outcomes by quantifying the independent contribution of trehalose utilization variants  
76 to infection severity across ribotypes, when controlling for all clinical factors independently associated  
77 with risk for severe infection.

78

## 79 METHODS

### 80 Study population

81 All subjects signed written consent to participate in this study. This study was approved by the  
82 University of Michigan Institutional Review Board (Study HUM33286). We considered a cohort of 1144  
83 cases of CDI from hospitalized adults diagnosed with CDI between October 2010 and January 2013 at  
84 the University of Michigan Health System[2]. The following predictors of severe outcome were noted in

85 the window 24-48 hours post CDI diagnosis: age, gender, metastatic cancer, concurrent antibiotic use,  
86 systolic blood pressure, creatinine, bilirubin, and white blood cell count. Of the 981 unique patients, 898  
87 had complete clinical information. CDI was classified as severe if any of the following outcomes  
88 attributable to CDI occurred within 30 days of diagnosis: admission to an intensive care unit,  
89 intraabdominal surgery, or death[1]. Ribotyping was performed and 137 ribotypes were identified  
90 (Supplementary Table 1).

## 91 **Data analysis**

92 Summary statistics, matching, modeling, and imputation were conducted in R version 3.5.0[6]. The R  
93 code is available at: [github.com/katiesaund/clinical\\_cdiffficile\\_trehalose\\_variants](https://github.com/katiesaund/clinical_cdiffficile_trehalose_variants).

94 **Severe Outcome Risk Score Matching** Unique subjects with complete clinical information (N=898) were  
95 assigned a severe outcome risk score based on a logistic regression model with severe CDI outcome as  
96 the response variable and the following predictors: age (years), female gender, metastatic cancer,  
97 concurrent antibiotic use, systolic blood pressure (mm Hg), creatinine (>1.5 mg/dL), bilirubin (>1.2  
98 mg/dL), and white blood cell count (cells/ $\mu$ L)[7]. Where possible, isolates were sorted into strata, each  
99 with exactly one case (severe CDI outcome) and at least one and up to four controls (non-severe CDI  
100 outcome), with control scores within  $\pm 0.10$  of case score. The matching process identified N=323 CDI  
101 cases isolates, of which only N=247 *C. difficile* isolates were successfully grown, isolated, and  
102 sequenced. Due to growth or sequencing failures only N=235 isolates were from N=59 complete strata  
103 (strata with at least one control and one case).

104 **Conditional logistic regression model for matched samples** Logistic regressions were modeled with  
105 severe CDI outcome as the response variable, conditioned on strata, and a trehalose variant as the  
106 predictor. Bonferroni corrected P-values were reported.

## 107 **Genomic Analysis**

108 Genomic DNA extracted with the MoBio PowerMag Microbial DNA Isolation Kit (Qiagen) from *C.*  
109 *difficile* isolates (N=247) was prepared for sequencing using the Illumina Nextera DNA Flex Library  
110 Preparation Kit. Sequencing was performed on either an Illumina HiSeq 4000 System at the University of  
111 Michigan Advanced Genomics Core or on an Illumina MiSeq System at the University of  
112 Michigan Microbial Systems Molecular Biology Laboratories. Quality of reads was assessed with

113 FastQC[8]. Adapter sequences and low-quality bases were removed with Trimmomatic[9]. Details on  
114 sequenced strains are available in Table S1. Sequence data are available under Bioproject  
115 PRJNA561087. For variant calling details see the Supplementary Methods. Pangenome analysis was  
116 performed with roary[10] and the insertion putatively conferring enhanced trehalose utilization was  
117 inferred based on the presence of four genes: *treA2*, *ptsT*, *treX*, and *treR2* (Supplementary File 1).

