1 **Flagellar switch inverted repeat sequence variation impacts *Clostridioides difficile* RT027/MLST1**
2 **virulence.**

3 Nguyen T. Q. Nhu[a,b,#], Huaiying Lin[b], Jonathan K. Sia[c], Evan S. Snitkin[d], Vincent Young[d,e], Mini Kamboj[f], Eric
4 G. Pamer[a,b], Phoebe A. Rice[g], Qiwen Dong[a,b,#].

5 [a]Department of Medicine, University of Chicago, Chicago, Illinois, USA

6 [b]Duchossois Family Institute, University of Chicago, Chicago, Illinois, USA

7 [c]Immunology Program, Memorial Sloan Kettering Cancer Center, New York City, New York, USA

8 [d]Division of Infectious Diseases, Department of Internal Medicine, University of Michigan, Ann Arbor,
9 Michigan, USA

10 [e]Department of Microbiology & Immunology, University of Michigan, Ann Arbor, MI, USA

11 [f]Infection Control, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New
12 York, USA

13 [g]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL, USA

14 Running Head: Flagellar switch inverted repeats impact *C. difficile* ST1 virulence.

15

16 [#]Address correspondence to Qiwen Dong, dongq@uchicago.edu; Nguyen T. Q. Nhu, nhun@umich.edu

17 *Present address: Nguyen T. Q. Nhu, Division of Infectious Diseases, Department of Internal Medicine,
18 University of Michigan, Ann Arbor, Michigan, USA

19

20 **ABSTRACT** The impact of *Clostridioides difficile* strain diversity on the severity of intestinal infection and
21 diarrhea remains unclear. The RT027/ST1 strain is highly prevalent and has been associated with more
22 severe disease and increased toxin production. Previous studies using antibiotic-treated mice, however,
23 have demonstrated that the severity of diarrhea and colonic inflammation caused by clinical RT027/ST1
24 *C. difficile* isolates cultured from different patients is isolate-specific and varies from avirulent to lethal.
25 Herein we demonstrate that increased flagellar expression is associated with increased in vivo virulence.
26 While sequencing the flagellar operon did not identify differences between high and low virulence
27 isolates, high virulence was associated with strains encoding higher proportions of the phase variable
28 flagellar switch in the ON mode. We found that the proportion of bacteria with the switch in the ON mode
29 correlates with maximum weight loss, a readout for virulence in mice. Further analysis of flagellar switch
30 sequences identified variant inverted repeat (IR) sequences, with 40% of isolates having lost one A or T in
31 the IR compared to the genome of *C. difficile* R20291. The isolates with the R20291 IR sequence have
32 increased virulence compared to those with variant IR ($p = 3.2e-05$) and these isolates have markedly
33 reduced ability to invert the flagellar switch. Our results suggest that a restricted capacity to invert the
34 flagellar switch during infection is associated with reduced virulence of *C. difficile* RT027/ST1 strains and
35 may account for some of the inconsistent associations between *C. difficile* infection and disease severity
36 in patients.

37

38 **IMPORTANCE** *C. difficile* RT027/ST1 has been associated with high virulence. However, variable virulence
39 among ST1 isolates has also been observed. Recent studies have shed light on the impacts of the gut
40 microbiota and host immunity during *Clostridioides difficile* infection (CDI). Beyond that, determining the
41 genetic factors of *C. difficile* regulating virulence is essential for a successful CDI diagnosis and treatment.
42 Here, we report a strong association between the flexibility of the flagellar switch and the virulence of *C.*
43 *difficile*. Specifically, the inverted repeats of the flagellar switch correlate with switch flexibility. Overall,

44 our study explores variations in *C. difficile* virulence and highlights the potential of using flagellar switch
45 inverted repeats as an easily accessible genetic trait to predict pathogen virulence.

46 **KEYWORDS** *Clostridioides difficile*, RT027, virulence, flagellar switch, inverted repeats, recombinase

47

48 **INTRODUCTION**

49 *Clostridioides difficile* (*C. difficile*) is the leading cause of hospital-acquired infection in the U.S., with an
50 estimated incidence of approximately 224.000 cases per year (1). The incidence of community-acquired
51 *C. difficile* infection (CDI) is also increasing, with up to 51.2 cases /100.000 population (2). Antibiotic-
52 mediated depletion of commensal bacterial species increases the risk of CDI and prolonged hospitalization
53 increases exposure to *C. difficile* spores. The spectrum of disease associated with *C. difficile* infection
54 ranges from asymptomatic colonization to mild/moderate diarrhea to severe pseudomembranous colitis
55 and death. It has been estimated that CDI causes approximately 12.800 deaths and costs the US Health
56 care system one billion U.S. dollars per year (1).

57 *C. difficile* virulence is mediated by two glucosylating toxins, TcdA and TcdB (3). These toxins are
58 internalized by epithelial cells by endocytosis (4-6) and are released into the cytosol where they
59 glucosylate and inactivate intracellular GTPases (7-9). These activities can disrupt cell signaling pathways
60 and cytoskeletal structure, leading to cell death (10). The impact of TcdA and TcdB on disease progression
61 and severity can be varied, depending on the strain of *C. difficile* and the host's innate and adaptive
62 immune defenses (11, 12).

63 The extent to which *C. difficile* strains vary in virulence is unclear. Because every patient with *C. difficile*
64 infection has a distinct residual microbiota and constellation of risk and resistance factors that impact the
65 severity of disease, it has been challenging to determine the impact of the pathogen's intrinsic virulence

66 on disease course (13, 14). The multi-locus sequence type 1 (ST1) strain, a.k.a. as NAP1/B1/ribotype 027,
67 is highly transmissible and has increased multidrug resistance (15, 16) and was initially reported to be
68 hyper-virulent (17, 18). Subsequent studies, however, suggested its virulence was similar to other strains
69 (19, 20). It emerged in the early 2000s in the USA and Canada then spread to European and South
70 American countries (21). Today, ST1 can be found worldwide. The virulence of ST1 has been attributed to
71 higher levels of toxin production (22). Subsequent studies, however, reported heterogeneity in toxin
72 production and virulence of ST-1 strains (23, 24).

73 In addition to toxin expression, bacterial motility mediated by flagella has been implicated in *C. difficile*
74 virulence. Similar to other gut-bacterial intestinal pathogens such as *Salmonella typhi* and *Escherichia coli,*
75 *C. difficile* is a peritrichous bacterium. In fact, the flagellar expression has been shown to contribute to
76 virulence by enhancing host cell invasion, adhesion to epithelial cells, and systemic inflammation of *S.*
77 *enterica* and *E. coli* (25-27). Deletion of flagellar structural genes or inactivation of the flagellar motor
78 resulted in distinct impacts on *in vivo* virulence, depending on the *C. difficile* strain (28). In *C. difficile*, the
79 deletion of *fli*C, which encodes flagellin, reduces the virulence of *C. difficile* R20291, an ST1 strain, in a
80 mouse model (29). In contrast, in *C. difficile* strain CD630, the deletion of *fli*C did not decrease virulence
81 in the hamster model of CDI (30). Recent studies demonstrated that flagellar expression is coupled with
82 TcdA/TcdB production, suggesting that toxin-mediated virulence is coupled to flagellin expression (31).
83 Whether flagellar expression contributes to the relative virulence of different *C. difficile* ST1 strains is
84 unclear.

85 In this study, we investigated the heterogeneity of *C. difficile* ST1 strains isolated by fecal culture from
86 patients with CDI. Testing of 22 ST1 clinical isolates in mice revealed marked differences in virulence. We
87 find that sequence differences in the flagellar switch region contribute to variations in ST1 strain virulence.
88 Specifically, we demonstrate strain diversity in the inverted repeat (IR) region of the flagellar switch
89 among ST1 strains. The proportion of ON versus OFF in the flagellar switch is determined by the sequence

90 of the IR. The published IR sequence in *C. difficile* R20291 enhances flexibility in switching between ON
91 and OFF but this IR sequence is not shared by all ST1 strains, many of which have reduced ability to alter
92 flagellar expression. Our results demonstrate that flexibility in flagellar switching mediated by the IR
93 contributes to heterogeneity in *C. difficile* ST1 virulence.

94 **RESULTS**

95 **Highly virulent isolates express more flagella and demonstrate higher motility.**

96 Previous studies characterizing *in vivo* virulence of 23 *C. difficile* ST1 isolates revealed disease severities in
97 infected C57BL/6 mice that ranged from mortality to absence of any weight loss or diarrhea (Dong et al.
98 in press). While the avirulence of 2 ST1 strains was attributed to a deletion in the cdtR gene (Dong et al.
99 in press), the pathogenicity and CDT loci that encode TcdA and TcdB glucosylating toxins and the CDT
100 binary toxin CDT were identical in the remaining 21 ST1 strains. To begin to identify additional mechanisms
101 for variable virulence among ST1 strains, we selected two strains that induced >10% weight loss (ST1-12
102 and ST1-53) and two strains that resulted in < 10% weight loss (ST1-6 and ST1-27) for further study (Fig.
103 S1). Transcriptomic profiling of these strains in cecal contents obtained 24 hours following infection of
104 antibiotic treated C57BL/6 mice demonstrated similar transcription of both glucosylating toxin and binary
105 toxin genes (Fig. 1 & Fig. S3). Similar *in vivo* toxin transcription among these strains was supported by toxin
106 profiling of fecal samples (Fig. S2).

107 Transcription of flagellar-related genes, on the other hand, was increased in highly virulent compared to
108 less virulent isolates (Fig. 1A), which was further validated by RT-qPCR (Fig. S3). To determine whether
109 flagellar-related gene transcription correlates with flagellar expression, we obtained transmission
110 electron microscopy (TEM) images of the four ST1 *C. difficile* strains grown on hard agar plates. High
111 virulence ST1 strains were found to be flagellated, with 4-5 peritrichous flagella, while low virulence ST1
112 isolates lacked detectable flagella (Fig. 1B).

113 To test the motility of ST1 *C. difficile* strains, we cultured bacteria in a hydrogel environment and measured
114 the spread of *C. difficile* colonies on 0.3% agar buffered with BHIS pH 7.0. Highly virulent ST1-12 and ST1-
115 53 strains produced larger colony sizes after one day of incubation than the reduced-virulence ST1-6 and
116 ST1-27 strains (Fig. 1C). We did not detect mutations in genes encoding flagellar proteins in high- or low-
117 virulence ST1 isolates.

