

Supporting information to: Prophage-stimulated toxin production in *Clostridium difficile* NAP1/027 lysogens

Ognjen Sekulovic, Mathieu Meessen-Pinard and Louis-Charles Fortier*

Département de microbiologie et d'infectiologie,

Faculté de médecine et des sciences de la santé,

Université de Sherbrooke, Québec, Canada, J1H 5N4,

* Corresponding author. Mailing address: Département de microbiologie et d'infectiologie,
Faculté de médecine et des sciences de la santé, Université de Sherbrooke, 3001 12e Ave Nord,
Sherbrooke, Québec, Canada, J1H 5N4. Phone: 819-564-5322. Fax: 819-564-5392. E-mail:

Louis-Charles.Fortier@USherbrooke.ca

Table S1. Primers used in this study

| Target | Strand | Primer | Sequence (5'- 3') | Product size (bp) | Reference |
|--------------------|--------|---------|-------------------------------------|-------------------|------------|
| <i>tcdA</i> | Fwd | LCF 398 | AGCTTTCGCTTTAGGCAGTGTTATAAATCAAGC | 139 | This study |
| | Rev | LCF 399 | TATCAGACTCTATGGCTGGGTTAAGGTG | | |
| <i>tcdB</i> | Fwd | LCF 400 | AGCATATGCGGCAGCTTATCAAGA | 158 | This study |
| | Rev | LCF 401 | GCATCGTCAAATGACCATAAGCTAGCCATT | | |
| <i>tcdC</i> | Fwd | LCF 444 | AGCTTGTTCTGAAGACCATGAGGAGG | 112 | This study |
| | Rev | LCF 445 | ACCATGGTTCAGCATCAGACAATTTGC | | |
| <i>tcdR</i> | Fwd | LCF 448 | TCAAAGTAAGTCTGTTTTTGAGGAA | 153 | (2) |
| | Rev | LCF 449 | TGCTCTATTTTTAGCCTTATTAACAGC | | |
| <i>tcdE</i> | Fwd | LCF 402 | AAATATGTGCTTATGTGGATTACCAGTACC | 70 | This study |
| | Rev | LCF 403 | TTCATCCTTAGCATTTCATTTTCATCTGTCATTGC | | |
| 16S <i>rRNA</i> | Fwd | LCF 408 | GGGAGACTTGAGTGCCAGGAG | 120 | (1) |
| | Rev | LCF 409 | GTGCCTCAGCGTCAGTTACA | | |
| Probe A | Fwd | LCF 320 | GGGTAGTAAGAAACCAGCTAACCC | 1,207 | This study |
| | Rev | LCF 411 | GCGTGGTTGACCCTGTATGTTGTT | | |
| Probe B | Fwd | LCF 225 | GTCAGGAAGCCAAACAATAG | 2,846 | This study |
| | Rev | LCF 249 | TCCGCCCAACTCATTAATC | | |

Table S2: Host range analysis of ϕ CD38-2

| PCR ribotype * | No. isolates | No. sensitive to ϕ CD38-2 | PCR ribotype * | No. isolates | No. sensitive to ϕ CD38-2 |
|----------------|--------------|--------------------------------|----------------|--------------|--------------------------------|
| 1 | 2 | 1 | 22 | 5 | |
| 2 | 4 | | 23 | 5 | 1 |
| 3 | 1 | | 24 | 13 | 2 |
| 4 | 1 | | 25 | 2 | |
| 5 | 1 | | 26 | 1 | |
| 6 | 3 | 2 | 27 | 89 | 79 |
| 7 | 2 | 2 | 28 | 3 | |
| 8 | 1 | | 29 | 3 | |
| 9 | 3 | | 30 | 2 | |
| 10 | 1 | | 31 | 2 | |
| 11 | 1 | | 32 | 1 | |
| 12 | 5 | | 33 | 1 | |
| 13 | 2 | | 34 | 1 | |
| 14 | 22 | 6 | 35 | 2 | 2 |
| 15 | 2 | | 36 | 3 | |
| 16 | 3 | 2 | 37 | 2 | |
| 17 | 6 | | 38 | 3 | |
| 18 | 1 | | 39 | 1 | |
| 19 | 1 | 1 | 40 | 1 | |
| 20 | 3 | 1 | 41 | 1 | |
| 21 | 1 | | | | |

* Except for type 027 that corresponds to the hypervirulent strain BI/NAP1/027, all other ribotype numbers were assigned arbitrarily according to our internal nomenclature.

