1   **Putative conjugative plasmids with *tcdB* and *cdtAB* genes in clinical *Clostridium difficile* strains**

2   **from MLST clades C-I, 2 and 4**

3

4   Gabriel Ramírez-Vargas, César Rodríguez*

5   Facultad de Microbiología, CIET, LIBA, Universidad de Costa Rica, Costa Rica

6

7   ***Address for correspondence**: César Rodríguez, Universidad de Costa Rica, Facultad de

8   Microbiología. 11501-2060. San José. Costa Rica.

9

10   **Abstract**: The major toxins of *Clostridium difficile* (TcdA, TcdB, CDT) are encoded chromosomally in

11   nearly all known strains. Following up on a previous report, we found five new examples of a family of

12   putative conjugative plasmids with *tcdB* and *cdtAB* in clinical *C. difficile* isolates from MLST Clades

13   C-I, 2, and 4.

14

15   **Keywords**: *C. difficile*, MGEs, extrachromosomal elements, TcdB, CDT

16

17   **Text**: Upon gut microbial dysbiosis, ingested *Clostridium difficile* spores may differentiate in the

18   human colon into vegetative cells and release one or two large clostridial cytotoxins (i.e. TcdA, TcdB)

19   and/or a binary toxin with ADP-ribosyltransferase activity (CDT) to cause colitis and diarrhea. This

20   pathogen has gained notoriety in the last decade and is now considered a growing public health threat

21   and burden on healthcare systems (*1*).

22   Not all *C. difficile* strains are toxigenic, but when present, genes for TcdA, TcdB and CDT are nearly

23   without exception encoded by two separate chromosomal loci known as PaLoc and CdtLoc (*2*). This

24   paradigm has been challenged by the rather recent finding of Clade C-I strains SA10-050, CD10-165 in

25   France (*3*) and HSJD-312, HMX-152 in Costa Rica (*4*), which carry a monotoxin *tcdB*⁺ PaLoc next to a

26   full Cdtloc on extrachromosomal molecules that resemble conjugative plasmids (*4*). In this regard,

27   Clade C-I strains have been traditionally considered to be of environmental origin (*5*) and non-

28   toxigenic due to lack of perfect PaLoc integration sites in their chromosomes (*6*).

29   The Research Laboratory for Anaerobic Bacteriology (LIBA) has been isolating and typing *C. difficile*

30   in Costa Rica for nearly a decade and thereby generated an isolate collection with over 800 records. A

31   search of mobile genetic elements (MGEs) among Illumina whole genome sequences from 150 of those

32   bacteria, led to the discovery of five new $tcdA^-/tcdB^+/cdtAB^+$ extrachromosomal DNA molecules among

33   isolates that were recovered between 2013 and 2018 from patients that developed diarrhea at three

34   Costa Rican hospitals (LIBA-6656, LIBA-7194, LIBA-7602, LIBA-7678, LIBA-7697). Raw

35   sequencing data for isolate LIBA-6656 can be retrieved from the European Nucleotide Archive (Run

36   ERR467623). In turn, reads for the other four isolates are available at the MicrobesNG platform

37   (https://microbesng.com/portal/projects/FB43968C-E9EF-4270-9D1A-054457CC9B54/).

38   As indicated by a tree of aligned, concatenated, *C. difficile* MLST allele combinations deposited in

39   pubMLST (Figure 1a), these putative plasmid sequences were found not only in Clade C-I isolates

40   (LIBA-7194, LIBA-7602, LIBA-7678), but also in isolates from clades 2 (LIBA-6656) and 4 (LIBA-

41   7697). This result expands the host range previously reported by us for this type of MGEs to *C. difficile*

42   clades of more common association with humans, raising stimulating questions about their role in

43   human disease.

44   BWA mapping of reads from isolates LIBA-6656, LIBA-7194, LIBA-7602, LIBA-7678 and LIBA-

45   7697 to a plasmid sequence obtained by PacBio sequencing of strain HSJD-312 (Figure 1b) revealed

46   that the new plasmid sequences are related, yet not identical to each other (92-98% coverage) or to the

47   circular extrachromosomal sequences previously obtained for Clade C-I strains HSJD-312 and HMX-

48   152 (91-98% coverage). Whereas their toxin loci, agr locus, and potential conjugation machinery were

49   conserved, mapping gaps corresponded to putative virulence factors (i.e. putative lectin- or cell wall-

50   binding proteins), hypothetical proteins, and MGEs, such as a class 2 intron and transposases (Figure

51 1b). This data matches the anticipated mobility potential of the *C. difficile* toxin plasmids (*4*) and

52 supports the notion that they belong to a family of chimeric molecules undergoing non-homologous

53 recombination (*4*).