118 Although low virulence ST1 strains cells have reduced motility and flagellar expression, we were able to
119 select strain variants that had upregulated flagellin expression by selectively culturing bacteria from the
120 periphery of swim colonies, which we refer to as E strains (i.e., ST1-6E and ST1-27E). TEM images
121 demonstrate that ST1-6E and ST1-27E strains are highly flagellated (Fig. 2A). Cecal contents of mice
122 infected with the E strains had higher Tcd toxins compared to the parental isolates (Fig. 2B).

123 The shift in flagellar expression and toxin production directed us to examine the flagellar switch, which is
124 in the upstream region of *flg*B operon (31) and has a length of approximately 200bp with a 154bp switch
125 region and two flanking inverted repeats. The switch can flip between the ON and OFF orientation but the
126 factors regulating its position remain unclear. We used the reads obtained by Illumina sequencing to
127 identify ON and OFF sequences in ST1 strains and discovered that both are present in all of the clinical
128 isolates. We quantified the fraction of ON and OFF sequences in *C. difficile* liquid culture more precisely
129 by flagellar switch qPCR (31) and found that high-virulence isolates have 70-80% flagellar switches in the
130 ON position, whereas less virulent isolates either have >90% in either the ON or OFF position. A higher
131 proportion of flagellar switches in the ON position weakly correlates with maximum weight loss by linear
132 regression (R^2 = 0.39) (Fig. 2C), suggesting that alternative factors regulate the virulence.

133 **The virulence has a strong association with the IR type.**

134 As we expanded the alignment to all 68 isolates in our collection, we found that 65 isolates have the
135 flagellar switch sequence identical to *C. difficile* R20291 while the remaining three isolates have the

136 flagellar switch identical to *C. difficile* CD630. To examine whether the local genetic structure might impact
137 the flagellar switch directions, we compared the left and right IR of the flagellar switch. While most isolates
138 have IRs identical to the reference strain *C. difficile* R20291, 40% of isolates has at least one less A or T in
139 either IRs. We identified four types of IR flanking regions in our strain collection and named them as
140 Common1 (C1-59.38% 6A/6T), Common2 (C2-32.81%, 5A/6T or 6A/5T), Rare1 (R1-6.25%, 6A/5T or 5A/6T),
141 and Rare2 (R2-1.56%, 5A/5T) (Fig. 3A). More specifically, C1 has the sequence as seen in *C. difficile* R20291
142 with 6A on the left IR (LIR) and 6T on the right IR (RIR) (Fig. 4A, sequence in the box). The number of A and
143 T on both IRs is the same when C1 switches between OFF and ON. On the other hand, C2 has LIR-5A and
144 RIR-6T when OFF and LIR-6A and RIR-5T when ON. R1 has LIR-6A and RIR-5T when OFF and LIR-5A and
145 RIR-6T when ON. R2 has LIR-5A and RIR-5T either when it is ON or OFF. Depending on the biased direction
146 of the flagellar switch, we can further classify C2 and R1 types as C2-ON and C2-OFF, R1-ON and R1-OFF.
147 R2 appears in only one strain, ST1-67, when it is in the OFF direction. Comparing to other IR types, we
148 found a strong correlation between IR type C1 and more severe weigh loss in infected mice ($p = 3.2e-05$)
149 (Fig. 3B). We speculate that the IR type C1 may contribute more to the virulence of *C. difficile* than the
150 other types.

151 **The inverted repeat type affects the flexibility of the flagella switch.**

152 To further investigate whether different IR types bias the flagellar switch direction, we expanded the
153 flagellar switch qPCR analysis to all our isolates, 64 of which yielded a successful amplification. 3-4
154 biological replicates were examined per isolate. To contrast the proportion of ON and OFF switch in each
155 culture, we determined the ΔCt value from qPCR results.

156 $\Delta Ct = |Ct_{ON} - Ct_{OFF}|$

157 in which Ct_{ON} is the cycle threshold from qPCR amplifying the ON sequence, and Ct_{OFF} is the cycle threshold
158 from qPCR amplifying the OFF sequence.

159 We found that isolates with IR type C1 have the lowest average ΔCt, with 2.09 (0.54 to 5.00) cycles,
160 indicating a flexible flagellar switch that can flip between ON and OFF positions. When calculating the
161 proportion, the C1 culture is biased towards ON fraction, with approximately 72.62% (54.61-99.18) (Fig.
162 4B). On the other hand, IR type C2 has an average of ΔCt 8.88 (7.08 to 15.30) cycles when it is at OFF
163 direction. The difference is significant larger when the C2 switch is ON, with 14.87 (14.46 -15.30) cycles.
164 Other rarer IR types like R1 and R2 have significantly larger ΔCt (Fig. 4A), with 15.36 (13.62-16.57) cycles
165 for R1-OFF and 16.28 for R1-ON. The rarest type, R2-OFF with only one isolate, has the largest ΔCt, with
166 20.99 cycles. Because of the large ΔCt, the proportion of ON and OFF in isolates with C2, R1, and R2 are
167 very skewed toward one direction. In average, the biased-OFF isolates have 95.58% (88.71-100.00) OFF
168 genomes, whereas the biased-ON isolates have 96.67% (92.60-100.00) ON genomes.

169 To test if the division of ON and OFF fraction of each isolate displays a similar pattern in an *in vivo* model,
170 we performed qPCR from DNA extracted from fecal pellets collected from mice 24 hours after infection
171 with *C. difficile* isolates. Isolates from each switch type, including *C. difficile* R20291 yielded that same
172 proportions of ON and OFF detected in *in vitro* cultures, with IR type C1 having the lowest ΔCt 1.10 (0.09-
173 3.00) for three distinct strains (Fig. 4A) with 76.69% ON (65.62-85.86%) and 23.31% OFF (Fig. 4B). The
174 other switch types maintained larger ΔCt which resulted in biased to ON or OFF status. These results
175 suggest that the C1 switch type enables more flexible switching of flagellar expression which may
176 contribute to *in vivo* virulence of *C. difficile*.

177 **The IR type C1 shows more variation in the switch direction in *in vivo* model.**

178 We infected mice with *C. difficile* R20291 and clinical isolates with different switch types and monitored
179 the ON and OFF fractions in fecal samples and cecal contents over the course of eight days (Fig. 5). Isolates
180 with IR type C2, R1, and R2 remain largely locked in terms of the proportion of ON and OFF over the course
181 of eight days. In contrast, we observed more variation in switch status in mice infected with IR type C1

182 strains. In some mice, C1 isolates are predominantly in the OFF status, with switching occurring
183 predominantly in the recovery phase after resolution of diarrhea and weight loss.

184 ***C. difficile* ST1 strains in the US have variable IR sequences.**

185 To further confirm IR diversity, we screened 1359 ST1 isolates with whole-genome sequences downloaded
186 from NCBI Biosamples. These samples are from the CDC HAI-Seq *C. difficile* collection (PRJNA629351) (32),
187 Texas and Michigan medical centers (PRJNA595724, PRJNA561087). The remaining samples come from
188 more recent publications (33-36). All *C. difficile* samples were collected from 1988 to 2020. Most
189 (1318/1359) are from Continental Europe, UK, Ireland and the U.S., with a small number (20/1359) of
190 samples from East Asia and Australia. Most *C. difficile* ST1 samples have flagellar switch with IR type C1,
191 accounting for 96.47% (1311/1359) of samples. Type C2 appears with only 2.43% (33/1359) of samples,
192 of which 27/33 originate from the U.S between 2007 and 2020. Other IR types are present at reduced
193 frequency of only 0.66% of samples (Fig. 6A). Focusing on the more recent CDC HAI-Seq collection, which
194 were predominantly submitted from Minnesota between 2009 and 2020, IR type C1 remains dominant,
195 but C2 accounts for 30.16% (19/63) samples, similar to our collection (32.80%). Interestingly, the
196 proportion of C2 increased in recent years, and accounts for up to 52.94% (9/17) of samples collected in
197 2020 (Fig. 6B). Our analysis on the prevalence of different IR types, especially in the U.S., suggests that the
198 heterogeneity in ST1 virulence maybe related to diversity of IR types.

199 **Proposed model of flagellar switch inversion**

200 RecV, a tyrosine recombinase, has been reported as necessary for the inversion of several switches in *C.*
201 *difficile* (37). However, the preferred DNA sequence for RecV binding had not been identified, nor was it
202 known how RecV avoids recombining incorrect pairs of sites – e.g., the ends of two different invertible
203 regions. Here we propose the mode of action of recV on the flagellar switch sequence.

204 The general mechanism for tyrosine recombinase-mediated recombination has been well-studied (38, 39)
205 and can be used as a basis for understanding RecV action. We propose that the inverted repeats which
206 flank each invertible switch are themselves degenerate inverted repeats consisting of two copies of a
207 RecV-binding motif flanking a central 6-bp spacer region. Recombination would occur within a tetramer
208 of RecV that pairs two such sites, with cleavage and strand exchange occurring at the outer edges of the
209 spacers. Recombinase binding does not depend on spacer sequence. However, because the spacer
210 segments become heteroduplex in the product (1 strand from each partner site), recombination of two
211 binding sites can only proceed to completion if spacers have identical sequences.

212 First, we used AlphaFold2 to predict the structure of a recV tetramer (Fig. 7A). The predicted structure is
213 highly similar to those of the catalytic domains of other tyrosine recombinases such as Cre (which
214 recombines loxP DNA sites), including the C-terminal helix that mediates protein – protein contacts. (38,
215 39). However, RecV lacks the N-terminal DNA binding domain found in many tyrosine recombinases. We
216 speculate that RecV's catalytic domain may bind more tightly and specifically than that of many tyrosine
217 recombinases, alleviating the need for an additional domain.

218 Second, we compared the inverted repeats of all six switches catalyzed at least in part by RecV in *C. difficile*
219 R20291, taking into account the approximate DNA cleavage points previously identified (37). We were
220 able to identify a conserved ~13bp potential RecV binding motif that is found in inverted repeat flanking
221 a 6bp spacer, as expected for canonical tyrosine recombinase binding sites (Fig. 7B). Searching the entire
222 *C. difficile* R20291 genome for a pattern based on this,
223 `AWAGTWNCCNTTWNNNNNNWAANGGNWACTWT`, (W= A or T; N = any base; one mismatch allowed per 10
224 bp) produced 17 hits, including both ends of *C. difficile* inversions (cdi) 2-5 and one end each of cdis 1 and
225 6 (the other ends contain similar sequences but too many mismatches) and 7 additional sites that may be
226 false positives. In the future, more sophisticated position-weighted searches would be likely to yield more
227 accurate results. In further support of this motif, previous work has reported that inversion in *C. difficile*

228 R20291 is locked when the conserved nucleotides in the RIR are deleted, suggesting that the activity of
229 recV was impaired (40).

