

UNIVERSIDAD DE COSTA RICA  
SISTEMA DE ESTUDIOS DE POSGRADO

**IDENTIFICACIÓN PANGENÓMICA DE SECUENCIAS FIRMA PARA LA  
DETECCIÓN DE CEPAS DEL CLADO C-I DE *CLOSTRIDIODES DIFFICILE* MEDIANTE  
qPCR**

Tesis sometida a la consideración de la Comisión del  
Programa de Estudios de Posgrado en Ciencias Biomédicas  
para optar al grado y título de Maestría Académica en  
Ciencias Biomédicas con Énfasis en Genómica

CAMILO MONGE CASCANTE

Ciudad Universitaria Rodrigo Facio, Costa Rica  
2021

## **Dedicatoria.**

A mi esposa, por su apoyo incondicional.

## **Agradecimientos**

En primer lugar, al Dr. César Rodríguez Sánchez, por su acompañamiento, tutelaje, y en particular por su franca y siempre amable actitud a lo largo de todo el proceso. Realmente el mundo sería mejor con más profesores y científicos como usted.

A la Dra. Rebeca Campos Sánchez y el Dr. Esteban Chaves Olarte, en calidad de lectores, cuyos valiosos comentarios y observaciones convirtieron este en un mejor trabajo.

Al personal del Laboratorio de Biología Molecular del Hospital Clínica Bíblica, en particular a la Dra. Karla Gutiérrez González, por permitirme realizar ensayos en sus equipos, las explicaciones brindadas y en general toda su valiosa ayuda.

Al personal del Centro de Investigación en Enfermedades Tropicales y el Laboratorio de Investigación en Bacteriología Anaerobia de la Universidad de Costa Rica, por su labor y asistencia durante las fases experimentales del proyecto.

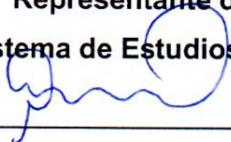
A los profesores y personal administrativo de la Maestría, por su paciencia, comprensión y ardua labor, sin la cual este trabajo no hubiera sido posible.

Esta Tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Ciencias Biomédicas de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Académica en **Ciencias Biomédicas con Énfasis en Genómica**



Ph.D. Marietta Flores Díaz

**Representante del Decano  
Sistema de Estudios de Posgrado**



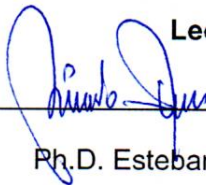
Ph.D. César Rodríguez Sánchez

**Profesor Guía**



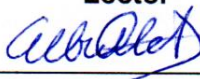
Ph.D. Rebeca Campos Sánchez

**Lectora**



Ph.D. Esteban Chaves Olarte

**Lector**



Ph.D. Alberto Alape Girón

**Representante  
Programa de Posgrado en Ciencias Biomédicas**



Camilo Monge Cascante

**Sustentante**

## Tabla de Contenidos

<b>DEDICATORIA.....</b>	<b>II</b>
<b>AGRADECIMIENTOS .....</b>	<b>II</b>
<b>HOJA DE APROBACIÓN.....</b>	<b>III</b>
<b>TABLA DE CONTENIDOS .....</b>	<b>IV</b>
<b>RESUMEN .....</b>	<b>VI</b>
<b>LISTA DE CUADROS .....</b>	<b>VII</b>
<b>LISTA DE FIGURAS .....</b>	<b>VIII</b>
<b>LISTA DE ABREVIATURAS.....</b>	<b>IX</b>
<b>CAPÍTULO I: INTRODUCCIÓN.....</b>	<b>1</b>
GENERALIDADES: .....	1
FACTORES DE RIESGO: .....	3
PRESENTACIÓN CLÍNICA: .....	3
IMPORTANCIA CLÍNICA DE LAS CDI: .....	3
DIAGNÓSTICO Y TIPIFICACIÓN:.....	4
PREVENCIÓN Y CONTROL DE LAS CDI A NIVEL HOSPITALARIO: .....	6
EPIDEMIOLOGÍA DE LAS CDI:.....	9
EPIDEMIOLOGÍA DE LAS CDI EN COSTA RICA: .....	10
PANGENÓMICA Y SECUENCIAS FIRMA: .....	11
APORTES DE LA GENÓMICA AL ESTUDIO Y MANEJO EN ENFERMEDADES INFECCIOSAS: .....	12
<b>CAPÍTULO II: JUSTIFICACIÓN, HIPÓTESIS Y OBJETIVOS .....</b>	<b>14</b>
JUSTIFICACIÓN: .....	14
HIPÓTESIS: .....	15
OBJETIVO GENERAL: .....	15
OBJETIVOS ESPECÍFICOS:.....	15
<b>CAPÍTULO III: METODOLOGÍA.....</b>	<b>16</b>
GENOMAS: .....	16
GENÓMICA COMPARATIVA:.....	17
IDENTIFICACIÓN DE SECUENCIAS FIRMA: .....	17
DISEÑO DE OLIGONUCLEÓTIDOS CEBADORES:.....	21
ESTANDARIZACIÓN Y VALIDACIÓN DE LA PCR: .....	22
<b>CAPÍTULO IV: RESULTADOS.....</b>	<b>28</b>
GENÓMICA COMPARATIVA Y PANGENÓMICA: .....	28
IDENTIFICACIÓN DE SECUENCIAS FIRMA: .....	35
DISEÑO Y VALIDACIÓN DE ENSAYO (Q)PCR:.....	37
<b>CAPÍTULO V: DISCUSIÓN .....</b>	<b>55</b>
GENÓMICA COMPARATIVA Y PANGENÓMICA: .....	55
IDENTIFICACIÓN DE SECUENCIAS FIRMA: .....	57
DISEÑO Y VALIDACIÓN DEL ENSAYO DE PCR: .....	60
PCR DE PUNTO FINAL: .....	61
qPCR USANDO CÉLULAS COMPLETAS (CULTIVOS PUROS): .....	62
qPCR CON ADN EXTRAÍDO A PARTIR DE HECES: .....	63
qPCR CON ADN EXTRAÍDO A PARTIR DE MUESTRAS INOCULADAS:.....	64
qPCR CON CEBADORES PARA EL GEN TCDB: .....	65
PROPUESTA TEMPORAL DE ALGORITMO DIAGNÓSTICO: .....	67
CONSIDERACIONES FINALES Y LIMITACIONES: .....	69
<b>CAPÍTULO VI: CONCLUSIONES:.....</b>	<b>73</b>
<b>BIBLIOGRAFÍA.....</b>	<b>74</b>

<b>ANEXO I. ST, RIBOTIPO DE PCR, IDENTIFICADOR DE AISLAMIENTO, TOXIGENICIDAD, CLADO, FUENTE Y FECHA DE AISLAMIENTO Y NUMERO DE CORRIDA DE SECUENCIACIÓN DE LAS CEPAS INCLUIDAS EN EL DISEÑO EXPERIMENTAL.....</b>	<b>87</b>
<b>ANEXO II. CARACTERÍSTICAS DETERMINADAS PARA LOS ENSAMBLAJES UTILIZADOS EN EL ANÁLISIS BIOINFORMÁTICO. ....</b>	<b>91</b>
<b>ANEXO III. PRINCIPALES CARACTERÍSTICAS PREDICHAS Y MEJORES HITS DE BLASTN DE LAS SECUENCIAS FIRMA DEL CLADO C-I IDENTIFICADAS Y DETERMINADAS POR NEPTUNE.....</b>	<b>94</b>
<b>ANEXO IV. CONGLOMERADOS DE CDSS IDENTIFICADOS POR SCOARY Y TREEWAS COMO EXCLUSIVOS DE LAS CEPAS DEL CLADO C-I (SENSIBILIDAD Y ESPECIFICIDAD= 100%, P CON CORRECCIÓN DE BENJAMINI HOCHBERG =1.21962E-11). ....</b>	<b>95</b>
<b>ANEXO V. CONGLOMERADOS DE GENES IDENTIFICADOS DEL BIN EXCLUSIVO DE LAS CEPAS DEL CLADO C-I IDENTIFICADO POR ANVI'O.....</b>	<b>110</b>
<b>ANEXO VI. CARACTERÍSTICAS DE LAS SECUENCIAS FIRMA IDENTIFICADAS POR DBGWAS DE MÁS DE 100PB PRESENTES EN LOS PRIMEROS 60 NODOS ANALIZADOS SEGÚN VALOR DE Q, Y SUS RESULTADOS AL ALINEARLAS CON LA COLECCIÓN DE NUCLEÓTIDOS DEL NCBI POR BLASTN.....</b>	<b>114</b>
<b>ANEXO VII. RESULTADOS DE PCR PUNTO FINAL PARA LOS PARES DE CEBADORES. ....</b>	<b>122</b>

## Resumen

*Clostridioides difficile* es un importante patógeno humano. Las cepas de esta bacteria han sido organizadas en cinco clados comunes, numerados del 1 al 5, y tres divergentes, conocidos como C-I, C-II y C-III. La infección con cepas patogénicas de *C. difficile* (CDI) se manifiesta clínicamente como diarrea en el paciente. Estos efectos son mediados principalmente por el efecto de tres toxinas (TcdA, TcdB y CDT). El diagnóstico de las CDI se realiza usualmente mediante inmunoensayos para las toxinas A y B, los cuales son rápidos, pero poco sensibles, o con cultivo toxigénico y ensayos de neutralización de citotoxicidad celular, ambos demandantes desde un punto de vista técnico y económico. Como alternativa, se han desarrollado ensayos de amplificación de ácidos nucleicos, dado que tienen alta sensibilidad, especificidad, y son relativamente rápidos y accesibles. Existen reportes de cepas toxigénicas del Clado C-I que escapan los métodos de diagnóstico utilizados rutinariamente en la mayoría de los centros de salud. Esta situación es preocupante, y motivó la realización del presente proyecto, el cual pretendía identificar secuencias firma nucleotídicas de cepas del Clado C-I susceptibles a ser amplificadas por qPCR, para emplearlas en un método de diagnóstico molecular. Se incluyeron en el análisis 77 genomas de cepas de *C. difficile* calidad *draft* (10 de cepas C-I, 67 de cepas de Clados 1-5) los cuales fueron ensamblados, verificada su calidad y anotados. Para identificar las posibles secuencias firma se utilizaron tres técnicas: estudios de asociación de genoma completo (DBGWAS, TreeWAS y Roary/Scoary), procesos automatizados basados en referencia y pareo de *k-mer* (Neptune) y análisis pangenómicos (Roary y Anvi'o). Los resultados de estos métodos fueron consolidados, y se eliminaron aquellas secuencias que tuvieran alta identidad con otras presentes en otros microorganismos. Se utilizó Primer-BLAST para diseñar juegos de oligonucleótidos cebadores para las secuencias firma identificadas, los cuales fueron depurados según contenido GC, autocomplementariedad, largo del producto, Tm. Los mejores pares fueron probados inicialmente en ensayos de PCR punto final. Estos cebadores, y uno adicional diseñado para el gen *tcdB* presente en las cepas del Clado C-I, fueron probados adicionalmente en qPCR usando ADN extraído a partir de cultivos bacterianos, de heces humanas negativas para *C. difficile*, de heces humanas inoculadas con ADN extraído a partir de cepas del Clado C-I y de heces humanas inoculadas con cultivos de *C. difficile* del Clado C-I y de otros clados. Se calcularon las sensibilidades y especificidades demostradas por los cebadores en cada ensayo. Finalmente, se aproximó el límite de detección de los dos mejores cebadores para las secuencias firma identificadas y el diseñado para el gen *tcdB*. El análisis pangenómico con Roary reveló que las cepas del Clado C-I comparten 890 CDSs que no pertenecen al genoma *core* determinado para la especie. Once conglomerados de secuencias codificantes fueron identificados como exclusivos de las cepas C-I por Anvi'o, Roary/Scoary y Treewas. Se generaron pares de cebadores para ocho de estas secuencias firma, los cuales fueron depurados, para sintetizar cuatro, los cuales fueron probados en laboratorio. Dos de los pares de cebadores mostraron 100% de sensibilidad y especificidad en los ensayos de qPCR con ADN extraído de cultivos líquidos y de heces inoculadas con cultivos, mientras que el par diseñado para el gen *tcdB* mostró 100% de sensibilidad y 72.7% de especificidad en los ensayos de qPCR con ADN extraído de cultivos en medio líquido, y 66% de sensibilidad y 100% de especificidad en los realizados con heces inoculadas con cultivos. A partir de estos resultados se propuso un algoritmo para determinar la presencia de ADN de cepas de *C. difficile* del Clado C-I, toxigénicas o no, en heces de pacientes diarreicos con sospecha de CDI, utilizando estos tres pares de cebadores. Estos hallazgos confirman que las cepas del Clado C-I presentan secuencias firma susceptibles a ser amplificadas por qPCR, y que los cebadores diseñados detectan la presencia de ADN de dichas cepas en cultivo y heces con 100% de especificidad y sensibilidad.

## Lista de cuadros

CUADRO I. GENES USUALMENTE PRESENTES EN EL PALOC DE <i>C. DIFFICILE</i> .....	2
CUADRO II. RESUMEN DE LA FORMA DE INTERPRETACIÓN DE LOS ENSAYOS DE qPCR .....	26
CUADRO III. PRINCIPALES CARACTERÍSTICAS PREDICHAS DE LAS SECUENCIAS FIRMA DEL CLADO C-I IDENTIFICADAS EN CONJUNTO POR ROARY/SOARY, TREEWAS Y ANVI'O Y DETERMINADAS POR ROARY Y ANVI'O. TODAS LAS SECUENCIAS TUVIERON LONGITUDES MAYORES A 300 PB PARA POSIBILITAR SU DETECCIÓN POR qPCR. ....	36
CUADRO IV. CARACTERÍSTICAS PREDICHAS DE LOS PARES DE CEBADORES IDENTIFICADOS POR PRIMERBLAST PARA LAS SECUENCIAS FIRMA SELECCIONADAS.....	39
CUADRO V. PROPIEDADES TERMODINÁMICAS DEL AMPLICÓN ESPERADO Y POSIBLES DÍMEROS DE CEBADOR ANTICIPADOS PARA LOS 7 MEJORES PARES DE CEBADORES CANDIDATOS. ....	44
CUADRO VI. CARACTERÍSTICAS PREDICHAS DE LOS PARES DE CEBADORES IDENTIFICADOS POR PRIMERBLAST PARA LAS SECUENCIAS DEL GEN <i>TCDB</i> DE LAS CEPAS DE <i>C. DIFFICILE</i> DEL CLADO C-I. ....	46
CUADRO VII. VALORES DE Ct Y RESULTADOS OBTENIDOS POR qPCR UTILIZANDO ADN EXTRAÍDO A PARTIR DE CULTIVOS DE DIVERSAS BACTERIAS Y LOS CEBADORES DISEÑADOS PARA LAS SECUENCIAS FIRMA DEL CLADO C-I. ....	47
CUADRO VIII. VALORES DE Ct OBTENIDOS POR qPCR UTILIZANDO ADN EXTRAÍDO A PARTIR DE HECES DE PACIENTES DIARREICOS Y LOS CEBADORES DISEÑADOS PARA LAS SECUENCIAS FIRMA DEL CLADO C-I. ....	48
CUADRO IX. VALORES DE Ct Y RESULTADOS OBTENIDOS POR qPCR UTILIZANDO ADN EXTRAÍDO A PARTIR DE HECES INOCULADAS CON ADN GENÓMICO DE CEPAS DEL CLADO C-I Y LOS CEBADORES DISEÑADOS PARA LAS SECUENCIAS FIRMA DEL CLADO C-I. ....	49
CUADRO X. VALORES DE Ct Y RESULTADOS OBTENIDOS POR qPCR UTILIZANDO ADN EXTRAÍDO A PARTIR DE HECES INOCULADAS CON DILUCIONES SERIADAS DE CULTIVOS DE CEPAS DEL CLADO C-I Y LOS CEBADORES DISEÑADOS PARA LAS SECUENCIAS FIRMA DEL CLADO C-I. ....	50
CUADRO XI. VALORES DE Ct OBTENIDOS EN EL ENSAYO DE qPCR UTILIZANDO ADN EXTRAÍDO A PARTIR DE CULTIVOS DE DIVERSAS BACTERIAS Y LOS CEBADORES DISEÑADOS PARA EL GEN <i>TCDB</i> PRESENTE EN LAS CEPAS DEL CLADO C-I. ....	51
CUADRO XII. VALORES DE Ct OBTENIDOS EN EL ENSAYO DE qPCR UTILIZANDO ADN EXTRAÍDO A PARTIR DE HECES INOCULADAS CON DILUCIONES SERIADAS DE CULTIVOS DE CEPAS DEL CLADO C-I Y LOS CEBADORES DISEÑADOS PARA EL GEN <i>TCDB</i> PRESENTE EN LAS CEPAS DEL CLADO C-I. ....	52
CUADRO XIII. RESUMEN DE LAS SENSIBILIDADES Y ESPECIFICIDADES DETERMINADAS PARA LOS DIFERENTES PARES DE CEBADORES ENSAYADOS. ....	53

## Lista de figuras

FIGURA 1. ESTRUCTURA HABITUAL DEL PALOC DE <i>C.DIFFICILE</i> . .....	1
FIGURA 2. ESQUEMA RESUMEN DE LA METODOLOGÍA UTILIZADA PARA IDENTIFICAR LAS POSIBLES SECUENCIAS FIRMA.. .....	20
FIGURA 3. ESQUEMA RESUMEN DE LAS PRUEBAS REALIZADAS PARA VALIDAR LA SENSIBILIDAD Y ESPECIFICIDAD DE LOS CEBADORES DISEÑADOS PARA LAS SECUENCIAS FIRMA IDENTIFICADAS. ....	27
FIGURA 4. ESQUEMA RESUMEN DE LAS PRUEBAS REALIZADAS PARA VALIDAR LA SENSIBILIDAD Y ESPECIFICIDAD DE LOS CEBADORES DISEÑADOS PARA EL GEN <i>TCDB</i> PRESENTE EN LAS CEPAS DEL CLADO C-I.....	27
FIGURA 5. ÁRBOL FILOGENÉTICO CONSTRUIDO A PARTIR DE UN ALINEAMIENTO DE SITIOS VARIABLES IDENTIFICADOS POR ROARY EN TODOS LOS GENOMAS ANALIZADOS. ....	29
FIGURA 6. DIAGRAMA DE CAJAS-BIGOTES PARA COMPARAR EL ANI DETERMINADO POR FASTANI PARA CEPAS DE CLADO C-I (N=10), CLADOS 1-5 (N=67) Y ENTRE AMBOS GRUPOS. ....	30
FIGURA 7. NÚMERO ABSOLUTO (A) Y RELATIVO EN PORCENTAJE (B) DE SECUENCIAS EN EL GENOMA <i>CORE</i> (PRESENTES EN AL MENOS 99% DE LAS CEPAS ANALIZADAS, N=76), <i>SOFT-CORE</i> (PRESENTES ENTRE EL 95% Y 99% DE LAS CEPAS, N=73 Y 76 CEPAS), <i>SHELL</i> (PRESENTES ENTRE EL 15% Y 95% DE LAS CEPAS, N= 11 Y 73 CEPAS) Y <i>CLOUD</i> (PRESENTES EN MENOS DEL 15% DE LAS CEPAS, N< 11) DE LAS CEPAS DEL CLADO C-I, LAS CEPAS DE LOS CLADOS 1-5 Y TODAS LAS CEPAS ANALIZADAS.....	32
FIGURA 8. MATRIZ DE PRESENCIA/AUSENCIA DE CDS CONTEXTUALIZADA FILOGENÉTICAMENTE (ROARY)..	33
FIGURA 9. CONGLOMERADOS DE GENES PRESENTES EN EL <i>BIN</i> EXCLUSIVO DE LAS CEPAS C-I IDENTIFICADO POR ANVI'0, CLASIFICADOS SEGÚN SU FUNCIÓN COG ANTICIPADA. ....	34
FIGURA 10. GRÁFICO SEMILOGARÍTMICO DE FACTOR DE DILUCIÓN VS CT OBTENIDO EN EL ENSAYO DE LÍMITE DE DETECCIÓN DEL ENSAYO DE PCR TIEMPO REAL UTILIZANDO ADN EXTRAÍDO A PARTIR DE UN CULTIVO DE LA CEPA LIBA-7602 Y LOS CEBADORES 114 (ROJO), 161 (VERDE) Y DISEÑADOS PARA EL GEN <i>TCDB</i> (AZUL) PRESENTE EN LAS CEPAS DEL CLADO C-I. ....	54
FIGURA 11. ALGORITMO PROPUESTO PARA DETERMINAR LA PRESENCIA DE CEPAS DE <i>C. DIFFICILE</i> DEL CLADO C-I, TOXIGÉNICAS O NO, EN HECES DE PACIENTES DIARREICOS CON SOSPECHA DE CDI. ....	67
FIGURA 12. AMPLIFICACIÓN DE LAS CEPAS UTILIZADAS CON LOS CEBADORES 110 (PANEL A), 161 (PANEL B) Y 114 (PANEL C), ASÍ COMO LAS CEPAS 6822, 7602 (PANEL D), NAP1 Y A-B+ (PANEL E) CON LOS CEBADORES 110, 114, 161 Y 195 .....	122

## Lista de abreviaturas

- CDI** “*Clostridioides difficile* infection” (infección por *Clostridioides difficile*)
- GTP** Guanosina trifosfato
- ADP** Adenosina difosfato
- PaLoc** Locus de patogenicidad
- TC** “*toxigenic culture*” (cultivo toxigénico)
- CCNA** “*cell cytotoxicity neutralization assay*” (ensayos de neutralización de citotoxicidad celular)
- EIA** “*enzyme immunoassay*” (inmunoensayos enzimáticos)
- NAAT** “*nucleic acid amplification test*” (ensayos de amplificación de ácidos nucleicos)
- PFGE** “*pulsed-field gel electrophoresis*” (electroforesis de campo pulsante)
- MLST** “*multilocus sequence typing*” (tipificación por secuencias multilocus)
- NGS** “*next generation sequencing*” (secuenciación de siguiente generación)
- PFGE** “*pulsed-field electrophoresis*” (electroforesis de campo pulsante)
- ESCMID** “*European Society of Clinical Microbiology and Infectious Diseases*” (Sociedad Europea de Microbiología Clínica y Enfermedades Infecciosas)
- IDSA** “*Infectious Diseases Society of America*” (Sociedad Americana de Enfermedades Infecciosas)
- SHEA** “*Society for Healthcare Epidemiology of America*” (Sociedad para la Epidemiología de la Atención Médica de América)
- ST** “*sequence type*” (secuencia tipo de MLST)
- PCR** “*polymerase chain reaction*” (reacción en cadena de la polimerasa)
- CA** California, E.E.U.U.
- qPCR** “*quantitative polymerase chain reaction*” (reacción en cadena de la polimerasa cuantitativa)
- GWAS** “*genome-wide association studies*” (estudio de asociación del genoma completo)
- BLAST** “*Basic Local Alignment Search Tool*” (herramienta de búsqueda por alineamiento local básico)
- COG** “*clusters of orthologous groups of proteins*” (conglomerados de grupos ortólogos de proteínas).
- NCBI** “*National Center for Biotechnology Information*” (Centro Nacional para Información en Biotecnología, E.E.U.U.)
- T<sub>m</sub>** “*melting temperature*” (temperatura de fusión)
- C<sub>t</sub>** “*cycle threshold*” (umbral del ciclo)
- VP** verdadero positivo
- VN** verdadero negativo
- FP** falso positivo
- FN** falso negativo



UNIVERSIDAD DE  
COSTA RICA

SEP Sistema de  
Estudios de Posgrado

**Autorización para digitalización y comunicación pública de Trabajos Finales de Graduación del Sistema de Estudios de Posgrado en el Repositorio Institucional de la Universidad de Costa Rica.**

Yo, Camilo Monye Casante, con cédula de identidad 113690025, en mi condición de autor del TFG titulado Identificación pan-genómica de secuencias firma para la detección de copes del clado C-I de Clostridioides difficile mediante qPCR

Autorizo a la Universidad de Costa Rica para digitalizar y hacer divulgación pública de forma gratuita de dicho TFG a través del Repositorio Institucional u otro medio electrónico, para ser puesto a disposición del público según lo que establezca el Sistema de Estudios de Posgrado. SI  NO \*

\*En caso de la negativa favor indicar el tiempo de restricción: 1 año (s).

Este Trabajo Final de Graduación será publicado en formato PDF, o en el formato que en el momento se establezca, de tal forma que el acceso al mismo sea libre, con el fin de permitir la consulta e impresión, pero no su modificación.

Manifiesto que mi Trabajo Final de Graduación fue debidamente subido al sistema digital Kerwá y su contenido corresponde al documento original que sirvió para la obtención de mi título, y que su información no infringe ni violenta ningún derecho a terceros. El TFG además cuenta con el visto bueno de mi Director (a) de Tesis o Tutor (a) y cumplió con lo establecido en la revisión del Formato por parte del Sistema de Estudios de Posgrado.

**INFORMACIÓN DEL ESTUDIANTE:**

Nombre Completo: Camilo Monye Casante

Número de Carné: A93943 Número de cédula: 113690025

Correo Electrónico: cmoneyc@gmail.com

Fecha: 19-1-2021 Número de teléfono: 7019-9811

Nombre del Director (a) de Tesis o Tutor (a): César Rodríguez Sánchez

**FIRMA ESTUDIANTE**

Nota: El presente documento constituye una declaración jurada, cuyos alcances aseguran a la Universidad, que su contenido sea tomado como cierto. Su importancia radica en que permite abreviar procedimientos administrativos, y al mismo tiempo genera una responsabilidad legal para que quien declare contrario a la verdad de lo que manifiesta, puede como consecuencia, enfrentar un proceso penal por delito de perjurio, tipificado en el artículo 318 de nuestro Código Penal. Lo anterior implica que el estudiante se vea forzado a realizar su mayor esfuerzo para que no sólo incluya información veraz en la Licencia de Publicación, sino que también realice diligentemente la gestión de subir el documento correcto en la plataforma digital Kerwá.

## Capítulo I: Introducción

### Generalidades:

Según el Manual de Bacteriología Sistemática de Bergey (Paul De Vos, 2009) *Clostridium difficile*, recientemente reclasificada como *Clostridioides difficile* (Lawson, Citron, Tyrrell, & Finegold, 2016) es una bacteria Gram-positiva, peritríca, esporulante y anaerobia estricta. Crece entre los 25°C y 45°C, de forma óptima a 37°C. Se le considera un patógeno oportunista que provoca infecciones intestinales cuando ocurre disbiosis de la microbiota residente, principalmente durante la atención en centros de salud (Samarkos, Mastrogianni, & Kampourpoulou, 2018). Este microorganismo fue originalmente descrito en 1935 como componente normal de la microbiota fecal de los recién nacidos, no obstante, fue posteriormente encontrado en heces de animales y en diversos tipos de muestras ambientales (Burnham & Carroll, 2013).

*C. difficile* alterna entre una forma vegetativa y una endospora resistente a condiciones ambientales adversas. Para que se produzcan las infecciones por *C. difficile* (CDI), pacientes con factores que los predisponen a adquirir la enfermedad (ver abajo) deben entrar en contacto con esporas de cepas toxigénicas de la bacteria, las cuales persisten sobre superficies por largos periodos de tiempo (Eze, Balsells, Kyaw, & Nair, 2017).

La virulencia de *C. difficile* se relaciona principalmente con la expresión de toxinas codificadas en una secuencia de 19.6 kb denominada locus de patogenicidad (PaLoc) que incluye 5 genes: *tcdA* y *tcdB*, que codifican para las toxinas identificadas con la letra correspondiente, *tcdE*, que codifica para una proteína potencialmente implicada en su secreción de ambas toxinas; y *tcdR* y *tcdC*, que corresponden a reguladores positivos y negativos de la síntesis de toxinas (Figura 1, Cuadro 1).

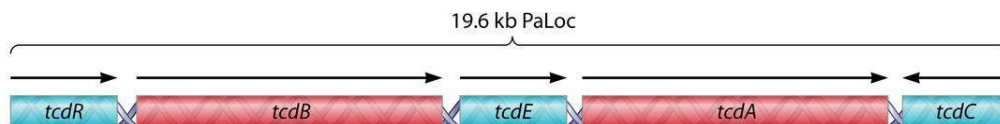


Figura 1. Estructura habitual del PaLoc de *C. difficile*. Tomado de (Hunt & Ballard, 2013)

Cuadro I. Genes usualmente presentes en el PaLoc de *C. difficile*

<b>Gen</b>	<b>Producto</b>	<b>Referencia</b>
<b><i>tcdA</i></b>	Toxina A	(Di Bella, Ascenzi, Siarakas, Petrosillo, & di Masi, 2016)
<b><i>tcdB</i></b>	Toxina B	(Di Bella et al., 2016)
<b><i>tdcE</i></b>	Holina para secreción de toxina	(Govind & Dupuy, 2012; Govind, Fitzwater, & Nichols, 2015)
<b><i>tcdR</i></b>	Regulador positivo de toxina	(Martin-Verstraete, Peltier, & Dupuy, 2016)
<b><i>tcdC</i></b>	Regulador negativo de toxina	(Martin-Verstraete et al., 2016)

Las toxinas TcdA y TcdB ejercen su efecto tóxico al inactivar proteínas Rho pequeñas y otras GTPasas de células eucariotas, interrumpiendo así vías metabólicas vitales en el hospedero (Burnham & Carroll, 2013).

Un grupo menor de cepas expresa otra toxina, denominada *C. difficile* transferasa (CDT) o toxina binaria. Esta enzima presenta actividad ADP-ribosiltransferasa y genera disrupción del citoesqueleto, promoviendo pérdida de fluido y eventualmente la muerte celular. Como su nombre lo indica, se compone de dos componentes: CdtA, que posee actividad catalítica, y CdtB, involucrada en la unión al sustrato. Los genes que codifican para estas proteínas se encuentran en el denominado CdtLoc (Elliott, Androga, Knight, & Riley, 2017), que se localiza en un sitio cromosomal diferente al PaLoc. Típicamente se considera que las cepas que no expresan TcdA o TcdB no causan CDI en humanos, sin embargo, como se verá más adelante, el papel de estas denominadas cepas no toxigénicas en la patogénesis aún no es claro.