118

## 119 RESULTS

120 Given the limited clinical data regarding the potential contribution of trehalose utilization variants  
121 to CDI severity, we set out to test for an association between trehalose utilization across all strains of *C.*  
122 *difficile* causing infection while comprehensively controlling for patient factors associated with severe  
123 outcome. Patient factors associated with CDI severity in our clinical cohort (N=1144) were age, female  
124 gender, metastatic cancer, concurrent antibiotic use, systolic blood pressure, creatinine, bilirubin, and  
125 white blood cell count[2]. To quantify the independent contribution of trehalose utilization variants to  
126 severe patient outcomes we implemented a retrospective, matched case-control study to control for these  
127 patient factors. Each *C. difficile* episode was assigned a severe outcome risk score utilizing variables  
128 available around the time of diagnosis, which is the patient's predicted probability of having severe CDI,  
129 based on a logistic regression model of CDI built from the eight patient factors. Unique patients were  
130 grouped into strata based on their risk score. Each stratum contained one case (N=59, severe CDI  
131 outcome) and up to four controls (N=176, non-severe CDI outcome; total N=235). The median number of  
132 controls per case was 3 (range, 1-4). There were 61 RT027 isolates (19/59 cases; 46/176 controls), 5  
133 RT078 isolates (0/59 cases; 5/176 controls), and only 1 RT015 (1/59 cases; 0/176 controls) in this  
134 matched data set.

135 All of the *C. difficile* isolates in the matched data set with complete strata were sequenced  
136 (N=235). We identified the presence of reported trehalose utilization variants: two amino acid  
137 substitutions in the transcriptional repressor, *treR*, (TreR C171S and TreR L172I) and a set of four  
138 accessory genes, called the four gene trehalose insertion, that contain an additional phosphotrehalase  
139 enzyme and transcriptional regulator[4,11]. Consistent with previous reports, we found TreR C171S in all  
140 4 RT017 isolates and 3 closely related ribotypes (Figure 1). Similarly, we found TreR L172I in all 61

141 RT027 isolates and in 7 closely related ribotypes. However, in contrast to Collins et al. who found the four  
142 gene trehalose insertion only in RT078, we identified the insertion in 25 isolates that were broadly  
143 distributed across the phylogeny[4].

144 To investigate if trehalose utilization variants were individually associated with severe CDI  
145 outcome, we performed unadjusted logistic regressions conditioned on severe outcome risk strata (Table  
146 S2). The presence of the TreR L172I trehalose utilization variant was associated with an increased, but  
147 not statistically significant, odds of severe CDI outcome (OR, 1.60; 95% CI, 0.84-3.05;  $P=0.52$ ). The  
148 presence of the four gene trehalose insertion was associated with decreased, but not statistically  
149 significant odds of a severe CDI outcome (OR, 0.35; 95% CI, 0.09-1.32;  $P=0.52$ ). None of the other four  
150 amino acid substitutions observed in TreR were significantly associated with severe CDI outcome (Table  
151 S2). Due to low prevalence ( $N=7$ ), we did not test for an association between the presence of the TreR  
152 C171S trehalose utilization variant and severe CDI outcome.

153 To evaluate the overall effect of the three previously described trehalose utilization variants on CDI  
154 outcome we combined the three variants into a single predictor of severe CDI outcome and performed a  
155 logistic regression conditioned on severe outcome risk strata. When combined, the presence of any of the  
156 three trehalose utilization variants does not significantly affect the odds of severe CDI outcome (OR, 1.10;  
157 95%CI, 0.61-1.99;  $P=0.86$ ). To evaluate whether these results translated to the original unmatched  
158 cohort, we also tested for associations between trehalose utilization variants and severe infections using  
159 severe risk score as a covariate rather than matching criterion. To determine trehalose utilization  
160 genotypes in the larger cohort for which genomic data was unavailable, we performed an imputation  
161 where the presence of trehalose utilization variants (TreR C171S, TreR L172I, and the four gene  
162 trehalose insertion) was assumed to have the same distribution across ribotypes as was observed in the  
163 matched cohort (see Supplementary Methods). We performed a logistic regression modeling severe  
164 outcome with presence of any of the three trehalose utilization variants as a predictor and severe risk  
165 score as a covariate. When combined, the presence of any of the three trehalose utilization variants does  
166 not significantly affect the odds of severe CDI outcome (Median OR, 0.99; Range, 0.87-0.99; Median  
167  $P=0.97$ ; Range, 0.64-0.99;  $N=586$ ).

168

## 169 DISCUSSION

170 The ability to predict the clinical outcome of *C. difficile* infection based on patient and pathogen  
171 characteristics could help guide therapy for this important nosocomial infection. An enhanced ability to  
172 utilize trehalose was shown to be associated with increased virulence in mice, prompting us to evaluate  
173 this relationship using a case-control study performed with a large clinical data set. Our controlled  
174 analysis failed to detect a significant association between any of the previously described trehalose  
175 utilization variants and severe CDI outcome. Our results from patients infected with a diversity of  
176 ribotypes are consistent with Eyre et al.'s observed lack of association between the four gene trehalose  
177 insertion and 30-day all-cause mortality specifically in patients infected with RT015. We also observed  
178 that the four-gene trehalose insertion is not exclusive to RT078, but rather is found throughout the *C.*  
179 *difficile* phylogeny.