230 Further examination of the ends of the 6 switches reveals how the system avoids aberrant recombination
231 between incorrectly paired repeats (Fig. 7C). For each individual switch, the spacer sequences within the
232 RecV dimer-binding sites at each end are identical. However, each switch has its own "spacer code."
233 Therefore, even if a RecV tetramer were to pair two ends of different switch regions, it would be difficult
234 to complete recombination between them.

235 The length of this variable spacer region is similar to the length of the spacer in other tyrosine recombinase
236 systems, and we propose that variations in its length explain our findings. We predict the model for recV
237 cleavage as in Figure 7D in which the tyrosine cleaves the inverted repeat at the 5' end of the last
238 conserved nucleotide. After the cleavage, two strands exchange the spacer region and form a Holliday
239 structure to initiate the inversion of the switch. In our collection, type C1 has the spacer sequence as LIR-
240 tacaaa(6)/RIR-tttgta(6). The other types have less bases on either left or right of the spacer region. For
241 example, C2 has sequence as LIR-tacaa(5)/RIR-tttgta(6), R1 as LIR-tacaaa(6)/RIR-ttgta(5), and R2 as LIR-
242 tacaa(5)/RIR-ttgta(5). Previous work on xerCD/ dif site system has reported a less efficient recombination
243 when there is a change in number of bps in the spacer region (41). With our model, C1 has the most
244 efficient spacer sequence, allowing it to switch back and forth easier. The flexibility helps it to maintain
245 an adequate proportion of both statuses of the switch. Other IR types have unequal or less bps in the
246 spacer sequences which may result in a stuck Holliday structure. Further research focusing on this spacer
247 should be done to validate the model.

248 **DISCUSSION**

249 *C. difficile* ST1 has attracted public attention as one of the hypervirulent and highly antibiotic-resistant
250 strains. However, its virulence relative to other *C. difficile* strains remains controversial. Our findings

251 suggest that the flagellar switch, specifically the number of A and T in the left and right IRs, account for
252 variability in *in vivo* virulence.

253 *C. difficile* virulence is mediated by the secretion of two glucosylating toxins TcdA and TcdB, and the binary
254 toxin CDT. In our collection, we did not find an association between toxin expression in fecal samples and
255 the degree of virulence. In fact, several studies reported that the toxin profile from stool samples is not
256 associated with severe CDI (23, 42). Toxin expression is partly regulated by the flagellar switch. We found
257 a weak correlation between the flagellar switch direction and virulence in the mouse model. The OFF
258 isolates secrete significantly more toxin when the flagellar switch inverts to ON. However, this increment
259 does not always increase the virulence at the strain level (data not shown). This result suggests that factors
260 in addition to toxin secretion contribute to virulence of *C. difficile* ST1 isolates.

261 The published flagellar switch for the R20291 ST1 strain contains a 154-bp switching region and a 21-bp
262 IR (31). The IR has a high content of A and T. Interestingly, we found four types of IRs with different
263 numbers of A and T among our clinical isolates. Community strains with different numbers of bp in the IR
264 were reported previously in ribotype 12 (43). In our collection, the most prevalent IR, C1, is present in
265 over half of the isolates. We found that the IR type C1 correlates with more severe weight loss and the
266 other IR types cause milder CDI symptoms. The difference in genotype between C1 and other IR types
267 may explain the heterogeneity in virulence of *C. difficile* ST1 (44, 45).

268 The strong association between C1 and high virulence suggest that this IR have some unique features. C1
269 has high percentage of ON status in *in vitro* and *in vivo* conditions. However, the higher proportion of ON
270 is unlikely to account for increased virulence as the highly biased ON isolates have reduced *in vivo*
271 virulence. In the mouse model, the IR type C1 has increased flexibility in switching between ON and OFF
272 which may contribute to virulence. In some mice, C1 isolates switch toward more OFF cells during the
273 recovery phase of CDI. On the other hand, the extremely biased ON isolates maintain the same percentage

274 over time. These results suggest that the capacity to switch between ON and OFF and maintain adequate
275 proportions of ON and OFF cells and may increase *C. difficile* virulence. Flagella production is a trade-off.
276 Other bacteria like *Pseudomonas aeruginosa* and *Vibrio cholera* use different strategies to switch between
277 sessile and motile phenotypes to reduce the metabolic cost of flagellar expression (46-48). In fact, in more
278 than 1359 ST1 isolates worldwide, more than 95% have type C1, a flexible type of switch. In *C. difficile*
279 ST1, ability to turn on and off the flagellar switch may reduce fitness costs and thereby impact virulence.

280 It is interesting that C2-OFF is the second most common type whereas it has a strong association with
281 reduced virulence. More specifically, the percentage of C2-OFF is higher in the U.S comparing to other
282 areas, potentially resulting from higher rates of testing and *C. difficile* cultivation in the U.S. In our
283 collection and in the CDC HAI-Seq *C. difficile* collection, C2-OFF accounts for approximately 30% of all
284 isolates. C2-OFF isolates are mostly non-flagellated. In the mouse model, C2-OFF remains mostly OFF over
285 the course of eight observation days. Flagellar synthesis and motility is costly (49). In addition, flagella are
286 targets for mucosal innate immune response. Saving the metabolic cost and evading the immune system
287 might be a strategy to increase persistence of *C. difficile*.

288 Our study correlates the flagellar switch and IR with the virulence of ST1 *C. difficile* clinical isolates. We
289 report here different types of IR in *C. difficile* ST1 and their association with disease severity in the mouse
290 model of CDI and propose a model to explain different ratios of ON and OFF status between IR types.
291 Further research on the flagellar switch regions of clinical ST1 and non-ST1 strains may provide insights
292 into the wide range of disease severities in patients infected with *C. difficile*.

293

294 **METHODS**

295 **Sample collection**

296 Clinical *C. difficile* isolates were collected from 2013 to 2017 at Memorial Sloan Kettering Cancer Center
297 (MSKCC) from patients receiving bone marrow transplants and cancer chemotherapy (50). *C. difficile*
298 isolates were classified using multilocus sequence typing (MLST) method (51). The isolates were
299 sequenced whole genome in an Illumina Hiseq platform (52) and circularized by Oxford Nanopore
300 Technologies.

301 **Bacterial growth and spore collection**

302 Frozen stock of *C. difficile* was streak on Brain heart infusion (BHI) agar plus 0.1% (w/v) sodium
303 taurocholate hydrate (Cat. 86339, Sigma-Aldrich). A single colony was picked and subcultured in BHI
304 medium (BD, 237500) with yeast extract and 0.1% L-Cysteine (BHIS) at 37°C in anaerobic chamber (Coy
305 lab).

306 To collect spore for mouse infection, *C. difficile* was either incubated in BHIS broth for approximately two
307 months or on 70:30 agar for 4-5 days to encourage sporulation (53). Next, spore is separated from cell
308 debris by gradient centrifuge with 20% (wt/vol) HistoDenz (D2158, Sigma-Aldrich) or 50%
309 (wt/vol) sucrose, then wash five times in sterile water. Spore solution was further incubated at 65°C for
310 20 minutes to omit vegetative cells. Spore purity is confirmed by plating on BHIS plate, checking for the
311 absence of vegetative cell growth.

312 **Mouse experiment**

313 C57BL/6 female mice from six- to eight-week-old were purchased from Jackson laboratory, room MP14.
314 Mice were randomized and administered *ad libitum* with a cocktail of metronidazole (0.25g/l), neomycin
315 (0.25g/l), and vancomycin (0.25g/l) for three days then rest for two days before being intraperitoneally
316 injected with 200µg clindamycin. After 24 hours, mice were oral gavage with approximately 200 spores of
317 *C. difficile*. All mice infected with *C. difficile* were single housed then monitored for weight loss and clinical
318 sign of disease for seven days.

319 CFU/ml counting, and toxin titer measurement was done from fecal pellet collected one day after
320 infection. Toxin assay was performed as described elsewhere (24).

321 **RNA sequencing**

322 Mice were infected with spore from four selected isolates (ST1-6, ST1-12, ST1-27, and ST1-53). After 24
323 hours, mice were sacrificed to collect cecal content for RNA extraction (Qiagen RNeasy PowerMicrobiome
324 kit). Total RNA was sent to Genewiz for rRNA removal and library preparation before sequencing. Adapters
325 were trimmed off from the raw reads, and their quality was assessed and controlled using Trimmomatic
326 (v.0.39) (54), then human genome was identified and removed by kneaddata (v0.7.10,
327 https://github.com/biobakery/kneaddata), while ribosomal RNA was removed by aligned the clean non-
328 host reads to silva database (138.1 SSURef). The remaining reads from each sample were mapped to their
329 corresponding circularized genome using bowtie2(55), and reads counts of each gene were obtained by
330 running featureCounts from Subread (v2.0.1) (56) ,and the core gene counts were normalized by DESeq2
331 (57).

332 **Swim plates**

333 *C. difficile* was grown in BHIS broth to late log phase then diluted to the same OD_{600nm} ~0.5. A 10µl tip is
334 immerged into diluted *C. difficile* culture and then stabbed into 0.3% agarose buffered with BHIS, pH 7.0.
335 The swim plates were then incubated at 37°C. After 24 hours, the colony size was measured and compared
336 between isolates.

337 **Transmission electron microscope (TEM)**

338 *C. difficile* was streak on BHIS agar plate and incubate overnight. To collect cells for TEM, 10µl of distilled
339 water was drop onto a colony. After two minutes, a copper grid (CF400-Cu, EMS) was put on top of the
340 soaked colony so that vegetative cells were passively transferred to the grid. To stain the cells, one drop

341 of uranyl acetate 1% was added to the grid and incubate for 30 seconds. TEM pictures of the stained cells
342 were then taken at 2900X1.4 or 5900X1.4 using FEI Tecnai F30 microscope at University of Chicago
343 Advanced Electron Microscopy Facility.