Además de las toxinas TcdA, TcdB y CDT, se han descrito otros factores de virulencia en esta especie, tales como la proteína de capa S, proteínas de flagelos, fimbrias y proteínas de unión a fibronectina (Awad, Johanesen, Carter, Rose, & Lyras, 2014).

### Factores de riesgo:

Una serie de factores de riesgo han sido asociados al desarrollo de las CDI, incluyendo admisión a centros de salud (y tiempo de internamiento), edad avanzada (relacionada con senescencia inmune), presencia de comorbilidades y, en particular, exposición a agentes antimicrobianos (Elliott et al., 2017), ya que estas drogas inducen cambios inmediatos y profundos en la microbiota intestinal endógena y reducen su diversidad. Esta disbiosis disminuye la resistencia a la colonización de *C. difficile* y modifica el metabolismo de las sales biliares (Winston & Theriot, 2016). Prácticamente todos los agentes antimicrobianos de uso clínico conocido han sido implicados con las CDI, sin embargo, metaanálisis recientes han encontrado diferencias importantes en la magnitud de asociación al disgregar los datos según factores, como sitio geográfico o el contexto hospitalario de los casos. Así, en comparación con controles adiarreicos, la administración hospitalaria de clindamicina, cefalosporinas, carbapenémicos, fluoroquinolonas y sulfonamidas se asocia con un riesgo al menos dos veces mayor de CDI. En casos comunitarios dicho riesgo aumenta entre 8 y 20 veces para clindamicina, y entre 3 y 5 veces para cefalosporinas y quinolonas (Eze et al., 2017).

### Presentación clínica:

Las CDI presenta un amplio espectro de manifestaciones clínicas, entre las que se incluyen diarrea (que puede desencadenar íleo severo), colitis pseudomembranosa y/o fulminante. Entre sus síntomas destacan dolor abdominal, anorexia, fiebre y malestar general. Las heces de los pacientes suelen ser líquidas y poseer un olor desagradable y característico (Bartlett & Gerding, 2008). Las complicaciones de esta patología incluyen internamiento en unidades de cuidado intensivo, colectomía e incluso la muerte (Pacheco & Johnson, 2013).

### Importancia clínica de las CDI:

A pesar de haber sido reconocido como un patógeno importante en el desarrollo de colitis pseudomembranosa desde 1970 (George et al., 1978), la incidencia y severidad de los casos de CDI ha aumentado en los últimos 15-20 años (He et al., 2013). Este aumento es atribuible, al menos en parte, a la aparición de nuevas cepas hipervirulentas (Knight,

Elliott, Chang, Perkins, & Riley, 2015; Rodriguez, Van Broeck, Taminiou, Delmée, & Daube, 2016).

La infección por este patógeno se ha convertido en un considerable problema de salud pública en numerosos países, incluido Costa Rica. Prueba de lo anterior es que en agosto del año anterior fue catalogado por los Centros para Control y Prevención de Enfermedades de E.E.U.U como una de las más grandes amenazas, otorgándole un nivel de peligro urgente.

En 2009-2010 se reportó un brote importante de CDI en nuestro país (Quesada-Gomez et al., 2010; Wong-McClure et al., 2012), y en años recientes se ha detectado la emergencia o predominancia de cepas con alta resistencia a antibióticos (López-Ureña et al., 2016).

#### Diagnóstico y tipificación:

La detección temprana de *C. difficile* y sus toxinas es crucial, pues permite un tratamiento más temprano y así disminuir significativamente la morbilidad y mortalidad de las infecciones y los costos médicos asociados a atenciones hospitalarias prolongadas (Peng et al., 2018).

Además de considerar los síntomas clínicos anteriormente mencionados, el diagnóstico de CDI requiere de resultados positivos de pruebas de laboratorio que persiguen detectar el patógeno y/o sus toxinas. Sin embargo, no existe un ensayo que posea sensibilidad, especificidad y capacidad predictiva óptimas (Pacheco & Johnson, 2013).

Entre los ensayos más comunes destacan:

- a) el cultivo toxigénico (TC, del inglés “*toxigenic culture*”), que detecta el efecto de toxinas presentes en suspensiones de heces sobre cultivos celulares. Su sensibilidad y especificidad son altas, pero en ciertos casos puede tardar días para generar resultados. Aunado a esto, el cultivo celular no se realiza en todos los laboratorios clínicos, lo que limita el acceso a este tipo de ensayos.
- b) los ensayos de neutralización de citotoxicidad celular (CCNA, del inglés “*cell cytotoxicity neutralization assay*”), que detectan toxinas presentes en heces mediante la observación de su efecto sobre cultivos celulares en presencia y ausencia de antitoxina(s). Estos ensayos poseen ventajas y desventajas similares a las del cultivo toxigénico.

- c) los inmunoensayos para toxina A y B (EIAs, del inglés “*enzyme immunoassay*”), que tienen bajo costo, requieren poco tiempo y son fáciles de realizar, pero son poco sensibles.
- d) los ensayos de amplificación de ácidos nucleicos (NAAT, del inglés “*nucleic acid amplification test*”), que normalmente tienen como blanco fragmentos de *tcdA*, *tcdB* y/o de los genes para la toxina binaria en combinación o solos.

Si bien se ha considerado al TC y a los CCNA como los dos estándares para diagnóstico, no se utilizan de forma rutinaria en la clínica debido a su alta demanda de tiempo y mano de obra calificada. Por otro lado, los NAAT poseen una serie de ventajas, tales como sensibilidad y especificidad > 80% y hasta 99% (Burnham & Carroll, 2013; Peng et al., 2018), baja complejidad (una vez que el ensayo se ha estandarizado) y poco tiempo de respuesta, además de ser altamente costo-efectivos.

A pesar de sus ventajas, existen algunos cuestionamientos con respecto a la utilidad de los NAAT en la clínica. Por ejemplo, si bien la detección de los genes usualmente correlaciona con la presencia de enfermedad, esto no es siempre cierto porque las toxinas no necesariamente se expresan. Además, existen reportes de cepas virulentas que carecen de *tcdA* (B. Wang, Peng, Zhang, & Su, 2018) o *cdtAB* (Marvaud, Quevedo-Torres, Eckert, Janoir, & Barbut, 2019), y dado que el tipo de cepas causantes varía en el tiempo y el espacio, es necesario revalidar los NAAT constantemente.

*C. difficile* es una especie genéticamente heterogénea y su eficiente tipificación es crítica para el análisis epidemiológico, particularmente durante los brotes (Awad et al., 2014). Este tipo de análisis también es requerido para diferenciar entre cepas circulantes, cepas nuevas y variantes.

Se han aplicado una serie de métodos de tipificación a *C. difficile*, sin que alguno predomine como estándar de oro, a saber:

- a) La electroforesis de campo pulsante (PFGE, del inglés “*pulsed-field gel electrophoresis*”), que fue uno de los primeros métodos de tipificación disponibles, sin embargo, presenta poca capacidad de discriminación.
- b) El ribotipado, que se basa en la identificación de patrones de secuencias localizadas entre los operones ribosomales y que ha sido particularmente utilizado en Europa y muestra una alta capacidad discriminatoria, pero poca reproducibilidad.

- c) La tipificación por secuencias multilocus (MLST, del inglés “*multilocus sequence typing*”), en la que se secuencian seis o siete genes conservados y los alelos se comparan contra los depositados en bases de datos curadas (Peng et al., 2018).
- d) Más recientemente, el bajo costo por base de los nuevos métodos de secuenciación masiva (NGS, del inglés “*next generation sequencing*”) y avances computacionales ha permitido la comparación de miles de genomas y la consecuente clasificación de *C. difficile* en 8 clados (Janezic, Potocnik, Zidaric, & Rupnik, 2016).

#### Prevención y control de las CDI a nivel hospitalario:

Las estrategias de manejo de las CDI se centran en la prevención primaria o de recurrencia (Khanna & Gerding, 2019). El documento para la prevención de CDI en escenarios de atención médica aguda elaborado por la Sociedad Europea de Microbiología Clínica y Enfermedades Infecciosas (ESCMID, del inglés “*European Society of Clinical Microbiology and Infectious Diseases*”) establece que restricciones en el uso de antibióticos reducen la incidencia de CDI en condiciones de brote y/o endémicas (Tschudin-Sutter et al., 2018). De forma similar, las guías de práctica clínica para CDI en adultos y niños elaboradas por la Sociedad Americana de Enfermedades Infecciosas (IDSA, del inglés “*Infectious Diseases Society of America*”) y la Sociedad para la Epidemiología de la Atención Médica de América (SHEA, del inglés “*Society for Healthcare Epidemiology of America*”) recomiendan minimizar la frecuencia y duración del uso de antibióticos de alto riesgo, así como el número de agentes antibióticos prescritos, para reducir el riesgo de CDI (McDonald et al., 2018).

Es probable que la resistencia a ciertos antibióticos que algunas cepas de *C. difficile* han adquirido sea importante para explicar la ocurrencia de brotes a nivel local, y se relacione con el buen o mal uso de los antibióticos en dichas regiones, como se evidencia con el caso del ribotipo 027 en Inglaterra (Wieczorkiewicz et al., 2015). De esta manera, es importante que los programas de vigilancia del uso de antibióticos dispongan de los recursos necesarios para llevar un adecuado control, para eventualmente prever la aparición de cepas resistentes de *C. difficile* que podrían causar brotes de CDI. Los gastos incurridos en la vigilancia se pueden justificar a nivel de salud pública, pues el mantenimiento de una buena vigilancia tiene efectos positivos como reducción en efectos

adversos, incremento en susceptibilidad y mejora en la resolución de los pacientes (Barlam et al., 2016).

Múltiples recomendaciones propuestas por las guías anteriormente mencionadas se relacionan con evitar la propagación del agente causal en los centros de salud. Las manos del personal contaminadas con esporas, así como la contaminación ambiental, son probablemente los medios que facilitan este proceso, por lo que la limpieza y precauciones por contacto ("*contact precautions*") son de suma importancia para mitigar el riesgo de contagio en los centros hospitalarios (McDonald et al., 2018). De esta forma, es recomendable que los pacientes con CDI se mantengan en cuartos privados con, de ser posible, un baño independiente. El equipo de protección personal (guantes y batas desechables) representa una barrera crítica para prevenir la transmisión de esporas entre pacientes y el personal de salud, por lo que se recomienda su uso desde antes de entrar en contacto con pacientes con CDI o su ambiente (Napolitano & Edmiston, 2017). Además de lo anterior, la buena higiene de manos se considera una piedra angular en la prevención de la transmisión de la CDI, recomendándose el lavado con agua y jabón, dado que las esporas de *C. difficile* son altamente resistentes al alcohol (Tschudin-Sutter et al., 2018). El equipo utilizado para el tratamiento del paciente y el espacio físico donde este se encuentra, es frecuentemente contaminado con esporas de *C. difficile*, por lo que se recomienda el uso de equipo desechable y la desinfección profunda del equipo reutilizable y cuartos donde se trata a los pacientes con CDI (Napolitano & Edmiston, 2017).

Los pacientes sintomáticos para CDI son más propensos a contaminar el medio y a otras personas que los asintomáticos, por lo tanto, el tratamiento temprano es de suma importancia para controlar la transmisión de esta enfermedad (Tschudin-Sutter et al., 2018). Dado que el tratamiento debe ser administrado a los pacientes una vez confirmado el diagnóstico (Debast, Bauer, & Kuijper, 2014), es importante contar con métodos de diagnóstico rápidos, certeros y accesibles que permitan una intervención temprana.

El control de la CDI también se logra a través de una vigilancia epidemiológica estricta, con el objetivo de comparar las tasas de incidencia y detectar posibles brotes de la enfermedad en sus estados iniciales, de forma tal que se puedan redoblar los esfuerzos descritos anteriormente para controlarlos de manera oportuna (Tschudin-Sutter et al., 2018). Las guías de E.E.U.U recomiendan llevar una vigilancia mínima reportando los casos de CDI de advenimiento en el centro de salud ("*healthcare facility-onset*"), definidos

como aquellos eventos suscitados luego del tercer día de internamiento. También es recomendable el monitoreo de los casos con advenimiento en la comunidad asociados a centros de salud, es decir, aquellos que ocurren en el transcurso de 28 días luego de la salida de un centro hospitalario (McDonald et al., 2018). Es crucial que las tasas de incidencia sean comunicadas efectivamente a los directores o coordinadores, para que ante un eventual aumento se asegure el acatamiento de las medidas de control de infección, o el incremento de estas.

En Costa Rica, el Sistema de Vigilancia Epidemiológica monitorea la incidencia de las Infecciones Asociadas a la Atención en Salud, que incluyen la CDI. Las medidas de control para brotes de esta enfermedad son similares a las descritas en las guías anteriormente citadas, e incluyen medidas de aislamiento por cohorte, restricción de personal y visitas en las zonas de aislamiento, uso estricto de equipo de protección personal, limpieza y desinfección de la zona de aislamiento, desinfección de ropa contaminada y manejo apropiado de desechos bioinfecciosos y restricción de medicamentos (Mairena-Morera, 2014).

El manejo clínico inicial de la CDI normalmente implica la discontinuación del uso de antibióticos (McDonald et al., 2018). Los antibióticos continúan siendo el tratamiento usual para la CDI. Típicamente se ha utilizado vancomicina y metronidazol como métodos de tratamiento. Más recientemente se ha añadido el uso de fidaxomicina como alternativa al uso de vancomicina. En casos de recurrencia se suele utilizar vancomicina si la infección anterior fue tratada con metronidazol, mientras que, si se utilizó vancomicina en la primera infección, para la segunda se suele utilizar vancomicina por pulsos o fidaxomicina (Khanna & Gerding, 2019).

Existen otras terapias para el tratamiento de CDI, sin embargo, su uso es limitado a ciertos casos. Por ejemplo, anticuerpos monoclonales anti-toxina B (bezlotoxumab) han demostrado ser útiles para prevenir la recurrencia de CDI en pacientes de riesgo (mayores de 65 años, historial previo de CDI, inmunocomprometidos, etc) (Gerding et al., 2018). Además, también existen trasplantes de microbiota fecal para restaurar la microflora normal intestinal en pacientes con CDI, sin embargo, su uso se suele restringir a pacientes con recurrencia múltiple (McDonald et al., 2018; Peng et al., 2018).

### Epidemiología de las CDI:

En los últimos 10-15 años, las principales cepas responsables de brotes de CDI han sido las del ribotipo 027/patrón de electroforesis NAP1 (Robinson, Auchtung, Collins, & Britton, 2014). Su mayor capacidad epidémica se ha asociado a su resistencia aumentada a las fluoroquinolonas (Spigaglia et al., 2008), así como a su capacidad de sobreproducción de toxinas (McDonald et al., 2005) y esporas (Akerlund et al., 2008), entre otros factores.

El primer análisis genético de la población y relación filogenética de *C. difficile* fue realizado por Lemee y colaboradores en 2004 (Lemee, Dhalluin, Pestel-Caron, Lemeland, & Pons, 2004), usando MLST (ST, del inglés “*Sequence Type*”). En este estudio se identificaron 34 secuencias tipo de MLST entre 72 cepas, las cuales se distribuían a su vez en tres clados, que no se asociaban a huéspedes ni zonas geográficas específicas. Estudios adicionales de comparación de genoma completo han confirmado esta organización cladística, y a la fecha se cuentan 5 clados numerados del uno al cinco, y tres clados crípticos denominados C-I, C-II y C-III.

El clado más heterogéneo es el Clado 1, que incluye al menos 200 STs. Este clado incluye además varias cepas de importancia clínica en Europa, como los ribotipos 014 (ST-2), 002 (ST-8), 001 (ST-3), entre otros. El Clado 2 es el segundo más heterogéneo, con 61 STs, e incluye los ribotipos epidémicos 027 (ST-1), 176 (ST-1) y 244 (ST-41). El Clado 3 incluye 13 STs diferentes, de los cuales el más representativo es el 023 (ST-22). El Clado 4 se asocia con cepas de ribotipo 017 (ST-37), las cuales poseen la particularidad de poseer toxina B, pero no así la A. El miembro más conocido del Clado 5 es el ribotipo 078 (ST-11), el cual en los últimos 10 años ha emergido como causante de CDI en humanos, a pesar de anteriormente ser reconocido como un patógeno de animales (Janezic, Garneau, & Monot, 2018).

Recientemente se han descrito cepas toxigénicas en el Clado C-I (Monot et al., 2015; Ramírez-Vargas et al., 2018). Los Clados C-II y C-III fueron definidos al analizar muestras ambientales, y sólo aparecen esporádicamente en muestras clínicas, por lo que se supone que el hábitat natural de sus miembros es el ambiente (Janezic et al., 2018). Es pertinente agregar además que, dada la notable divergencia de estos grupos, se ha sugerido que podrían corresponder a nuevas especies (Knight et al., 2020).

En los últimos años se ha reportado un aumento en la incidencia de CDI en la población general, con un incremento en los casos de advenimiento comunitario de la enfermedad (Thornton et al., 2018). Este aumento podría estar relacionado con cambios a nivel de

ambiente comunitario que favorecen la emergencia de *C. difficile* en la población general (Ogielska et al., 2015). Los animales de granja, productos cárnicos y otros relacionados podrían ser un reservorio de cepas toxigénicas de *C. difficile* (Brown & Wilson, 2018). Se han identificado cepas toxigénicas en caballos, cerdos, ganado vacuno, así como en animales domésticos como perros y gatos (Weese, 2020) principalmente del ribotipo 078, así como de otros ribotipos relacionados con epidemias en seres humanos (014, 027) (Rabold et al., 2018). Esta evidencia sugiere que existe un importante potencial de transmisión zoonótica de la CDI, aunque aún es necesario realizar más estudios para comprobar esta hipótesis (Knight & Riley, 2019).

#### Epidemiología de las CDI en Costa Rica:

Entre diciembre de 2009 y abril de 2010 se presentó un brote nosocomial de diarrea asociada a *C. difficile* en el Hospital San Juan de Dios, reportado inicialmente por un aumento de casos de 0.3 y 0.6 a 2.4 por cada 10000 pacientes/día. Durante el periodo total se identificaron 84 casos, para una incidencia de entre 3.1 y 6.7 casos por cada 10000 pacientes/día. El brote fue finalmente declarado controlado en julio de 2010, gracias a la aplicación de diversas medidas de higiene intrahospitalarias y restricción de uso de antibióticos. En su reporte, los autores mencionan que la mayor limitación del estudio fue la poca sensibilidad del método diagnóstico utilizado, e indican además que el sistema de vigilancia nosocomial debe mejorar la sensibilidad de los métodos diagnósticos de *C. difficile* para prevenir casos futuros (Quesada-Gomez et al., 2010; Wong-McClure et al., 2012).

En un estudio inicial con 37 aislamientos obtenidos a partir de muestras de este brote se identificaron nueve patrones de electroforesis de campo pulsante, todos positivos para *tcdA* y *tcdB*, susceptibles a vancomicina y metronidazol. Más de la mitad de estos mostraron el patrón de macrorestricción NAP1, los cuales poseían, además, alta resistencia a fluoroquinolonas. También se encontraron aislamientos con patrones NAP2 y NAP9 (Quesada-Gomez et al., 2010). En un estudio subsecuente se identificaron 68 aislamientos de *C. difficile* obtenidos a lo largo de dos años en los cuales no se presentaron brotes de CDI. Interesantemente, en este periodo predominó un genotipo local designado como NAP<sub>CR1</sub> (26%) en conjunto con otras 14 cepas típicamente asociadas a infecciones intrahospitalarias (i.e. NAP2, NAP4 y NAP6), así como 8 genotipos nuevos. En este periodo no se aislaron cepas NAP1, y se encontraron 7 cepas

no toxigénicas a partir de las muestras de heces de los pacientes analizados (López-Ureña et al., 2016).

Más recientemente se ha demostrado que aislamientos de la cepa NAP<sub>CR1</sub> se caracterizan por ser multirresistentes a los antibióticos. Esta resistencia puede ser al menos parcialmente explicada por la presencia de mutaciones en *gyrA* y *rpoB* y la presencia de genes adquiridos en diversos elementos móviles (Ramírez-Vargas et al., 2017).

Un último hallazgo reciente con respecto a *C. difficile* en nuestro país fue la identificación de tres aislamientos pertenecientes al Clado C-I en muestras de heces de pacientes humanos diarreicos. Estas cepas escapan pruebas diagnósticas basadas en PCR e inmunológicas comúnmente utilizadas en laboratorios clínicos, además de que representan nuevos ribotipos, tipos de electroforesis de campo pulsante y STs. Contienen, además, PaLocs atípicos, localizados en moléculas extracromosomales circulares, y variantes en las secuencias de CdtA y CdtB con respecto a la cepa de referencia R20291. A pesar de lo anterior, uno de estos aislamientos mostró efectos citotóxicos y mortalidad en hámsteres similares a los causados de una cepa NAP1 epidémica (Ramírez-Vargas et al., 2018).

#### Pangenómica y secuencias firma:

El llamado pangenoma de una especie bacteriana está compuesto por el genoma central (*core genome*), que incluye los genes responsables de funciones básicas de la biología de la especie y es compartido por todas las cepas conocidas de la misma, y el genoma accesorio o dispensable, que contiene genes adaptativos para su nicho y es específico de una cepa o grupo de cepas (Chaudhari, Gupta, & Dutta, 2016). *C. difficile* se caracteriza por poseer un pangenoma amplio y abierto, con una gran cantidad de genes accesorios y un componente central pequeño (Elliott et al., 2017). Así, las cepas de *C. difficile* pueden diferir de forma sustancial en su contenido génico.

Las secuencias firma (en inglés “*signature sequences*”) son secuencias de longitud variable específicas para una especie, cepa, o grupo particular de estas (Karimi & Hajdu, 2016). Su presencia correlaciona con la identidad del taxón de interés, por lo que tienen un valor diagnóstico. Este concepto ha sido explotado en el desarrollo de métodos de detección basados en amplificación o hibridación de ADN, o bien, el reconocimiento de

péptidos o proteínas con anticuerpos o aptámeros para bacterias de interés clínico (Hung, Nagamine, Li, & Lo, 2012; Nagamine, Hung, Li, & Lo, 2015; Phillippy et al., 2007).

*Aportes de la genómica al estudio y manejo en enfermedades infecciosas:*

La disminución del costo de la secuenciación de genomas y la disponibilidad de genomas completos en bases de datos públicas han revolucionado el campo de las enfermedades infecciosas. La genómica microbiana ha permitido el diseño de ensayos diagnósticos basados en herramientas moleculares como la PCR. Estos ensayos no solo permiten muchas veces la detección e identificación de patógenos directamente en matrices complejas, sin necesidad de realizar cultivo (Logares et al., 2012), sino que también posibilitan la selección de diferentes blancos según su objetivo.

La detección de microorganismos por PCR en especímenes clínicos ha sido utilizada por un tiempo considerable, en parte por su capacidad de identificar microorganismos difíciles de cultivar en el laboratorio. Si bien inicialmente los blancos para estos ensayos se determinaban empíricamente, hoy en día es factible diseñar ensayos específicos para secuencias presentes en un microorganismo o grupo particular, gracias a la gran cantidad de genomas de especies bacterianas disponibles en bases de datos públicas y la accesibilidad de las aplicaciones bioinformáticas necesarias para lograr tal fin.

Actualmente existen varios ensayos y tecnologías comercialmente disponibles para la detección de microorganismos patógenos basados en PCR, tales como SeptiFast (Roche®) (Korber et al., 2017), el cual permite la identificación simultánea de una importante variedad de microorganismos en sangre, y GeneXpert, que ha sido recomendado por la Organización Mundial de la Salud para el diagnóstico de tuberculosis (Saeed et al., 2017), solamente para citar dos ejemplos.

Además del diseño de nuevos métodos de detección de patógenos, la disminución en el costo de la secuenciación de genoma completo ha posibilitado el desarrollo y aplicación de métodos de genotipado basados en esta metodología, siendo el análisis de MLST el más comúnmente utilizado actualmente. El sitio web PubMLST.com lista alelos y STs de más de 100 especies microbianas, muchas de importancia clínica, como *C. difficile*, *Escherichia coli* y *Staphylococcus aureus*, permitiendo la caracterización genotípica de aislamientos de interés a partir de los datos de secuenciación (Jolley, Bray, & Maiden, 2018).

La secuenciación de genoma completo posibilita también la detección de marcadores genéticos de interés clínico, tales como factores de virulencia y genes relacionados con resistencia a antibióticos, lo cual puede ser relevante para el manejo de un paciente o de posibles brotes (Fournier, Dubourg, & Raoult, 2014).

El reciente aumento en la cantidad de genomas completos depositados en bases de datos públicas ha permitido la ejecución de estudios de pangenómica, que permiten, por ejemplo, determinar cuáles genes del genoma se comparten entre distintas cepas de interés y cuáles están presentes únicamente en algunos grupos. Este tipo de estudio aplicado a microorganismos de interés clínico ha repercutido en la identificación de nuevos blancos para el desarrollo de vacunas, terapias y métodos de diagnóstico (Anani, Zgheib, Hasni, Raoult, & Fournier, 2020).

## Capítulo II: Justificación, hipótesis y objetivos

### Justificación:

El uso de métodos de diagnóstico sensibles, sencillos y rápidos es vital para iniciar el tratamiento y evitar la diseminación de *C. difficile* dentro y fuera de los centros hospitalarios. Los métodos moleculares existentes están diseñados principalmente para detectar las cepas típicamente descritas como virulentas (p. ej. NAP1/027/ST01) y se enfocan primordialmente en elementos del PaLoc, lo que excluye una gran parte de la población de *C. difficile* que podría estar presente en algunos pacientes. Un ejemplo de ello son las cepas Clado C-I con potencial patogénico que se han detectado por cultivo en pacientes costarricenses con CDI, ya que sus genomas y genes del PaLoc son altamente divergentes (Ramírez-Vargas & Rodríguez, 2020).

Este proyecto pretende identificar secuencias firma de cepas del Clado C-I mediante genómica comparativa y usarlas en el diseño de un método de identificación por PCR cuantitativa en tiempo real (qPCR) más exhaustivo que los disponibles hoy en día comercialmente. Este abordaje se escogió porque el qPCR es ampliamente usado en los laboratorios clínicos y porque su implementación podría simplemente implicar la adición de un par adicional de oligonucleótidos a las reacciones usuales.

Además de la relevancia diagnóstica de este esfuerzo, la identificación de secuencias firma contribuiría al poco conocimiento que se tiene sobre las cepas del Clado C-I, ya que potencialmente permitiría identificar genes relacionados con su anticipada inusual biología y facilitar su estudio epidemiológico.

Se pretende que el método detecte *C. difficile* de Clado C-I en heces y cultivo, por lo que podría utilizarse tanto en la clínica como fuera de centros hospitalarios, como por ejemplo, en estudios de veterinaria o microbiología ambiental para indagar sobre los casos comunitarios y el posible componente zoonótico de esta enfermedad (Janezic et al., 2016; Knetsch et al., 2018).

Hipótesis:

Las cepas de *Clostridioides difficile* que pertenecen al Clado C-I poseen secuencias nucleotídicas exclusivas susceptibles de ser amplificadas diferencialmente por qPCR a partir de cultivos o heces.

Objetivo general:

- Diseñar un método de diagnóstico molecular para cepas del Clado C-I de *Clostridioides difficile* basado en secuencias firma identificadas mediante pangenómica.

Objetivos específicos:

- I. Utilizar pangenómica para identificar secuencias firma en el Clado C-I de *C. difficile*.
- II. Diseñar un método basado en qPCR que posibilite la detección de cepas del Clado C-I de *C. difficile* en cultivo y heces.

### Capítulo III: Metodología

#### Genomas:

En el presente proyecto se incluyeron 77 genomas calidad “draft”, distribuidos en un grupo de inclusión compuesto por 10 cepas de *C. difficile* del Clado C-I y un grupo de exclusión compuesto por 67 genomas de *C. difficile* pertenecientes a los Clados 1-5.

Específicamente, 40 del Clado 1, 7 del Clado 2, 6 del Clado 3, 8 del Clado 4 y 7 del Clado 5 (ver Anexo 1). Cinco de los genomas del grupo de inclusión fueron secuenciados y ensamblados en el Laboratorio de Investigación en Bacteriología Anaerobia (LIBA) de la Universidad de Costa Rica, los cuales corresponden a cepas toxigénicas del Clado C-I aisladas en Costa Rica entre los años 2014 y 2017 a partir de heces de pacientes diarreicos (Ramírez-Vargas et al., 2018). Los 72 genomas restantes fueron ensamblados a partir de *reads* de Illumina pareados descargados usando el NCBI SRA Toolkit (Leinonen, Sugawara, & Shumway, 2011). Estos genomas corresponden a aislamientos del Reino Unido y Australia obtenidos entre los años 2005 y 2012 y reportados por Dingle y colaboradores en un estudio de historia evolutiva del PaLoc (Dingle et al., 2014).

Se empleó KmerGenie ver. 1.7049 (Chikhi & Medvedev, 2014) para evaluar la mejor longitud de *k-mer* teórica para ensamblar *de novo* cada juego de *reads*, usando subgrupos de 5 genomas según la longitud de *read*. Este software estima la longitud de *k-mer* ideal para un ensamblaje *de novo* (Chikhi & Medvedev, 2014). Los *reads* fueron ensamblados usando SPAdes (3.12.0) (Bankevich et al., 2012) en el modo predeterminado para genomas procariotas y usando la longitud de *k-mer* calculada usando KmerGenie, en ambos casos habilitando el método de corrección de *reads* BayesHammer con *scaffolding*. Este software utiliza gráficos de Bruijn multilongitudinales, además, posee un paso de ensamblaje iterativo, de forma tal que permite aproximar la longitud de *k-mer* más apropiada de forma automática. Incluye además un módulo de corrección de *reads* de Illumina (BayesHammer) y un corrector de errores de apareamiento. Es apropiado para el ensamblaje de genomas pequeños y se ha demostrado que genera N50 superiores a otros ensambladores manteniendo una alta cobertura genómica (Bankevich et al., 2012). En este contexto, el valor N50 es una mediana ponderada, de forma tal que 50% del ensamblaje total está contenido en *contigs* o *scaffolds* de tamaño igual o mayor a este valor.