TABLE S3. Putative coding sequences from ϕ CD38-2 and bioinformatics analyses of the deduced proteins

| ORF | Start | End | Size (bp) | Size (aa) | Ribosome binding site (RBS) | Predicted fonction ^a | BLAST hit accession number [source organism] ^b | Alignment (%ID) | Size (aa) ^c | E-Value |
|-----|-------|-------|-----------|-----------|------------------------------------|--|---|-----------------------------------|------------------------|-----------------|
| 1 | 29 | 736 | 708 | 235 | gaaaggagTTG | Terminase (small) | ZP_05399284 [C. difficile QCD-37x79] YP_001308765.1 [C. beijerinckii NCIMB 8052] | 235/235 (100%) 98/236 (41%) | 235 260 | 6E-133 6E-35 |
| 2 | 723 | 2045 | 1320 | 440 | agaagaaatagaggcagaatatATG | Terminase (large) | ZP_05399283 [C. difficile QCD-37x79] ZP_02640253 [C. perfringens CPE str. F4969] | 440/440 (100%) 287/440 (65%) | 440 443 | 0E+00 3E-166 |
| 3 | 2046 | 3458 | 1413 | 470 | aagagg _{tgactaa} ATG | Minor capsid protein | ZP_05399282 [C. difficile QCD-37x79] NP_695169 [Lactobacillus phage phig1e] | 470/470 (100%) 111/434 (25%) | 470 508 | 0E+00 3E-35 |
| 4 | 3459 | 4691 | 1233 | 410 | gacaaggg _{tggtatgtag} ATG | Minor capsid protein | ZP_05399281 [C. difficile QCD-37x79] YP_001974361 [Streptococcus phage PH15] | 410/410 (100%) 64/274 (23%) | 410 381 | 0E+00 1E-05 |
| 5 | 4684 | 4968 | 285 | 94 | aaaggagag _{aaaa} ATG | Phage-related protein | ZP_05399280 [C. difficile QCD-37x79] ZP_01821446 [Streptococcus pneumoniae SP6-BS73] | 94/94 (100%) 37/88 (42%) | 94 87 | 6E-48 1E-10 |
| 6 | 5098 | 5670 | 573 | 190 | aggagg _{ttttt} ATG | Scaffolding protein | ZP_05399279 [C. difficile QCD-37x79] NP_469454 [Listeria innocua Clip11262] | 190/190 (100%) 61/196 (31%) | 190 189 | 6E-101 9E-09 |
| 7 | 5681 | 6754 | 1074 | 357 | aggagg _{tagaaa} ATG | Major capsid protein | ZP_05399278 [C. difficile QCD-37x79] YP_001391656 [C. botulinum F str. Langeland] | 357/357 (100%) 95/345 (27%) | 357 321 | 0E+00 3E-23 |
| 8 | 6758 | 7090 | 333 | 110 | gagaggg _{taggct} ATG | | ZP_05399277 [C. difficile QCD-37x79] ZP_02640259 [C. perfringens CPE str. F4969] | 110/110 (100%) 47/114 (41%) | 110 114 | 5E-57 1E-12 |
| 9 | 7092 | 7526 | 435 | 144 | gagg _{tgtttagt} GTG | | ZP_05399276 [C. difficile QCD-37x79] ZP_02616048 [C. botulinum Bf] | 144/144 (100%) 58/143 (40%) | 144 136 | 2E-76 2E-21 |