344 **Quantitative PCR of the flagellar switch**

345 *C. difficile* was grown in BHIS broth to late log phase (OD_{600nm} ~ 0.5) with three to four replicates per
346 isolate. *C. difficile* DNA was extracted by Qiamp PowerFecal Pro DNA kit (Cat. 51804). Quantitative real
347 time PCR was done using 20ng template and flagellar switch primers (31). *C. difficile* R20291 mutated
348 strains with flagellar switch locked ON and OFF were used as the control (40). Adenosine kinase (*adk)* gene
349 is used as a reference gene. qPCR was set up using PowerUp™ SYBR™ Green Master Mix (Cat. A25742)
350 and run by QuantStudio Real-Time PCR Systems. Percentage of the switch direction was calculated using
351 the *ΔCt* method.

352 **Flagella switch alignment and IR type survey in public databases**

353 *C.difficile* isolates (N=131) from BioProject PRJNA595724 and PRJEB2318 (N=1198) were downloaded
354 from NCBI and assembled into contigs using SPAdes (58). Eleven of those isolates were lost during the
355 downloading and assembling process. In addition, a collection of 66 *C.difficile* genomes from Patric (date:
356 Mar. 15 2022) { https://www.bv-brc.org/} were also downloaded. MLST was determined on those contigs
357 by mlst {Seemann T, mlst Github https://github.com/tseemann/mlst}. Flagellar switch region and 50 bp
358 upstream and downstream of 68 isolates in our collection of both the ON and OFF sequences were used
359 as query to BLAST (59) against the assembled contigs, and hits with at least 85% identity and 85% coverage
360 of the query are considered a valid match.

361 **RNA extraction, reverse transcription and RT-qPCR**

362 Cecal RNA was extracted using Rneasy PowerMicrobiome Kit (Qiagen) according to the manufacturer's
363 instructions. Complementary DNA was generated using the QuantiTect reverse transcriptase kit (Qiagen)
364 according to the manufacturer's instructions. Quantitative PCR was performed on complementary DNA
365 using primers with PowerTrack SYBR Green Master Mix (Thermo Fisher). Reactions were run on a
366 QuantStudio 6 pro (Thermo Fisher). Relative abundance was normalized by ΔΔCt. TcdA_qFor 5'-
367 GTATGGATAGGTGGAGAAGTCA-3'; TcdA_qRev 5'-CTCTTCCTCTAGTAGCTGTAATGC-3' {Babakhani 2012 J
368 Antimicrobial Chemotherapy}; TcdB_qFor 5'-AGCAGTTGAATATAGTGGTTTAGTTAGAGTTG-3'; TcdB_qRev
369 5'-CATGCTTTTTTAGTTTCTGGATTGAA-3' {Wroblewski 2009 J Clin Microbiol.}; FliS1_qFor 5'-
370 TGCAGGACAATGGGCAAAGG-3'; FliS1_qRev 5'- CAGGCAACACATTATCTATTACCTGG-3'; FliE_qFor 5'-
371 AGGCGAAGATGTTTCTATGCA-3'; FliE_qRev 5'- ACCTTATTCATTTCTTGATATGCATCA-3'

372 **Structure prediction and conserved sequence analysis**

373 The RecV tetramer structure was predicted by Alphafold-multimer using a web-based Colab
374 implementation, and default parameters (60-62). Sequences of the inverted repeats flanking the switch
375 regions were located in the *Clostridium difficile* R20291 genome (accession number CP029423) using
376 Snapgene (https://www.snapgene.com/). After manual alignment, the sequence logo shown in Fig. 7B
377 was created using https://weblogo.berkeley.edu/logo.cgi (63) An approximation of the proposed
378 consensus sequence for binding of a RecV dimer was based on that logo
379 (AWAGTWNCCNTTWNNNNNNWAANGGNWACTWT) and was tested by searching the full genome in
380 Snapgene for related sequences with 1 mismatch allowed per 10bp.

381 **Statistical analysis**

382 Kruskal-Wallis or Two Sample T tests were performed to test the difference in maximum weight loss, swim
383 plate and ΔCt results. Statistical significance was determined by using a P value of <0.05. Linear regression
384 analysis was used to estimate the correlation between IR type and weight loss.

385 **ACKNOWLEDGEMENT**

386 We would like to thank Dr. Rita Tamayo for providing the primers and reference strains for the flagellar
387 switch qPCR. This work was supported by the National Institutes of Health R01 AI095706 and by the
388 Duchossois Family Institute of the University of Chicago.. The funders had no role in study design, data
389 collection and interpretation, or the decision to submit the work for publication.

390 **AUTHOR CONTRIBUTION**

391 N.T.Q.N., Q.D. and E.G.P. conceived the project. Q.D., N.T.Q.N., H.L. and Q.D. analyzed the data. N.T.Q.N.,
392 Q.D., and J.K.S., performed experiments. M.K. isolated *C. difficile* isolates. V.B.Y. and E.S.S. sequenced
393 clinical isolates. N.T.Q.N., H.L., Q.D., P.A.R. and E.G.P. interpreted the results and wrote the manuscript.

394

395 **REFERENCES**

396

397 1. CDC. 2019. Antibiotic Resistance Threats in the United States Atlanta GUSDoHaHS, CDC; 2019.,
398 2. CDC. 2020. 2020 Annual Report.
399 3. Lyerly DM, Krivan HC, Wilkins TD. 1988. Clostridium difficile: its disease and toxins. Clin Microbiol
400 Rev 1:1-18.
401 4. Ciesielski-Treska J, Ulrich G, Rihn B, Aunis D. 1989. Mechanism of action of Clostridium difficile
402 toxin B: role of external medium and cytoskeletal organization in intoxicated cells. Eur J Cell Biol
403 48:191-202.
404 5. Florin I, Thelestam M. 1983. Internalization of Clostridium difficile cytotoxin into cultured human
405 lung fibroblasts. Biochim Biophys Acta 763:383-92.
406 6. Henriques B, Florin I, Thelestam M. 1987. Cellular internalisation of Clostridium difficile toxin A.
407 Microb Pathog 2:455-63.
408 7. Just I, Selzer J, Hofmann F, Green GA, Aktories K. 1996. Inactivation of Ras by Clostridium sordellii
409 lethal toxin-catalyzed glucosylation. J Biol Chem 271:10149-53.
410 8. Just I, Selzer J, Wilm M, von Eichel-Streiber C, Mann M, Aktories K. 1995. Glucosylation of Rho
411 proteins by Clostridium difficile toxin B. Nature 375:500-3.
412 9. Just I, Wilm M, Selzer J, Rex G, von Eichel-Streiber C, Mann M, Aktories K. 1995. The enterotoxin
413 from Clostridium difficile (ToxA) monoglucosylates the Rho proteins. J Biol Chem 270:13932-6.
414 10. Just I, Fritz G, Aktories K, Giry M, Popoff MR, Boquet P, Hegenbarth S, von Eichel-Streiber C. 1994.
415 Clostridium difficile toxin B acts on the GTP-binding protein Rho. J Biol Chem 269:10706-12.
416 11. Schnizlein MK, Young VB. 2022. Capturing the environment of the Clostridioides difficile infection
417 cycle. Nat Rev Gastroenterol Hepatol 19:508-520.

418 12. Vedantam G, Clark A, Chu M, McQuade R, Mallozzi M, Viswanathan VK. 2012. Clostridium difficile
419 infection: toxins and non-toxin virulence factors, and their contributions to disease establishment
420 and host response. Gut Microbes 3:121-34.
421 13. Lesniak NA, Schubert AM, Flynn KJ, Leslie JL, Sinani H, Bergin IL, Young VB, Schloss PD. 2022. The
422 Gut Bacterial Community Potentiates Clostridioides difficile Infection Severity. mBio
423 13:e0118322.
424 14. Loo VG, Bourgault AM, Poirier L, Lamothe F, Michaud S, Turgeon N, Toye B, Beaudoin A, Frost EH,
425 Gilca R, Brassard P, Dendukuri N, Beliveau C, Oughton M, Brukner I, Dascal A. 2011. Host and
426 pathogen factors for Clostridium difficile infection and colonization. N Engl J Med 365:1693-703.
427 15. Tenover FC, Tickler IA, Persing DH. 2012. Antimicrobial-resistant strains of Clostridium difficile
428 from North America. Antimicrob Agents Chemother 56:2929-32.
429 16. Wieczorkiewicz JT, Lopansri BK, Cheknis A, Osmolski JR, Hecht DW, Gerding DN, Johnson S. 2016.
430 Fluoroquinolone and Macrolide Exposure Predict Clostridium difficile Infection with the Highly
431 Fluoroquinolone- and Macrolide-Resistant Epidemic C. difficile Strain BI/NAP1/027. Antimicrob
432 Agents Chemother 60:418-23.
433 17. Hubert B, Loo VG, Bourgault AM, Poirier L, Dascal A, Fortin E, Dionne M, Lorange M. 2007. A
434 portrait of the geographic dissemination of the Clostridium difficile North American pulsed-field
435 type 1 strain and the epidemiology of C. difficile-associated disease in Quebec. Clin Infect Dis
436 44:238-44.
437 18. Labbe AC, Poirier L, Maccannell D, Louie T, Savoie M, Beliveau C, Laverdiere M, Pepin J. 2008.
438 Clostridium difficile infections in a Canadian tertiary care hospital before and during a regional
439 epidemic associated with the BI/NAP1/027 strain. Antimicrob Agents Chemother 52:3180-7.
440 19. Morgan OW, Rodrigues B, Elston T, Verlander NQ, Brown DF, Brazier J, Reacher M. 2008. Clinical
441 severity of Clostridium difficile PCR ribotype 027: a case-case study. PLoS One 3:e1812.
442 20. Sirard S, Valiquette L, Fortier LC. 2011. Lack of association between clinical outcome of Clostridium
443 difficile infections, strain type, and virulence-associated phenotypes. J Clin Microbiol 49:4040-6.
444 21. Valiente E, Cairns MD, Wren BW. 2014. The Clostridium difficile PCR ribotype 027 lineage: a
445 pathogen on the move. Clin Microbiol Infect 20:396-404.
446 22. Warny M, Pepin J, Fang A, Killgore G, Thompson A, Brazier J, Frost E, McDonald LC. 2005. Toxin
447 production by an emerging strain of Clostridium difficile associated with outbreaks of severe
448 disease in North America and Europe. Lancet 366:1079-84.
449 23. Carlson PE, Jr., Walk ST, Bourgis AE, Liu MW, Kopliku F, Lo E, Young VB, Aronoff DM, Hanna PC.
450 2013. The relationship between phenotype, ribotype, and clinical disease in human Clostridium
451 difficile isolates. Anaerobe 24:109-16.
452 24. Qiwen Dong HL, Marie-Maude Allen, Julian R Garneau, Jonathan K Sia, Rita C Smith, Fidel Haro,
453 Tracy McMillen, Rosemary L Pope, Carolyn Metcalfe, Victoria Burgo, Che Woodson, Nicholas
454 Dylla, Claire Kohout, Anitha Sundararajan, Evan S Snitkin, Vincent B Young, Louis-Charles Fortier,
455 Mini Kamboj, Eric G Pamer. 2023. Virulence and genomic diversity among clinical isolates of ST1
456 (BI/NAP1/027) Clostridioides difficile. bioRxiv doi:doi: 10.1101/2023.01.12.523823.
457 25. La Ragione RM, Sayers AR, Woodward MJ. 2000. The role of fimbriae and flagella in the
458 colonization, invasion and persistence of Escherichia coli O78:K80 in the day-old-chick model.
459 Epidemiol Infect 124:351-63.
460 26. Eaves-Pyles T, Murthy K, Liaudet L, Virag L, Ross G, Soriano FG, Szabo C, Salzman AL. 2001.
461 Flagellin, a novel mediator of Salmonella-induced epithelial activation and systemic inflammation:
462 I kappa B alpha degradation, induction of nitric oxide synthase, induction of proinflammatory
463 mediators, and cardiovascular dysfunction. J Immunol 166:1248-60.
464 27. Horstmann JA, Lunelli M, Cazzola H, Heidemann J, Kuhne C, Steffen P, Szefs S, Rossi C, Lokareddy
465 RK, Wang C, Lemaire L, Hughes KT, Uetrecht C, Schluter H, Grassl GA, Stradal TEB, Rossez Y, Kolbe