La calidad de los genomas ensamblados fue analizada utilizando la herramienta QUAST (Gurevich et al., 2013) con las cepas *C. difficile* 630, R20291, VL-0391, M68, y M120 como referencias para las cepas pertenecientes a los Clados 1, 2, 3, 4 y 5, respectivamente. Se eliminaron ensamblajes con una longitud menor a 3,8 Mbp, pues esta corresponde a la menor longitud esperable para *C. difficile* y se eligió el mejor ensamblaje para cada cepa según mayor N50 y menor L50, donde L50 corresponde a la cantidad de secuencias necesarias para alcanzar el N50. Los valores de N50 obtenidos fueron comparados con los reportados originalmente en el trabajo de Dingle y colaboradores (Dingle et al., 2014), obteniéndose solo 4 ensamblajes con menor valor. Estos ensamblajes fueron aún así conservados para mantener la homogeneidad en el tamaño del set de datos.

Los ensamblajes escogidos fueron anotados con Prokka (1.14.6) (Seemann, 2014) usando una base de datos de proteínas pertenecientes a las anotadas en el genoma de la cepa *C. difficile* 630 (AM180355), dado que estas han sido curadas manualmente por el Instituto Sanger. Prokka coordina varias herramientas que permiten anotar secuencias codificantes para proteínas y ARN no codificantes, incluyendo Prodigal, RNAmmer, Aragorn, SignalP e Infernal (Seemann, 2014).

#### Genómica comparativa:

Con el fin de verificar la designación cladística de las cepas incluidas en el análisis, se realizó un alineamiento de genoma *core* con Roary (Page et al., 2015), del cual se extrajeron los sitios variables para construir un alineamiento Clustal Omega y un dendrograma con SeaView (Gouy, Guindon, & Gascuel, 2010), el cual fue visualizado y editado en TreeGraph 2 (Stöver & Müller, 2010). Asimismo, se determinaron los valores de identidad nucleotídica promedio (ANI) por pares entre todas las cepas de ambos grupos usando FastANI (Jain, Rodriguez-R, Phillippy, Konstantinidis, & Aluru, 2018).

#### Identificación de secuencias firma:

Para determinar las secuencias firma de las cepas Clado C-I se utilizaron tres técnicas: a) GWAS, b) procesos automatizados basados en referencia y pareo de *k-mer* y c) análisis pangenómicos.

Para llevar a cabo los estudios de asociación genética o GWAS (Falush, 2016) se utilizaron los programas DBGWAS, TreeWAS y Roary/Scoary. DBGWAS realiza GWAS basado en *k*-mer extendidos y gráficos de Bruijn compactados. Durante el proceso, los nodos del gráfico se prueban por asociación con la característica de interés (en este caso, pertenecer al grupo de inclusión). La salida consiste en regiones del gráfico que contienen nodos significativamente asociados con la característica de interés, a decir, pertenencia al Clado C-1) (Jaillard et al., 2018). Dada la sobre-representación de cepas del Clado 1 y la cantidad de *k*-mer significativos esperados, se realizó un primer análisis exploratorio con opciones predeterminadas, y uno posterior usando *SFF* (*significant features filter*)=70, *nh* (*neighbourhood*)=3, *k*=47 y *maf* (*minor allele frequency*)=0,0014, pretendiendo que los componentes resultantes tuvieran menos nodos y fueran más sencillos de visualizar.

Roary es una herramienta que construye pangenomas e identifica genes accesorios a partir de las secuencias codificantes de un grupo de genomas. Usando BLASTp, CD-HIT y agrupamiento por MCL, este software construye un gráfico con la relación de grupos de genes ortólogos y parálogos y una tabla de presencia/ausencia de estos genes en los genomas analizados (Page et al., 2015). Esta información es usada por Scoary para hacer un “pan-GWAS”, en donde cada gen candidato en el genoma accesorio se puntea según su correlación aparente con una característica (en este caso, pertenecer al grupo de inclusión). Aquellos genes que pasan el primer cribado son reanalizados incorporando la información de la estructura filogenética de la muestra, infiriendo la estructura poblacional a partir de los datos (Brynildsrud, Bohlin, Scheffer, & Eldholm, 2016).

La tabla de presencia/ausencia generada por Roary fue utilizada para realizar un tercer análisis con TreeWAS (Collins & Didelot, 2018), el cual utiliza un acercamiento al GWAS basado en árboles filogenéticos, midiendo la asociación estadística entre una característica de interés y el genotipo de todos los *loci*, identificando asociaciones significativas y corrigiendo por el efecto de población clonal. Roary, Scoary y TreeWAS fueron empleados con las opciones predeterminadas usando los mismos grupos de inclusión y exclusión.

Adicionalmente, para identificar las secuencias firma de forma automatizada se utilizó la herramienta Neptune, que identifica *k*-mers conservados en el grupo de inclusión y ausentes en el de exclusión. Las secuencias encontradas son evaluadas y filtradas realizando un análisis de especificidad utilizando alineamientos en parejas. Aquellas secuencias que pasan este filtro son clasificadas según su puntuación de especificidad y

sensibilidad, usando rangos de confianza (Marinier et al., 2017). Se utilizaron los grupos de inclusión y exclusión mencionados anteriormente, especificando que las secuencias extraídas estuvieran presentes en todos los genomas de inclusión, ausentes en todos los genomas de exclusión, y que fueran al menos de 200 pares de bases.

Para la identificación de secuencias mediante análisis pangenómico se utilizó el flujo de trabajo para pangenómica de Anvi'o (Delmont & Eren, 2018; Eren et al., 2015). Este procedimiento genera una base de datos genómica que almacena las secuencias nucleotídicas y aminoacídicas, así como las anotaciones funcionales para cada gen en el genoma de ambos grupos incorporados en el análisis, para luego computar un pangenoma a partir de conglomerados de genes. El pangenoma obtenido permite visualizar los conglomerados de genes presentes o ausentes en los distintos grupos de genomas. Anvi'o se usó habilitando la opción de evaluación de función COG (del inglés "*clusters of orthologous groups of proteins*") (Tatusov, Galperin, Natale, & Koonin, 2000). Se corrió, además, un análisis de enriquecimiento funcional entre los grupos de inclusión y exclusión, y manualmente se identificó un *bin* con conglomerados de genes presentes exclusivamente en el grupo de inclusión.

Se determinó manualmente la concordancia entre las secuencias firma identificadas con Scoary y TreeWAS, verificando sus valores de  $p$ , sensibilidad y especificidad.

Posteriormente las secuencias identificadas con ambos métodos fueron revisadas contra las identificadas en el *bin* exclusivo reportado por Anvi'o, extrayendo manualmente la secuencia aminoacídica de Anvi'o, encontrando la secuencia en el genoma anotado, extrayendo su identificador único, y verificando en la salida consolidada de Scoary y TreeWAS.

El flujo de trabajo comentado en esta sección se resume en la Figura 2.

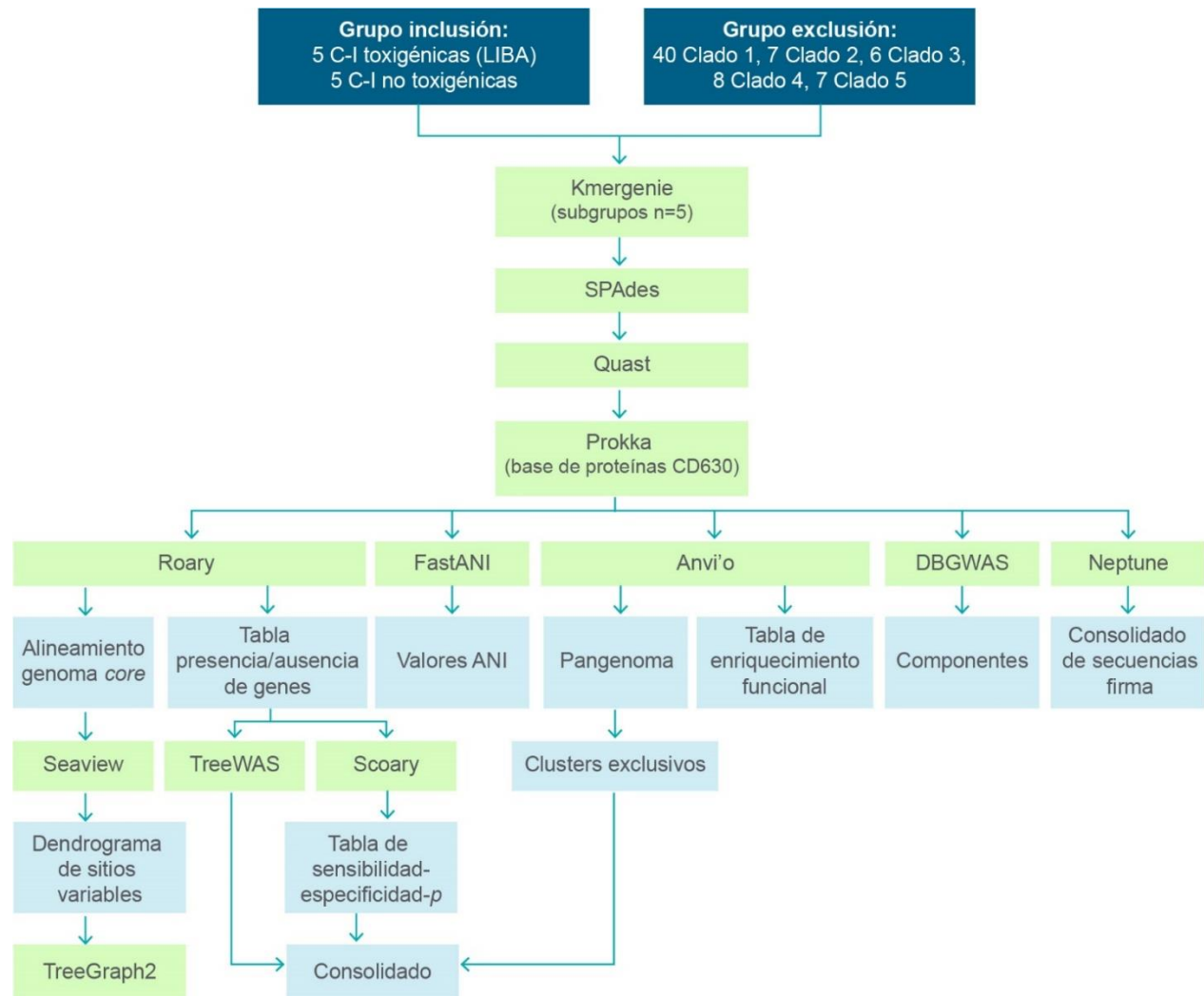


Figura 2. Esquema resumen de la metodología utilizada para identificar las posibles secuencias firma. Se representan en azul los grupos de inclusión y exclusión, en verde los programas empleados, y en celeste sus salidas.

Diseño de oligonucleótidos cebadores:

Las secuencias firma identificadas con los métodos anteriormente descritos fueron comparadas según especificidad, sensibilidad, y cantidad de métodos que las identificaron. Las mejores secuencias fueron filtradas usando la base no redundante del NCBI con BLASTn, eliminando aquellas que tenían alta identidad con secuencias de ADN de otras especies o con secuencias de cepas de *C. difficile* no pertenecientes al Clado C-I ( $E\text{-value} < 1$ ,  $bit\ score > 40$ ).

Se utilizó Primer-BLAST (Ye et al., 2012) para diseñar juegos de oligonucleótidos cebadores que amplificaran por PCR 11 presuntas secuencias firma del grupo de inclusión, identificadas tal y como se detalló en el apartado anterior. Se conservaron los parámetros predeterminados de esta aplicación, salvo aquellos relacionados con especificidad de los pares de oligonucleótidos, para los cuales se usó la base de datos Refseq con genomas de *Homo sapiens*, Clostridia (NCBI taxid:186801) y Firmicutes (NCBI taxid:1239), considerando que se realizaron ampliaciones en presencia de ADN humano y en heces humanas. Se procuró que estas secuencias fueran compatibles con el sistema multiplex descrito por Miller y colaboradores (2010), el cual es recomendado para el diagnóstico de CDI a partir de muestras de heces. Este método utiliza ADN extraído a partir de cultivos de *C. difficile* aislados a partir de heces previamente en dos ensayos de PCR multiplex con cebadores para los genes *cdtB*, *tcdA*, *tcdB*, *tcdC* y *tpi*. En este contexto, mientras que los cebadores para *tcdA*, *tcdB* y *cdtB* permiten categorizar las cepas de acuerdo a las toxinas que contienen, los cebadores para el gen *tpi* amplifican un fragmento específico presente en el genoma de todas las cepas conocidas de *C. difficile*, por lo que permiten detectar cepas no toxigénicas (Lemee et al., 2004).

Los pares de oligonucleótidos cebadores sugeridos por el software fueron clasificados según los siguientes parámetros: a)  $T_m$  más cercano a 60 °C, b) largo del producto entre 70 y 170 pares de bases, c) contenido GC entre 40% y 60%, d) ausencia de autocomplementariedad, particularmente en el extremo 3' (en este caso la autocomplementariedad representa la tendencia de los cebadores a alinearse con el otro, sin necesariamente favorecer amplificación del dímero, mientras que la autocomplementariedad 3' representa la tendencia de que se generen dímeros de cebadores que se amplifiquen). Posteriormente, para los siete mejores pares se determinó *in silico* la posible existencia de estructuras secundarias en el amplicón y sus

características termodinámicas (Markham & Zuker, 2005), así como las de su hibridación y posible heterogeneidad en sus sitios de unión al ADN de las cepas del grupo de inclusión (mediante alineamiento Clustal Omega). A partir de esta información se seleccionaron los cuatro mejores pares, los cuales se sintetizaron comercialmente (servicio de Macrogen, Korea).

Dado que los oligonucleótidos del párrafo anterior no tienen como blanco ninguna de las toxinas requeridas para causar enfermedad, también se diseñaron pares de oligonucleótidos cebadores para los alelos del gen *tcdB* detectados entre las cepas toxigénicas del grupo de inclusión, según lo descrito para los diseñados para las secuencias firma. El mejor par fue también sintetizado en Macrogen, Korea.

#### Estandarización y validación de la PCR:

Los cebadores para amplificar las secuencias firma fueron inicialmente probados en un ensayo de PCR punto final. Para tal fin se usó utilizando ADN genómico extraído con el kit DNEasy tissue kit (Qiagen) a partir de cultivos de cuatro cepas de *C. difficile* pertenecientes al Clado C-I (identificadas como LIBA-6622, LIBA-6625, LIBA-6822 y LIBA-7602), ocho cepas de otros clados de *C. difficile*, identificadas como NAP7, NAP1, A-B+, dos Clado 1 identificadas como LIBA-5755 y LIBA-6291, una Clado 2 identificada como LIBA-5757, una Clado 3 identificada como LIBA-7110 y una Clado 4 identificada como LIBA-6279), dos de *C. perfringens* (identificadas como SM101 y LIBA-8327), una de *C. filamentosum* (ATCC 25785), una de *C. botulinum* (identificada como A2), y una de *Bacteroides thetaiotaomicron* (ATCC 29741). La amplificación se realizó con una concentración de ADN entre 0,065 ng/μL y 1,17 ng/μL, 0,3 μl de oligonucleótido cebador *forward* y reverso en concentración final de 0,15 μM, 10 μl de PowerUp™ SYBR® Green Master Mix (Applied Biosystems) y agua libre de nucleasas para PCR, para un volumen final de 20 μl. El perfil de amplificación incluyó los siguientes pasos: a) preincubación de 2 minutos a 50 °C seguidos de 2 minutos a 95 °C; b) 40 ciclos de amplificación, cada uno con 15 segundos a 95 °C, 15 segundos a 60 °C y 1 minuto a 72 °C. Luego de la amplificación, 5 μl del producto de cada reacción fueron cargados con 2 μl de buffer de carga con GelRed® en geles de agarosa al 1,5% en TAE 0,5X, y corridos por 45 minutos a 100 V.

Luego de comprobar que los oligonucleótidos diseñados amplificaban las secuencias firma en las cepas del grupo de inclusión, se procedió a realizar pruebas en un equipo

de qPCR LightCycler480 (Roche®). Para un primer ensayo se utilizaron cuatro cepas de *C. difficile* del Clado C-I (LIBA-6622, LIBA-6822, LIBA-7602 y LIBA-7194), cuatro cepas *C. difficile* pertenecientes a otros clados (LIBA-5755, LIBA-5755, LIBA-6291 y LIBA-7110), y una cepa de *C. perfringens* (LIBA-8327). En estos primeros ensayos se utilizaron volúmenes iguales a los presentados en la PCR anteriormente presentada. Se empleó el siguiente perfil térmico: a) preincubación de 5 minutos a 95 °C; b) 45 ciclos de amplificación, cada uno con 10 segundos a 95 °C, 15 segundos a 60 °C y 20 segundos a 72 °C con adquisición sencilla; c) análisis de *melting* de 65 °C a 97 °C.

Seguidamente se realizaron ensayos de qPCR con los mismos oligonucleótidos cebadores y ADN extraído de heces humanas de pacientes diarreicos negativos para *C. difficile*, con el fin de comprobar que la ausencia de amplificación en ausencia de ADN de cepas del grupo de inclusión. Para tal fin, se tomó la cantidad de heces que quedaba unida a un hisopo estéril y se disolvió en 1 ml de medio amies líquido (Hubbard, Pellar, & Emanuel, 2011). Se extrajo el ADN genómico total a partir de 200 µl de las heces diluidas usando un instrumento MagNA pure-12 según las instrucciones del fabricante (Roche®), con un volumen de eluido de 50 µl, con y sin pretratamiento con proteinasa K, e incubación a 65 °C sin agitación por 10 minutos. El ADN extraído fue cuantificado por fluorometría en un equipo Qubit (Invitrogen) y utilizado para un ensayo de PCR tiempo real según se describió anteriormente, modificando los volúmenes de ADN agregado de forma tal que el volumen final de la reacción tuviera una concentración final de ADN genómico igual a 1,5 ng/µl.

También se realizó un ensayo para comprobar si la adición de ADN de cepas del Clado C-I a heces de pacientes diarreicos negativos por *C. difficile* daba lugar a la amplificación de los productos esperados en otro ensayo de qPCR. Esto con los mismos oligonucleótidos cebadores y condiciones descritas. Para tal fin se tomó 1 ml de cuatro muestras de heces de pacientes diarreicos colectadas y diluidas (ver arriba), a las cuales se le agregaron diferentes volúmenes de ADN genómico extraído a partir de cultivos de cepas Clado C-I (LIBA-7194, LIBA-6822, LIBA-7602, LIBA-6622), de forma que cada muestra de heces recibiera 9-12 ng de ADN. Posteriormente se realizó una extracción de ADN genómico de las heces inoculadas en un instrumento MagNA pure 12 (Roche®), sin realizar pretratamiento. Este ADN extraído fue cuantificado y utilizado para un ensayo de qPCR utilizando los oligonucleótidos cebadores en el mismo instrumento LightCycler480 (Roche®), el perfil de temperatura arriba descrito y

ajustando los volúmenes de ADN añadido de forma tal que cada reacción tuviera una concentración de ADN de entre 1,15 ng/μl y 1,25 ng/μl.

Como última prueba para determinar la especificidad y sensibilidad de los oligonucleótidos diseñados para detectar cepas del Clado C-I en matrices complejas, se inocularon heces de un paciente diarreico con cultivos en medio líquido de tres cepas del Clado C-I (MX-149, MX-152 y SJ312) y dos de otros clados de *C. difficile* (LIBA-6533 y LIBA-6538). Para tal fin se mezclaron 750 μl de heces y un volumen igual del cultivo de alguna de las cepas, sea puro, diluido 1:100 ó 1:500 en solución salina, para un total de 15 muestras, además de una sin inocular, mezclada con solución salina en la misma proporción. A partir de cada muestra se tomaron 500 μl de mezcla, los cuales fueron tratados con 40 μl de proteinasa K, para, acto seguido, extraer el ADN genómico a partir 400 μl de cada una en un equipo MagCore Plus II (RBC Bioscience) con un volumen de eluido de 40 μl. El ADN extraído fue cuantificado utilizando un fluorómetro Qubit según las instrucciones del fabricante. Se realizó un ensayo de qPCR para cada muestra de ADN extraído utilizando los oligonucleótidos cebadores en un instrumento LightCycler480 (Roche®) utilizando el perfil de temperatura descrito pero con 45 ciclos de amplificación y ajustando los volúmenes de ADN añadido de forma tal que cada reacción tuviera entre 0,97 ng/μl y 3,23 ng/μl de ADN.

Los oligonucleótidos cebadores para el gen *tcdB* fueron probados por qPCR usando ADN genómico de cuatro cepas toxigénicas de *C. difficile* del clado C-I (LIBA-6622, LIBA-6625, LIBA-6822 y LIBA-7194), seis cepas de otros clados de *C. difficile* (NAP1, A-B+, LIBA-5755, LIBA-6291, LIBA-7110 y LIBA-6279), dos de *C. perfringens* (SM101 y LIBA-8327), una de *C. filamentosum* (ATCC 25785), una de *C. botulinum* (identificada como A2), y una de *Bacteroides thetaiotaomicron* (ATCC 29741), en concentraciones de entre 0,054 ng/μl y 2,34 ng/μl de ADN, así como las heces inoculadas con cultivos mencionadas anteriormente. La amplificación se realizó con 1 μl de ADN, 0,5 μl de oligonucleótido cebador *forward* y reverso en concentración 10 μM, 10 μl de PowerUp™ SYBR® Green Master Mix (Applied Biosystems) y completando el volumen final de 20 μl por reacción con agua para PCR. Se empleó un instrumento LightCycler480 (Roche®) utilizando las concentraciones de ADN y el perfil de temperatura descritos para los otros pares de oligonucleótidos cebadores, pero con 40 ciclos de amplificación.

En ensayos de qPCR, el valor de Ct (o Cq) corresponde al número de ciclos de amplificación transcurridos cuando la curva de amplificación alcanza la línea del umbral, esto es, el momento en que la fluorescencia se hace detectable (Kralik & Ricchi, 2017). Se seleccionó Ct=35 como punto de corte dado que la presencia de una única copia de la secuencia blanco en la reacción generaría este resultado (asumiendo eficiencia de amplificación del 90%-100%). A partir de los resultados obtenidos en cada ensayo de PCR, excepto el realizado con ADN extraído a partir de heces de pacientes diarreicos, se calcularon la especificidad y sensibilidad de los pares de cebadores para clasificar correctamente una cepa en cultivo puro según pertenezca o no al Clado C-I, o en determinar correctamente la presencia o ausencia de una cepa del Clado C-I en una muestra de heces diarreica.

Estos valores se determinaron según las ecuaciones (Kralik & Ricchi, 2017):

$$\text{Sensibilidad} = \frac{VP}{VP + FN}$$

$$\text{Especificidad} = \frac{VN}{VN + FP}$$

Donde VP representa la cantidad de verdaderos positivos, VN la cantidad de verdaderos negativos, FP la cantidad de falsos positivos y FN la cantidad de falsos negativos.

En los resultados de los ensayos realizados con ADN extraído a partir de cultivos puros se consideró: i) como verdadero positivo la amplificación antes del punto de corte de una reacción que contenía ADN extraído a partir de una cepa del Clado C-I, ii) como verdadero negativo la ausencia de amplificación antes del punto de corte de una reacción que contenía ADN extraído a partir de una cepa que no pertenece al Clado C-I, iii) como falso positivo la amplificación antes del punto de corte de una reacción que contenía ADN extraído a partir de una cepa que no pertenece al Clado C-I y iv) como falso negativo la ausencia de amplificación antes del punto de corte de una reacción que contenía ADN extraído a partir de una cepa del Clado C-I.

Por su parte, en los resultados de los ensayos realizados con ADN extraído de muestras de heces se consideró: i) como verdadero positivo la amplificación antes del punto de corte de una reacción que contenía ADN extraído a partir de heces que contenían *C. difficile* del Clado C-I, ii) como verdadero negativo la ausencia de amplificación antes del punto de corte de una reacción que contenía ADN extraído a

partir de heces que no contenían *C. difficile* del Clado C-I, iii) como falso positivo la amplificación antes del punto de corte de una reacción que contenía ADN extraído a partir de heces que no contenían *C. difficile* del Clado C-I y iv) como falso negativo la ausencia de amplificación antes del punto de corte de una reacción que contenía ADN extraído a partir de heces con *C. difficile* del Clado C-I.

Cuadro II. Resumen de la forma de interpretación de los ensayos de qPCR

	Amplificación antes del punto de corte	Ausencia de amplificación antes del punto de corte
Muestra con ADN de <i>C. difficile</i> Clado C-I	VP	FN
Muestra con ADN de <i>C. difficile</i> no Clado C-I	FP	VN

Para estimar el límite de detección de los dos mejores cebadores diseñados para las secuencias firma encontradas y el gen *tcdB*, se prepararon nueve diluciones decimales seriadas de ADN extraído a partir de un cultivo líquido de una cepa del Clado C-I que generó amplificación por qPCR utilizando los métodos descritos arriba (LIBA-7602). Se realizó un ensayo de qPCR para cada una de estas soluciones y pares de cebadores, empleando un instrumento LightCycler480 (Roche®) y los perfiles de temperatura descritos, pero con 35 ciclos de amplificación y agregando 1 µL de cada solución de ADN, para generar concentraciones seriadas en las reacciones de entre 0,165 ng/µl y 1,65x10<sup>-9</sup> ng/µl de ADN.

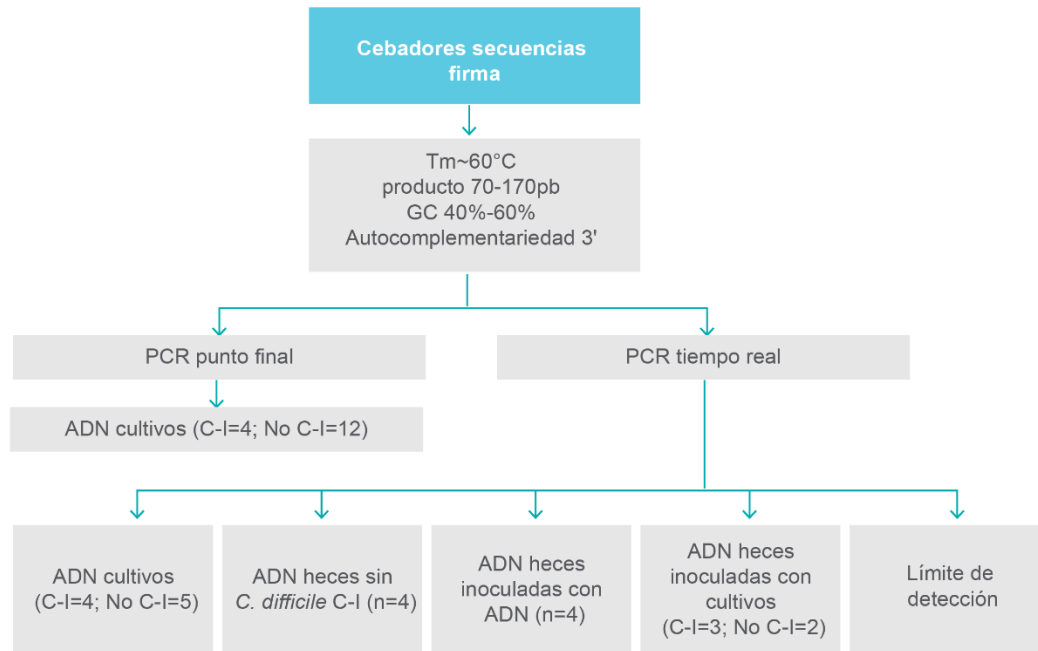


Figura 3. Esquema resumen de las pruebas realizadas para validar la sensibilidad y especificidad de los cebadores diseñados para las secuencias firma identificadas.

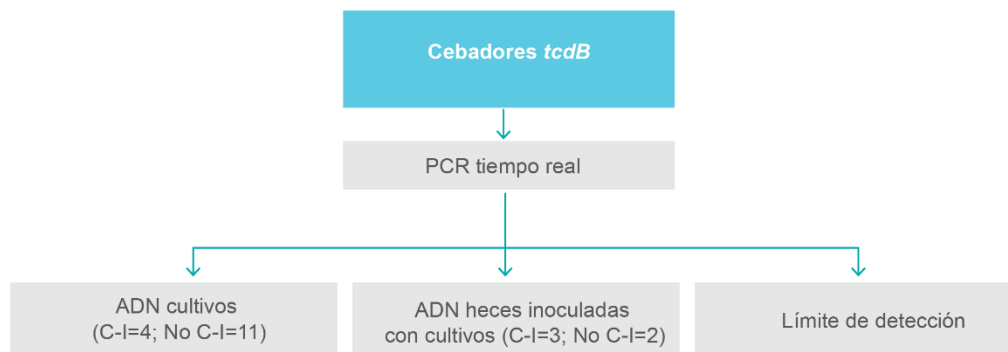


Figura 4. Esquema resumen de las pruebas realizadas para validar la sensibilidad y especificidad de los cebadores diseñados para el gen *tcdB* presente en las cepas del Clado C-I.

## Capítulo IV: Resultados

### Genómica comparativa y pangenómica:

La Figura 5 confirma la distribución cladística reportada por Dingle y colaboradores (Dingle et al., 2014), así como la pertenencia de las cepas aisladas en el LIBA al Clado C-I. Este resultado confirma que la distribución de las cepas en los grupos de inclusión y exclusión es apropiada según el propósito de la investigación.

