Romo-Castillo Mariana[1*] and Torres Javier[2]

[1]CONACYT-IMSS, Unidad de Investigación Médica en Enfermedades Infecciosas y Parasitarias, UMAE Hospital de Pediatría. Centro Médico Nacional Siglo XXI, IMSS. [2] Unidad de Investigación Médica en Enfermedades Infecciosas y Parasitarias, UMAE Hospital de Pediatría. Centro Médico Nacional Siglo XXI, IMSS.

*mromo@conacyt.mx

Introduction

Clostridium difficile is a strict anaerobic bacterium, Gram positive and spore-forming that has become the leading cause of nosocomial diarrhea in the world . The recurrence rate reaches up to 20%, with a 5% mortality rate. It relates to the 30% of cases of antibiotic-associated diarrhea.

The main toxins involved in the development of ICD are the TcdA toxin and TcdB toxin. These toxins are encoded by the *tcdA* and *tcdB* genes, respectively; located on the pathogenicity island, called PaLoc. Within this island also are encoded three genes: *tcdR*, *tcdE* and *tcdC* (Figure 1) .

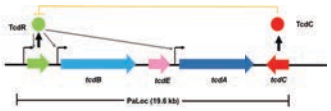

Fig 1. Schematic representations of PaLoc.

The *tcdC* gene codes for a small protein, of 231 aa, with a molecular weight of 25.7 kDa. Studies have identified that TcdC is associated with the membrane, and that it has a transmembrane domain between residues 31 to 51. Its maximum level of expression occurs during the exponential phase of growth. Previously, it was reported that some strains encoding a truncated version of the protein, due to a specific insertion that causes a modification in the reading frame, were able to produce a greater amount of toxins, and were called hypertoxigenic strains . However, a subsequent study showed that replacing the wildtype gene with the version that encodes the truncated protein (65 aa) does not increase toxin production, creating a controversy over the role that TcdC protein has during the pathogenesis. More than 30 different alleles have been described for the *tcdC* gene. The objective of this project is to identify different alleles from the genomic sequencing of clinical isolates; and analyze the dominance that could exist between them.

Results

44 strains of hospitalized patients who had infection caused by *C. difficile (CDI)* were obtained. Clinical patient information from which were obtained strains shown in Table 1. The strains were isolated and sequenced. Of the sequences obtained, only 27 presented the PaLoc island, or the tcdC gene in isolation (like strain 80, sequence 8855).

Table 1. Clinical information of the hospitalized patients with CDI

	Positive to *tcdC*	Negative to *tcdC*	TOTAL
FEMALE	16	8	24
MALE	11	9	20
ADULT	23	8	31
INFANT	4	9	13
	STRAINS SEQUENCED = 44		

From the analysis of the secondary structure (PSIPRED) two main domains could be identified (Figure 2). The first is made up of alpha helices (residues 1-149), while the second is made up of beta sheets (residues 150-231).

Fig 2. CdtC predicted secundary structure (PSIPRED)

A first alignment between the tcdC sequences allowed to identify the presence of 5 different alleles (Figure 6B) :

TcdC^{89}: Wildtype version of the gene . Present in sequence 8889.
TcdC^{86}: Present one deletion of 18 bases (region 320-337) called Δ18. Present in sequence 8886.
TcdC^{58}: In addition to the Δ18 deletion, it presents the N220K mutation. Present in sequence 8858.
TcdC^{54}: In addition to the Δ18 deletion, it has two point mutations, A35R and E57K. Present in sequence 8854.
TcdC^{16}: Presents the deletion of base 117, producing a change in the reading frame and coding for a truncated version of 65 aa. Present in the rest of the sequences.

With the help of the I-TASSER program, an *in silico* model of the structure of the five identified versions of the TcdC protein was obtained (Figure 3). The different alterations that they present with respect to the reference sequence obtained from strain *C. difficile* 630 (Gene ID: 4914044, Protein ID: YP_001087138.1) are indicated in red.

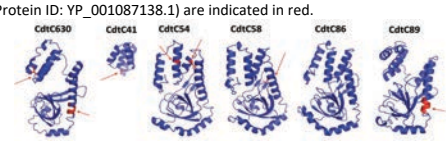

Fig 3. The CdtC tertiary structure resembles the 3D structures of the identified

When superimposing the structure of the protein designed with the sequence of strain 630 (pink), and the TcdC^{54} version (cyan), it can be seen that the sequences have a great structural similarity (Figure 4) , with an RMSD of 12.233 Å (TM-align server).

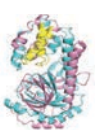

Fig 4. Structural alignment of CdtC^{630} (pink) and CdtC^{54} (cyan) gemerated using the TM-align server.The calculated root mean square veciation is 12.233 Å)

When analyzing whether the sequences had any phylogeographic relationship (Figure 5) using , it was observed that allelic sequences are grouped by their similarity, regardless of the geographical origin

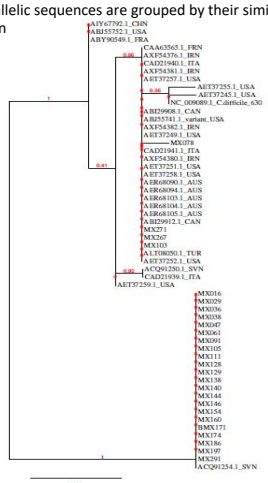

Fig 5. Phylogenetic tree of the sequence of mexican strains,compared with reported sequences of other countries.

Plasmids containing the different versions of the gene have been designed to analyze the effect that overproduction and competition on toxin production could have , which will be analyzed through the titration of toxins in Vero cells.

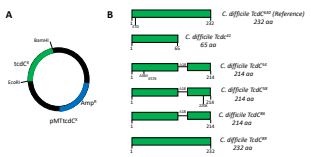

Fig 6. Plasmid construction. A. Backbone of the plasmid used to tranform C. difficile strains. B. TcdC versions identified in the present study, the corresponding genes are used to constructed the plasmids.

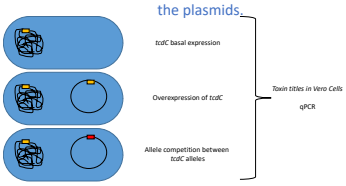

Fig 7. Analysis of allelic dominance

Preliminary results suggested that the overproduction of TcdC41 induces an increase in the total production of toxins (Figure 8), due to the fact that a higher titer of them was produced in the supernatant, when analyzing their effect in Vero cells.

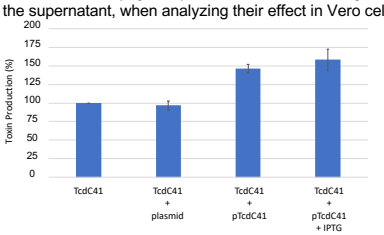

Fig 8. Effect of overproduction of TcdC41 over toxin title (p<0.01)

Finally, the results obtained so far suggest a dominance of TcdC41 over the other alleles identified. This is due to the fact that by co-expressing the Tcd41 allele in the strains, we can observe an increase in the production of toxins with respect to the basal toxin production of the strains (Figure 9).

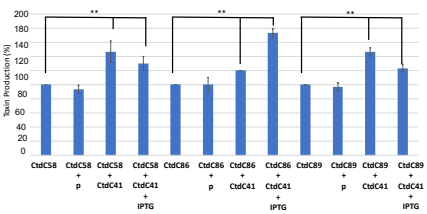

Fig 9. Toxin production of the strains in presence of TcdC41 (p<0.01)

Acknowledgements

CONACYT Consejo Nacional de Ciencia y Tecnología

IMSS