466 M, Erhardt M. 2020. Methylation of Salmonella Typhimurium flagella promotes bacterial adhesion
467 and host cell invasion. Nat Commun 11:2013.
468 28. Baban ST, Kuehne SA, Barketi-Klai A, Cartman ST, Kelly ML, Hardie KR, Kansau I, Collignon A,
469 Minton NP. 2013. The role of flagella in Clostridium difficile pathogenesis: comparison between a
470 non-epidemic and an epidemic strain. PLoS One 8:e73026.
471 29. Batah J, Kobeissy H, Bui Pham PT, Deneve-Larrazet C, Kuehne S, Collignon A, Janoir-Jouveshomme
472 C, Marvaud JC, Kansau I. 2017. Clostridium difficile flagella induce a pro-inflammatory response
473 in intestinal epithelium of mice in cooperation with toxins. Sci Rep 7:3256.
474 30. Dingle TC, Mulvey GL, Armstrong GD. 2011. Mutagenic analysis of the Clostridium difficile flagellar
475 proteins, FliC and FliD, and their contribution to virulence in hamsters. Infect Immun 79:4061-7.
476 31. Anjuwon-Foster BR, Tamayo R. 2017. A genetic switch controls the production of flagella and
477 toxins in Clostridium difficile. PLoS Genet 13:e1006701.
478 32. Paulick A, Adamczyk M, Anderson K, Vlachos N, Machado MJ, McAllister G, Korhonen L, Guh AY,
479 Halpin AL, Rasheed JK, Karlsson M, Lutgring JD, Gargis AS, Emerging Infections Program
480 Clostridioides difficile Pathogen G. 2021. Characterization of Clostridioides difficile Isolates
481 Available through the CDC & FDA Antibiotic Resistance Isolate Bank. Microbiol Resour Announc
482 10.
483 33. Dingle KE, Didelot X, Quan TP, Eyre DW, Stoesser N, Golubchik T, Harding RM, Wilson DJ, Griffiths
484 D, Vaughan A, Finney JM, Wyllie DH, Oakley SJ, Fawley WN, Freeman J, Morris K, Martin J, Howard
485 P, Gorbach S, Goldstein EJC, Citron DM, Hopkins S, Hope R, Johnson AP, Wilcox MH, Peto TEA,
486 Walker AS, Crook DW, Modernising Medical Microbiology Informatics G. 2017. Effects of control
487 interventions on Clostridium difficile infection in England: an observational study. Lancet Infect
488 Dis 17:411-421.
489 34. Eyre DW, Davies KA, Davis G, Fawley WN, Dingle KE, De Maio N, Karas A, Crook DW, Peto TEA,
490 Walker AS, Wilcox MH, Group ES. 2018. Two Distinct Patterns of Clostridium difficile Diversity
491 Across Europe Indicating Contrasting Routes of Spread. Clin Infect Dis 67:1035-1044.
492 35. Eyre DW, Didelot X, Buckley AM, Freeman J, Moura IB, Crook DW, Peto TEA, Walker AS, Wilcox
493 MH, Dingle KE. 2019. Clostridium difficile trehalose metabolism variants are common and not
494 associated with adverse patient outcomes when variably present in the same lineage.
495 EBioMedicine 43:347-355.
496 36. Stoesser N, Crook DW, Fung R, Griffiths D, Harding RM, Kachrimanidou M, Keshav S, Peto TE,
497 Vaughan A, Walker AS, Dingle KE. 2011. Molecular epidemiology of Clostridium difficile strains in
498 children compared with that of strains circulating in adults with Clostridium difficile-associated
499 infection. J Clin Microbiol 49:3994-6.
500 37. Sekulovic O, Mathias Garrett E, Bourgeois J, Tamayo R, Shen A, Camilli A. 2018. Genome-wide
501 detection of conservative site-specific recombination in bacteria. PLoS Genet 14:e1007332.
502 38. Grindley ND, Whiteson KL, Rice PA. 2006. Mechanisms of site-specific recombination. Annu Rev
503 Biochem 75:567-605.
504 39. Meinke G, Bohm A, Hauber J, Pisabarro MT, Buchholz F. 2016. Cre Recombinase and Other
505 Tyrosine Recombinases. Chem Rev 116:12785-12820.
506 40. Trzilova D, Warren MAH, Gadda NC, Williams CL, Tamayo R. 2022. Flagellum and toxin phase
507 variation impacts intestinal colonization and disease development in a mouse model of
508 Clostridioides difficile infection. Gut Microbes 14:2038854.
509 41. Blakely G, Sherratt D. 1996. Determinants of selectivity in Xer site-specific recombination. Genes
510 Dev 10:762-73.
511 42. Rao K, Micic D, Natarajan M, Winters S, Kiel MJ, Walk ST, Santhosh K, Mogle JA, Galecki AT, LeBar
512 W, Higgins PD, Young VB, Aronoff DM. 2015. Clostridium difficile ribotype 027: relationship to age,

513 detectability of toxins A or B in stool with rapid testing, severe infection, and mortality. Clin Infect
514 Dis 61:233-41.
515 43. Anjuwon-Foster BR, Maldonado-Vazquez N, Tamayo R. 2018. Characterization of Flagellum and
516 Toxin Phase Variation in Clostridioides difficile Ribotype 012 Isolates. J Bacteriol 200.
517 44. Gu T, Li W, Yang LL, Yang SM, He Q, He HY, Sun DL. 2022. Systematic review of guidelines for the
518 diagnosis and treatment of Clostridioides difficile infection. Front Cell Infect Microbiol 12:926482.
519 45. Johnson S, Lavergne V, Skinner AM, Gonzales-Luna AJ, Garey KW, Kelly CP, Wilcox MH. 2021.
520 Clinical Practice Guideline by the Infectious Diseases Society of America (IDSA) and Society for
521 Healthcare Epidemiology of America (SHEA): 2021 Focused Update Guidelines on Management of
522 Clostridioides difficile Infection in Adults. Clin Infect Dis 73:755-757.
523 46. Fernandez NL, Hsueh BY, Nhu NTQ, Franklin JL, Dufour YS, Waters CM. 2020. Vibrio cholerae
524 adapts to sessile and motile lifestyles by cyclic di-GMP regulation of cell shape. Proc Natl Acad Sci
525 U S A 117:29046-29054.
526 47. Lichtenberg M, Kragh KN, Fritz B, Kirkegaard JB, Tolker-Nielsen T, Bjarnsholt T. 2022. Cyclic-di-
527 GMP signaling controls metabolic activity in Pseudomonas aeruginosa. Cell Rep 41:111515.
528 48. Nadell CD, Bassler BL. 2011. A fitness trade-off between local competition and dispersal in Vibrio
529 cholerae biofilms. Proc Natl Acad Sci U S A 108:14181-5.
530 49. Schavemaker PE, Lynch M. 2022. Flagellar energy costs across the tree of life. Elife 11.
531 50. Kamboj M, McMillen T, Syed M, Chow HY, Jani K, Aslam A, Brite J, Fanelli B, Hasan NA, Dadlani M,
532 Westblade L, Zehir A, Simon M, Babady NE. 2021. Evaluation of a Combined Multilocus Sequence
533 Typing and Whole-Genome Sequencing Two-Step Algorithm for Routine Typing of Clostridioides
534 difficile. J Clin Microbiol 59.
535 51. Griffiths D, Fawley W, Kachrimanidou M, Bowden R, Crook DW, Fung R, Golubchik T, Harding RM,
536 Jeffery KJ, Jolley KA, Kirton R, Peto TE, Rees G, Stoesser N, Vaughan A, Walker AS, Young BC, Wilcox
537 M, Dingle KE. 2010. Multilocus sequence typing of Clostridium difficile. J Clin Microbiol 48:770-8.
538 52. Miles-Jay A, Young VB, Pamer EG, Savidge TC, Kamboj M, Garey KW, Snitkin ES. 2021. A multisite
539 genomic epidemiology study of Clostridioides difficile infections in the USA supports differential
540 roles of healthcare versus community spread for two common strains. Microb Genom 7.
541 53. Weldy M, Evert C, Dosa PI, Khoruts A, Sadowsky MJ. 2020. Convenient Protocol for Production
542 and Purification of Clostridioides difficile Spores for Germination Studies. STAR Protoc 1:100071.
543 54. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data.
544 Bioinformatics 30:2114-20.
545 55. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-
546 9.
547 56. Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning
548 sequence reads to genomic features. Bioinformatics 30:923-30.
549 57. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-
550 seq data with DESeq2. Genome Biol 15:550.
551 58. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes De Novo
552 Assembler. Curr Protoc Bioinformatics 70:e102.
553 59. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+:
554 architecture and applications. BMC Bioinformatics 10:421.
555 60. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R,
556 Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B,
557 Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M,
558 Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli
559 P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596:583-
560 589.