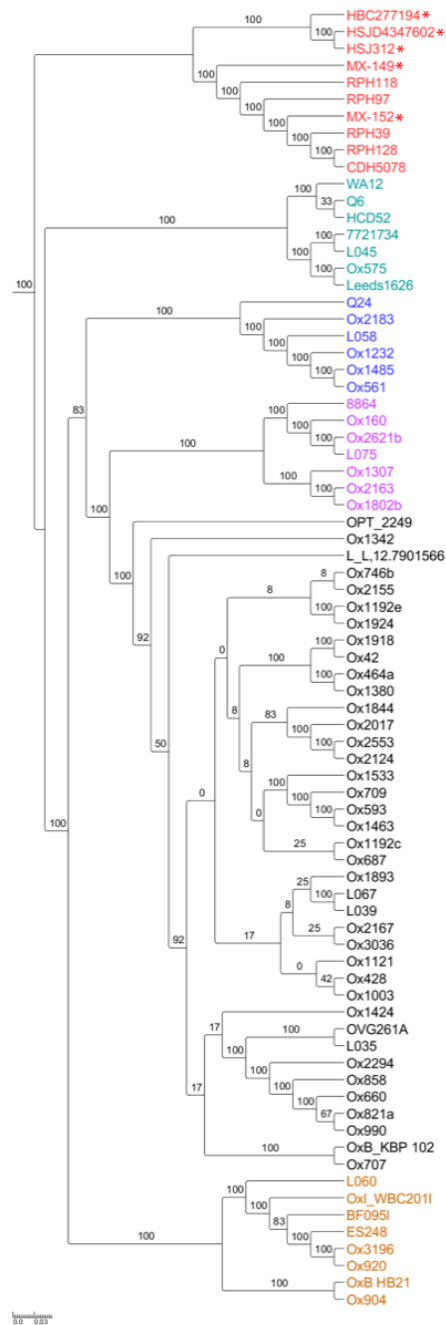


Figura 5. Árbol filogenético construido a partir de un alineamiento de sitios variables identificados por Roary en todos los genomas analizados. Las cepas C-I (rojo) pertenecen a un clado genéticamente divergente. Se muestran las cepas del Clado 1 en negro, del Clado 2 en morado, del Clado 3 en azul, del Clado 4 en naranja y del Clado 5 en verde, así como los valores de *bootstrap* (100 iteraciones) para cada nodo y la escala (sustituciones por sitio). Las cepas aisladas en el LIBA se marcan con un asterisco.

A pesar de ser diferentes entre sí, las cepas Clado C-I son genéticamente distintas a las cepas Clados 1-5 (Figura 6). Los valores de porcentaje de identidad nucleotídica promedio (ANI) obtenidos de la comparación de estos grupos sugieren que ambos representan especies distintas.

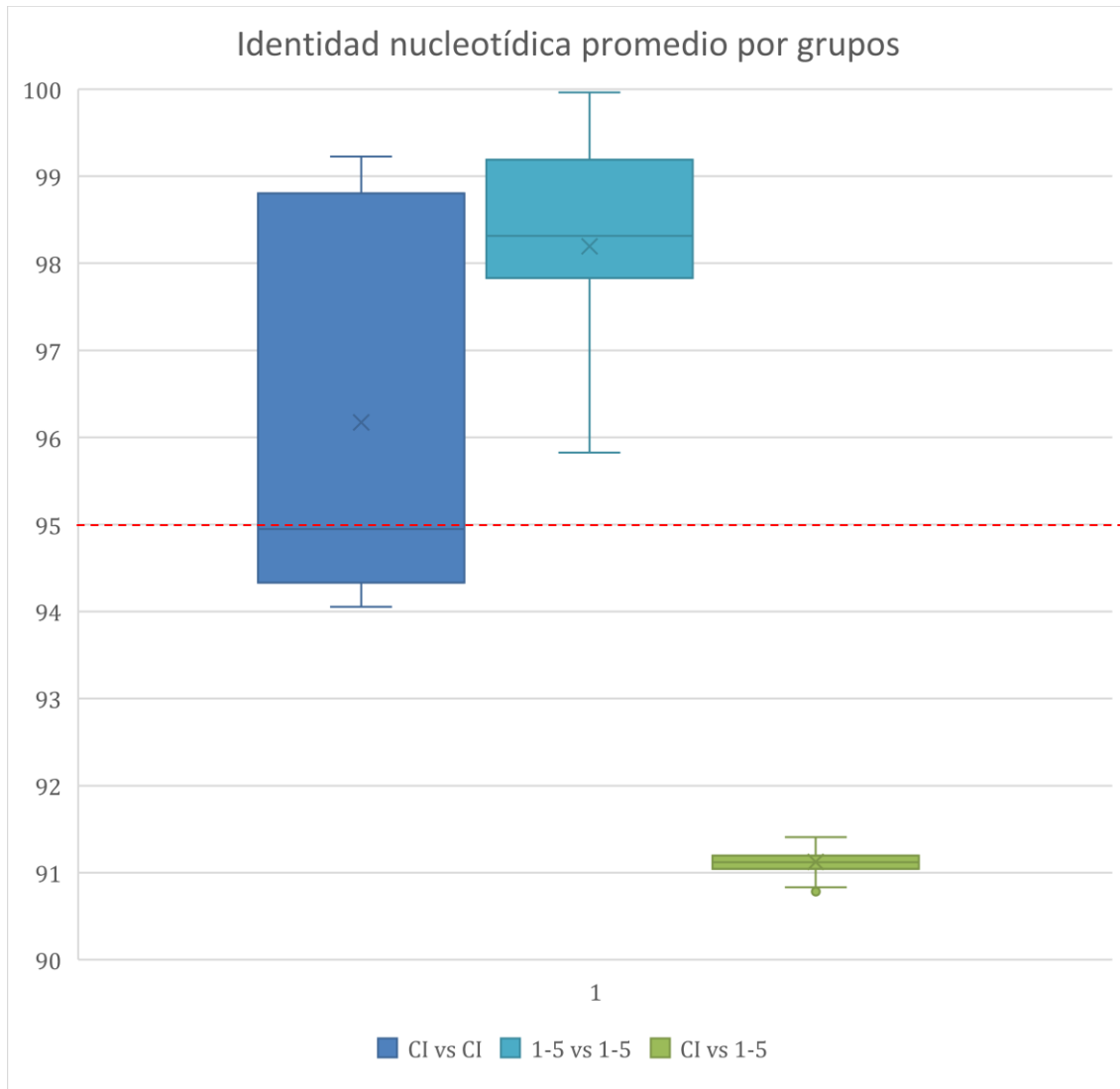


Figura 6. Diagrama de cajas-bigotes para comparar el ANI determinado por FastANI para cepas de Clado C-I (n=10), Clados 1-5 (n=67) y entre ambos grupos. La línea roja marca un 95% de ANI, el cual es el umbral aceptado para delimitar especies bacterianas (Jain, Rodriguez-R, Phillippy, Konstantinidis, & Aluru, 2018).

El alto grado de divergencia de las cepas del Clado C-I se confirmó a través de la observación de un incremento sustancial en el tamaño del pangenoma de *C. difficile* y

una concomitante reducción del tamaño de su genoma *core* luego de incluir cepas Clado C-I en la comparación (Figura 7). Un número pequeño, pero detectable, de secuencias codificantes son compartidas por todos los miembros del Clado C-I, pero no pertenecen al genoma *core* de *C. difficile* (n=890), por lo que son potencialmente secuencias firma y potencialmente aprovechables en el diagnóstico específico que se desea diseñar (Figura 7).

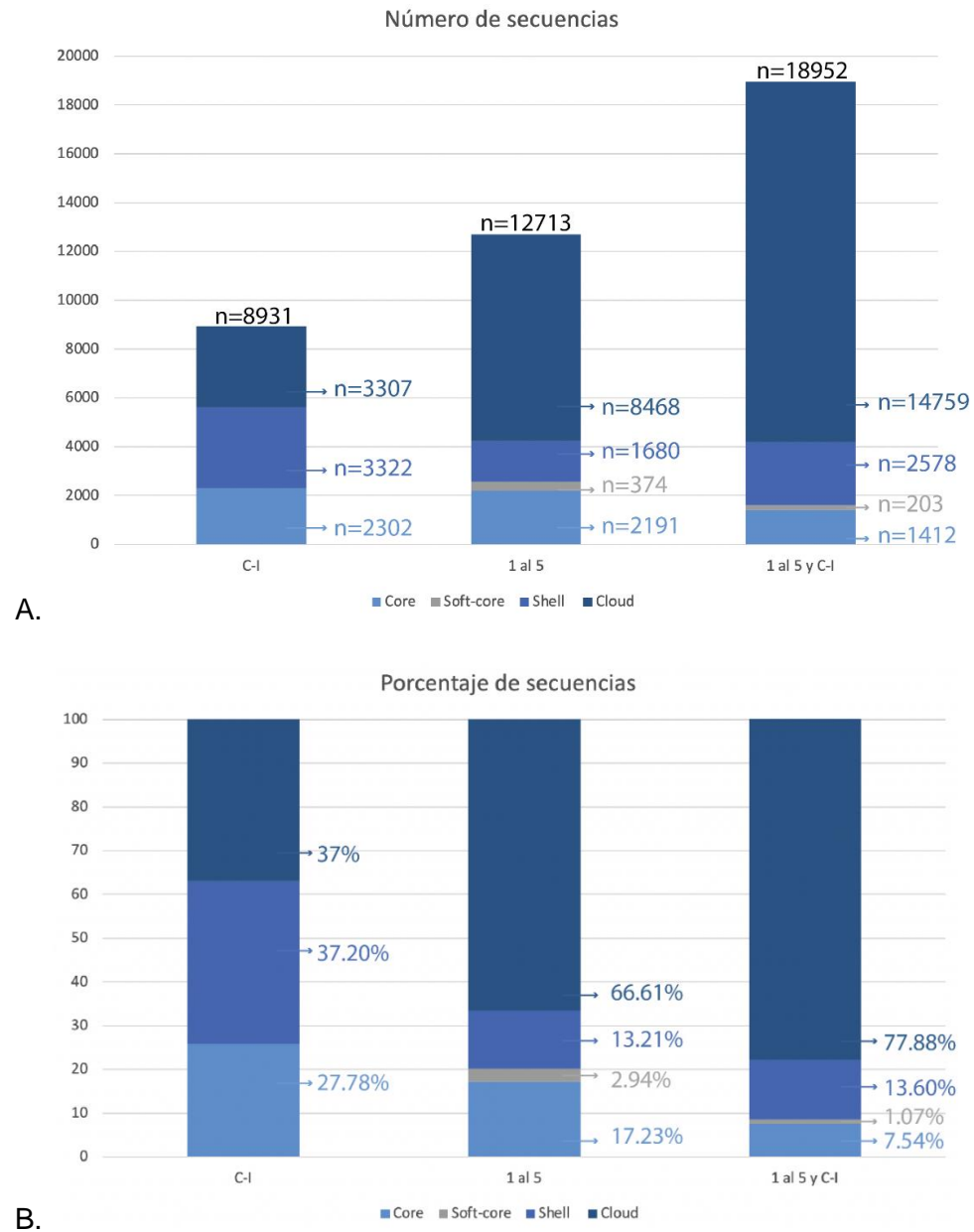


Figura 7. Número absoluto (A) y relativo en porcentaje (B) de secuencias en el genoma *core* (presentes en al menos 99% de las cepas analizadas,  $n=76$ ), *soft-core* (presentes entre el 95% y 99% de las cepas,  $n=73$  y 76 cepas), *shell* (presentes entre el 15% y 95% de las cepas,  $n= 11$  y 73 cepas) y *cloud* (presentes en menos del 15% de las cepas,  $n< 11$ ) de las cepas del Clado C-I, las cepas de los Clados 1-5 y todas las cepas analizadas.

En congruencia con el resultado anterior, una fracción del genoma *core* de las cepas Clado C-I no está presente en las cepas de los Clados 1-5 (Figura 8).

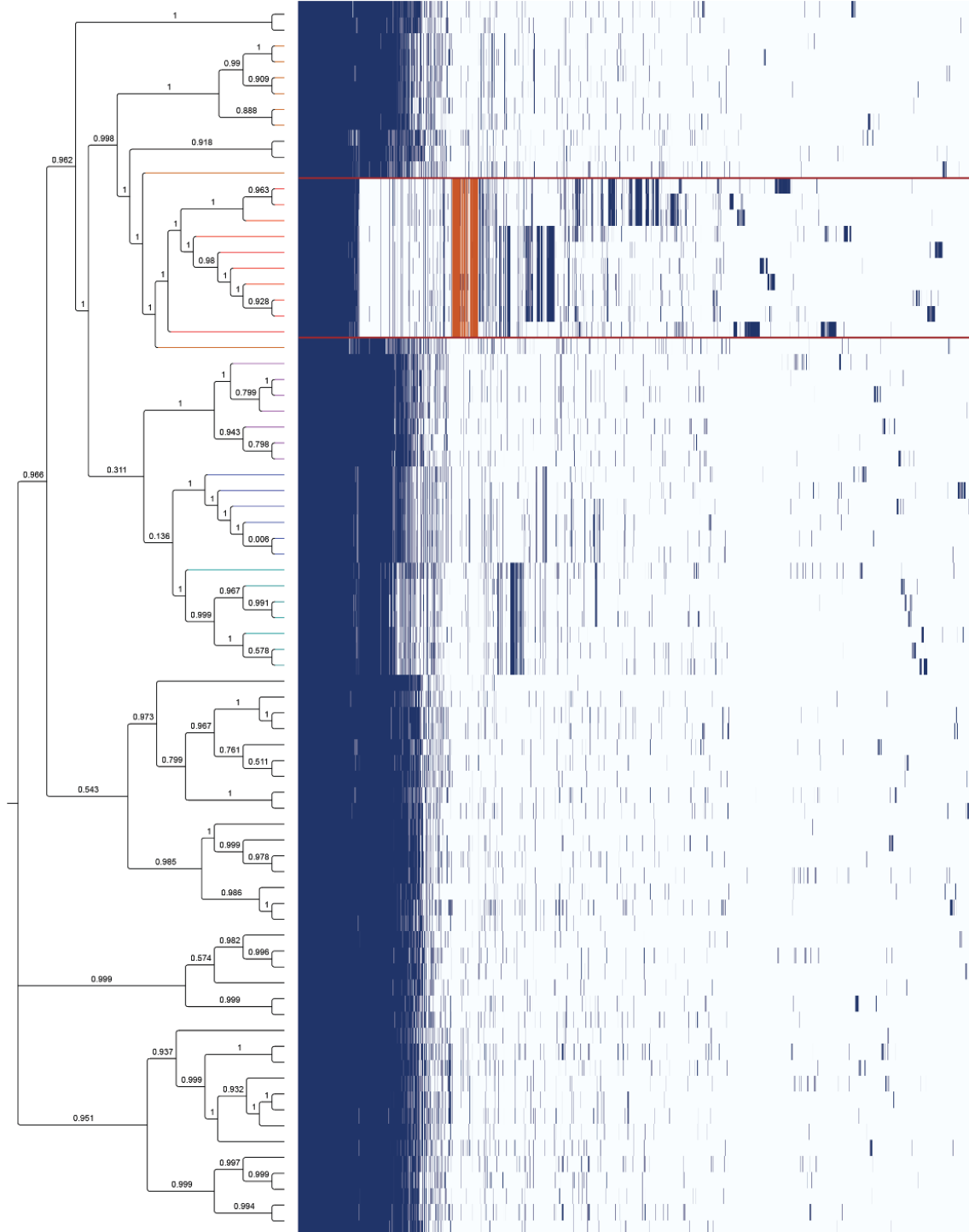


Figura 8. Matriz de presencia/ausencia de CDS contextualizada filogenéticamente (Roary). Las líneas verticales azules y naranja indican presencia de CDSs o presencia de CDS exclusiva en cepas del Clado C-I, respectivamente. En el dendrograma se muestran las cepas del Clado 1 en negro, del Clado 2 en morado, del Clado 3 en azul,

del Clado 4 en naranja y del Clado 5 en verde, así como los valores de soporte local (test de Shimodaira-Hasegawa con 1000 reemplazos) para cada nodo. Las cepas C-I se encuentran en el recuadro rojo.

De acuerdo con los resultados obtenidos con Scoary y treeWAS, la mayoría de los CDS universales, pero exclusivos de las cepas C-I, codifican para enzimas metabólicas y proteínas hipotéticas. En un menor grado se encontraron múltiples factores de transcripción, sistemas de transporte ABC, sistemas reguladores de dos componentes, entre otros (ver Anexo IV para la descripción completa). Por su parte, el *bin* exclusivo de las cepas C-I extraído por Anvi'o, contiene un importante número de conglomerados de genes de copia única relacionados con transcripción (K) (n=25), mecanismos de transducción de señales (T) (n=11) y metabolismo de aminoácidos y carbohidratos (E) (n=11) (Figura 9) (ver Anexo V para la descripción completa).

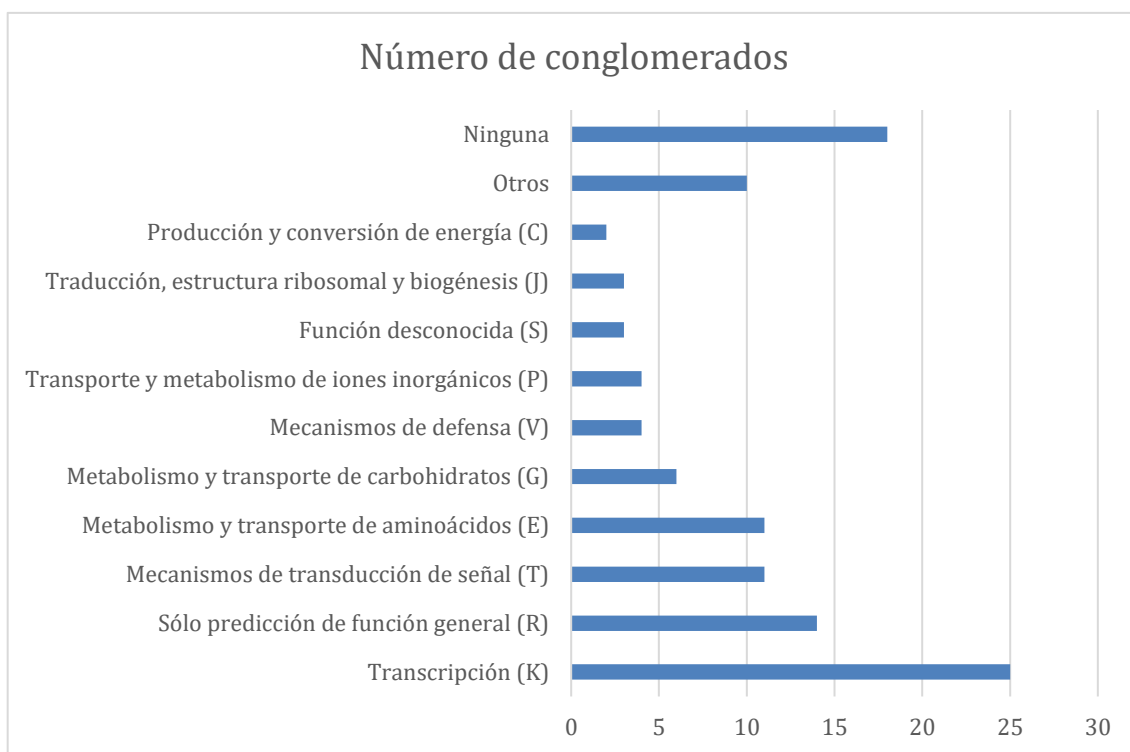


Figura 9. Conglomerados de genes presentes en el *bin* exclusivo de las cepas C-I identificado por Anvi'o, clasificados según su función COG anticipada.

Identificación de secuencias firma:

Utilizando Neptune se identificaron varias secuencias firma candidatas, sin embargo, ninguna de estas estaba totalmente ausente en el grupo de exclusión, y todas mostraron una considerable similitud con secuencias de otras especies o de cepas de *C. difficile* al alinearlas con BLASTn a la colección de nucleótidos del NCBI (Ver Anexo III). Por tal razón, se descartaron y no se consideraron para los análisis subsecuentes.

Al menos 11 conglomerados de CDS (cada uno conformado por una CDS de cada cepa del Clado C-I) fueron identificados por Anvi'o, Roary/Scoary y Treewas como exclusivos de las cepas C-I ( $p < 0.05$ ) con 100% de sensibilidad y especificidad (valores de  $p$  con corrección Benjamini–Hochberg de  $1,22 \times 10^{-11}$ ), y sin similitud significativa con secuencias nucleotídicas de organismos que no pertenecen al Clado C-I de *C. difficile*, tales como bacterias del género *Clostridium* (*C. perfringens*, *C. septicum*) *E. coli*, *Enterococcus faecalis*, las cuales son constituyentes de la flora intestinal normal del adulto (Cuadro III). Estas secuencias firma se utilizaron como candidatas para el diseño de cebadores para PCR.

Cuadro III. Principales características predichas de las secuencias firma del Clado C-I identificadas en conjunto por Roary/Scoary, TreeWAS y Anvi'o y determinadas por Roary y Anvi'o. Todas las secuencias tuvieron longitudes mayores a 300 pb para posibilitar su detección por qPCR.

ID Roary	Anotación funcional (Roary)	Identificador de conglomerado de Anvi'o	Índice de homogeneidad geométrica por Anvi'o	Función COG por Anvi'o	Categoría COG por Anvi'o*	Puntaje de enriquecimiento de Anvi'o	Core en C-I?
<i>IsrF</i>	3-hidroxi-5-fosfonooxipentano-2,4-diona tiolasa	GC_00004247	1	Fructosa bisfosfato aldolasa clase Ia	G	7,77	Si
<i>dhaK</i>	Dihidroxiacetona kinasa dependiente de PTS, subunidad de unión a dihidroxiacetona Dhak	GC_00004331	1	Dihidroxiacetona kinasa	G	7,77	Si
<i>crcB</i>	Transportador de ión fluoruro predicho CrcB	GC_00004262	0,9983871	Exportador de ión fluoruro CrcB/FEX	DJP	7,77	Si
grupo_5507	Proteína hipotética	GC_00004244	1	NA	NA	NA	NA
grupo_8205	Regulador transcripcional	GC_00004306	1	Regulador transcripcional de unión a ADN LsrR	K	NA	NA
grupo_5266	Proteína hipotética	GC_00004280	1	NA	NA	NA	NA
grupo_5459	Proteína hipotética	GC_00004274	1	Proteína de unión a metal predicha	S	NA	NA
grupo_7858	Proteína hipotética	GC_00004267	1	NA	NA	NA	NA
grupo_8076	Proteína hipotética	GC_00004294	0,998726115	Regulador transcripcional de unión al ADN.	K	NA	NA
grupo_3785	Regulador transcripcional, familia MerR	GC_00004253	0,9889393	Regulador transcripcional de unión a ADN	K	NA	NA
grupo_3695	Sistema de transporte tipo ABC, permeasa de familia cistina/aminoácido	GC_00004326	0,9465858	Sistema de transporte de aminoácido tipo ABC	E	NA	NA

\*Las categorías COG mostradas corresponden a: G=Metabolismo y transporte de carbohidratos; D=Control del ciclo celular y mitosis; P= Transporte y metabolismo de iones inorgánicos; K= Transcripción; E=Metabolismo y transporte de aminoácidos; S=Función desconocida. NA = indeterminado.

Diseño y validación de ensayo (q)PCR:

Ocho secuencias firma identificadas generaron pares de oligonucleótidos cebadores específicos utilizando Primer-BLAST. En el Cuadro IV se muestran algunas de sus características predichas, las cuales funcionaron como primer criterio de selección según se describe en la metodología. No se encontraron pares de oligonucleótidos cebadores específicos para los *clusters* identificados como *IsrF* y *dhaK*. Además, no se diseñaron oligonucleótidos cebadores para el grupo\_3695 debido a diferencias en el tamaño (pb) de los CDS que conforman dicho conglomerado (Cuadro IV).

Cuadro IV. Características predichas de los pares de cebadores identificados por PrimerBLAST para las secuencias firma seleccionadas.

Cluster	Largo producto (bp)		Hebra molde	Largo	Tm	GC%	Autocomplementariedad	Autocomplementariedad 3'
crcB	116	Forward	+	22	59,89	45,45	5	1
		Reverse	-	21	57,79	47,62	2	2
	118	Forward	+	23	60,25	43,48	7	1
		Reverse	-	22	57,65	45,45	3	2
	114	Forward	+	23	60,44	43,48	5	0
		Reverse	-	24	58,79	45,83	5	3
	123	Forward	+	23	59,23	43,48	7	0
		Reverse	-	24	58,79	45,83	3	1
	124	Forward	+	22	58,22	40,91	7	0
		Reverse	-	24	58,79	45,83	3	0
	110	Forward	+	24	59,96	41,67	2	0
		Reverse	-	25	59,93	44	7	2
	152	Forward	+	24	60,14	45,83	6	2
		Reverse	-	25	59,7	48	3	1
	100	Forward	+	22	57,8	45,45	2	0
		Reverse	-	25	59,7	44	7	1
5507	126	Forward	+	23	58,87	43,48	6	0
		Reverse	-	23	59,67	43,48	4	2
	115	Forward	+	24	59,66	41,67	4	2
		Reverse	-	22	58,33	45,45	4	2
	123	Forward	+	24	60,38	41,67	6	0
		Reverse	-	22	58,07	45,45	4	1
	159	Forward	+	22	57,52	45,45	6	0
		Reverse	-	24	60,02	37,5	6	2
111	Forward	+	25	59,99	40	4	2	

		Reverse	-	24	60,2	41,67	4	2
	128	Forward	+	25	60,16	40	6	0
		Reverse	-	24	60,2	41,67	6	2
	146	Forward	+	23	59,68	43,48	4	2
		Reverse	-	25	58,53	36	4	0
	124	Forward	+	25	59,7	40	6	0
		Reverse	-	24	60,56	45,83	4	2
	126	Forward	+	24	58,7	41,67	6	0
		Reverse	-	24	60,8	45,83	4	1
	120	Forward	+	23	57,51	39,13	6	0
		Reverse	-	24	59,01	41,67	4	2
8205	234	Forward	+	20	60,4	55	6	0
		Reverse	-	22	59,96	45,45	4	2
	260	Forward	+	24	59,53	45,83	3	1
		Reverse	-	21	59,99	52,38	4	0
	332	Forward	+	22	61,18	50	4	0
		Reverse	-	22	60,5	50	6	2
3785	101	Forward	+	22	58,48	45,45	4	3
		Reverse	-	21	57,62	42,86	6	3
	102	Forward	+	21	57,39	42,86	4	3
		Reverse	-	23	59,32	43,48	6	3
	104	Forward	+	23	59,13	43,48	4	1
		Reverse	-	22	58,07	40,91	6	3
	106	Forward	+	21	57,07	47,62	4	1
		Reverse	-	22	58,06	45,45	6	2
	165	Forward	+	22	58,07	40,91	6	2
		Reverse	-	21	57,03	47,62	5	2
	167	Forward	+	22	58,07	40,91	6	3
		Reverse	-	24	59,9	41,67	5	2

	196	Forward	+	23	59,07	43,48	6	2	
		Reverse	-	24	59,78	41,67	2	1	
	134	Forward	+	24	58,94	37,5	8	2	
		Reverse	-	22	58,47	45,45	5	2	
	195	Forward	+	24	59,67	41,67	6	0	
		Reverse	-	24	59,47	41,67	2	0	
	277	Forward	+	24	60,38	41,67	4	3	
		Reverse	-	24	59,47	41,67	2	0	
	5266	103	Forward	+	23	59,22	43,48	8	0
			Reverse	-	24	57,59	33,33	3	1
108		Forward	+	24	59,96	45,83	8	0	
		Reverse	-	25	58,18	32	3	1	
102		Forward	+	22	57,58	45,45	8	0	
		Reverse	-	25	57,2	36	3	0	
72		Forward	+	25	57,46	36	6	3	
		Reverse	-	24	58,62	37,5	4	0	
5459		434	Forward	+	23	57,05	39,13	4	2
			Reverse	-	24	59,36	37,5	3	2
7858	161	Forward	+	24	60,2	45,83	4	2	
		Reverse	-	23	59,93	43,48	4	0	
	170	Forward	+	24	59,48	41,67	6	2	
		Reverse	-	22	58,6	45,45	4	2	
	160	Forward	+	25	59,82	40	6	2	
		Reverse	-	21	57,33	47,62	5	1	
	151	Forward	+	25	59,35	40	4	3	
		Reverse	-	22	57,96	50	5	1	
	241	Forward	+	24	58,34	41,67	4	3	
		Reverse	-	23	58,94	43,48	4	2	
236	Forward	+	23	57,34	39,13	6	3		

		Reverse	-	24	60,08	45,83	5	1
	161	Forward	+	23	57,31	39,13	6	2
		Reverse	-	24	59,84	45,83	5	0
	156	Forward	+	24	58,17	37,5	4	3
		Reverse	-	23	58,69	43,48	4	3
	163	Forward	+	24	58,12	37,5	6	2
		Reverse	-	23	58,68	43,48	4	3
	166	Forward	+	25	59,06	40	6	2
		Reverse	-	22	57,65	40,91	4	2
8076	434	Forward	+	23	57,05	39,13	4	2
		Reverse	-	24	59,36	37,5	3	2

Luego de realizar una primera depuración según los criterios establecidos en la metodología, se continuó teniendo un número elevado de pares de cebadores candidatos, sin embargo, muchos de ellos amplificaban las mismas secuencias, difiriendo en pocos nucleótidos. Por esta razón se seleccionó un subgrupo de siete mejores candidatos para los cuales se determinaron las características termodinámicas de la estructura secundaria del amplicón esperado y la capacidad de hibridación de ambos oligonucleótidos del par (Cuadro V). Este paso adicional permitió descartar algunos pares de cebadores cuyos amplicones esperados tenían estructuras secundarias con valores de  $T_m$  elevado donde el extremo 3' no se encontraba libre, o bien, dímeros de cebador estables. Además, mediante alineamiento Clustal Omega de las secuencias blanco de los cebadores se determinó que todos los pares candidatos tenían algunos sitios variables (Cuadro V).