| 10 | 7533 | 7880 | 348 | 115 | ggaggcaga _{ataataggg} ATG | | ZP_05399275 [C. difficile QCD-37x79] ZP_02633339 [C. perfringens E str. JGS1987] | 115/115 (100%) 59/115 (51%) | 115 117 | 2E-61 1E-22 |
| 11 | 7877 | 8308 | 432 | 143 | ggaggtaga _{agggga} ATG | | ZP_05399274 [C. difficile QCD-37x79] ZP_02616050 [C. botulinum Bf] | 143/143 (100%) 70/141 (49%) | 143 143 | 2E-78 5E-37 |
| 12 | 8311 | 9111 | 800 | 266 | aaaggagaa _{atgtaaaa} ATG | Major structural protein | ZP_05399273 [C. difficile QCD-37x79] ZP_02640263 [C. perfringens CPE str. F4969] | 266/266 (100%) 73/156 (46%) | 266 168 | 2E-152 3E-29 |
| 13 | 9127 | 9489 | 363 | 120 | aagaggggg _{ctaaggaa} ATG | | ZP_05399272 [C. difficile QCD-37x79] ZP_02616053 [C. botulinum Bf] | 120/120 (100%) 49/123 (39%) | 120 114 | 3E-63 5E-14 |
| 14 | 9489 | 10094 | 606 | 201 | aacgtagag _{ctaaaaata} ATG | | ZP_05399271 [C. difficile QCD-37x79] ZP_02616054 [C. botulinum Bf] | 201/201 (100%) 119/202 (58%) | 201 197 | 3E-113 3E-63 |
| 15 | 10168 | 10809 | 642 | 213 | gggggggg _{attggt} ATG | | ZP_05332086 [C. difficile QCD-37x79] ZP_02949175 [C. butyricum 5521] | 213/213 (100%) 31/72 (43%) | 213 362 | 8E-119 1E-06 |
| 16 | 10920 | 16070 | 5151 | 1716 | aaaaggggg _{tgatag} ATG | Tail tape measure | ZP_05399269 [C. difficile QCD-37x79] ZP_02640311 [C. perfringens CPE str. F4969] | 1716/1716 (100%) 178/702 (25%) | 1716 933 | 0E+00 4E-62 |
| 17 | 16191 | 16925 | 735 | 244 | gagagggg _{gcataaagct} ATG | Major tail protein | ZP_05399268 [C. difficile QCD-37x79] YP_001787259 [C. botulinum A3 str. Loch Maree] | 244/244 (100%) 87/241 (36%) | 244 225 | 3E-137 4E-29 |
| 18 | 17285 | 19054 | 1770 | 589 | agaaagt _{ttggaa} TTG | Tail protein / putative RBP ^d | ZP_05399267 [C. difficile QCD-37x79] ZP_02621516 [C. botulinum C str. Eklund] | 589/589 (100%) 81/245 (33%) | 589 568 | 0E+00 1E-24 |
| 19 | 19081 | 19323 | 243 | 80 | aaaagggag _{aagaagt} ATG | | ZP_05399266 [C. difficile QCD-37x79] YP_699982 [C. phage phiSM101] | 80/80 (100%) 27/75 (36%) | 80 82 | 3E-38 0.022 |
| 20 | 19338 | 21800 | 2463 | 820 | aaaaggag _{tgataaat} TTG | Tail fiber protein | ZP_05399319 [C. difficile QCD-37x79] ZP_782154 [C. tetani E88] | 476/478 (99%) 45/149 (30%) | 478 660 | 0E+00 3E-05 |