54 The MGE-associated *tcdB* sequence of the clade 2 strain LIBA-6656 could not be fully assembled. In

55 the remaining four strains this gene was highly conserved and expected to encode variant TcdBs (99-

56 100% protein sequence identity) that would cause a "sordellii" cytophatic effect (*7*). Besides its

57 plasmidial *tcdB*, LIBA-6656 carries a different *tcdB* on a chromosomal PaLoc. This surprising finding

58 demonstrates that two *tcdB* alleles can coexist in a single strain. The contribution of each *tcdB* to

59 infection is unclear at this moment, yet the coexistence of two PaLocs within a host is compatible with

60 the suggested transition from ancient monotoxin PaLocs to modern bitoxin PaLocs (*3*). A high level of

61 sequence identity was also noted for *cdtA* (≥99%) and *cdtB* (≥98%) in all five putative plasmids.

62 As previously reported, the toxin genes of the new putative plasmid sequences are flanked by

63 recombinases and integrases (*4*). Other elements from this family of potential MGEs lack toxin genes

64 (*4*), indicating that the latter were likely gained through lateral gene transfer events. However, it is

65 difficult with such a small dataset to determine whether the noted conservation of toxin gene sequences

66 reflects stable coevolution or simply short evolutionary time after their acquisition.

67 Three of the five isolates that host new toxin plasmids would have remained undetected if we had not

68 attempted *C. difficile* cultivation from TcdB⁻ stool samples of patients with CDI symptoms and

69 sequenced isolates with negative results for *tcdC* and *tcdA* in a PCR-based screening (LIBA-7194,

70 LIBA-7602, LIBA-7678). Hence, we anticipate that the frequency of *C. difficile* isolates with toxin

71 plasmids has been underestimated and urge for refined diagnostic procedures. Moreover, our results

72 open avenues to explore whether plasmids from this group are present in species other than *C. difficile*

73 and account for undiagnosed cases of antibiotic-associated diarrhea.

74 **Acknowledgements**: Vicerrectoría de Investigación/UCR funded this work. Dr. Thomas Riedel and

75 Prof. Jörg Overmann generated the PacBio plasmid sequence of isolate HSJD-312 that was used as a

76 reference in the context of a BMBF-MICITT project.

77

78 **Conflicts of interest:** The authors have nothing to declare.

79

80 **References**:

81 (*1*) Balsells E, Shi T, Leese C, Lyell I, Burrows J, Wiuff C, et al. Global burden of *Clostridium difficile*

82 infections: a systematic review and meta-analysis. J Glob Health. 2019;9(1).

83 (*2*) Knight DR, Elliott B, Chang BJ, Perkins TT, Riley TV. Diversity and evolution in the genome of

84 *Clostridium difficile*. Clin Microbiol Rev. 2015;28(3):721-41.

85 (*3*) Monot M, Eckert C, Lemire A, Hamiot A, Dubois T, Tessier C, et al. *Clostridium difficile*: New

86 insights into the evolution of the Pathogenicity Locus. Sci Rep. 2015;8:5.

87 (*4*) Ramirez-Vargas G, Lopez-Urena D, Badilla A, Orozco-Aguilar J, Murillo T, Rojas P, et al. Novel