Cuadro V. Propiedades termodinámicas del amplicón esperado y posibles dímeros de cebador anticipados para los 7 mejores pares de cebadores candidatos.

Cluster	Largo producto	Cebador	Hebra molde	Parámetros termodinámicos de estructura secundaria del amplicón					Parámetros termodinámicos de posible dímero de cebadores				Número de sitios variables
				$\Delta G$ (kcal/mol)	$\Delta H$ (Kcal/mol)	$\Delta S$ (cal/mol/K)	Tm (°C)	Extremo 3'	$\Delta G$ (Kcal/mol)	$\Delta H$ (Kcal/mol)	$\Delta S$ (cal/mol/K)	Tm (°C)	
crcB	114	Forward	+	-7,16	-96,3	-287,41	61,9	libre	-8,2	-77,0	-222,0	37,8	2
		Reverse	-										1
	110	Forward	+	-9,25	-103,7	-304,53	67,4	libre	-7	-46,3	-126,6	31,0	4
		Reverse	-										2
	152	Forward	+	-15,92	-231,7	-695,73	59,9	libre	-4,2	-54,0	-160,4	17,1	4
		Reverse	-										1
grupo_55 07	115	Forward	+	-4,57	-126,8	-394,1	48,6	horquilla de 3 nucleótidos	-5,5	-39,0	-108,1	18,5	4
		Reverse	-										3
	111	Forward	+	-4,98	-156,2	-487,57	47,2	asa	-4,3	-29,6	-81,6	2,9	5
		Reverse	-										3
grupo_37 85	195	Forward	+	-12,22	-251,3	-770,85	52,9	libre	-2,2	-23,9	-69,8	-22,7	2
		Reverse	-										2
grupo_78 58	161	Forward	+	-8,16	-232,1	-722,04	48,3	horquilla	-3,7	-45,1	-133,5	10,3	3
		Reverse	-										0

Los 7 pares de cebadores cumplían con la mayoría de los criterios de inclusión establecidos en la metodología ( $T_m$  cercano a 60 °C, largo de producto entre 70 y 170 pares de bases, contenido CG entre 40% y 60%, baja estabilidad del posible dímero de cebador, estructura secundaria del amplicón con extremo 3' libre, o de baja estabilidad). Por el contrario, todos poseían cierto grado de autocomplementariedad y cantidad de sitios variables. Con base en esta información se decidió probar en el laboratorio los cebadores de largo de producto 110, 114, 161 y 195, designados de esta forma en lo subsecuente.

Con el fin de poder determinar el potencial toxigénico de las cepas detectables con los cebadores indicados arriba, se diseñaron 10 cebadores específicos para las secuencias de *tcdB* de las cepas de *C. difficile* del Clado C-I (Cuadro VI). De estos, se decidió sintetizar y ensayar el primer par, pues se apegaba mejor a los criterios de selección establecidos en la metodología.

Cuadro VI. Características predichas de los pares de cebadores identificados por PrimerBLAST para las secuencias del gen *tcdB* de las cepas de *C. difficile* del Clado C-I.

Largo producto	Cebador	Hebra molde	Largo	Tm	GC%	Autocomplementariedad	Autocomplementariedad 3'
166	Forward	+	21	59,99	47,62	4	2
	Reverse	-	25	57,43	36	4	2
84	Forward	+	23	58,55	39,13	5	2
	Reverse	-	23	57,5	39,13	4	3
98	Forward	+	24	59,1	37,5	5	2
	Reverse	-	25	60,52	40	5	1
196	Forward	+	24	58,38	33,33	7	2
	Reverse	-	24	60,87	45,83	8	0
194	Forward	+	24	58,19	33,33	7	1
	Reverse	-	22	57,12	45,45	8	0
96	Forward	+	24	59,1	37,5	5	2
	Reverse	-	25	59,11	40	5	2
160	Forward	+	25	59	36	6	2
	Reverse	-	22	57,08	40,91	5	0
101	Forward	+	25	59,63	36	5	2
	Reverse	-	25	59,11	40	5	0
194	Forward	+	25	58,76	32	7	1
	Reverse	-	24	58,75	41,67	6	0
86	Forward	+	24	58,57	37,5	5	0
	Reverse	-	24	57,93	37,5	4	2

En la discriminación de las cepas del Clado C-I por PCR punto final, los pares de cebadores 114 mostraron una sensibilidad del 100% y una especificidad del 80% ( $VP = 2$ ;  $FN = 0$ ;  $VN = 8$ ;  $FP = 2$ ;  $Sensibilidad = \frac{2}{2+0}$ ;  $Especificidad = \frac{8}{8+2}$ ), mientras que los pares de cebadores 110 y 161 mostraron sensibilidad de 100% y especificidad del 90% ( $VP = 2$ ;  $FN = 0$ ;  $VN = 9$ ;  $FP = 1$ ;  $Sensibilidad = \frac{2}{2+0}$ ;  $Especificidad = \frac{9}{9+1}$ ) (Anexo VII).

Al establecer como punto de corte  $Ct=35$  en el ensayo de qPCR, tal y como sugieren otros autores (Chang, Hur, & Park, 2020; Jordan & Durso, 2005; Sacramento de Oliveira et al., 2018), los pares de cebadores 114 y 161 discriminaron la presencia de ADN de cepas de *C. difficile* del Clado C-I con una sensibilidad y especificidad del 100% ( $VP = 4$ ;  $FN = 0$ ;  $VN = 5$ ;  $FP = 0$ ;  $Sensibilidad = \frac{4}{4+0}$ ;  $Especificidad = \frac{5}{5+0}$ ), mientras que el par 110 tuvo una sensibilidad del 100% y una especificidad del 60% ( $VP = 4$ ;  $FN = 0$ ;  $VN = 3$ ;  $FP = 2$ ;  $Sensibilidad = \frac{4}{4+0}$ ;  $Especificidad = \frac{3}{3+2}$ ). Ninguna de las cepas ensayadas generó amplificación utilizando el par de cebadores 195 (Cuadro VII).

Cuadro VII. Valores de Ct y resultados obtenidos por qPCR utilizando ADN extraído a partir de cultivos de diversas bacterias y los cebadores diseñados para las secuencias firma del Clado C-I.

Cepa	Clado	Concentración ADN en la reacción (ng/μl)	Cebadores	Ct	Interpretación del resultado
LIBA-5755	1	0,775	110	37,96	VN
LIBA-6291		0,445	110	32,66	FP
LIBA-5757	2	0,068	110	35,73	VN
LIBA-5757		0,068	114	35,21	
LIBA-5757		0,068	161	40	
LIBA-7110	3	0,51	110	30	FP
LIBA-6622	C-I	0,212	110	28,31	VP
LIBA-6622		0,212	114	20,79	
LIBA-6622		0,212	161	18,52	
LIBA-6822		0,054	110	19,4	
LIBA-6822		0,054	114	18,35	
LIBA-6822		0,054	161	21,37	
LIBA-7194		0,11	110	16,73	
LIBA-7194		0,11	114	16,36	
LIBA-7194		0,11	161	28,79	
LIBA-7602		0,165	110	20,3	
LIBA-7602		0,165	114	19,1	
LIBA-7602		0,165	161	23,81	

\* La cepa LIBA-8327 (*C. perfringens*) no amplificó con los cebadores 110. Las cepas LIBA-5755 y LIBA-6291 (clado 1), LIBA-7710 (clado 3) y LIBA-8327 no amplificaron con los cebadores 114 y 161. Las cepas LIBA-5755 y LIBA-6291 (clado 1), y LIBA-5757 (clado 2) no amplificaron con los cebadores 195. Todos estos corresponden a resultados verdaderos negativos. Las cepas LIBA-6822 y LIBA-7602 (Clado C-I) no amplificaron con los cebadores 195 (FN). Se muestran los demás identificadores de las cepas, el clado o especie al que pertenecen, las concentraciones de ADN usadas en las reacciones de PCR, el cebador utilizado y el Ct obtenido. El control negativo corresponde a mezcla de reacción a la que no se le adicionó ADN, y no generó amplificación con ninguno de los cebadores. Las cepas LIBA-6622, LIBA-6822, LIBA-7110, LIBA-8327 y LIBA-7194 no fueron probadas con los cebadores 195 dado que ya se había comprobado que estos no amplificaban el ADN de las otras cepas C-I en las mismas condiciones.

Para comprobar que los cebadores no amplificaban fragmentos de ADN extraídos a partir de heces diarreicas se realizó un ensayo de qPCR con cuatro muestras de este tipo libres de *C. difficile*. En este ensayo no se observó amplificación en ninguna muestra con ninguno de los cebadores, únicamente amplificaron los controles positivos (Cuadro VIII).

Cuadro VIII. Valores de Ct obtenidos por qPCR utilizando ADN extraído a partir de heces de pacientes diarreicos y los cebadores diseñados para las secuencias firma del Clado C-I.

Muestra	Concentración ADN en reacción (ng/μl)	Cebador	Ct
LIBA-6822	0,054	110	21,27
LIBA-7194	0,11		17,62
LIBA-6625	0,065		30
LIBA-6622	0,212		30
LIBA-6822	0,054	114	21,29
LIBA-7194	0,11		16,82
LIBA-6625	0,065		22,55
LIBA-6622	0,212		22,72
LIBA-6822	0,054	161	23,75
LIBA-7194	0,11		19,08
LIBA-6625	0,065		18,67
LIBA-6622	0,212		19,87

\* Las cepas LIBA-6822, LIBA-7194, LIBA-6625 y LIBA-6622 se utilizaron como control positivo y como control negativo mezcla de reacción a la que no se le adicionó ADN.

Ninguna de las muestras de heces probadas, con o sin pretratamiento (tal y como se describe en la metodología), ni el control negativo generaron amplificación con ninguno de los cebadores. Se muestran las demás muestras ensayadas, las concentraciones de ADN adicionadas a cada reacción, el cebador utilizado y el Ct obtenido.

En el ensayo de qPCR con ADN extraído a partir de heces inoculadas con ADN genómico de cepas del Clado C-I (estableciendo como punto de corte  $Ct=35$ ), los cebadores 110, 114 y 161 mostraron sensibilidades del 25% ( $VP = 1$ ;  $FN = 3$ ;  $Sensibilidad = \frac{1}{1+3}$ ) (Cuadro IX).

Cuadro IX. Valores de Ct y resultados obtenidos por qPCR utilizando ADN extraído a partir de heces inoculadas con ADN genómico de cepas del Clado C-I y los cebadores diseñados para las secuencias firma del Clado C-I.

Muestra	Cepa usada para la inoculación	Cantidad de ADN inoculado (ng)	Concentración ADN en la reacción (ng/ $\mu$ l)	Cebadores	Ct	Resultado
3	LIBA-7602	9,9	1,25	110	30	VP
LIBA-7602	-	-	0,165	110	23,38	-
3	LIBA-7602	9,9	1,25	114	30	VP
LIBA-7602	-	-	0,165	110	21,15	-
4	LIBA-6622	12,72	1,15	161	30	VP
LIBA-7602	-	-	0,165	161	24,91	-

\*Las muestras inoculadas con ADN de las cepas LIBA-7194, LIBA-6822 y LIBA-6622 no generaron amplificación con los cebadores 110 y 114. Las muestras inoculadas con ADN de las cepas LIBA-7602, LIBA-7194 y LIBA-6822 no generaron amplificación con los cebadores 161. Todos estos corresponden a resultados falsos negativos. Se utilizó ADN extraído a partir de un cultivo de la cepa LIBA-7602 como control positivo y como control negativo mezcla de reacción a la que no se le adicionó ADN. El control negativo no generó amplificación con ninguno de los cebadores. Se detallan las muestras utilizadas, la cepa de la cual se tomó ADN para inocular las heces, la cantidad de ADN inoculado, las concentraciones de ADN en la reacción para cada una, el cebador utilizado, el Ct obtenido y el resultado obtenido.

En la detección por qPCR de cepas del Clado C-I en heces inoculadas con diluciones seriadas de cultivos (estableciendo como punto de corte  $Ct=35$ ), los cebadores 114 y 161 discriminaron con un 100% de sensibilidad y especificidad ( $VP = 9$ ;  $FN = 0$ ;  $VN = 6$ ;  $FP = 0$ ;  $Sensibilidad = \frac{9}{9+0}$ ;  $Especificidad = \frac{6}{6+0}$ ) mientras que el par de cebadores

110 mostró una sensibilidad del 33% y una especificidad del 100% ( $VP = 3$ ;  $FN = 6$ ;  $VN = 6$ ;  $FP = 0$ ;  $Sensibilidad = \frac{3}{3+6}$ ;  $Especificidad = \frac{6}{6+0}$ ) (Cuadro X).

Cuadro X. Valores de Ct y resultados obtenidos por qPCR utilizando ADN extraído a partir de heces inoculadas con diluciones seriadas de cultivos de cepas del Clado C-I y los cebadores diseñados para las secuencias firma del Clado C-I.

Muestra	Clado	Concentración ADN en la reacción (ng/μl)	Cebadores	Ct	Resultado
Mx149 sin diluir	C-I	0,97	110	40	FN
Mx149 1:500		2,14		40	
Mx152 sin diluir		2,68		38,76	
Mx152 1:100		2,03		40	
SJ312 sin diluir		3,23		21,87	VP
SJ312 1:100		1,95		26,51	
SJ312 1:500		2,10		29,33	
LIBA-6538 1:100	1	2,15		40	VN
LIBA-6822	C-I	0,16	114	19,85	-
Mx149 sin diluir		0,97		30,66	VP
Mx149 1:100		2,43		31,35	
Mx149 1:500		2,14		33,64	
Mx152 sin diluir		2,68		25,17	
Mx152 1:100		2,03		30,86	
Mx152 1:500		2,22		34,19	
SJ312 sin diluir		3,23		20,4	
SJ312 1:100		1,95		24,22	
SJ312 1:500		2,10		27,38	
LIBA-6533 1:100	1	2,04		37,72	
LIBA-6533 1:500		2,42		40	
LIBA-6538 1:100		2,15		38,14	
Heces sin inocular	-	2,37		38,45	-
LIBA-6822	C-I	0,16	161	17,61	-
Mx149 sin diluir		0,97		28,11	VP
Mx149 1:100		2,43		28,91	
Mx149 1:500		2,14		31,41	
Mx152 sin diluir		2,68		21,96	
Mx152 1:100		2,03		28,06	
Mx152 1:500		2,22		31,6	
SJ312 sin diluir		3,23		23,33	
SJ312 1:100		1,95		27,03	
SJ312 1:500		2,10		30,28	
LIBA-6533 1:100	1	2,04		40	
LIBA-6533 1:500		2,42		40	

LIBA-6538 1:100		2,15		40	
LIBA-6538 1:500		1,95		40	
Heces sin inocular	-	2,37		40	-
LIBA-6822	C-I	0,16		20,57	-

\*Las muestras inoculadas con cultivos de las cepas Mx149 diluido 1:100 (FN), Mx152 diluido 1:500 (FN), LIBA-6533 en ninguna concentración (VN) y 6538 sin diluir (VN) y diluido 1:500 (VN) no generaron amplificación con los cebadores 110. Las muestras inoculadas con cultivos de las cepas LIBA-6533 sin diluir, LIBA-6538 sin diluir y diluido 1:500 no amplificaron con los cebadores 114 (VN). Las muestras inoculadas con cultivos de las cepas LIBA-6533 y LIBA-6538 sin diluir no amplificaron con los cebadores 161 (VN). Se utilizó ADN extraído a partir de un cultivo de la cepa LIBA-6822 como control positivo y como control negativo mezcla de reacción a la que no se le adicionó ADN. El control negativo no amplificó con ninguno de los cebadores. Se detallan las muestras utilizadas, el clado al que pertenecen las cepas utilizadas para la inoculación, las concentraciones de ADN total extraído en la reacción para cada una, el cebador utilizado y el Ct obtenido.

Al establecer como punto de corte  $Ct=35$ , los cebadores diseñados para el gen *tcdB* mostraron 100% de sensibilidad y 72,7% de especificidad ( $VP = 4$ ;  $FN = 0$ ;  $VN = 8$ ;  $FP = 3$ ;  $Sensibilidad = \frac{4}{4+0}$ ;  $Especificidad = \frac{8}{8+3}$ ) en el ensayo de qPCR con ADN extraído a partir de cultivos (Cuadro XI).

Cuadro XI. Valores de Ct obtenidos en el ensayo de qPCR utilizando ADN extraído a partir de cultivos de diversas bacterias y los cebadores diseñados para el gen *tcdB* presente en las cepas del Clado C-I.

Muestra	Clado	Concentración ADN en la reacción (ng/μl)	Ct	Resultado
LIBA-6291	1	0,89	34,31	FP
LIBA-7110	3	0,51	30	
A-B+	4	0,128	26,26	
LIBA-6625	C-I	0,065	17,24	VP
LIBA-6822		0,054	21,13	
LIBA-7194		0,22	20,05	
LIBA-6622		0,424	26,69	

\* Las cepas LIBA-5755 (Clado 1), NAP1 (Clado 2), LIBA-6279 (Clado 4), 29741 (*Bacterioides thetaiotaomicron*), A2 (*C. botulinum*), 25785 (*C. filamentosum*), 8327 y Sm101 (*C. perfringens*) no amplificaron. Todos estos corresponden a resultados

verdaderos negativos. Se utilizó mezcla de reacción a la que no se le adicionó ADN como control negativo, el cual no amplificó. Se indican las demás muestras utilizadas, el clado al que pertenecen, la concentración de ADN utilizada en la reacción y el Ct obtenido para cada una.

Estableciendo como punto de corte  $Ct=35$ , los cebadores diseñados para el gen *tcdB* discriminaron la presencia de cepas toxigénicas del Clado C-I en un ensayo de qPCR utilizando ADN extraído a partir de heces inoculadas con un 66% de sensibilidad y un 100% de especificidad ( $VP = 6$ ;  $FN = 3$ ;  $VN = 6$ ;  $FP = 0$ ; **Sensibilidad** =  $\frac{6}{6+3}$ ; **Especificidad** =  $\frac{6}{6+0}$ ) (Cuadro XII).

Cuadro XII. Valores de Ct obtenidos en el ensayo de qPCR utilizando ADN extraído a partir de heces inoculadas con diluciones seriadas de cultivos de cepas del Clado C-I y los cebadores diseñados para el gen *tcdB* presente en las cepas del Clado C-I.

Muestra	Clado/Especie	Concentración ADN en la reacción (ng/μl)	Ct	Resultado
Mx152 sin diluir	C-I	2,68	21,13	VP
Mx152 1:100		2,03	26,55	
Mx152 1:500		2,22	30,26	
SJ312 sin diluir		3,23	22,6	
SJ312 1:100		1,95	26,28	
SJ312 1:500		2,1	29,39	
Heces sin inocular	-	2,37	35	-

Se utilizó el mismo control negativo que en los casos anteriores. Las muestras inoculadas con cultivos de las cepas Mx149 (Clado C-I) (FN), LIBA-6533 (Clado 1) (VN) y LIBA-6538 (Clado 1) (VN) no amplificaron. Se utilizó mezcla de reacción a la que no se le adicionó ADN como control negativo, el cual no amplificó. Se detallan las muestras utilizadas, el clado al que pertenecen, la concentración de ADN total utilizada en la reacción y el Ct obtenido para cada una.

En general, los cebadores 114 y 161 obtuvieron sensibilidades del 100% y especificidades sobre el 80% en los ensayos realizados, exceptuando en el que se utilizaron heces inoculadas con ADN extraído a partir de cultivos. El par 110 mostró valores elevados de sensibilidad en los ensayos realizados con ADN extraído a partir de cepas (Cuadro XIII).

Cuadro XIII. Resumen de las sensibilidades y especificidades determinadas para los diferentes pares de cebadores ensayados.

Ensayo	Cebadores	Sensibilidad	Especificidad
PCR punto final	110	100%	90%
	114	100%	80%
	161	100%	90%
qPCR con ADN cepas	110	100%	60%
	114	100%	100%
	161	100%	100%
	<i>tcdB</i>	100%	72.7%
qPCR con ADN extraído de heces inoculadas con ADN	110	25%	-
	114	25%	-
	161	25%	-
qPCR con ADN extraído de heces inoculadas con cultivos	110	33%	100%
	114	100%	100%
	161	100%	100%
	<i>tcdB</i>	66%	100%

Utilizando ADN de la cepa C-I 7602 en un ensayo de qPCR, los cebadores 114, 161 y *tcdB* alcanzaron límites de detección de 0,0000165 (ng/μl) (dilución 1/10000), 0,000165 (ng/μl) (dilución 1/1000) y 0,0000165 (ng/μl) (dilución 1/10000), respectivamente (Figura 10).

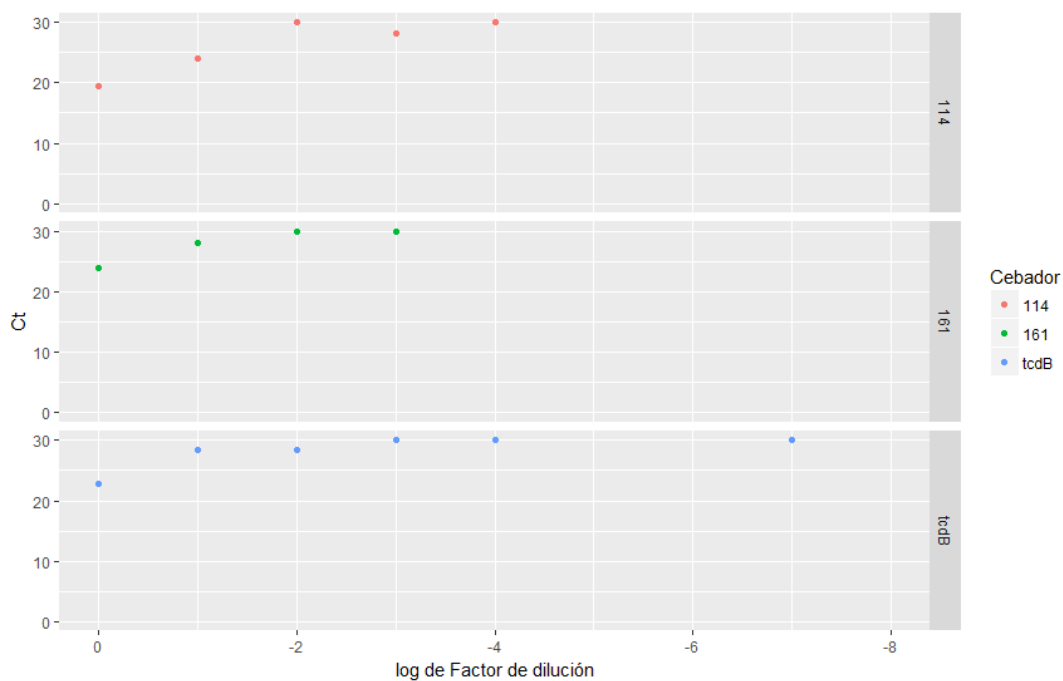


Figura 10. Gráfico semilogarítmico de factor de dilución vs Ct obtenido en el ensayo de límite de detección del ensayo de PCR tiempo real utilizando ADN extraído a partir de un cultivo de la cepa LIBA-7602 y los cebadores 114 (rojo), 161 (verde) y diseñados para el gen *tcdB* (azul) presente en las cepas del Clado C-I.

## Capítulo V: Discusión

### Genómica comparativa y pangenómica:

La Figura 5 muestra que las cepas del Clado C-I son genéticamente divergentes con respecto a las de los otros clados analizados. El presente análisis ubicó a estas cepas en dos subclados, uno integrado por tres cepas toxigénicas aisladas en Costa Rica, el otro por las 2 cepas aisladas en nuestro país restantes y las 5 reportadas por Dingle y colaboradores (Dingle et al., 2014). Estos resultados son similares a los reportados por otros grupos en estudios similares (Knight et al., 2015; Ramírez-Vargas et al., 2018; Stone et al., 2016). Esta divergencia también es evidenciada por los valores de ANI promedio mostrados en la Figura 6. En este caso, los valores obtenidos al comparar las cepas del Clado C-I con las de los otros clados fueron considerablemente menores a los generados a partir de las comparaciones dentro de estos dos grupos, lo cual revela un bajo nivel de similitud genética. De hecho, los valores obtenidos son menores al aceptado como umbral para la determinación de especie bacteriana utilizado por el NCBI (Ciuffo et al., 2018), lo que indica que cepas del Clado C-I deberían ser reclasificadas como una nueva genomoespecie del género *Clostridioides*, tal y como lo sugieren Knight y colaboradores (Knight et al., 2020).

El análisis realizado con Roary reveló un pangenoma amplio, con un total de 18952 CDSs, de las cuales únicamente 1412 (el 7,54%) pertenecen al genoma *core* (ver figura 9). Este pangenoma es más amplio que el reportado para otras especies de importancia clínica, como *Escherichia coli* (13000 genes) (Rasko et al., 2008) y *Staphylococcus aureus* (7411 genes) (Bosi et al., 2016). Dado que la riqueza del genoma accesorio sugiere adaptación a diferentes sitios u hospederos, esta observación indica que la muestra de cepas de *C. difficile* incluida en el análisis representa un grupo adaptado a una mayor diversidad de hábitats que los reportados para *E. coli* y *S. aureus*. El repertorio genético más amplio identificado para *C. difficile* sugiere también un metabolismo más versátil, así como una tendencia de la especie a adquirir material genético por transferencia horizontal (Blaustein et al., 2019).

Existen diferentes reportes de pangenomas determinados para *C. difficile*, el más antiguo reportado fue determinado utilizando tecnología de *microarray* y obtuvo un tamaño considerablemente menor (9640 secuencias codificantes) (Scaria et al., 2010), lo cual es

de esperarse dado que incluía una menor cantidad y diversidad de cepas de la especie. Allí se identificó un genoma *core* de entre 974 y 1033 genes, lo cual es paradójico, dado que se espera que a mayor diversidad de genomas, menos secuencias sean compartidas entre todos, y por tanto, el genoma *core* se haga más pequeño (Medini, Donati, Tettelin, Massignani, & Rappuoli, 2005). Es importante mencionar que dicho análisis es previo al uso de la clasificación cladística aceptada actualmente, y a la descripción de las cepas del Clado C-I, por lo que este factor puede ser importante para la explicación de dichos resultados.

Otros estudios han investigado el pangenoma de un subgrupo particular de esta especie (Knetsch et al., 2018; Knight et al., 2019; Knight, Squire, Collins, & Riley, 2017) siendo también más reducidos en tamaño, al representar subpoblaciones de la especie. En este caso el mejor punto de comparación es nuevamente el estudio realizado por Knight y colaboradores (2020) en el cual se obtuvo un pangenoma de 17470 genes y un genoma *core* de 2232 genes. Este resultado es similar al obtenido en el presente análisis, pero inesperado considerando que con mayor número de cepas ( $n=260$ ), todas de STs distintas, los autores obtuvieron un pangenoma menor y un genoma *core* mayor. Este hallazgo probablemente se deba a los sistemas bioinformáticos utilizados en los análisis. En el estudio mencionado se utilizaron dos aplicaciones bioinformáticas distintas para determinar el pangenoma: Roary, el mismo usado en la presente investigación, y Panaroo (Tonkin-Hill et al., 2020). Los autores indican que, utilizando su set de datos, Roary sobreestimó el tamaño del pangenoma (32802 genes), con una concomitante disminución en el porcentaje del genoma *core* (del 12.8% determinado con Panaroo a 3.9%). Esto es atribuible a que Panaroo identifica posibles errores de ensamblaje y anotación durante la generación del pangenoma, disminuyendo el porcentaje de genes por genoma. Es probable entonces que el pangenoma determinado por Roary en el presente estudio contenga algunos genes que en realidad representan anotaciones erróneas, o identificados a partir de errores de ensamblaje, generando un valor mayor al verdadero. La documentación de Panaroo fue publicada recientemente, por lo que no pudo ser utilizado en el presente proyecto.

Los resultados del análisis pangenómico por Roary también ilustran que las cepas del Clado C-I analizadas poseen una importante cantidad de secuencias que no se encuentran en las demás analizadas. Esto se evidencia en la reducción considerable en el tamaño absoluto (6239) del pangenoma de la especie que resulta de la sustracción de las

CDSs presentes en las cepas del Clado C-I del juego de datos completo. Este valor ilustra además la variabilidad presente en el clado y su divergencia con respecto a los otros incorporados en el análisis.

Como se ilustra en la Figura 9, las CDSs compartidas entre las cepas C-I, pero ausentes en las demás, tienen funciones relacionadas con el metabolismo, la transducción de señales y la transcripción. Todas estas funciones son típicas del genoma accesorio y se suelen relacionar con vías metabólicas que no son esenciales, pero otorgan a las cepas que las poseen ventajas, como la adaptación a nuevos nichos ecológicos o la colonización de nuevos hospederos (Medini et al., 2005). Esto es esperable considerando que *C. difficile* muestra especiación simpátrica, lo que implica adaptación a diferentes nichos ecológicos y hábitats, tales como agua, suelo, alimentos y el tracto gastrointestinal de diversas especies de animales. Es probable, entonces, que las secuencias identificadas le permitan a las cepas del Clado C-I adaptarse a nichos distintos a los ocupados por las de las cepas de los Clados 1-5; incluso se ha hipotetizado que las cepas de los Clados C-I a C-III podrían ser características de muestras ambientales (Janezic et al., 2016). Considerando que todas las cepas del Clado C-I analizadas en el presente trabajo corresponden a aislamientos realizados a partir de heces de pacientes diarreicos, y cinco de estos poseen potencial toxigénico, esto no implica que no sean capaces de colonizar o infectar huéspedes humanos, aunque quizás a través de mecanismos idiosincráticos que merecen ser explorados.