| ORF | Start | End | Size (bp) | Size (aa) | Ribosome binding site (RBS) | Predicted fonction ^a | BLAST hit accession number [source organism] ^b | Alignment (%ID) | Size (aa) ^c | E-Value |
|-----|-------|-------|-----------|-----------|--|---------------------------------|---|-----------------|------------------------|---------|
| 21 | 21800 | 23650 | 1851 | 616 | agaagagg _{tgag} ta ATG | Tail fiber protein | ZP_05399318 [C. difficile QCD-37x79] | 616/616 (100%) | 616 | 0E+00 |
| | | | | | | | YP_719584 [Haemophilus somnus 129PT] | 67/238 (28%) | 762 | 4E-13 |
| 22 | 23736 | 23999 | 264 | 87 | aggagg _{aaagt} ta ATG | Holin | ZP_05399317 [C. difficile QCD-37x79] | 87/87 (100%) | 87 | 3E-41 |
| | | | | | | | ADK37889 [C. difficile phage phiCD6356 ORF27] | 78/84 (93%) | 87 | 9E-42 |
| 23 | 24000 | 24806 | 807 | 268 | agaaagaagg _{tctt} aa ATG | Endolysine | ZP_05399316 [C. difficile QCD-37x79] | 268/268 (100%) | 268 | 1E-154 |
| | | | | | | | ADK37890 [C. difficile phage phiCD6356 ORF28] | 237/268 (89%) | 268 | 2E-143 |
| 24 | 25067 | 24855 | 213 | 70 | aagagagg _{tg} tgaaaa ATG | | ZP_05399315 [C. difficile QCD-37x79] | 70/70 (100%) | 70 | 5E-29 |
| | | | | | | | ZP_02086761 [C. bolteae ATCC BAA-613] | 21/68 (30%) | 70 | 4E-06 |
| 25 | 25246 | 26076 | 831 | 276 | aaggggg _{taag} caaga ATG | | ZP_05399314 [C. difficile QCD-37x79] | 276/276 (100%) | 276 | 1E-152 |
| | | | | | | | YP_699976 [Clostridium phage phiSM101] | 50/127 (39%) | 137 | 2E-13 |
| 26 | 26155 | 26661 | 507 | 168 | ggagga _{ttga} gt ATG | | ZP_05399313 [C. difficile QCD-37x79] | 168/168 (100%) | 168 | 7E-90 |
| | | | | | | | ZP_05399427 [C. difficile QCD-37x79] | 154/168 (91%) | 168 | 2E-81 |
| 27 | 26681 | 27151 | 471 | 156 | ggagga _{tata} aaaa ATG | | ZP_05331942 [C. difficile QCD-63q42] | 156/156 (100%) | 156 | 9E-84 |
| 28 | 27665 | 27384 | 282 | 93 | aaaaagggg _{ttg} gaaa ATG | | ZP_05399311 [C. difficile QCD-37x79] | 92/93 (98%) | 93 | 2E-46 |
| | | | | | | | YP_001088968 [C. difficile 630] | 31/85 (36%) | 491 | 2E-03 |
| 29 | 28141 | 27809 | 333 | 110 | agaaggggag _{tgtt} atag ATG | | ZP_05399310 [C. difficile QCD-37x79] | 106/110 (96%) | 110 | 3E-51 |
| 30 | 28951 | 28142 | 807 | 269 | gaggtggag _{gaaa} ATG | ParA/SOJ-like protein | ZP_05331939 [C. difficile QCD-37x79] | 265/269 (98%) | 269 | 1E-151 |
| | | | | | | | ADK37894 [C. difficile phage phiCD6356 ORF32] | 98/266 (37%) | 272 | 4E-49 |
| 31 | 29420 | 29238 | 183 | 60 | aaaaaggag _{cta} atga ATG | | ZP_05399308 [C. difficile QCD-37x79] | 60/60 (100%) | 60 | 3E-27 |
| | | | | | | | ADK37895 [C. difficile phage phiCD6356 ORF33] | 57/60 (95%) | 60 | 9E-31 |