561 61. Mirdita M, Schutze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making
562 protein folding accessible to all. Nat Methods 19:679-682.
563 62. Richard Evans PON, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin
564 Žídek, Russ Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal
565 Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain,
566 Ellen Clancy, Pushmeet Kohli, John Jumper, Demis Hassabis. 2022. Protein complex prediction
567 with AlphaFold-Multimer. bioRxiv doi:https://doi.org/10.1101/2021.10.04.463034.
568 63. Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator.
569 Genome Res 14:1188-90.

570

571 **FIGURE LEGENDS**

572 Figure 1. Highly virulent isolates have more flagella. A) Heat map of RNA sequencing results from four
573 ST1 isolates. The heat map is generated based on the number of transcripts per sample after
574 normalization. RNA was extracted from cecal content of mice after one day infected with *C. difficile* ST1
575 isolates (n= 5 mice per isolate). B) TEM pictures of four selected isolates. C) Swim plates results of four
576 isolates after 24 hours incubation (n = 3 replicates per isolate). Highly virulent isolates: pink. Less
577 virulent isolate: black

578 Figure 2. The role of flagellar switch in *C. difficile* ST1 virulence. A) The swimming cells (E strains)
579 collected from less virulent isolates have flagella and B) higher toxin production in fecal samples. C)
580 Summary of the proportion of ON and OFF of each isolates and its correlation with the maximum weight
581 loss. Data is generated from the isolates shown in Figure 1.

582 Figure 3. The flagellar-switch IR type has a strong association with the virulence. A) Flagellar switch
583 alignment of some representative isolates with the reference sequence of *C. difficile* R20291. B)
584 Comparison between IR type C1 and other non-C1 IR type regarding the maximum weight loss. Data is
585 generated based on the collection of isolates shown in Figure 1.

586 Figure 4. Flagellar-switch IR type C1 is more flexible than other IR types. A) The plot summarizes the ΔCt
587 of all IR types. ΔCt is calculated from flagellar-switch qPCR results. B) The percentage of ON genomes per
588 IR type. For *in vitro* condition, the qPCR was performed using genomic DNA extracted from BHIS culture
589 (each dot represents for one isolate, n = 3 to 4 replicates per isolate). For *in vivo* condition, genomic DNA
590 was extracted from day-one-post-infection fecal pellet (each dot represents for one mouse, n = 4 to 10
591 mice per isolate). *In vivo* results for IR type C1 are generated from R20291, ST1-12, and ST1-53 (18 mice),
592 C2-ON from ST1-19 (4 mice), C2-OFF from ST1-6, and ST1-27 (8 mice), R1-ON from ST1-51 (4 mice), R1-
593 OFF from ST1-13, ST1-68 (8 mice), and R2 from ST1-67 (4 mice).

594 Figure 5. Flagellar-switch IR type C1 has larger variation in terms of the percentage of ON genomes. Data
595 was generated from qPCR on DNA extracted from fecal pellets collected over the course of seven days.
596 Day-eighth data was from DNA extracted from cecal content. Each dot represents for one mouse.

597 Figure 6. The diversity of IR types is mostly seen in the U.S. A) The contribution of each IR type across
598 geographical areas. B) The contribution of each IR type in the U.S across years as comparing with our
599 collection (MSKCC) and C) the recent CDC HAI *C. difficile* collection.

600 Figure 7. The proposed model for recV binding to the flagellar switch. A) The tetramer structure of recV
601 when binding to the IRs of the flagellar switch. B) The conserved sequence in the IR of six switches
602 mediated by recV in *C. difficile* R20291. C) The proposed spacer sequence in the flagellar switch IRs
603 (underlined by arrows). We are showing here the sequence of the upper strand. D) The binding of
604 tetramer recV to the flagellar switch IRs initiates the inversion of the switch. The model is drawn and
605 licensed by BioRender.

606

607 **SUPPLEMENTAL FIGURE LEGENDS**

608 Figure S1. *C. difficile* ST1 isolates have variable virulence in antibiotic-treated mice. A) Boxplot showing
609 the CFU/g data from fecal pellet collected one day after infection (n = 2 to 13 mice per isolate). B) The
610 maximum weight loss percentage (n = 5 to 13 mice per isolate). In antibiotic-treated mice infected with
611 *C. difficile* ST1, the most severe weight loss happens at day two or day three after infection C) Toxin titer
612 from fecal pellets collected one day after infection (n = 2 to 13 mice per isolate, except for ST1-51). Each
613 black dot represents for a data point from one mouse. The red asterisk represents for an average value.
614 The dash line in panel B shows the 10% cutoff of weight loss.

615 Figure S2. Comparable toxin production and CFU between high and low virulence isolates. A) The toxin
616 titer and B) CFU/g generated from fecal content collected from mice one day post infection (n = 5 mice
617 per isolate).

618 Figure S3. RT_qPCR confirms the difference in flagellar gene expression between highly and less virulent
619 isolates. Fold change of flagellar genes fliE A), fliS1 B), and toxin genes tcdA C), and tcdB D) to adk to
620 ST1-6 (n=3 to 4 RT-qPCR for each isolate).

621

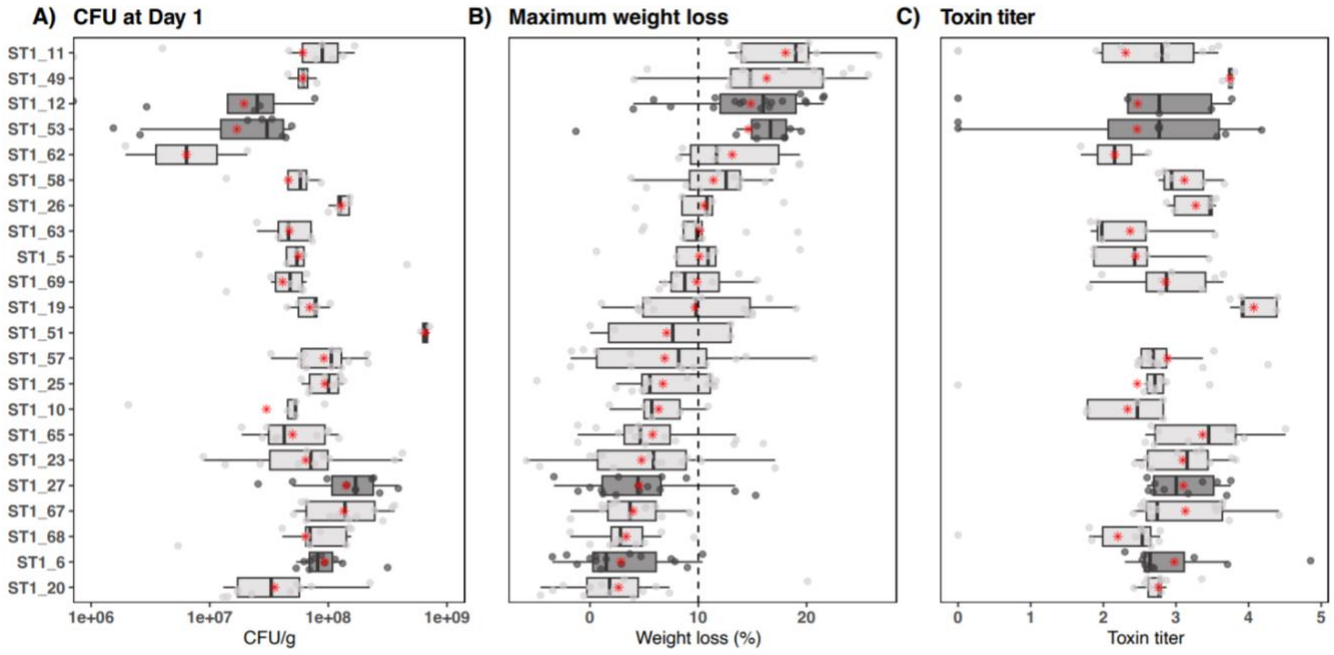

Figure S1. *C. difficile* ST1 isolates have variable virulence in antibiotic-treated mice.