#### Identificación de secuencias firma:

Todos los métodos utilizados para la identificación de secuencias firma encontraron múltiples secuencias nucleotídicas estadísticamente asociadas con el Clado C-I, sin embargo, no todas fueron compatibles con el propósito de la investigación.

DBGWAS permitió la identificación de 111 componentes con valor de  $q$  menor a 0,05. Al proceder con la extracción manual de los unitigs significativos presentes en todas las cepas del grupo de inclusión, ausentes en el grupo de exclusión y de al menos 100pb de longitud (considerando que pudieran ser amplificables por PCR), se encontró que generaban *hits* con altos puntajes con secuencias presentes en otras cepas de *C. difficile* (ver Anexo VI). DBGWAS generó además un componente de tamaño inmanejable por el módulo que genera las visualizaciones (289709 unitigs), por lo que no fue posible extraer la secuencia de los nodos significativos. Es probable que este componente contenga

secuencias características del grupo, dada su magnitud, lo cual representa una pérdida importante de secuencias que posiblemente podrían haber sido útiles. Es importante aclarar que sólo recientemente se agregó la generación de resultados en forma de texto, lo cual evita esta limitante y facilita considerablemente la extracción de resultados de forma automatizada.

Por otro lado, al correr Neptune con parámetros predeterminados no se identificó ninguna secuencia firma con 100% de especificidad y sensibilidad. A pesar de lo anterior, se procedió a realizar BLASTn contra la colección del NCBI mencionada de las mejores cinco secuencias nucleotídicas identificadas con base en el puntaje establecido por la aplicación (ver Anexo III), encontrando nuevamente *hits* que hacían match de alto puntaje con otras cepas de *C. difficile* y otras especies bacterianas.

Lo anterior indica que los métodos mencionados no son apropiados para identificar secuencias firma susceptibles a ser amplificadas por PCR en el set de datos utilizado. Esto puede deberse a que ambos sistemas de identificación están basados en *k-mer*, por lo que se especializan en detectar secuencias exactamente iguales en el grupo de inclusión y distintas en al menos 1pb a las presentes en el grupo de exclusión. Quizá dada la alta variabilidad genética presente en esta especie, estos métodos parecen haber identificado muchas secuencias que tienen pequeñas diferencias, como SNPs o inserciones y deleciones de nucleótido simple. A pesar de que estas variantes podrían ser de interés para otro tipo de estudio, tales como los de asociación de genoma completo, no son tan útiles para el diseño de métodos de diagnóstico molecular, pues es poco probable encontrar cebadores específicos para las mismas, que no generen amplificación inespecífica en las variantes presentes en el grupo de exclusión.

Al consolidar los resultados obtenidos mediante el uso de Anvi'o, Scoary y treeWAS, se identificaron múltiples CDSs presentes en todas las cepas del Clado C-I y ausentes en el grupo de exclusión, que no generaron *hits* significativos al compararlas por BLASTn con la base de datos de colección de nucleótidos del NCBI (ver cuadro III). Estas secuencias son entonces, teóricamente, apropiadas para diseñar un método de identificación molecular basado en ellas. De estas, *lsrF*, *dhaK* y *crcB* fueron identificadas como parte del genoma *core* del Clado C-I, y sus funciones COG se encontraron enriquecidas en este grupo, lo cual refuerza su importante relación con las cepas del clado. El hecho de que múltiples de estas secuencias posean altos índices de homogeneidad geométrica es apropiado,

considerando que el método de identificación por diseñar se basa en amplificación de ácidos nucleicos.

En *E. coli*, *lsrF* codifica para una tiolasa dependiente de coenzima A que cataliza el paso final de la degradación del autoinductor-2 de señal de *quorum sensing*, interfiriendo con la comunicación de otras especies (Marques et al., 2014). En *C. difficile*, la producción del autoinductor-2 se ha relacionado con la generación de biopelículas importantes para la adherencia y colonización intestinal (Slater, Frost, Jossi, Millard, & Unnikrishnan, 2019) así como con la expresión de genes de virulencia (Lee & Song, 2005). El hecho de que las cepas del Clado C-I codifiquen este gen conlleva la interesante posibilidad de que tengan la capacidad de interferir con el *quorum sensing* y la expresión de factores de virulencia de las cepas de los Clados 1-5. Esto podría ser verificado en ensayos de cocultivo de cepas de ambos grupos, comparando el crecimiento y producción de toxinas de aquellas de los Clados 1-5 en estas condiciones y en monocultivo. De observarse un efecto, podría comprobarse que fuera ocasionado por la eliminación del autoinductor-2 realizando otro experimento similar, pero con cepas de *C. difficile* mutantes para el gen *luxS*, las cuales son incapaces de sintetizar el autoinductor-2 y por tanto no deberían verse afectadas.

En *E. coli*, el gen *dhaK* codifica para una subunidad catalíticamente activa de una dihidroxiacetona quinasa que utiliza un sistema fosfoenolpiruvato carbohidrato fosfotransferasa como fuente de energía. Este gen es parte de un operón que contiene además el gen *dhaL*, que codifica para otra subunidad con la misma función, y el gen *dhaM* o *ptsD*, que codifica una proteína de sistema fosfotransferasa (PTS) de fusión tripartita (Deutscher, Francke, & Postma, 2006). Se ha observado que las subunidades DhaK y DhaL se unen al factor de transcripción DhaR, que interviene en el metabolismo de la dihidroxiacetona, funcionando como correpresor y coactivador respectivamente (Bächler, Schneider, Bähler, Lustig, & Erni, 2005). Roary identificó dos conglomerados de genes parálogos (uno con 7 secuencias, otro con 3) y una secuencia única anotados como *dhaL*, presentes únicamente en cepas del Clado C-I. Estos resultados, aunados al hecho de que la función COG dihidroxiacetona quinasa fue determinada como parte del genoma *core* de las cepas del Clado C-I, sugiere que las cepas C-I podrían tener un sistema regulador similar al descrito en *E. coli*.

El gen *crcB* fue inicialmente identificado en *E. coli* y se relacionó con resistencia al alcanfor y condensación cromosomal (Hu et al., 1996). Posteriormente, este gen se ha

implicado en resistencia al fluoruro en diversos organismos, hipotetizándose que la proteína que codifica se trata de un transportador de eflujo para este ion (Li et al., 2013). Más recientemente se ha demostrado tal efecto en *Enterobacter cloacae* FRM (Liu et al., 2017). La expresión de este gen en las cepas del Clado C-I podrían conferir tal resistencia, la cual sería una ventaja con respecto a las cepas de los Clados 1-5, pues el fluoruro es ubicuo en el ambiente y tiene efectos antibacterianos importantes (Nelson, Plummer, Blount, Ames, & Breaker, 2015).

Además de las secuencias anteriores, la presencia de reguladores transcripcionales, como LsrR y de familia MerR, y una proteína de sistema de transporte ABC, refuerzan la noción de que múltiples de las secuencias firma identificadas en las cepas del Clado C-I pueden representar adaptaciones para nichos ecológicos distintos, o ventajas con respecto a otras cepas de la misma especie.

#### Diseño y validación del ensayo de PCR:

Utilizando Primer-BLAST se encontraron múltiples pares de cebadores específicos para ocho de las secuencias firma reportadas anteriormente (ver cuadro IV). Según los criterios establecidos en la metodología se eliminaron aquellos pares de cebadores que: a) tuvieran contenido GC menor al 40%, dado que esto disminuye la interacción entre estos y la secuencia molde, b) tuvieran grandes diferencias de Tm entre ambos cebadores, dado que esto dificulta la selección de una temperatura de alineamiento en el ensayo de PCR que permita una unión simultánea de ambos oligonucleótidos del par a la secuencia molde, c) tuvieran Tms con diferencia de más de 2°C con respecto al procedimiento descrito por Miller y colaboradores (2010), dado que se pretendía que fueran utilizables como parte de un esquema multiplex y d) aquellos que tuvieran autocomplementariedad elevada, particularmente en el extremo 3', dado que esta favorece la amplificación de dímeros de cebador.

A pesar de esto, se mantuvo una alta cantidad de pares de cebadores factibles, pero similares entre sí. Por lo anterior se decidió seleccionar los mejores 7 candidatos que amplificaran secuencias considerablemente distintas, para hacer un segundo cribado basado en las características termodinámicas predichas de los posibles dímeros de cebador y del amplicón esperado (cuadro V). En términos generales, las posibles estructuras identificadas no representaban un detrimento importante para su aplicación en ensayos de PCR, dado que los posibles dímeros de cebador tenían Tms bajos y las

estructuras secundarias del amplicón esperado no bloqueaban su extremo 3' (sitio de unión de la ADN polimerasa) y/o tenían también Tms bajos.

PCR de punto final:

Los resultados de los ensayos de PCR punto final (ver Anexo VII) mostraron que los pares de cebadores identificados como 110, 114 y 161 amplifican fragmentos de ADN de los tamaños esperados en presencia de ADN genómico de cepas del Clado C-I, demostrando que podrían ser útiles para discriminar la pertenencia de un microorganismo a este grupo, basándose en una metodología sencilla, rápida y económica. Los valores de especificidad obtenidos en este ensayo, particularmente para los cebadores 110 y 114, fueron elevados, comparables incluso con los reportados para los inmunoensayos de la enzima glutamato deshidrogenasa utilizados como pruebas rápidas de diagnóstico de la CDI (Arimoto et al., 2016). Por otro lado, se obtuvo una sensibilidad del 100% para todos los cebadores ensayados, el cual es el máximo valor posible, pero determinado con una pequeña cantidad de cepas.

Llama la atención que la cepa LIBA-7110, perteneciente al Clado 3 de *C. difficile*, haya sido positiva para todos los cebadores seleccionados como específicos. Estos resultados podrían sugerir que la determinación de la cepa como miembro del Clado 3 es incorrecta, sin embargo, su designación fue confirmada utilizando la base de datos de PubMLST, la cual clasificó la cepa como ST5 del Clado 3. Es importante recalcar que si bien el análisis filogenético por MLST es ampliamente utilizado en esta especie (Griffiths et al., 2010) no es inequívoco (X. Wang et al., 2018), por lo que este resultado no implica necesariamente que la cepa 7110 no sea filogenéticamente más cercana al Clado C-I que otras cepas del Clado 3, compartiendo algunos genes enriquecidos en dicho clado. A raíz de lo anterior se realizaron BLASTn comparando las secuencias blanco de los cebadores probados contra el genoma completo de la cepa 7110, sin embargo, no se encontraron generaron hits significativos entre ambos, restando credibilidad a esta hipótesis. Como método adicional para aclarar las dos interrogantes anteriores, se realizó un nuevo análisis de Roary incluyendo los 77 genomas utilizados en el estudio y el genoma de la cepa 7110. Este análisis determinó nuevamente que las CDSs blanco sólo se encuentran en las cepas del Clado C-I. Podría entonces suponerse que la amplificación se debe a que existan en el genoma de la cepa 7110 sitios de unión para los pares de cebadores probados, que no corresponden a las secuencias identificadas como blanco en las cepas C-I. Para

comprobar esta hipótesis se mapearon las secuencias de los cebadores al genoma completo de la cepa 7110 utilizando la herramienta PrimerMap (Stothard, 2000), sin embargo, no se reportaron sitios de unión.

Dado lo anterior es probable que los resultados observados se deban a cierto nivel de unión inespecífica con secuencias similares a las que amplifican los cebadores. Lo anterior se refuerza por el hecho que en todos los ensayos la banda observada para la cepa 7110 es considerablemente más tenue que la del control positivo (ver anexo VI, paneles c, d y e), a pesar de que el ADN de la primera se utilizó a una concentración poco más de tres veces mayor.

*qPCR usando células completas (cultivos puros):*

El punto de corte de  $C_t=35$  (sobre el cual el resultado se considera negativo) para los ensayos de qPCR para todos los cebadores probados fue seleccionado con base en la premisa de que dicho resultado se obtendría para muestras con una única copia de la secuencia blanco (suponiendo una eficiencia de amplificación del 90%-100%). Dado que los ensayos se realizaron con SYBR<sup>®</sup> Green, el cual se une al ADN doble banda de forma inespecífica, es probable que los valores de  $C_t$  superiores al punto de corte observados al agregar ADN extraído a partir de cepas no C-I cultivadas en medio líquido (ver cuadro VII) correspondan a amplificación por formación de dímeros de primer o unión inespecífica, la cual es menos eficiente. Considerando que los controles negativos consistentemente no alcanzaron el punto de corte en este, y los otros ensayos de PCR tiempo real, probablemente se trate del segundo caso descrito. Esta idea es reforzada por el hecho que, en el mismo ensayo, a concentraciones similares o menores de ADN, las reacciones con ADN de las cepas C-I mostraron  $C_t$ s cercanos a 20, evidenciando una mejor amplificación.

En el ensayo mencionado anteriormente, los cebadores 114 y 161 mostraron una especificidad y sensibilidad del 100% utilizando el punto de corte establecido. Si bien es cierto que dichos valores fueron calculados a partir del ensayo con un número restringido de cepas, particularmente pertenecientes al Clado C-I ( $n=4$ ), sugieren que ambos juegos podrían ser utilizados como una herramienta sencilla y fidedigna para contribuir a la clasificación filogenética de aislamientos de *C. difficile*. Dado que el perfil de temperatura utilizado es similar al utilizado en el protocolo descrito por Miller y colaboradores (Miller et al., 2010), teóricamente podrían incluirse en alguna de las reacciones multiplex

presentadas por esos autores, con lo cual se ampliaría el rango dinámico de detección sin incrementar la complejidad logística en el costo del ensayo.

A pesar de mostrar una sensibilidad del 100% en el mismo ensayo, la especificidad calculada para los cebadores 110 fue apenas de un 60%, la cual es muy baja y los excluye de ser considerados como una herramienta diagnóstica. A la luz de lo discutido sobre el ensayo de PCR de punto final para la cepa 7110, la selección del punto de corte, y la cantidad de ADN utilizado en la reacción, es posible suponer que los valores de Ct observados para las cepas 7110 y 6291 (ver cuadro VII) corresponden también a amplificación inespecífica, por lo que no invalidan el supuesto de que la secuencia utilizada para diseñar los cebadores (*crcB*) es firma de las cepas C-I. Lo anterior se evidencia en que, por ejemplo, el ADN de la cepa 7110 se utilizó a una concentración en reacción de 0,51 ng/μl y generó un Ct de 30, mientras que la cepa 6822 se empleó a una concentración en reacción de 0,054 ng/μl mostrando un Ct de 19,4, lo cual muestra una mejor amplificación en el segundo caso. Es probable que exista entonces alguna secuencia lo suficientemente complementaria al par de cebadores en las cepas 7110 y 6291 que genere la amplificación observada.

No se observó amplificación en ninguna de las reacciones con ADN de cepas del Clado C-I usando los cebadores 195. Por esta razón, aunado el hecho de que los otros cebadores diseñados para las secuencias firma identificadas mostraron valores elevados de sensibilidad y especificidad, se decidió no proseguir con más ensayos para los primeros. Es probable que estos resultados negativos se deban a variables analíticas más que de diseño, tomando en cuenta que los cebadores 195 amplificaron los productos esperados en el ensayo de PCR de punto final (ver figura 9). Es por tanto probable que modificando las condiciones de la qPCR se podría estandarizar su uso como prueba de identificación de *C. difficile* del Clado C-I, a pesar de que en el presente no se realizó, a la luz de limitaciones de recurso económico y temporal.

#### qPCR con ADN extraído a partir de heces:

El Cuadro VIII demuestra que los cebadores diseñados no generan amplificación en ensayos de PCR tiempo real al analizar heces de pacientes diarreicos negativos para *C. difficile*. Esto es relevante para los objetivos del proyecto, ya que demuestra que no se generan resultados falsos positivos en el ensayo, lo cual da pie a su uso para identificar la presencia de cepas de *C. difficile* del Clado C-I en esta matriz compleja. Además, al no

requerirse el cultivo, los cebadores podrán ser utilizados como parte de un sistema diagnóstico en una mayor cantidad de laboratorios clínicos, dado que el aislamiento de bacterias anaerobias es complejo y requiere de equipo y personal que muchas veces no se encuentra disponible en clínicas u hospitales del país. Esto implica también que el costo de implementar y utilizar el ensayo en centros de salud donde se cuenta con equipo de qPCR sería reducido.

qPCR con ADN extraído a partir de muestras inoculadas:

En el ensayo de qPCR usando ADN extraído a partir de heces inoculadas con ADN de cepas del Clado C-I (ver cuadro IX), los cebadores 110, 114 y 161 mostraron sensibilidades mucho menores que en los ensayos anteriores. Considerando que las cepas utilizadas para la inoculación dieron lugar a resultados positivos utilizando ADN en cultivo, no sería lógico suponer que estos resultados se tratan de falsos negativos por errores de diseño de los cebadores, sino que estos se generaron a raíz de las diferentes condiciones experimentales. Es probable que, durante el paso de extracción, el ADN inoculado al encontrarse libre en el medio se haya degradado considerablemente, al exponerse al buffer de extracción y altas temperaturas. Cabe aclarar que este ensayo no refleja realmente las condiciones en que se pretende utilizar los cebadores, dado que la idea es detectar ADN que se encuentra en las esporas o células vegetativas de *C. difficile* y no ADN libre en las heces. Por lo anterior, estos resultados no invalidan la utilidad del sistema propuesto.

En el ensayo con ADN extraído a partir de heces inoculadas con cultivos de *C. difficile* (ver cuadro X) se observaron Cts por debajo del punto de corte de 35 para las muestras inoculadas con cepas del Clado C-I. Esto demuestra que a concentraciones de ADN similares a las utilizadas en los ensayos anteriores sí es posible observar amplificación lo que sugiere que el factor de peso es la presencia o ausencia de la secuencia blanco en la reacción. En este grupo de ensayos, los cebadores 114 y 161 obtuvieron sensibilidades y especificidades del 100%, los cuales son incluso superiores a los reportados para ensayos disponibles comercialmente que identifican otras cepas de la misma especie en heces (Shin et al., 2012).

En el mismo ensayo, el par de cebadores 110 mostró una sensibilidad bastante menor que los otros dos mencionados (33%). Por sí solos, estos resultados podrían llevar a pensar que lo anterior se debe a que no existía suficiente ADN de la cepa C-I en las

reacciones para que la señal se reconociera, dado que este no fue cuantificado directamente, sino que se cuantificó el ADN total extraído a partir de la muestra de heces inoculada. Esto es poco probable, dado que los otros dos cebadores ensayados si generaron amplificación a Ct menor al punto de corte utilizando las mismas muestras inoculadas con cultivos de cepas del Clado C-I, e incluso el par 114 amplifica una secuencia en el mismo gen (*crcB*) que el par 110. Probablemente estos resultados se deban a que las cepas que no generaron amplificación antes de alcanzar el punto de corte tengan variantes en la secuencia blanco para el par de cebadores 110, provocando una menor interacción entre ambos, y una menor amplificación. Como se observa en el cuadro VI, el cebador forward del par 110 se une a una secuencia que, comparando las secuencias de las 10 cepas analizadas para la identificación de las secuencias firma, posee 4 sitios variables, lo cual podría explicar la diferencia observada. A partir de los resultados anteriores, y de forma similar a lo propuesto anteriormente con respecto a su uso para clasificación de aislamientos, no se recomienda utilizar este par de cebadores para identificar la presencia de cepas de *C. difficile* del Clado C-I en heces de pacientes diarreicos.

#### qPCR con cebadores para el gen *tcdB*:

En el ensayo de qPCR con ADN extraído a partir de cultivos, los cebadores diseñados para el gen *tcdB* mostraron un 100% de sensibilidad (ver cuadro XI), lo cual es apropiado para los fines metodológicos planteados. Al contarse con otros dos cebadores que mostraron una especificidad del 100% en la detección de la presencia de ADN de cepas del Clado C-I, los cebadores para el gen *tcdB* cumplen la función de determinar si las cepas del clado presentes en una muestra tienen o no potencial toxigénico, y por tanto podrían causar CDI. Para tal propósito es útil disponer de un ensayo altamente sensible, que identifique la mayor cantidad posible de ensayos positivos, dado que los falsos positivos serían identificables al corroborar que los cebadores 114 o 161 no generan amplificación, como sucedió con las cepas 6291, A-B+ y 7110, las cuales podrían suponerse como *C. difficile* toxigénicas de otros clados. Al respecto, es importante mencionar que PrimerMap no mostró sitios de unión para los cebadores diseñados, lo cual, en conjunto con los valores altos de Ct obtenidos para las cepas 6291 y 7110, podrían sugerir que la amplificación observada corresponde a unión inespecífica y

descarta el uso de estos cebadores como método para identificar la presencia del gen *tcdB* de otros clados de *C. difficile* en una muestra, sea de heces o cultivo puro.

Los valores de Ct obtenidos al utilizar ADN extraído de la cepa A-B+ fueron comparables a los de algunas cepas del Clado C-I, sin embargo, no fue posible verificar la presencia de sitios de unión en su genoma para los cebadores dado que este no está disponible. No obstante, algunos alelos de *tcdB* de las cepas Clado C-I codifican toxinas variantes como las sintetizadas por las cepas A-B+ (Ramírez-Vargas et al., 2018), por lo que podría existir cierto nivel de reactividad cruzada.

Al ensayarse con ADN extraído a partir de heces inoculadas con cultivos de *C. difficile*, los cebadores diseñados para el gen *tcdB* mostraron una sensibilidad del 66% (ver cuadro XII). Interesantemente, no se observó amplificación en ninguna de las reacciones con ADN extraído a partir de la cepa MX-149 del Clado C-I. Esta cepa posee un gen *tcdB* con una secuencia notablemente distinta (datos no mostrados), por lo que posiblemente escapa a la detección. Esta es una limitante importante del método propuesto, dado que es posible que otras cepas de este clado presenten secuencias de *tcdB* divergentes lo que generaría falsos negativos en cuanto a la predicción de capacidad toxigénica. Estos resultados se presentarían con una frecuencia incierta, dado que se desconoce cuán abundantes son estos alelos variantes en la población. Sería posible verificar si las cepas identificadas con los cebadores 114 y/o 161 son toxigénicas mediante cultivo toxigénico y análisis de neutralización de citotoxicidad celular, sin embargo, esto incurriría en costos y tiempos de espera mayores.

El límite de detección para ADN extraído a partir de cultivos puros observado en el ensayo de PCR tiempo real realizado con diluciones seriadas para los cebadores 114 y los diseñados para el gen *tcdB* fue de 0,0000165 ng/ $\mu$ l, lo cual corresponde a 3,56 copias de ADN cromosomal completo (asumiendo un largo de 4,3 Mpb) por microlitro de reacción (ver Figura 10). Estos valores son cercanos a los reportados para ensayos similares de identificación de otras especies y medios a partir de cultivo. Es importante aclarar que este valor es una aproximación, dado que para ser estadísticamente válido se deben realizar al menos 6 réplicas de cada concentración (Kralik & Ricchi, 2017), lo cual no pudo realizarse debido a limitaciones de recurso económico y tiempo. De hecho, es probable que la amplificación observada a Ct 30 (*late call*) en la reacción utilizando ADN diluido a razón 1/10000000 con los cebadores para el gen *tcdB* se trate de un resultado falso

irrepetible estadísticamente, dada la muy baja concentración de ADN que se encuentra en la reacción.

Propuesta temporal de algoritmo diagnóstico:

A partir de los resultados obtenidos en los ensayos de PCR tiempo real y su discusión, se propone el siguiente algoritmo para determinar la presencia de *C. difficile* del Clado C-I en heces de pacientes diarreicos.

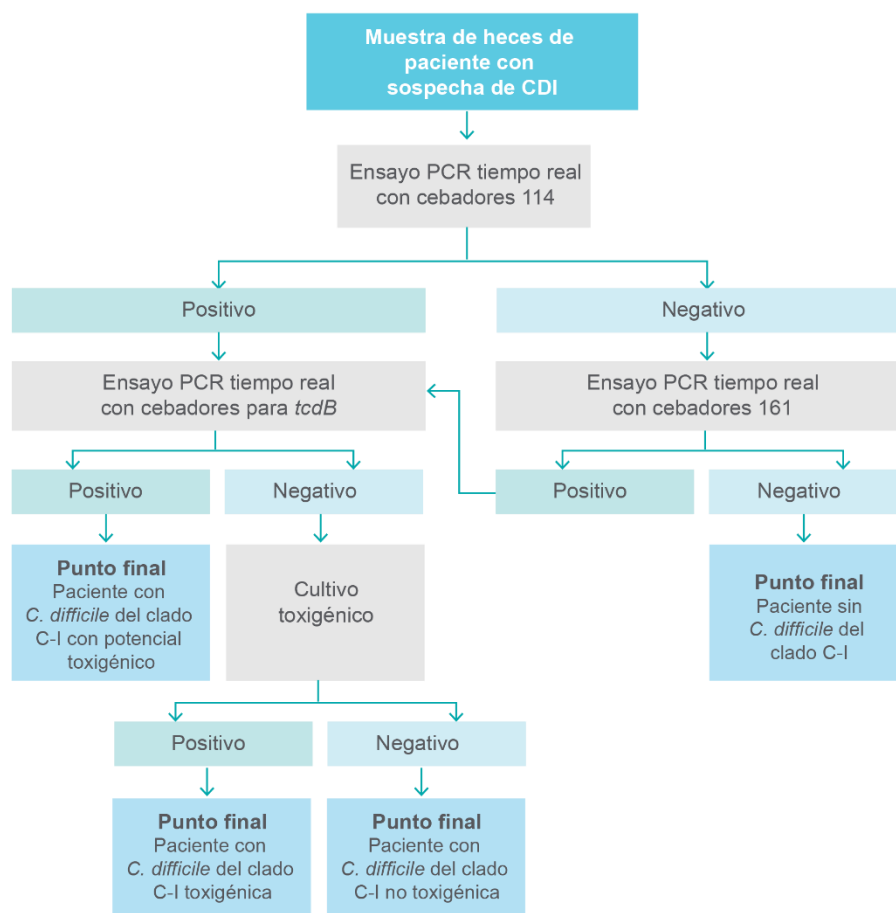


Figura 11. Algoritmo propuesto para determinar la presencia de cepas de *C. difficile* del Clado C-I, toxigénicas o no, en heces de pacientes diarreicos con sospecha de CDI. Se representan los métodos a utilizar en gris y los puntos finales en azul.

El algoritmo propuesto permite discernir la presencia de cepas del Clado C-I en muestras de heces utilizando uno o dos de los pares de cebadores diseñados para secuencias firma identificadas en el presente proyecto. Se recomienda utilizar primero uno de los dos cebadores, en lugar de ambos en paralelo, pues mostraron sensibilidades y especificidades del 100% en este ensayo, por lo que el resultado positivo para uno de los cebadores sería suficiente para confirmar que la cepa pertenece a este clado. Se establece como primer paso utilizar el par 114 dado que mostró un menor límite de detección. Se recomienda utilizar el segundo par como un mecanismo alternativo de confirmación, en el caso hipotético de que la presunta cepa C-I presente en las heces tenga algún(os) SNPs en el gen *crcB* con respecto a las cepas analizadas que impida su amplificación y por ende genere un resultado falso negativo.

De forma adicional a los objetivos planteados se incluye la determinación del potencial toxigénico de la cepa en el ensayo, mediante una reacción de PCR en tiempo real con los cebadores diseñados para el gen *tcdB*. Este paso extra le da un mayor potencial de uso a nivel clínico al algoritmo propuesto, dado que podría llevar a la identificación de casos de CDI por cepas del Clado C-I toxigénicas, lo cual, a la fecha, no es posible utilizando los métodos disponibles en laboratorio clínico. Sólo en caso de que este cebador mostrase resultados negativos en pacientes con sospecha de CDI y presencia de *C. difficile* del Clado C-I determinada por alguno de los otros dos cebadores presentados sería necesario utilizar cultivo toxigénico o un ensayo de neutralización de citotoxicidad para determinar la expresión de las toxinas por parte de la cepa del Clado C-I y así confirmar el caso. Como se mencionó anteriormente, también existe la posibilidad de diseñar otro juego de cebadores para detectar el gen *tcdB* de la cepa MX-149 y similares, lo cual permitiría identificar estos resultados falsos positivos, disminuyendo la necesidad de realizar otro tipo de ensayos.

Es importante recalcar que la presencia de *C. difficile* con potencial toxigénico en las heces de un paciente, detectada por NAATs, no necesariamente implica que se trate de un caso de CDI, dado que en estos no se detecta directamente la producción de la toxina (Peng et al., 2018). Así mismo, existe cierto porcentaje de individuos colonizados por *C. difficile* que no presentan CDI, los cuales tampoco podrían ser distinguibles por el ensayo. Por esta razón, al igual que con otras pruebas NAAT para este patógeno, es importante su uso únicamente en pacientes en los que se sospecha la enfermedad, de forma tal que

no se sobrediagnostique la condición e incurra en costos innecesarios por tratamiento u hospitalización (Polage et al., 2015).

Además del uso anterior, el algoritmo descrito puede ser aplicado con los mismos pasos, pero a partir del ADN extraído de cultivos puros de *C. difficile* para determinar si la cepa en cuestión pertenece al Clado C-I y es toxigénica. Esta posible aplicación permitiría tal tipificación por medio de un ensayo relativamente más sencillo y económico que el ribotipado o MLST.

#### Consideraciones finales y limitaciones:

Este trabajo valida el uso de pangenómica para identificar secuencias firma para su posterior uso en métodos de identificación de microorganismos específicos en una matriz compleja o su caracterización filogenética.

Utilizando un procedimiento relativamente sencillo y accesible se identificaron múltiples secuencias candidatas, para las cuales se diseñaron cebadores específicos que las amplifican de forma exitosa. A pesar de que no se diseñaron cebadores para todas las secuencias posibles, y de que solamente se ensayó un subgrupo de los cebadores identificados en ensayos de PCR tiempo real, se obtuvo un algoritmo funcional y operativo, sin que esto implique que sea el óptimo. Existe aún en los resultados un reservorio de opciones disponibles para, en un futuro caso hipotético, diseñar un nuevo método diagnóstico con objetivos similares. De hecho, el procedimiento bioinformático aquí seguido podría ser usado para identificar cualquier otro tipo de secuencias firma en cualquier bacteria.