88 Clade C-I *Clostridium difficile* strains escape diagnostic tests, differ in pathogenicity potential and

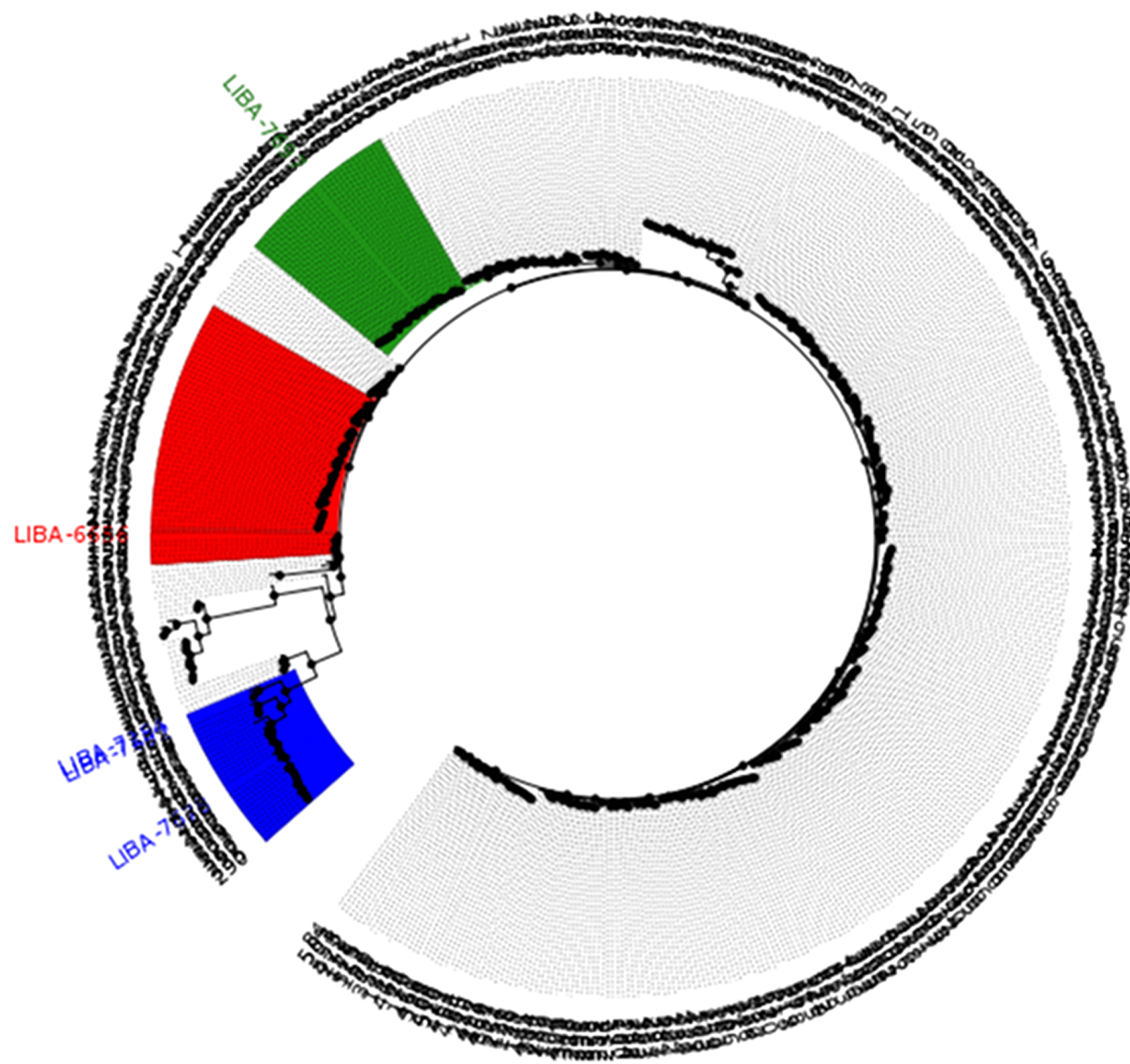89 carry toxins on extrachromosomal elements. Sci Rep. 2018;17:8.

90 (*5*) Janezic S, Potocnik M, Zidaric V, Rupnik M. Highly divergent *Clostridium difficile* strains isolated

91 from the environment. PLoS One. 2016;23:11(11).

92 (*6*) Dingle KE, Elliott B, Robinson E, Griffiths D, Eyre DW, Stoesser N, et al. Evolutionary history of

93 the *Clostridium difficile* Pathogenicity Locus. Genome Biol Evol. 2014;6(1):36-52.

94 (*7*) Rupnik M, Janezic S. An update on *Clostridium difficile* toxinotyping. J Clin Microbiol.

95 2016;54(1):13-8.

96 **Figure legends**

97 **Figure 1.** MLST-based classification (A) and diversity of extrachromosomal circular sequences (B) of

98 *C. difficile* strains with plasmid-encoded toxins. The FastTree tree shown in (A) was derived from a

99 MUSCLE alignment of concatenated MLST alleles from all *C. difficile* sequence types (STs) deposited

100 in the PubMLST database. Tip labels represent STs or strain names. Strains from clades C-I, 2, and 4

101 are highlighted in blue, red, and green, respectively. Panel B shows short reads from strains LIBA-7194,

102 LIBA-7678, LIBA-7602 (Clade C-I, blue), LIBA-6656 (Clade 2, red), and LIBA-7697 (Clade 4, green)

103 mapped to the plasmid sequence of strain HSJD-312, which was obtained through long-read PacBio

104 sequencing and therefore used as a reference (145 kb, bottom). Arrows in the reference sequence

105 represent annotated CDS. Different colors were used to show genes for toxins (red), transposases,

106 integrases, and recombinases (blue), and proteins from a putative conjugation machinery (green).

LIBA-7___

LIBA-6556

LIBA-7___

LIBA-7___

0.02