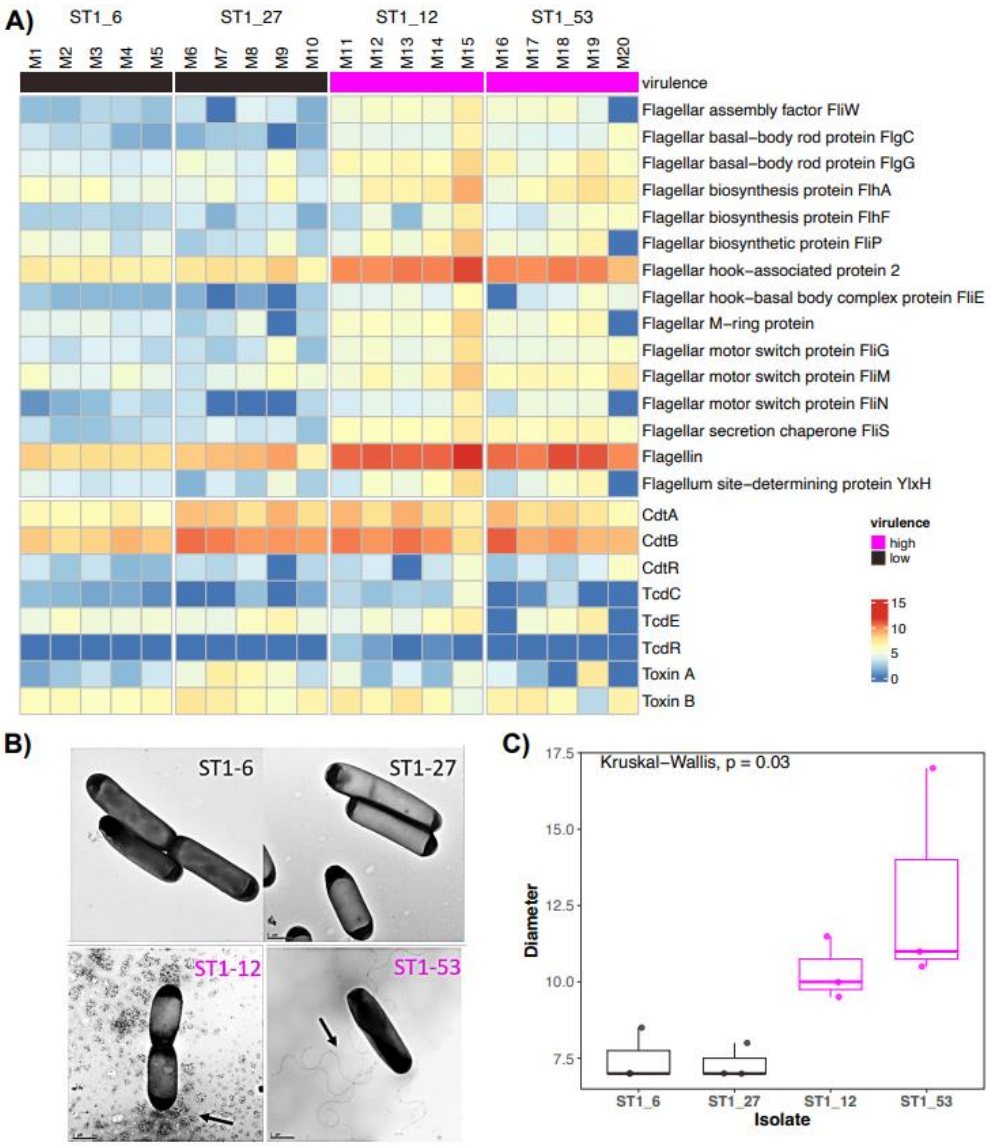

Figure 1. Highly virulent isolates have more flagella.

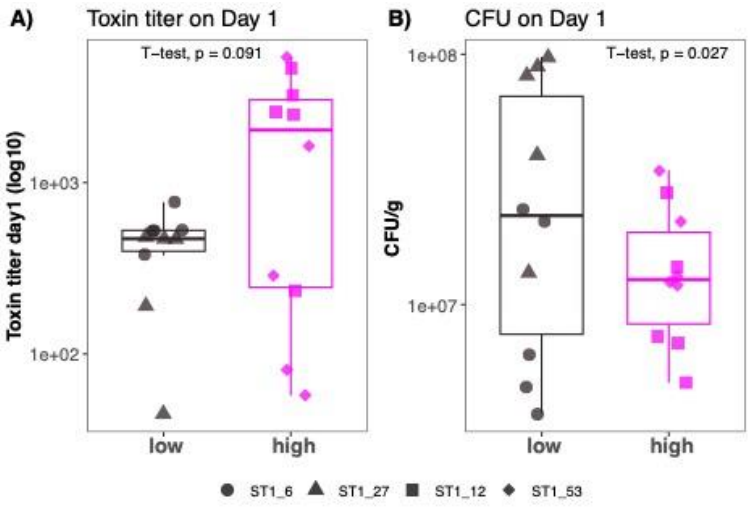

Figure S2. Comparable toxin production and CFU between high and low virulence isolates.

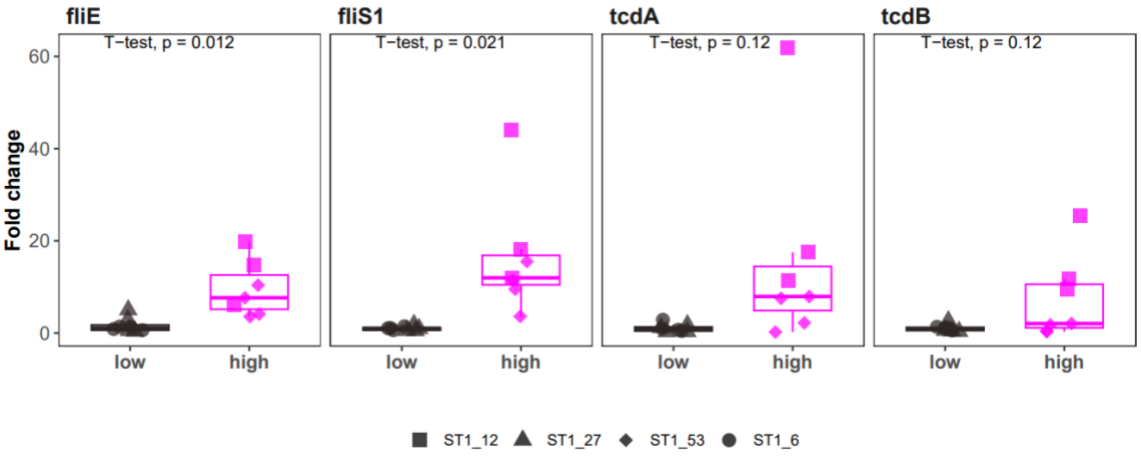

Figure S3. RT_qPCR confirms the difference in flagellar gene expression between highly and less virulent isolates.

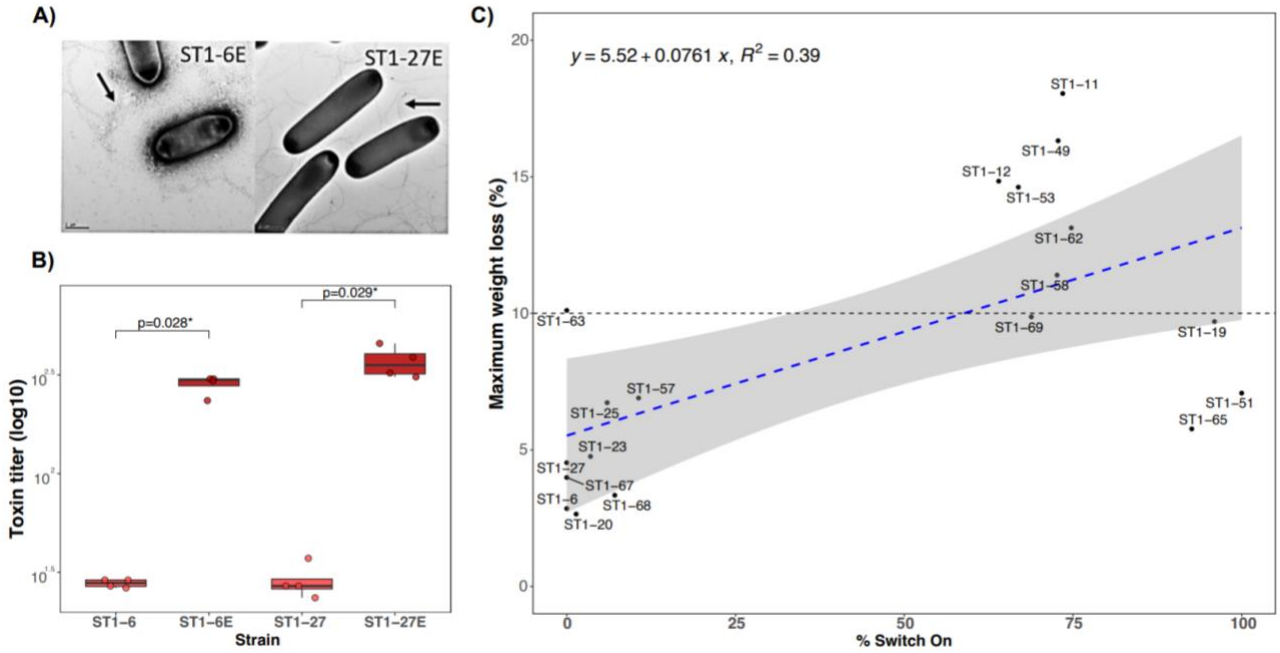

Figure 2. The role of flagellar switch in *C. difficile* ST1 virulence.

A)

Isolate	Virulence	Type	Left inverted repeat	Central region	Right inverted repeat
R20291	High	C1	AAGTTACTATTTTACAAAAAA	AGGCAA…….AAGTCA	TTTTTTGTAATATAGCAACTT
ST1-12	High	C1	AAGTTACTATTTTACAAAAAA	AGGCAA…….AAGTCA	TTTTTTGTAATATAGCAACTT
ST1-53	High	C1	AAGTTACTATTTTACAAAAAA	AGGCAA…….AAGTCA	TTTTTTGTAATATAGCAACTT
ST1-6	Low	C2-OFF	AAGTTACTATTTTAC AAAAA	TGACTT…….. TTGCCT	TTTTTTGTAATATAGCAACTT
ST1-27	Low	C2-OFF	AAGTTACTATTTTAC AAAAA	TGACTT…….. TTGCCT	TTTTTTGTAATATAGCAACTT
ST1-19	Intermediate	C2-ON	AAGTTACTATTTTACAAAAAA	AGGCAA…….AAGTCA	TTTTT GTAATATAGCAACTT
ST1_65	Low	C2-ON	AAGTTACTATTTTACAAAAAA	AGGCAA…….AAGTCA	TTTTT GTAATATAGCAACTT
ST1-68	Low	R1-OFF	AAGTTACTATTTTACAAAAAA	TGACTT…….. TTGCCT	TTTTT GTAATATAGCAACTT
ST1-51	Low	R1-ON	AAGTTACTATTTTAC AAAAA	AGGCAA…….AAGTCA	TTTTTTGTAATATAGCAACTT
ST1-67	Low	R2-OFF	AAGTTACTATTTTAC AAAAA	TGACTT…….. TTGCCT	TTTTT GTAATATAGCAACTT

B)

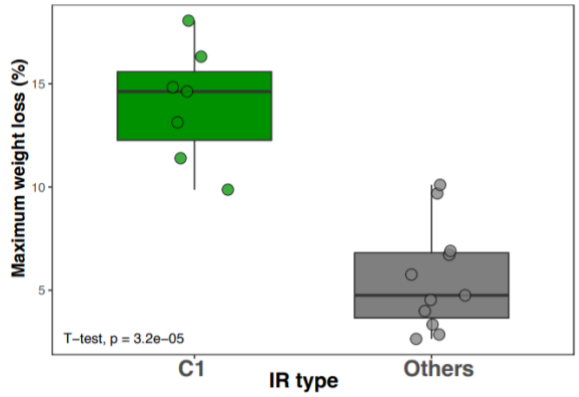

Figure 3. The flagellar-switch IR type has a strong association with the virulence.