| 32 | 29788 | 29660 | 129 | 42 | agaaaggagga _{aac} ggga ATG | Membrane protein ^e | ZP_05399307 [C. difficile QCD-37x79] | 42/42 (100%) | 42 | 8E-13 |
| 33 | 30391 | 30239 | 153 | 50 | aaagggagg _{tgat} gaagt ATG | Membrane protein ^e | ZP_05399306 [C. difficile QCD-37x79] | 50/50 (100%) | 50 | 5E-20 |
| | | | | | | | ADK37896 [C. difficile phage phiCD6356 ORF34] | 46/48 (96%) | 50 | 3E-23 |
| 34 | 30432 | 30617 | 183 | 61 | gggagg _{ata} atctt TTG | Secreted protein ^e | ZP_05399305 [C. difficile QCD-37x79] | 61/61 (100%) | 61 | 5E-24 |
| 35 | 30911 | 33076 | 2163 | 721 | ggaggg _{ata} aaag ATG | DNAse/Helicase | ZP_05399304 [C. difficile QCD-37x79] | 721/721 (100%) | 721 | 0E+00 |
| | | | | | | | YP_659583 [C. difficile 630] | 375/703 (53%) | 869 | 0E+00 |
| 36 | 33123 | 33470 | 348 | 115 | aaggag _{ata} aaaagg ATG | | ZP_05399303 [C. difficile QCD-37x79] | 115/115 (100%) | 115 | 4E-57 |
| 37 | 33683 | 33793 | 111 | 36 | agaaaacagagg _{tta} TTG | | ZP_05331922 [C. difficile QCD-63q42] | 36/36 (100%) | 36 | 2E-10 |
| 38 | 34150 | 34245 | 93 | 31 | aggggg _{at} gattgg TTG | | ZP_05399302 [C. difficile QCD-37x79] | 31/31 (100%) | 31 | 3E-07 |
| 39 | 34643 | 34290 | 354 | 117 | gaaaggag _t tttaa ATG | CI phage repressor | ZP_05399301 [C. difficile QCD-37x79] | 117/117 (100%) | 117 | 1E-60 |
| | | | | | | | ADK37900 [C. difficile phage phiCD6356 ORF38] | 26/64 (41%) | 122 | 2E-09 |
| 40 | 34816 | 34998 | 183 | 60 | aggagg _{ttgt} ttct ATG | | ZP_05399300 [C. difficile QCD-37x79] | 60/60 (100%) | 60 | 3E-24 |
| 41 | 35065 | 35490 | 426 | 141 | aaggagg _{ctg} gaaa ATG | | ZP_05399299 [C. difficile QCD-37x79] | 141/141 (100%) | 141 | 5E-74 |
| 42 | 35565 | 35969 | 405 | 134 | agaaggggg _{gtt} taaaag ATG | Replication terminator | ZP_05399298 [C. difficile QCD-37x79] | 134/134 (100%) | 134 | 1E-70 |
| | | | | | | | YP_535158 [Lactobacillus phage Sal2] | 42/120 (35%) | 138 | 5E-10 |
| 43 | 35992 | 36705 | 714 | 237 | aaaaaggag _{tgt} ttaat ATG | | ZP_05399297 [C. difficile QCD-37x79] | 237/237 (100%) | 237 | 1E-133 |
| | | | | | | | NP_782698 [C. tetani E88] | 120/228 (52%) | 238 | 1E-55 |
| 44 | 36761 | 36925 | 165 | 54 | aaaaaggag _{tgt} ttatt TTG | | ZP_05399296 [C. difficile QCD-37x79] | 54/54 (100%) | 54 | 5E-22 |
| | | | | | | | ADK37907 [C. difficile phage phiCD6356 ORF45] | 16/52 (31%) | 58 | 4E-03 |
| 45 | 36922 | 37233 | 312 | 103 | aggagg _{ctg} ggca ATG | | ZP_05399295 [C. difficile QCD-37x79] | 103/103 (100%) | 103 | 2E-50 |
| | | | | | | | ADK37908 [C. difficile phage phiCD6356 ORF46] | 48/106 (46%) | 121 | 4E-15 |