Dado que los ensayos basados en NAAT son costo-efectivos, este procedimiento podría ser útil para diseñar nuevos ensayos que mejoren la capacidad diagnóstica o de identificación de agentes causales de enfermedades infecciosas que afectan a países o poblaciones con menos recursos económicos que los requeridos para aplicar otros métodos de mayor costo. Los cebadores diseñados también podrían ser utilizados para mejorar métodos diagnósticos ya existentes, lo cual favorece su transferencia desde el laboratorio de investigación al laboratorio clínico, eludiendo la generación de un procedimiento completamente nuevo y los costos que esto conlleva. Sería posible, además, que este tipo de metodología sea usada para desarrollar métodos de detección de patógenos, u otros organismos de interés, en otras matrices complejas, como agua o

suelo, ajustando la especificidad de los cebadores diseñados y los métodos de extracción de ADN.

La metodología utilizada, y por tanto los resultados obtenidos en el presente proyecto, están sujetos a ciertas limitaciones. En primer lugar, se encuentra la escasa disponibilidad de genomas para incluirse en la investigación. Si bien para el 30 de septiembre de 2020 la base de datos de genomas del NCBI incluía 2345 ensamblajes de *C. difficile*, pocos de ellos se acompañaban de información sobre su clasificación filogenética. En el presente estudio se incluyeron 10 cepas del Clado C-I, un número limitado, pero el materialmente posible considerando únicamente aquellas cepas debidamente reportadas y clasificadas cladísticamente en estudios previos. Alternativamente se podría haber buscado genomas de esta especie cuya clasificación cladística fuera desconocida y realizar la clasificación utilizando métodos bioinformáticos, con el fin de encontrar genomas adicionales del Clado C-I. Esta opción hubiera sido altamente demandante en términos de tiempo y poder computacional, y no garantiza que genomas con estas características se hubieran encontrado, dado que podrían no existir en esta base de datos. Dada esta limitante, es posible que existan cepas del Clado C-I que no tengan en su genoma las secuencias identificadas como firma en el presente análisis, lo cual generaría resultados falsos negativos. A pesar de lo anterior, debido a los elevados valores de ANI obtenidos entre las cepas del Clado C-I (ver figura 6) y a que el gen blanco para el cebador 114 (*crcB*) fue identificado como *core* en este grupo, esta opción es poco probable.

En lo referente a las pruebas en “*wet lab*”, los cebadores 114 y 161 mostraron valores del 100% de especificidad y sensibilidad para determinar la presencia de cepas de *C. difficile* del Clado C-I en heces de pacientes diarreicos, así como en la capacidad de discernir si una cepa aislada en cultivo pertenece o no al mismo clado. Los cebadores para el gen *tcdB* diseñados también mostraron una buena capacidad en discernir si dichas cepas, sea en heces o en cultivo, tienen potencial toxigénico o no. Lo anterior sugiere que podrían utilizarse en la práctica clínica para ayudar en el diagnóstico de la CDI, en aquellos casos donde el cuadro clínico genera sospechas de la enfermedad, pero los ensayos de laboratorio normalmente utilizados para su diagnóstico muestran resultados negativos o contradictorios. Aún así, los resultados representan una “prueba de concepto”, y sería necesario complementar con varios pasos antes de que los productos de la investigación puedan ser aplicados en la práctica clínica.

Primeramente, a pesar de que se aproximó el límite de detección en el ensayo de PCR en tiempo real, esto se hizo utilizando ADN extraído a partir de cultivos puros, por lo que fueron expresados como ng/μl. La aplicación de los cebadores para realizar la detección de la bacteria en heces a nivel clínico requiere de la determinación del límite de detección del ensayo expresado como UFC/volumen inoculado, de forma que se compruebe si es posible registrar la presencia de la bacteria en cantidades similares a las que se espera que tengan los pacientes afectados por CDI. Este factor es reconocido como una fuente potencial de resultados falsos negativos en ensayos comercialmente disponibles (Murad et al., 2016), por lo que mostrar un límite de detección “aceptable” sería clave para confirmar la utilidad del ensayo en el contexto clínico.

Otro aspecto que podría mejorarse con respecto a la metodología descrita es que a pesar de que en el proyecto se utilizó SYBR® Green para la detección, sería recomendable utilizar sondas Taqman® para los amplicones esperados. Este cambio permitiría discernir la amplificación de las secuencias esperadas de otras inespecíficas, de forma tal que sea posible distinguir ambos casos y evitar resultados falsos positivos. El uso de sondas Taqman posibilitará además determinar la amplificación del producto del cebador 114 y el diseñado para el gen *tcdB* en una sola reacción de PCR tiempo real, disminuyendo el tiempo de espera del algoritmo propuesto. Por último, el uso de esta tecnología permitiría cuantificar la cantidad de ADN de *C. difficile* toxigénico o no, lo cual correlaciona con la cantidad de bacterias presentes en heces. Este factor puede ser clínicamente relevante en ciertos casos de CDI.

Finalmente, los ensayos para determinar la sensibilidad y especificidad realizados tienen un número bajo de cepas presuntamente positivas (4) y un número mayor, pero aun así limitado, de aquellas presuntamente negativas (>10). Tal número es considerablemente menor que el generalmente utilizado para validar una metodología diagnóstica. Esta limitante existe a pesar de que el estudio se realizó con el apoyo del LIBA, el cual es referencia en el aislamiento, cultivo y tipificación de *C. difficile* a nivel nacional, y de que en nuestro país se han aislado varias cepas pertenecientes al Clado C-I. Sería necesario validar los resultados obtenidos con un número considerablemente mayor de cepas, particularmente no pertenecientes al Clado C-I, y de muestras de heces para asegurar el 100% de sensibilidad y especificidad en los ensayos, o para determinar su verdadero valor y si son aceptables para los fines planteados. Este tipo de ensayo a gran escala posee un costo elevado, que difícilmente puede ser asumido por una institución

académica, por lo que su realización probablemente requeriría del establecimiento de alianzas con empresas u organismos similares, interesados en aplicar el ensayo comercialmente.

**Capítulo VI: Conclusiones:**

- Las cepas del Clado C-I de *C. difficile* pertenecen a un grupo genéticamente divergente con respecto a las cepas de los Clados 1-5.
- Las 10 cepas del Clado C-I de *C. difficile* contempladas en este estudio poseen al menos 11 secuencias firma que se encuentran ausentes en cepas de los Clados 1-5 de la misma especie.
- Los cebadores identificados como 114 y 161 detectan por qPCR las cepas del Clado C-I de *C. difficile* en cultivo y heces con un 100% de especificidad y sensibilidad.

## Bibliografía

- Akerlund, T., Persson, I., Unemo, M., Noren, T., Svenungsson, B., Wullt, M., & Burman, L. G. (2008). Increased Sporulation Rate of Epidemic *Clostridium difficile* Type 027/NAP1. *Journal of Clinical Microbiology*, *46*(4), 1530–1533. <https://doi.org/10.1128/JCM.01964-07>
- Anani, H., Zgheib, R., Hasni, I., Raoult, D., & Fournier, P.-E. (2020). Interest of bacterial pangenome analyses in clinical microbiology. *Microbial Pathogenesis*, *149*, 104275. <https://doi.org/https://doi.org/10.1016/j.micpath.2020.104275>
- Arimoto, J., Horita, N., Kato, S., Fuyuki, A., Higurashi, T., Ohkubo, H., ... Nakajima, A. (2016). Diagnostic test accuracy of glutamate dehydrogenase for *Clostridium difficile*: Systematic review and meta-analysis. *Scientific Reports*, *6*, 29754. <https://doi.org/10.1038/srep29754>
- Awad, M. M., Johanesen, P. A., Carter, G. P., Rose, E., & Lyras, D. (2014). *Clostridium difficile* virulence factors: Insights into an anaerobic spore-forming pathogen. *Gut Microbes*, *5*(5), 579–593. <https://doi.org/10.4161/19490976.2014.969632>
- Bächler, C., Schneider, P., Bähler, P., Lustig, A., & Erni, B. (2005). *Escherichia coli* dihydroxyacetone kinase controls gene expression by binding to transcription factor DhaR. *The EMBO Journal*, *24*(2), 283–293. <https://doi.org/10.1038/sj.emboj.7600517>
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., ... Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Barlam, T. F., Cosgrove, S. E., Abbo, L. M., MacDougall, C., Schuetz, A. N., Septimus, E. J., ... Trivedi, K. K. (2016). Executive Summary: Implementing an Antibiotic Stewardship Program: Guidelines by the Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, *62*(10), 1197–1202. <https://doi.org/10.1093/cid/ciw217>
- Bartlett, J. G., & Gerding, D. N. (2008). Clinical Recognition and Diagnosis of *Clostridium difficile* Infection. *Clinical Infectious Diseases*, *46*(s1), S12–S18. <https://doi.org/10.1086/521863>
- Blaustein, R. A., McFarland, A. G., Ben Maamar, S., Lopez, A., Castro-Wallace, S., & Hartmann, E. M. (2019). Pangenomic Approach To Understanding Microbial Adaptations within a Model Built Environment, the International Space Station,

- Relative to Human Hosts and Soil. *MSystems*, 4(1), e00281-18.  
<https://doi.org/10.1128/mSystems.00281-18>
- Bosi, E., Monk, J. M., Aziz, R. K., Fondi, M., Nizet, V., & Palsson, B. Ø. (2016). Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proceedings of the National Academy of Sciences of the United States of America*, 113(26), E3801-9.  
<https://doi.org/10.1073/pnas.1523199113>
- Brown, A. W. W., & Wilson, R. B. (2018). Clostridium difficile colitis and zoonotic origins-a narrative review. *Gastroenterology Report*, 6(3), 157–166.  
<https://doi.org/10.1093/gastro/goy016>
- Brynildsrud, O., Bohlin, J., Scheffer, L., & Eldholm, V. (2016). Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biology*, 17(1), 238. <https://doi.org/10.1186/s13059-016-1108-8>
- Burnham, C.-A. D., & Carroll, K. C. (2013). Diagnosis of Clostridium difficile Infection: an Ongoing Conundrum for Clinicians and for Clinical Laboratories. *Clinical Microbiology Reviews*, 26(3), 604–630. <https://doi.org/10.1128/CMR.00016-13>
- Chang, M. C., Hur, J., & Park, D. (2020). Interpreting the COVID-19 Test Results: A Guide for Physiatrists. *American Journal of Physical Medicine & Rehabilitation*, 99(7), 583–585. <https://doi.org/10.1097/PHM.0000000000001471>
- Chaudhari, N. M., Gupta, V. K., & Dutta, C. (2016). BPGA- an ultra-fast pan-genome analysis pipeline. *Sci Rep*, 6, 24373. <https://doi.org/10.1038/srep24373>
- Chikhi, R., & Medvedev, P. (2014). Informed and automated k-mer size selection for genome assembly. *Bioinformatics*, 30(1), 31–37.  
<https://doi.org/10.1093/bioinformatics/btt310>
- Ciufo, S., Kannan, S., Sharma, S., Badretdin, A., Clark, K., Turner, S., ... DiCuccio, M. (2018). Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *International Journal of Systematic and Evolutionary Microbiology*, 68(7), 2386–2392. <https://doi.org/10.1099/ijsem.0.002809>
- Collins, C., & Didelot, X. (2018). A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. *PLOS Computational Biology*, 14(2), e1005958.  
<https://doi.org/10.1371/journal.pcbi.1005958>
- Debast, S. B., Bauer, M. P., & Kuijper, E. J. (2014). European Society of Clinical Microbiology and Infectious Diseases: Update of the Treatment Guidance Document

- for *Clostridium difficile* Infection. *Clinical Microbiology and Infection*, 20, 1–26. <https://doi.org/10.1111/1469-0691.12418>
- Delmont, T. O., & Eren, A. M. (2018). Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ*, 6, e4320. <https://doi.org/10.7717/peerj.4320>
- Deutscher, J., Francke, C., & Postma, P. W. (2006). How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiology and Molecular Biology Reviews: MMBR*, 70(4), 939–1031. <https://doi.org/10.1128/MMBR.00024-06>
- Di Bella, S., Ascenzi, P., Siarakas, S., Petrosillo, N., & di Masi, A. (2016). *Clostridium difficile* toxins A and B: Insights into pathogenic properties and extraintestinal effects. *Toxins*, 8(5), 134. <https://doi.org/10.3390/toxins8050134>
- Dingle, K. E., Elliott, B., Robinson, E., Griffiths, D., Eyre, D. W., Stoesser, N., ... Didelot, X. (2014). Evolutionary History of the *Clostridium difficile* Pathogenicity Locus. *Genome Biology and Evolution*, 6(1), 36–52. <https://doi.org/10.1093/gbe/evt204>
- Elliott, B., Androga, G. O., Knight, D. R., & Riley, T. V. (2017). *Clostridium difficile* infection: Evolution, phylogeny and molecular epidemiology. *Infection, Genetics and Evolution*, 49, 1–11. <https://doi.org/10.1016/j.meegid.2016.12.018>
- Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*, 3, e1319. <https://doi.org/10.7717/peerj.1319>
- Eze, P., Balsells, E., Kyaw, M. H., & Nair, H. (2017). Risk factors for *Clostridium difficile* infections – an overview of the evidence base and challenges in data synthesis. *Journal of Global Health*, 7(1), 1–9. <https://doi.org/10.7189/jogh.07.010417>
- Falush, D. (2016). Bacterial genomics: Microbial GWAS coming of age. *Nature Microbiology*, 1(5), 16059. <https://doi.org/10.1038/nmicrobiol.2016.59>
- Fournier, P.-E., Dubourg, G., & Raoult, D. (2014). Clinical detection and characterization of bacterial pathogens in the genomics era. *Genome Medicine*, 6(11), 114. <https://doi.org/10.1186/s13073-014-0114-2>
- George, R. H., Symonds, J. M., Dimock, F., Brown, J. D., Arabi, Y., Shinagawa, N., ... Burdon, D. W. (1978). Identification of *Clostridium difficile* as a cause of pseudomembranous colitis. *British Medical Journal*, 1(6114), 695. Retrieved from <http://www.bmj.com/content/1/6114/695.abstract>
- Gerding, D. N., Kelly, C. P., Rahav, G., Lee, C., Dubberke, E. R., Kumar, P. N., ... Dorr, M.

- B. (2018). Bezlotoxumab for Prevention of Recurrent *Clostridium difficile* Infection in Patients at Increased Risk for Recurrence. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 67(5), 649–656.  
<https://doi.org/10.1093/cid/ciy171>
- Gouy, M., Guindon, S., & Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, 27(2), 221–224.  
<https://doi.org/10.1093/molbev/msp259>
- Govind, R., & Dupuy, B. (2012). Secretion of *Clostridium difficile* toxins A and B requires the holin-like protein tcdE. *PLoS Pathogens*, 8(6), 1–14.  
<https://doi.org/10.1371/journal.ppat.1002727>
- Govind, R., Fitzwater, L., & Nichols, R. (2015). Observations on the role of TcdE isoforms in *Clostridium difficile* toxin secretion. *Journal of Bacteriology*, 197(15), 2600–2609.  
<https://doi.org/10.1128/JB.00224-15>
- Griffiths, D., Fawley, W., Kachrimanidou, M., Bowden, R., Crook, D. W., Fung, R., ... Dingle, K. E. (2010). Multilocus sequence typing of *Clostridium difficile*. *Journal of Clinical Microbiology*, 48(3), 770–778. <https://doi.org/10.1128/JCM.01796-09>
- He, M., Miyajima, F., Roberts, P., Ellison, L., Pickard, D. J., Martin, M. J., ... Lawley, T. D. (2013). Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nature Genetics*, 45(1), 109–113. <https://doi.org/10.1038/ng.2478>
- Hu, K. H., Liu, E., Dean, K., Gingas, M., DeGraff, W., & Trun, N. J. (1996). Overproduction of Three Genes Leads to Camphor Resistance and Chromosome Condensation in *Escherichia coli*. *Genetics*, 143(4), 1521 LP – 1532. Retrieved from <http://www.genetics.org/content/143/4/1521.abstract>
- Hubbard, K., Pellar, G., & Emanuel, P. (2011). Suitability of Commercial Transport Media for Biological Pathogens under Nonideal Conditions. *International Journal of Microbiology*, 2011, 463096. <https://doi.org/10.1155/2011/463096>
- Hung, G.-C., Nagamine, K., Li, B., & Lo, S.-C. (2012). Identification of DNA Signatures Suitable for Use in Development of Real-Time PCR Assays by Whole-Genome Sequence Approaches: Use of *Streptococcus pyogenes* in a Pilot Study. *Journal of Clinical Microbiology*, 50(8), 2770–2773. <https://doi.org/10.1128/JCM.01155-12>
- Jaillard, M., Lima, L., Tournoud, M., Mahé, P., van Belkum, A., Lacroix, V., & Jacob, L. (2018). A fast and agnostic method for bacterial genome-wide association studies: Bridging the gap between k-mers and genetic events. *PLOS Genetics*, 14(11),



problem: global phylogenomic analysis uncovers three ancient, toxigenic, genomospecies. *BioRxiv*, 2020.09.21.307223.

<https://doi.org/10.1101/2020.09.21.307223>

Knight, D. R., Kullin, B., Androga, G. O., Barbut, F., Eckert, C., Johnson, S., ... Riley, T. V. (2019). Evolutionary and Genomic Insights into *Clostridioides difficile* Sequence Type 11: a Diverse Zoonotic and Antimicrobial-Resistant Lineage of Global One Health Importance. *MBio*, 10(2), e00446-19.

<https://doi.org/10.1128/mBio.00446-19>

Knight, D. R., & Riley, T. V. (2019). Genomic Delineation of Zoonotic Origins of *Clostridium difficile*. *Frontiers in Public Health*, Vol. 7, p. 164. Retrieved from

<https://www.frontiersin.org/article/10.3389/fpubh.2019.00164>

Knight, D. R., Squire, M. M., Collins, D. A., & Riley, T. V. (2017). Genome Analysis of *Clostridium difficile* PCR Ribotype 014 Lineage in Australian Pigs and Humans Reveals a Diverse Genetic Repertoire and Signatures of Long-Range Interspecies Transmission. *Frontiers in Microbiology*, 7, 2138.

<https://doi.org/10.3389/fmicb.2016.02138>

Korber, F., Zeller, I., Grünstäudl, M., Willinger, B., Apfalter, P., Hirschl, A. M., & Makristathis, A. (2017). SeptiFast versus blood culture in clinical routine - A report on 3 years experience. *Wiener Klinische Wochenschrift*, 129(11–12), 427–434.

<https://doi.org/10.1007/s00508-017-1181-3>

Kralik, P., & Ricchi, M. (2017). A basic guide to real time PCR in microbial diagnostics: Definitions, parameters, and everything. *Frontiers in Microbiology*, 8(FEB), 1–9.

<https://doi.org/10.3389/fmicb.2017.00108>

Lawson, P. A., Citron, D. M., Tyrrell, K. L., & Finegold, S. M. (2016). Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prévot 1938.

*Anaerobe*, 40, 95–99. <https://doi.org/10.1016/j.anaerobe.2016.06.008>

Lee, A. S. Y., & Song, K. P. (2005). LuxS/autoinducer-2 quorum sensing molecule regulates transcriptional virulence gene expression in *Clostridium difficile*.

*Biochemical and Biophysical Research Communications*, 335(3), 659–666.

<https://doi.org/10.1016/j.bbrc.2005.07.131>

Leinonen, R., Sugawara, H., & Shumway, M. (2011). The Sequence Read Archive.

*Nucleic Acids Research*, 39(Database), D19–D21.

<https://doi.org/10.1093/nar/gkq1019>

Lemee, L., Dhalluin, A., Pestel-Caron, M., Lemeland, J.-F., & Pons, J.-L. (2004). Multilocus

- sequence typing analysis of human and animal *Clostridium difficile* isolates of various toxigenic types. *Journal of Clinical Microbiology*, 42(6), 2609–2617.  
<https://doi.org/10.1128/JCM.42.6.2609-2617.2004>
- Li, S., Smith, K. D., Davis, J. H., Gordon, P. B., Breaker, R. R., & Strobel, S. A. (2013). Eukaryotic resistance to fluoride toxicity mediated by a widespread family of fluoride export proteins. *Proceedings of the National Academy of Sciences*, 110(47), 19018 LP – 19023. <https://doi.org/10.1073/pnas.1310439110>
- Liu, X., Tian, J., Liu, L., Zhu, T., Yu, X., Chu, X., ... Fan, Y. (2017). Identification of an operon involved in fluoride resistance in *Enterobacter cloacae* FRM. *Scientific Reports*, 7(1), 6786. <https://doi.org/10.1038/s41598-017-06988-1>
- Logares, R., Haverkamp, T. H. A., Kumar, S., Lanzén, A., Nederbragt, A. J., Quince, C., & Kausrud, H. (2012). Environmental microbiology through the lens of high-throughput DNA sequencing: synopsis of current platforms and bioinformatics approaches. *Journal of Microbiological Methods*, 91(1), 106–113.  
<https://doi.org/10.1016/j.mimet.2012.07.017>
- López-Ureña, D., Quesada-Gómez, C., Montoya-Ramírez, M., del Mar Gamboa-Coronado, M., Somogyi, T., Rodríguez, C., & Rodríguez-Cavallini, E. (2016). Predominance and high antibiotic resistance of the emerging *Clostridium difficile* genotypes NAPCR1 and NAP9 in a Costa Rican hospital over a 2-year period without outbreaks. *Emerging Microbes & Infections*, 5(5), e42–e42.  
<https://doi.org/10.1038/emi.2016.38>
- Mairena-Morera, J. (2014). Infección hospitalaria por *Clostridium* está dentro del comportamiento habitual. Retrieved from <https://www.ccss.sa.cr/noticia?infeccion-hospitalaria-por-clostridium-esta-dentro-del-comportamiento-habitual>
- Marinier, E., Zaheer, R., Berry, C., Weedmark, K. A., Domaratzki, M., Mabon, P., ... Van Domselaar, G. (2017). Neptune: a bioinformatics tool for rapid discovery of genomic variation in bacterial populations. *Nucleic Acids Research*, 45(18), e159–e159.  
<https://doi.org/10.1093/nar/gkx702>
- Markham, N. R., & Zuker, M. (2005). DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Research*, 33(Web Server issue), W577-81.  
<https://doi.org/10.1093/nar/gki591>
- Marques, J. C., Oh, I. K., Ly, D. C., Lamosa, P., Ventura, M. R., Miller, S. T., & Xavier, K. B. (2014). LsrF, a coenzyme A-dependent thiolase, catalyzes the terminal step in processing the quorum sensing signal autoinducer-2. *Proceedings of the National*

- Academy of Sciences of the United States of America*, 111(39), 14235–14240.  
<https://doi.org/10.1073/pnas.1408691111>
- Martin-Verstraete, I., Peltier, J., & Dupuy, B. (2016). The regulatory networks that control *Clostridium difficile* toxin synthesis. *Toxins*, 8(5), 1–24.  
<https://doi.org/10.3390/toxins8050153>
- Marvaud, J. C., Quevedo-Torres, S., Eckert, C., Janoir, C., & Barbut, F. (2019). Virulence of new variant strains of *Clostridium difficile* producing only toxin A or binary toxin in the hamster model. *New Microbes and New Infections*, 32(toxinotype VIII), 100590.  
<https://doi.org/10.1016/j.nmni.2019.100590>
- McDonald, L. C., Gerding, D. N., Johnson, S., Bakken, J. S., Carroll, K. C., Coffin, S. E., ... Wilcox, M. H. (2018). Clinical Practice Guidelines for *Clostridium difficile* Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clinical Infectious Diseases*, 66(7), e1–e48. <https://doi.org/10.1093/cid/cix1085>
- McDonald, L. C., Killgore, G. E., Thompson, A., Owens, R. C., Kazakova, S. V., Sambol, S. P., ... Gerding, D. N. (2005). An Epidemic, Toxin Gene–Variant Strain of *Clostridium difficile*. *New England Journal of Medicine*, 353(23), 2433–2441.  
<https://doi.org/10.1056/NEJMoa051590>
- Medini, D., Donati, C., Tettelin, H., Massignani, V., & Rappuoli, R. (2005). The microbial pan-genome. *Current Opinion in Genetics & Development*, 15(6), 589–594.  
<https://doi.org/10.1016/j.gde.2005.09.006>
- Miller, M., Gravel, D., Mulvey, M., Taylor, G., Boyd, D., Simor, A., ... Kelly, S. (2010). Health Care–Associated *Clostridium difficile* Infection in Canada: Patient Age and Infecting Strain Type Are Highly Predictive of Severe Outcome and Mortality. *Clinical Infectious Diseases*, 50(2), 194–201. <https://doi.org/10.1086/649213>
- Monot, M., Eckert, C., Lemire, A., Hamiot, A., Dubois, T., Tessier, C., ... Dupuy, B. (2015). *Clostridium difficile*: New Insights into the Evolution of the Pathogenicity Locus. *Sci Rep*, 5, 15023. <https://doi.org/10.1038/srep15023>
- Murad, Y. M., Perez, J., Ybazeta, G., Mavin, S., Lefebvre, S., Weese, J. S., ... Nokhbeh, R. (2016). False Negative Results in *Clostridium difficile* Testing. *BMC Infectious Diseases*, 16(1), 430. <https://doi.org/10.1186/s12879-016-1741-6>
- Nagamine, K., Hung, G.-C., Li, B., & Lo, S.-C. (2015). DNA Sequence Signatures for Rapid Detection of Six Target Bacterial Pathogens Using PCR Assays. *Microbiology Insights*, 8, MBI.S29736. <https://doi.org/10.4137/MBI.S29736>

- Napolitano, L. M., & Edmiston, C. E. (2017). Clostridium difficile disease: Diagnosis, pathogenesis, and treatment update. *Surgery*, *162*(2), 325–348.  
<https://doi.org/https://doi.org/10.1016/j.surg.2017.01.018>
- Nelson, J. W., Plummer, M. S., Blount, K. F., Ames, T. D., & Breaker, R. R. (2015). Small molecule fluoride toxicity agonists. *Chemistry & Biology*, *22*(4), 527–534.  
<https://doi.org/10.1016/j.chembiol.2015.03.016>
- Ogielska, M., Lanotte, P., Le Brun, C., Valentin, A. S., Garot, D., Tellier, A.-C., ... Bernard, L. (2015). Emergence of community-acquired Clostridium difficile infection: the experience of a French hospital and review of the literature. *International Journal of Infectious Diseases : IJID : Official Publication of the International Society for Infectious Diseases*, *37*, 36–41. <https://doi.org/10.1016/j.ijid.2015.06.007>
- Pacheco, S. M., & Johnson, S. (2013). Important clinical advances in the understanding of Clostridium difficile infection. *Curr Opin Gastroenterol*, *29*(1), 42–48.  
<https://doi.org/10.1097/MOG.0b013e32835a68d4>
- Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., ... Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, *31*(22), 3691–3693. <https://doi.org/10.1093/bioinformatics/btv421>
- Paul De Vos, editors. (2009). *Bergey's manual of systematic bacteriology. Volume three, The firmicutes*. Retrieved from  
<https://search.library.wisc.edu/catalog/9910108350802121>
- Peng, Z., Ling, L., Stratton, C. W., Li, C., Polage, C. R., Wu, B., & Tang, Y. W. (2018). Advances in the diagnosis and treatment of Clostridium difficile infections. *Emerg Microbes Infect*, *7*(1), 15. <https://doi.org/10.1038/s41426-017-0019-4>
- Phillippy, A. M., Mason, J. A., Ayanbule, K., Sommer, D. D., Taviani, E., Huq, A., ... Salzberg, S. L. (2007). Comprehensive DNA Signature Discovery and Validation. *PLoS Computational Biology*, *3*(5), e98. <https://doi.org/10.1371/journal.pcbi.0030098>
- Polage, C. R., Gyorke, C. E., Kennedy, M. A., Leslie, J. L., Chin, D. L., Wang, S., ... Cohen, S. H. (2015). Overdiagnosis of Clostridium difficile Infection in the Molecular Test Era. *JAMA Internal Medicine*, *175*(11), 1792–1801.  
<https://doi.org/10.1001/jamainternmed.2015.4114>
- Quesada-Gomez, C., Rodriguez, C., Gamboa-Coronado, M. d. M., Rodriguez-Cavallini, E., Du, T., Mulvey, M. R., ... Boza-Cordero, R. (2010). Emergence of Clostridium difficile NAP1 in Latin America. *Journal of Clinical Microbiology*, *48*(2), 669–670.  
<https://doi.org/10.1128/JCM.02196-09>