180 The lack of association between trehalose utilization variants and severe CDI outcome in  
181 hospitalized patients emphasizes the need to incorporate clinical results earlier into the genetic variant  
182 discovery pipeline. Indeed, a more relevant hypothesis generating process could begin by identifying  
183 genetic loci of interest first through comparative genomic analysis of clinical isolates followed by validation  
184 in lab strains with molecular and animal studies. A critical component of this discovery work is to control  
185 for both patient factors and strain background as both may confound analysis, which we did above in  
186 separate analyses using two distinct modeling strategies. Given the association between ribotype, recent  
187 acquisition, and CDI outcome, controlling for patient factors is critical for identifying the genetic variants  
188 associated with severe CDI outcome[12].

189 While these clinical data cannot rule out a potential role of trehalose utilization variants in the  
190 success of hypervirulent *C. difficile* lineages, they do emphasize the difference between human infection  
191 and murine models. More broadly, generating hypotheses in controlled laboratory systems and having  
192 them generalize to genetically and clinically heterogeneous human populations will limit insights to only  
193 the most penetrant phenotypes. With increasing availability of electronic health record and microbial  
194 genomic data, it is now becoming feasible to flip the script and generate hypotheses through analysis of  
195 human data, which can subsequently be tested in appropriate animal and *in vitro* systems with *a priori*  
196 knowledge of relevance to human disease.

197

## 198 FUNDING

199 This work was supported by the National Institutes of Health [1U01AI124255]. KS was supported by the  
200 National Institutes of Health [T32GM007544].

201

## 202 ACKNOWLEDGEMENTS

203 We thank Ali Pirani for bioinformatics support. We thank Arianna Miles-Jay for her helpful suggestions.

204

## 205 REFERENCES

- 206 1. Clifford McDonald L, Coignard B, Dubberke E, Song X, Horan T, Kutty PK. Recommendations for  
207 Surveillance of Clostridium difficile–Associated Disease. *Infect Control Hosp Epidemiol* **2007**; 28.
- 208 2. Rao K, Micic D, Natarajan M, et al. Clostridium difficile Ribotype 027: Relationship to Age,  
209 Detectability of Toxins A or B in Stool with Rapid Testing, Severe Infection, and Mortality. *Clin*  
210 *Infect Dis* **2015**; 61:233–241. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25828993>.  
211 Accessed 10 November 2016.
- 212 3. Walker AS, Eyre DW, Wyllie DH, et al. Relationship between bacterial strain type, host  
213 biomarkers, and mortality in clostridium difficile infection. *Clin. Infect. Dis.* 2013; 56:1589–1600.  
214 Available at: [https://oup.silverchair-](https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/cid/56/11/10.1093_cid_cit127/2/cit127.pdf?Expires=1500592769&Signature=btVMYEpwYmM7E08K3gthf-e6qsNyxMnPaDF8UHR3eTBis0v4jD0zfPQobzNDM7c604DF7gHqUQv-HzZswHvKcQ1HmGjiMR01OWGQT~M~4Tcy1XBCWIXs)  
215 [cdn.com/oup/backfile/Content\\_public/Journal/cid/56/11/10.1093\\_cid\\_cit127/2/cit127.pdf?Expires=](https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/cid/56/11/10.1093_cid_cit127/2/cit127.pdf?Expires=1500592769&Signature=btVMYEpwYmM7E08K3gthf-e6qsNyxMnPaDF8UHR3eTBis0v4jD0zfPQobzNDM7c604DF7gHqUQv-HzZswHvKcQ1HmGjiMR01OWGQT~M~4Tcy1XBCWIXs)  
216 [1500592769&Signature=btVMYEpwYmM7E08K3gthf-](https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/cid/56/11/10.1093_cid_cit127/2/cit127.pdf?Expires=1500592769&Signature=btVMYEpwYmM7E08K3gthf-e6qsNyxMnPaDF8UHR3eTBis0v4jD0zfPQobzNDM7c604DF7gHqUQv-HzZswHvKcQ1HmGjiMR01OWGQT~M~4Tcy1XBCWIXs)  
217 [e6qsNyxMnPaDF8UHR3eTBis0v4jD0zfPQobzNDM7c604DF7gHqUQv-](https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/cid/56/11/10.1093_cid_cit127/2/cit127.pdf?Expires=1500592769&Signature=btVMYEpwYmM7E08K3gthf-e6qsNyxMnPaDF8UHR3eTBis0v4jD0zfPQobzNDM7c604DF7gHqUQv-HzZswHvKcQ1HmGjiMR01OWGQT~M~4Tcy1XBCWIXs)  
218 [HzZswHvKcQ1HmGjiMR01OWGQT~M~4Tcy1XBCWIXs](https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/cid/56/11/10.1093_cid_cit127/2/cit127.pdf?Expires=1500592769&Signature=btVMYEpwYmM7E08K3gthf-e6qsNyxMnPaDF8UHR3eTBis0v4jD0zfPQobzNDM7c604DF7gHqUQv-HzZswHvKcQ1HmGjiMR01OWGQT~M~4Tcy1XBCWIXs). Accessed 19 July 2017.
- 219 4. Collins J, Robinson C, Danhof H, et al. Dietary trehalose enhances virulence of epidemic  
220 Clostridium difficile. *Nature* **2018**; 553:291–294. Available at:  
221 <https://www.nature.com/articles/nature25178.pdf>. Accessed 12 April 2019.
- 222 5. Eyre DW, Didelot X, Buckley AM, et al. Clostridium difficile trehalose metabolism variants are  
223 common and not associated with adverse patient outcomes when variably present in the same  
224 lineage. *EBioMedicine* **2019**; Available at:



- 225 <https://linkinghub.elsevier.com/retrieve/pii/S2352396419302774>.
- 226 6. R Core Team. R: A language and environment for statistical computing. 2018; Available at:  
227 <https://www.r-project.org/>.
- 228 7. Rao K, Micic D, Natarajan M, et al. Clostridium difficile Ribotype 027: Relationship to Age,  
229 Detectability of Toxins A or B in Stool with Rapid Testing, Severe Infection, and Mortality. Clin  
230 Infect Dis **2015**; 61:233–241.
- 231 8. Andrews Si. FastQC: a quality control tool for high throughput sequence data. 2010;
- 232 9. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data.  
233 Bioinformatics **2014**; 30:2114–2120. Available at:  
234 <http://www.usadellab.org/cms/index.php?page¼trimmomatic>. Accessed 1 March 2018.
- 235 10. Page AJ, Cummins CA, Hunt M, et al. Roary: Rapid large-scale prokaryote pan genome analysis.  
236 Bioinformatics **2015**; 31:3691–3693.
- 237 11. Collins J, Danhof H, Britton RA. The role of trehalose in the global spread of epidemic Clostridium  
238 difficile. Gut Microbes **2018**; :1–6. Available at:  
239 <http://www.tandfonline.com/action/journalInformation?journalCode=kgmi20>. Accessed 21 August  
240 2018.
- 241 12. Martin JSH, Eyre DW, Fawley WN, et al. Patient and strain characteristics associated with  
242 clostridium difficile transmission and adverse outcomes. Clin Infect Dis **2018**; 67:1379–1387.  
243 Available at: <https://academic.oup.com/cid/article-abstract/67/9/1379/4969186>. Accessed 1 May  
244 2019.
- 245
- 246
- 247
- 248
- 249
- 250
- 251
- 252

253

254 **FIGURE LEGEND**

255 Figure 1. Comparative analysis of trehalose genetic variants in *Clostridium difficile*. A maximum-likelihood  
256 phylogeny was constructed (N=235) using variants identified in the genome (scale in mutations per site in  
257 whole genome). Trehalose variants are either point mutations in *treR* (C171S empty orange circle, L172I  
258 empty blue circle) or presence of the four gene trehalose insertion (empty black circle). A severe infection  
259 outcome is indicated by a red triangle. Ribotype is indicated by the filled circles.

- TreR L172I
- TreR C171S
- Four Gene Insertion
- ▲ Severe Infection Outcome
- RT014
- RT015
- RT017
- RT027
- RT053
- RT078
- Other Ribotype