| ORF | Start | End | Size (bp) | Size (aa) | Ribosome binding site (RBS) | Predicted fonction ^a | BLAST hit accession number [source organism] ^b | Alignment (%ID) | Size (aa) ^c | E-Value |
|-----|-------|-------|-----------|-----------|---------------------------------|---------------------------------|---|-----------------|------------------------|---------|
| 46 | 37279 | 37428 | 150 | 49 | aggaggctgggca ATG | | ZP_05399294 [C. difficile QCD-37x79] | 49/49 (100%) | 49 | 1E-19 |
| | | | | | | | ADK37909 [C. difficile phage phiCD6356 ORF47] | 33/47 (71%) | 58 | 1E-15 |
| 47 | 37459 | 37632 | 174 | 57 | gaatggaggatataaat ATG | | ZP_05399293 [C. difficile QCD-37x79] | 57/57 (100%) | 57 | 2E-25 |
| | | | | | | | ADK37911 [C. difficile phage phiCD6356 ORF49] | 43/57 (76%) | 57 | 2E-23 |
| 48 | 37632 | 37748 | 117 | 38 | gaaaggaatgacgaaatata ATG | | ZP_05399292 [C. difficile QCD-37x79] | 38/38 (100%) | 38 | 5E-13 |
| | | | | | | | ZP_05401366 [C. difficile QCD-23m63] | 22/38 (57%) | 40 | 6E-05 |
| 49 | 37745 | 38071 | 327 | 108 | aagggggagggtggaa ATG | SSB ^f | ZP_05399291 [C. difficile QCD-37x79] | 108/108 (100%) | 108 | 7E-56 |
| | | | | | | | ADK37912 [C. difficile phage phiCD6356 ORF50] | 97/106 (92%) | 135 | 1E-54 |
| 50 | 38163 | 38528 | 363 | 121 | aagtaggtgattaa ATG | | ZP_05399290 [C. difficile QCD-37x79] | 121/121 (100%) | 121 | 6E-65 |
| | | | | | | | YP_002290935 [C. difficile phage phiCD27] | 49/122 (40%) | 121 | 2E-10 |
| 51 | 38512 | 38685 | 174 | 57 | aagtaggtgattaa ATG | | ZP_05399289 [C. difficile QCD-37x79] | 57/57 (100%) | 57 | 3E-23 |
| | | | | | | | ADK37917 [C. difficile phage phiCD6356 ORF55] | 27/53 (51%) | 56 | 2E-11 |
| 52 | 38697 | 39269 | 573 | 190 | aggaagtgaattt ATG | Sigma70/SigmaF-like | ZP_05399288 [C. difficile QCD-37x79] | 190/190 (100%) | 190 | 4E-103 |
| | | | | | | | ADK37918 [C. difficile phage phiCD6356 ORF56] | 61/145 (43%) | 153 | 1E-27 |
| 53 | 39447 | 40055 | 609 | 202 | gagaaggtgaatagaa GTG | Integrase | ZP_05399287 [C. difficile QCD-37x79] | 202/202 (100%) | 202 | 3E-113 |
| | | | | | | | YP_001885139 [C. botulinum B str. Eklund 17B] | 122/198 (61%) | 212 | 6E-63 |
| 54 | 40265 | 40372 | 108 | 35 | ggtgaagcttaaggac ATG | | ZP_05332102 [C. difficile QCD-63q42] | 43/43 (100%) | 43 | 4E-16 |
| | | | | | | | ZP_05399286 [C. difficile QCD-37x79] | 34/35 (97%) | 43 | 2E-11 |
| 55 | 40521 | 41090 | 567 | 189 | agtaaggggtgtatttt ATG | Resolvase | ZP_05399285 [C. difficile QCD-37x79] | 189/189 (100%) | 189 | 2E-107 |
| | | | | | | | YP_001090102 [C. difficile 630] | 114/185 (61%) | 189 | 1E-58 |

^a Inferred from BLASTp analyses on protein sequences in GenBank and ACLAME databases, and by the presence of conserved domains found through searches in the conserved domains database (CDD) at NCBI and by InterProScan analysis

^b The best hit is shown along with another relevant hit when available (excluding hits from the QCD-63q42 strain)

^c Size in amino acids (aa) of the BLAST hit (which may be used to calculate the overall %ID)

REFERENCES

1. **Denève, C., S. Bouttier, B. Dupuy, F. Barbut, A. Collignon, and C. Janoir.** 2009. Effects of subinhibitory concentrations of antibiotics on colonization factor expression by moxifloxacin-susceptible and moxifloxacin-resistant *Clostridium difficile* strains. *Antimicrob Agents Chemother* **53**:5155-5162.
2. **Govind, R., G. Vedyappan, R. D. Rolfe, B. Dupuy, and J. A. Fralick.** 2009. Bacteriophage-mediated toxin gene regulation in *Clostridium difficile*. *J Virol* **83**:12037-12045.