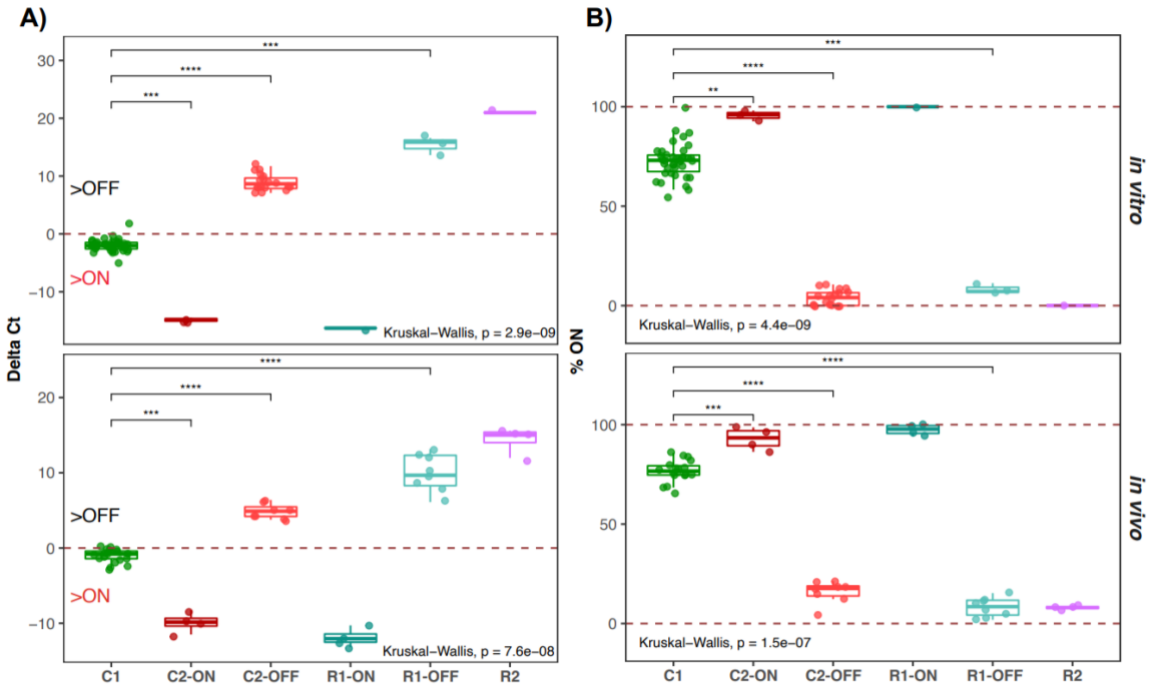

Figure 4. Flagellar-switch IR type C1 is more flexible than other IR types.

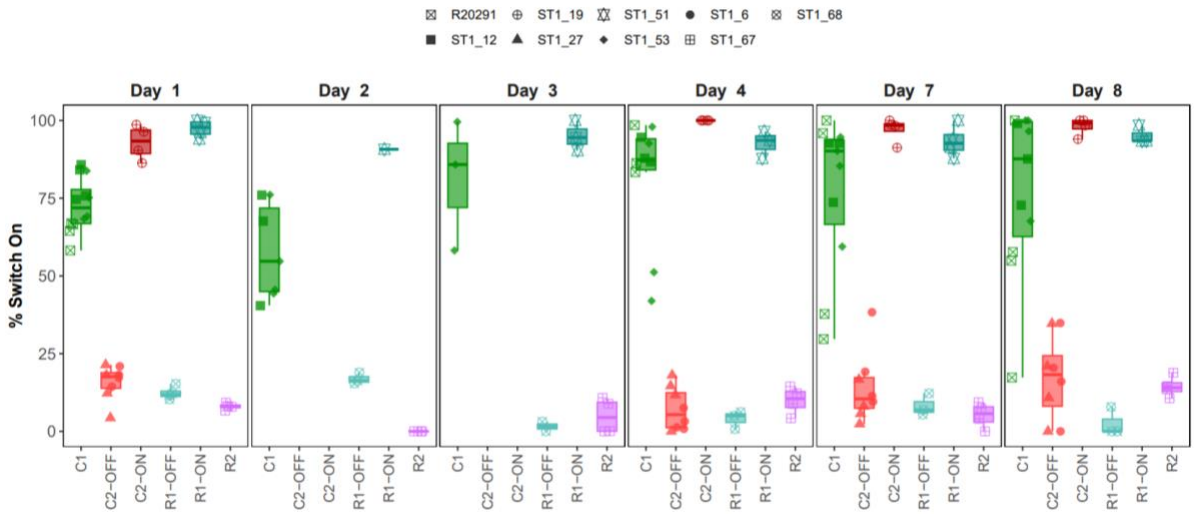

Figure 5. Flagellar-switch IR type C1 has larger variation in terms of the percentage of ON genomes.

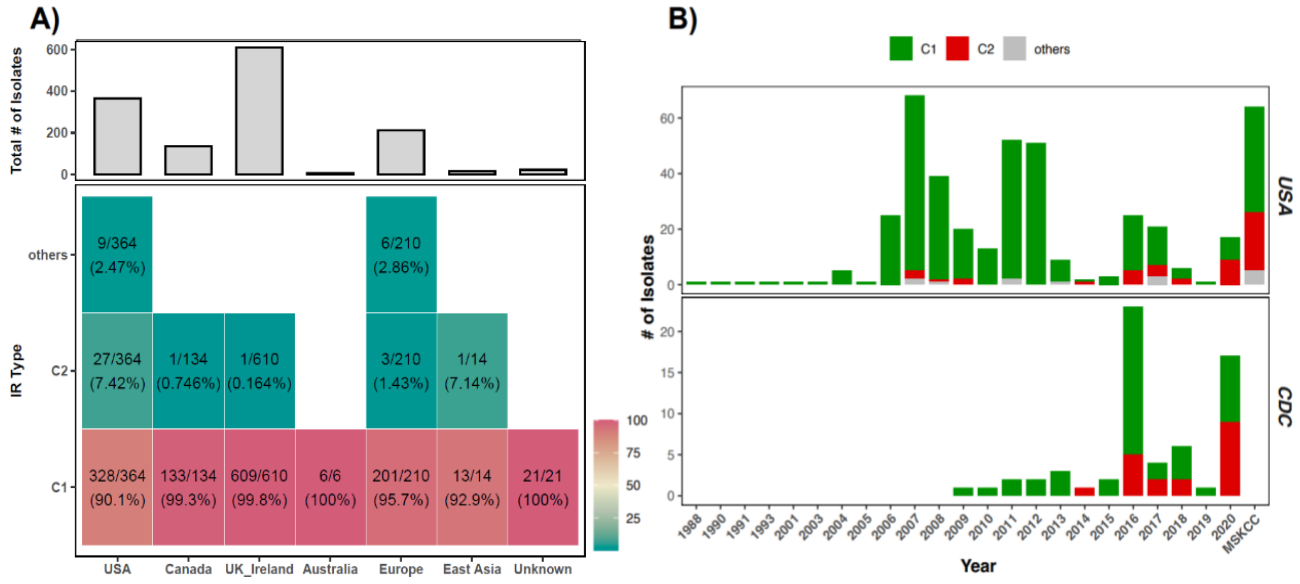

Figure 6. The diversity of IR types is mostly seen in the U.S.

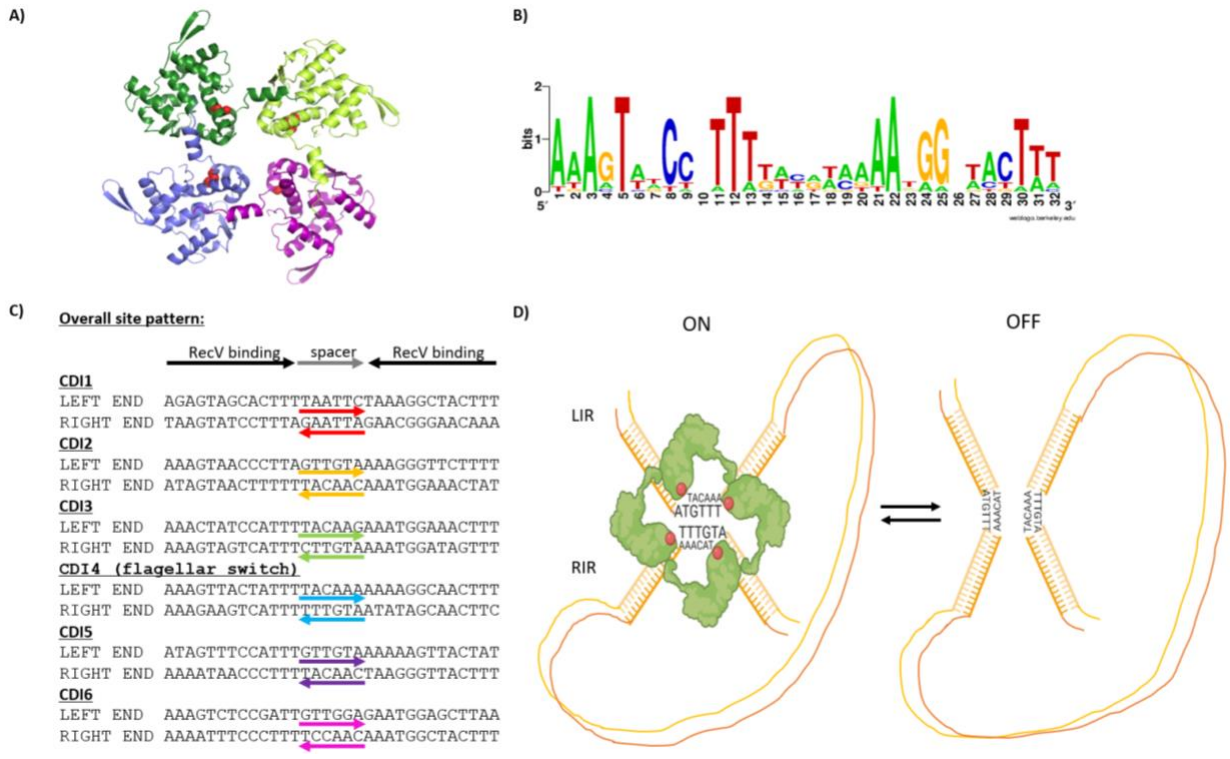

Figure 7. The proposed model for recV binding to the flagellar switch.