- Rabold, D., Espelage, W., Abu Sin, M., Eckmanns, T., Schneeberg, A., Neubauer, H., ... Lübke-Becker, A. (2018). The zoonotic potential of *Clostridium difficile* from small companion animals and their owners. *PLoS One*, *13*(2), e0193411. <https://doi.org/10.1371/journal.pone.0193411>
- Ramírez-Vargas, G., López-Ureña, D., Badilla, A., Orozco-Aguilar, J., Murillo, T., Rojas, P., ... Rodríguez, C. (2018). Novel Clade C-I *Clostridium difficile* strains escape diagnostic tests, differ in pathogenicity potential and carry toxins on extrachromosomal elements. *Scientific Reports*, *8*(1), 13951. <https://doi.org/10.1038/s41598-018-32390-6>
- Ramírez-Vargas, G., Quesada-Gómez, C., Acuña-Amador, L., López-Ureña, D., Murillo, T., del Mar Gamboa-Coronado, M., ... Rodríguez, C. (2017). A *Clostridium difficile* Lineage Endemic to Costa Rican Hospitals Is Multidrug Resistant by Acquisition of Chromosomal Mutations and Novel Mobile Genetic Elements. *Antimicrobial Agents and Chemotherapy*, *61*(4). <https://doi.org/10.1128/AAC.02054-16>
- Ramírez-Vargas, G., & Rodríguez, C. (2020). Putative Conjugative Plasmids with *tcdB* and *cdtAB* Genes in *Clostridioides difficile*. *Emerging Infectious Diseases*, *26*(9), 2287–2290. <https://doi.org/10.3201/eid2609.191447>
- Rasko, D. A., Rosovitz, M. J., Myers, G. S. A., Mongodin, E. F., Fricke, W. F., Gajer, P., ... Ravel, J. (2008). The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *Journal of Bacteriology*, *190*(20), 6881–6893. <https://doi.org/10.1128/JB.00619-08>
- Robinson, C. D., Auchtung, J. M., Collins, J., & Britton, R. A. (2014). Epidemic *Clostridium difficile* Strains Demonstrate Increased Competitive Fitness Compared to Nonepidemic Isolates. *Infection and Immunity*, *82*(7), 2815–2825. <https://doi.org/10.1128/IAI.01524-14>
- Rodriguez, C., Van Broeck, J., Taminiau, B., Delmée, M., & Daube, G. (2016). *Clostridium difficile* infection: Early history, diagnosis and molecular strain typing methods. *Microb Pathog*, *97*, 59–78. <https://doi.org/10.1016/j.micpath.2016.05.018>
- Sacramento de Oliveira, A. C. do, Rosa, M. C., Borchardt, J. L., Menegon, Y. A., Menegon, Y. A., Andrade Fernandes, M. M., ... de Moraes, C. M. (2018). Validating the Efficiency of a Simplex PCR and Quantitative SYBR Green qPCR for the Identification of *Salmonella* spp. DNA. *Journal of Food: Microbiology, Safety & Hygiene*, *03*(01). <https://doi.org/10.4172/2476-2059.1000130>
- Saeed, M., Ahmad, M., Iram, S., Riaz, S., Akhtar, M., & Aslam, M. (2017). GeneXpert

- technology. A breakthrough for the diagnosis of tuberculous pericarditis and pleuritis in less than 2 hours. *Saudi Medical Journal*, 38(7), 699–705.  
<https://doi.org/10.15537/smj.2017.7.17694>
- Samarkos, M., Mastrogianni, E., & Kampouropoulou, O. (2018). The role of gut microbiota in *Clostridium difficile* infection. *Eur J Intern Med*, 50, 28–32.  
<https://doi.org/10.1016/j.ejim.2018.02.006>
- Scaria, J., Ponnala, L., Janvilisri, T., Yan, W., Mueller, L. A., & Chang, Y.-F. (2010). Analysis of ultra low genome conservation in *Clostridium difficile*. *PLoS One*, 5(12), e15147. <https://doi.org/10.1371/journal.pone.0015147>
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
- Shin, S., Kim, M., Kim, M., Lim, H., Kim, H., Lee, K., & Chong, Y. (2012). Evaluation of the Xpert *Clostridium difficile* assay for the diagnosis of *Clostridium difficile* infection. *Annals of Laboratory Medicine*, 32(5), 355–358.  
<https://doi.org/10.3343/alm.2012.32.5.355>
- Slater, R. T., Frost, L. R., Jossi, S. E., Millard, A. D., & Unnikrishnan, M. (2019). *Clostridioides difficile* LuxS mediates inter-bacterial interactions within biofilms. *Scientific Reports*, 9(1), 9903. <https://doi.org/10.1038/s41598-019-46143-6>
- Soehn, F., Wagenknecht-Wiesner, A., Leukel, P., Kohl, M., Weidmann, M., von Eichel-Streiber, C., & Braun, V. (1998). Genetic rearrangements in the pathogenicity locus of *Clostridium difficile* strain 8864 – implications for transcription, expression and enzymatic activity of toxins A and B. *Molecular and General Genetics MGG*, 258(3), 222–232. <https://doi.org/10.1007/s004380050726>
- Spigaglia, P., Barbanti, F., Mastrantonio, P., Brazier, J. S., Barbut, F., Delmee, M., ... R. Poxton, I. (2008). Fluoroquinolone resistance in *Clostridium difficile* isolates from a prospective study of *C. difficile* infections in Europe. *Journal of Medical Microbiology*, 57(6), 784–789. <https://doi.org/10.1099/jmm.0.47738-0>
- Stone, N. E., Sidak-Loftis, L. C., Sahl, J. W., Vazquez, A. J., Wiggins, K. B., Gillece, J. D., ... Wagner, D. M. (2016). More than 50% of *Clostridium difficile* Isolates from Pet Dogs in Flagstaff, USA, Carry Toxigenic Genotypes. *PLOS ONE*, 11(10), e0164504. Retrieved from <https://doi.org/10.1371/journal.pone.0164504>
- Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques*, 28(6), 1102, 1104.  
<https://doi.org/10.2144/00286ir01>

- Stöver, B. C., & Müller, K. F. (2010). TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics*, *11*(1), 7. <https://doi.org/10.1186/1471-2105-11-7>
- Tatusov, R. L., Galperin, M. Y., Natale, D. A., & Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, *28*(1), 33–36. <https://doi.org/10.1093/nar/28.1.33>
- Thornton, C. S., Rubin, J. E., Greninger, A. L., Peirano, G., Chiu, C. Y., & Pillai, D. R. (2018). Epidemiological and genomic characterization of community-acquired *Clostridium difficile* infections. *BMC Infectious Diseases*, *18*(1), 443. <https://doi.org/10.1186/s12879-018-3337-9>
- Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J. A., ... Parkhill, J. (2020). Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biology*, *21*(1), 180. <https://doi.org/10.1186/s13059-020-02090-4>
- Tschudin-Sutter, S., Kuijper, E. J., Durovic, A., Vehreschild, M. J. G. T., Barbut, F., Eckert, C., ... van Broeck, J. (2018). Guidance document for prevention of *Clostridium difficile* infection in acute healthcare settings. *Clinical Microbiology and Infection*, *24*(10), 1051–1054. <https://doi.org/10.1016/j.cmi.2018.02.020>
- Wang, B., Peng, W., Zhang, P., & Su, J. (2018). The characteristics of *Clostridium difficile* ST81, a new PCR ribotype of toxin A- B+ strain with high-level fluoroquinolones resistance and higher sporulation ability than ST37/PCR ribotype 017. *FEMS Microbiology Letters*, *365*(17). <https://doi.org/10.1093/femsle/fny168>
- Wang, X., Holzbauer, S., Pung, K., Bye, M., Adamczyk, M., Paulick, A. L., ... Workgroup, the E. I. P. C. D. I. (2018). 473. Molecular Typing of *Clostridium difficile*: Concordance Between PCR-Ribotyping and Multilocus Sequence Typing (MLST). *Open Forum Infectious Diseases*, *5*(Suppl 1), S176–S176. <https://doi.org/10.1093/ofid/ofy210.482>
- Weese, J. S. (2020). *Clostridium* (*Clostridioides*) *difficile* in animals. *Journal of Veterinary Diagnostic Investigation : Official Publication of the American Association of Veterinary Laboratory Diagnosticians, Inc*, *32*(2), 213–221. <https://doi.org/10.1177/1040638719899081>
- Wieczorkiewicz, J. T., Lopansri, B. K., Cheknis, A., Osmolski, J. R., Hecht, D. W., Gerding, D. N., & Johnson, S. (2015). Fluoroquinolone and Macrolide Exposure Predict *Clostridium difficile* Infection with the Highly Fluoroquinolone- and Macrolide-Resistant Epidemic *C. difficile* Strain BI/NAP1/027. *Antimicrobial Agents and Chemotherapy*, *60*(1), 418–423. <https://doi.org/10.1128/AAC.01820-15>

- Winston, J. A., & Theriot, C. M. (2016). Impact of microbial derived secondary bile acids on colonization resistance against *Clostridium difficile* in the gastrointestinal tract. *Anaerobe*, *41*, 44–50. <https://doi.org/10.1016/j.anaerobe.2016.05.003>
- Wong-McClure, R. A., Guevara-Rodríguez, M., Abarca-Gómez, L., Solano-Chinchilla, A., Marchena-Picado, M., O’Shea, M., & Badilla-Vargas, X. (2012). *Clostridium difficile* outbreak in Costa Rica: control actions and associated factors. *Rev Panam Salud Publica*, *32*(6), 413–418. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/23370184>
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*(847 Pt 1), 134. <https://doi.org/10.1186/1471-2105-13-134>

**Anexo I. ST, ribotipo de PCR, identificador de aislamiento, toxigenicidad, clado, fuente y fecha de aislamiento y numero de corrida de secuenciación de las cepas incluidas en el diseño experimental.**

ST	Ribotipo de PCR	ID de aislamiento	ID de genoma	Toxigenicidad	Clado	Fuente del aislamiento	Fecha de aislamiento	Número de Run	Cepa
10	015	Ox660	C00000096	+	1	Clínica de Oxford	11.09.2007	ERR126247	Ox660
2	ND	Ox709	C00000098	+	1	Clínica de Oxford	24.09.2007	ERR126248	Ox709
8	002	Ox1192c	C00000117	+	1	Clínica de Oxford	14.04.2008	ERR126249	Ox1192c
17	018	Ox1192e	C00000143	+	1	Clínica de Oxford	14.04.2008	ERR126250	Ox1192e
13	129	Ox1533	C00000149	+	1	Clínica de Oxford	01.10.2008	ERR126251	Ox1533
14	014	Ox1463	C00000160	+	1	Clínica de Oxford	27.08.2008	ERR126252	Ox1463
42	106	Ox687	C00000175	+	1	Clínica de Oxford	19.09.2007	ERR126253	Ox687
35	284	Ox1121	C00000182	+	1	Clínica de Oxford	09.03.2008	ERR126254	Ox1121
77	011	Ox2017	C00000197	+	1	Clínica de Oxford	22.07.2009	ERR126256	Ox2017
44	015	Ox990	C00000477	+	1	Clínica de Oxford	21.01.2008	ERR126258	Ox990
58	056	Ox707	C00000509	+	1	Clínica de Oxford	26.09.2007	ERR126259	Ox707
43	054	Ox1844	C00000516	+	1	Clínica de Oxford	28.03.2009	ERR126260	Ox1844
3	262	Ox1003	C00000551	-	1	Clínica de Oxford	23.01.2008	ERR126263	Ox1003
12	003	Ox1424	C00000562	+	1	Clínica de Oxford	01.08.2008	ERR126265	Ox1424
59	316	Ox1893	C00001463	+	1	Clínica de Oxford	04.05.2009	ERR126267	Ox1893
63	053	Ox1918	C00002482	+	1	Clínica de Oxford	25.05.2009	ERR126279	Ox1918
66	022	Ox2124	C00002975	+	1	Clínica de Oxford	11.09.2009	ERR126283	Ox2124
16	029	Ox858	C00004139	+	1	Clínica de Oxford	30.11.2007	ERR126284	Ox858
36	011	Ox2553	C00004350	+	1	Clínica de Oxford	07.05.2010	ERR126285	Ox2553
15	010	Ox1342	C00005172	-	1	Clínica de Oxford	10.06.2008	ERR126287	Ox1342
45	454	Ox2294	C00005740	+	1	Clínica de Oxford	15.12.2009	ERR126288	Ox2294
3	305	Ox428	C00006019	+	1	Clínica de Oxford	14.06.2007	ERR126290	Ox428
49	014	Ox593	C00006325	+	1	Clínica de Oxford	14.08.2007	ERR126292	Ox593
48	038	Ox2167	C00006353	-	1	Clínica de Oxford	12.10.2009	ERR126294	Ox2167

44	ND	Ox821a	C00000041	+	1	Clínica de Oxford	15.11.2007	ERR142058	Ox821a
3	ND	Ox746b	C00000059	+	1	Clínica de Oxford	16.10.2007	ERR142075	Ox746b
6	005	Ox2155	C00002831	+	1	Clínica de Oxford	29.09.2009	ERR142456	Ox2155
7	ND	Ox464a	C00005171	-	1	Clínica de Oxford	05.07.2007	ERR232376	Ox464a
7	ND	Ox1380	C00004164	+	1	Clínica de Oxford	08.07.2008	ERR232377	Ox1380
26	039	L039	C00005982	-	1	Aislamiento de librería de ribotipo de PCR	07.06.2008 [a]	ERR232378	L039
27	067	L067	C00007149	-	1	Aislamiento de librería de ribotipo de PCR	07.06.2008 [a]	ERR232379	L067
29	ND	OxB_KBP 102	C00005945	-	1	Infante de Oxford	25.01.2009	ERR232380	OxB_KBP_102
40	035	L035	C00006347	-	1	Aislamiento de librería de ribotipo de PCR	07.06.2008 [a]	ERR232384	L035
54	012	Ox42	C00000483	+	1	Clínica de Oxford	04.10.2006	ERR232385	Ox42
60	336	Ox1924	C00006256	+	1	Clínica de Oxford	30.05.2009	ERR232386	Ox1924
103	ND	Ox3036	C00007799	+	1	Clínica de Oxford	09.07.2011	ERR232388	Ox3036
107	ND	OVG261A	C00011314	-	1	Infante de Oxford	06.09.2011	ERR232389	OVG261A
122	ND	OPT_2249	C00010940	+	1	Clínica Canadiense	16.04.2009	ERR232391	OPT_2249
152	ND	L_L,12.7901566	C00011733	+	1	Clínica de Leeds	16.02.2012	ERR232392	L_L127901566
1	027	Ox160	C00000090	+	2	Clínica de Oxford	29.01.2007	ERR126246	Ox160
41	106	Ox1307	C00000224	+	2	Clínica de Oxford	30.05.2008	ERR126257	Ox1307
41	321	Ox1802b	C00002439	+	2	Clínica de Oxford	28.02.2009	ERR126273	Ox1802b
95	075	L075	C00002460	+	2	Aislamiento de librería de ribotipo de PCR	07.06.2008 [a]	ERR126276	L075
67	019	Ox2163	C00002493	+	2	Clínica de Oxford	07.10.2009	ERR126281	Ox2163
47	375	Ox2621b	C00005984	+	2	Clínica de Oxford	17.06.2010	ERR126289	Ox2621b
62	036	8864	C00012856	+	2	Clínica de humano (Soehn et al., 1998)	1980s	ERR232387	8864
22	023	Ox561	C00001506	+	3	Clínica de Oxford	07.08.2007	ERR126268	Ox561

<b>96</b>	058	L058	C00002452	+	3	Aislamiento de librería de ribotipo de PCR	07.06.2008 [a]	ERR126275	L058
<b>5</b>	069	Ox2183	C00005010	+	3	Clínica de Oxford	18.10.2009	ERR126286	Ox2183
<b>25</b>	023	Ox1232	C00002723	+	3	Clínica de Oxford	27.04.2007	ERR142362	Ox1232
<b>5</b>	023	Ox1485	C00004983	+	3	Clínica de Oxford	11.09.2008	ERR232375	Ox1485
<b>162</b>	ND	Q24	C00006439	+	3	Clínica Australiana	03.12.2008	ERR232393	Q24
<b>37</b>	017	Ox920	C00000089	A-B+	4	Clínica de Oxford	26.12.2007	ERR126245	Ox920
<b>23</b>	138	Ox904	C00006282	-	4	Clínica de Oxford	18.12.2007	ERR126291	Ox904
<b>30</b>	ND	OxB_HB21	C00006340	-	4	Infante de Oxford	12.11.2007	ERR232381	OxB_HB21
<b>38</b>	060	L060	C00006345	-	4	Aislamiento de librería de ribotipo de PCR	07.06.2008 [a]	ERR232382	L060
<b>39</b>	ND	OxI_WBC201I	C00007671	-	4	Infante de Oxford	02.05.2011	ERR232383	OxI_WBC201I
<b>109</b>	ND	BF095I	C00011286	-	4	Infante de Oxford	23.12.2011	ERR232390	BF095I
<b>170</b>	ND	ES248	C00006473	A-B+	4	Clínica Australiana	2010	ERR232397	ES248
<b>37</b>	ND	Ox3196	C00011267	A-B+	4	Clínica de Oxford	31.01.2012	ERR266408	Ox3196
<b>11</b>	078	Ox575	C00001729	+	5	Clínica de Oxford	14.08.2007	ERR126270	Ox575
<b>11</b>	193	Leeds 16/26	C00002387	+	5	Clínica de Leeds	19.10.2006	ERR126271	Leeds1626
<b>11</b>	066	7721734	C00002471	+	5	Clínica de Leeds	24.11.2009	ERR126277	7721734
<b>11</b>	045	L045	C00002490	+	5	Aislamiento de librería de ribotipo de PCR	07.06.2008 [a]	ERR126280	L045
<b>164</b>	ND	HCD52	C00006462	-	5	Clínica Australiana	1980s	ERR232394	HCD52
<b>167</b>	ND	Q6	C00006451	-	5	Clínica Australiana	03.05.2007	ERR232395	Q6
<b>168</b>	239	WA12	C00006522	-	5	Clínica Australiana (Elliott et al. 2009)	11.12.2005	ERR232396	WA12
<b>177</b>	ND	C. grifficile H5078	C00007672	-	C-I	Clínica de Oxford	18.01.2011	ERR232398	CDH5078
<b>178</b>	125	RPH118	C00006475	-	C-I	Clínica Australiana	11.04.2007	ERR232399	RPH118
<b>179</b>	289	RPH97	C00006510	-	C-I	Clínica Australiana	05.01.2007	ERR232400	RPH97
<b>180</b>	290	RPH128	C00006499	-	C-I	Clínica Australiana	10.07.2007	ERR232401	RPH128
<b>181</b>	127	RPH39	C00009015	-	C-I	Clínica Australiana	11.10.2006	ERR232402	RPH39

<b>ND</b>	ND	HBC-277194	ND	A-B+	C-I	LIBA	2016	ND	HBC-277194
<b>360</b>	ND	MX-152	ND	A-B+	C-I	LIBA	2015	ND	MX-152
<b>361</b>	ND	HSJD-312	ND	A-B+	C-I	LIBA	2014	ND	HSJD-312
<b>ND</b>	ND	HSJD-4347602	ND	A-B+	C-I	LIBA	2017	ND	HSJD-4347602
<b>359</b>	ND	MX-149	ND	A-B+	C-I	LIBA	2015	ND	MX-149

## Anexo II. Características determinadas para los ensamblajes utilizados en el análisis bioinformático.

Cepa	Largo del read	Cobertura teórica	Fracción de genoma (%)	Número de contigs	Número de contigs >= 500bp	Contig más largo	Tamaño total	N50	L50	Número de errores de ensamblaje	Número de contigs con errores de ensamblaje
Ox660	<a href="#">51</a>	46x	87.56	110	110	272149	4111188	91263	15	44	28
Ox709	<a href="#">51</a>	77x	88.688	71	71	400254	3980276	130580	10	26	17
Ox1192c	<a href="#">51</a>	70x	87.5	75	75	343852	4214795	135352	10	50	24
Ox1192e	<a href="#">51</a>	85x	87.513	80	80	244800	4265456	120007	13	41	24
Ox1533	<a href="#">51</a>	94x	89.088	75	75	400090	4206164	108273	12	31	21
Ox1463	<a href="#">51</a>	79x	89.143	86	86	398452	4250972	106748	12	36	22
Ox687	<a href="#">51</a>	62x	88.018	69	69	233222	4077384	124899	12	46	27
Ox1121	<a href="#">51</a>	102x	90.267	90	90	264645	4097831	134818	11	44	21
Ox2017	<a href="#">51</a>	74x	87.581	98	98	411717	4130749	85907	15	35	23
Ox990	<a href="#">108</a>	141x	89.091	85	85	392419	4413997	146303	9	47	20
Ox707	<a href="#">108</a>	122x	87.979	59	59	413038	4074903	235826	7	37	17
Ox1844	<a href="#">108</a>	144x	88.271	50	50	299368	4271475	194094	9	50	21
Ox1003	<a href="#">108</a>	17x	88.722	187	187	167596	4030303	41365	31	39	32
Ox1424	<a href="#">108</a>	137x	87.806	44	44	527871	4100943	232049	7	47	19
Ox1893	<a href="#">99</a>	69x	88.815	176	176	137894	4109284	45367	29	37	26
Ox1918	<a href="#">155</a>	268x	90.5	61	61	457414	3967686	137723	9	23	13
Ox2124	<a href="#">100</a>	36x	88.282	126	126	218554	4223891	72860	18	37	23
Ox858	<a href="#">100</a>	98x	88.407	77	77	428366	4333678	114537	10	41	19
Ox2553	<a href="#">100</a>	98x	88.261	50	50	308142	4099856	182047	9	45	18
Ox1342	<a href="#">100</a>	82x	87.423	49	49	703660	4165837	332155	4	52	14
Ox2294	<a href="#">100</a>	74x	88.283	70	70	333541	4292906	158392	10	53	23
Ox428	<a href="#">100</a>	115x	87.715	32	32	559352	4061767	240214	6	49	17
Ox593	<a href="#">100</a>	99x	89.429	73	73	506327	4195179	199012	8	30	16
Ox2167	<a href="#">100</a>	131x	85.628	792	792	46880	3937559	6821	165	34	29
Ox821a	<a href="#">51</a>	64x	88.646	139	139	276540	4104760	80356	16	39	24

<b>Ox746b</b>	<a href="#">51</a>	93x	88.118	138	138	173521	4250859	63788	20	43	27
<b>Ox2155</b>	<a href="#">100</a>	33x	88.25	100	100	260756	4298696	93397	16	42	24
<b>Ox464a</b>	<a href="#">100</a>	83x	89.406	66	66	476144	4420106	184826	8	48	18
<b>Ox1380</b>	<a href="#">100</a>	82x	88.61	66	66	449016	4036327	157711	9	46	20
<b>L039</b>	<a href="#">100</a>	91x	88.408	53	53	561062	4233476	193689	7	70	20
<b>L067</b>	<a href="#">100</a>	96x	89.502	56	56	636671	4248840	217825	6	64	20
<b>OxB_KBP_102</b>	<a href="#">100</a>	97x	89.216	94	94	302465	4072891	133573	11	40	23
<b>L035</b>	<a href="#">100</a>	102x	84.06	635	635	61756	3989526	9331	117	26	24
<b>Ox42</b>	<a href="#">108</a>	75x	94.68	85	85	361099	4234340	161034	9	12	9
<b>Ox1924</b>	<a href="#">100</a>	80x	88.937	37	37	553695	4138856	264086	6	41	15
<b>Ox3036</b>	<a href="#">100</a>	178x	88.883	43	43	479486	4084091	252510	6	37	15
<b>OVG261A</b>	<a href="#">100</a>	101x	87.068	48	48	760660	4242620	210236	6	56	20
<b>OPT_2249</b>	<a href="#">100</a>	110x	86.82	53	53	424454	4058439	159377	9	55	22
<b>L_L12790_1566</b>	<a href="#">100</a>	115x	87.389	64	64	257085	4109217	165767	11	50	26
<b>Ox160</b>	<a href="#">51</a>	130x	97.087	92	92	248351	4207996	108491	13	7	5
<b>Ox1307</b>	<a href="#">51</a>	84x	92.191	157	157	199569	4175753	63895	21	25	20
<b>Ox1802b</b>	<a href="#">99</a>	197x	91.369	163	163	164126	4133806	56980	22	30	23
<b>L075</b>	<a href="#">99</a>	193x	93.043	81	81	345532	4060645	128750	10	27	16
<b>Ox2163</b>	<a href="#">99</a>	186x	91.616	108	108	210748	4051327	94526	15	29	20
<b>Ox2621b</b>	<a href="#">100</a>	108x	93.617	67	67	332953	4218575	171369	9	21	9
<b>8864</b>	<a href="#">151</a>	90x	94.225	46	46	730431	4021097	189689	7	20	13
<b>Ox561</b>	<a href="#">99</a>	84x	97.433	232	232	147403	4169335	32845	40	0	0
<b>L058</b>	<a href="#">99</a>	171x	97.506	81	81	281771	4114233	114294	13	0	0
<b>Ox2183</b>	<a href="#">100</a>	82x	99.319	73	73	341447	3975005	141566	9	0	0
<b>Ox1232</b>	<a href="#">100</a>	50x	97.605	94	94	269519	4156099	102669	14	0	0
<b>Ox1485</b>	<a href="#">100</a>	102x	97.521	58	58	336078	4174555	140875	10	0	0
<b>Q24</b>	<a href="#">100</a>	104x	97.385	76	76	345312	4088995	99523	12	0	0
<b>Ox920</b>	<a href="#">51</a>	96x	95.682	65	65	299261	4122002	168237	9	1	1
<b>Ox904</b>	<a href="#">100</a>	74x	88.12	92	92	220027	4218879	95132	17	46	29

<b>OxB_HB2 1</b>	<a href="#">100</a>	97x	84.954	777	777	74328	4107222	7325	158	26	24
<b>L060</b>	<a href="#">100</a>	73x	83.794	871	871	75452	3867303	6141	183	33	32
<b>OxI_WBC 201I</b>	<a href="#">100</a>	101x	91.063	47	47	491498	4143635	204792	7	29	14
<b>BF095I</b>	<a href="#">100</a>	104x	91.408	41	41	499227	4155714	214487	7	28	17
<b>ES248</b>	<a href="#">100</a>	110x	90.856	43	43	485413	4063748	160629	9	27	18
<b>Ox3196</b>	<a href="#">100</a>	88x	96.676	33	33	600926	4229769	256979	6	9	6
<b>Ox575</b>	<a href="#">99</a>	78x	97.146	135	135	164476	3938943	50766	24	2	2
<b>Leeds162 6</b>	<a href="#">99</a>	187x	94.146	123	123	215730	3962127	63957	19	4	4
<b>7721734</b>	<a href="#">99</a>	182x	94.203	94	94	181414	3983504	81237	17	11	8
<b>L045</b>	<a href="#">99</a>	68x	93.242	367	367	107551	4318185	28640	45	11	10
<b>HCD52</b>	<a href="#">100</a>	108x	90.143	56	56	379095	3933621	167315	8	43	18
<b>Q6</b>	<a href="#">100</a>	109x	90.636	53	53	374077	4070913	185315	8	41	21
<b>WA12</b>	<a href="#">100</a>	114x	89.126	79	79	375656	4014078	114092	11	52	24
<b>CDH5078</b>	<a href="#">100</a>	58x	ND	37	37	508295	4223384	276272	6	ND	ND
<b>RPH118</b>	<a href="#">100</a>	115x	ND	53	53	312953	4029548	162315	9	ND	ND
<b>RPH97</b>	<a href="#">100</a>	82x	ND	84	84	418389	4396193	146899	9	ND	ND
<b>RPH128</b>	<a href="#">100</a>	107x	ND	124	124	476107	4196179	244077	6	ND	ND
<b>RPH39</b>	<a href="#">100</a>	78x	ND	57	57	312959	4029418	150650	10	ND	ND
<b>HBC- 277194</b>	ND	ND	ND	45	45	690001	4646655	228792	6	ND	ND
<b>MX-152</b>	ND	ND	ND	47	47	685047	4405448	219616	7	ND	ND
<b>HSJD-312</b>	ND	ND	ND	3	3	4348004	4613922	4348004	2	ND	ND
<b>HSJD- 4347602</b>	ND	ND	ND	67	67	624331	4545301	22165	7	ND	ND
<b>MX-149</b>	ND	ND	ND	51	51	548772	4308731	229199	7	ND	ND

**Anexo III. Principales características predichas y mejores hits de BLASTn de las secuencias firma del Clado C-I identificadas y determinadas por Neptune.**

NEPTUNE					BLASTn				
Identificador de la secuencia	Puntaje	Frecuencia en el grupo de inclusión	Frecuencia en el grupo de exclusión	Longitud	Mejor hit	Puntaje	Esperado	Identities	Gaps
<b>0.46</b>	0.9523	0.9694	0.0171	2208	<i>Clostridium estertheticum</i> subsp. estertheticum strain DSM 8809, complete genome	892 bits (988)	0	1243/1728 (72%)	37/1728 (2%)
<b>6.1993</b>	0.9516	0.9537	0.0021	162	<i>Clostridium butyricum</i> strain CFSA3989 chromosome, complete genome	156 bits (172)	3.00E-34	131/161 (81%)	0/161 (0%)
<b>7.3614</b>	0.95	0.9698	0.0198	2278	<i>Romboutsia ilealis</i> strain CRIB genome assembly, chromosome: chr1	820 bits (908)	0	1528/2214 (69%)	65/2214 (2%)
<b>9.2737</b>	0.9464	0.9775	0.0311	4219	<i>Fusobacterium ulcerans</i> ATCC 49185 isolate Fusobacterium ulcerans 81A6 genome assembly, chromosome: 1	1114 bits (1235)	0	1390/1884 (74%)	47/1884 (2%)
<b>2.186</b>	0.9452	0.9704	0.0252	2389	<i>Clostridioides difficile</i> strain FDAARGOS_723 chromosome, complete genome	96.0 bits (105)	2.00E-14	78/95 (82%)	0/95 (0%)

**Anexo IV. Conglomerados de CDSs identificados por Scoary y TreeWAS como exclusivos de las cepas del Clado C-I (sensibilidad y especificidad= 100%, p con corrección de Benjamini Hochberg =1.21962e-11).**

Gen	Nombre de gen no único	Anotación
blaI_2		Regulador transcripcional, represor de penicilinas
crcB		Transportador de ion fluoruro predicho CrcB
csfT_1		factor sigma de función extracitoplasmática csfT
csfU_1		factor sigma de función extracitoplasmática csfU
dhaK		Dihidroxiacetona quinasa dependiente de PTS, subunidad de unión a dihidroxiacetona DhaK
grupo_1175		Proteína canal de ion cloruro
grupo_14521		Regulador de respuesta de dos componentes
grupo_14523	<i>fumA</i>	Fumarato hidratasa clase I, subunidad A
grupo_14526		Antiportador de droga/sodio predicho, familia MATE
grupo_14568		Regulador transcripcional, familia GntR
grupo_14602		Lípido quinasa predicha
grupo_14632		Proteína hipotética
grupo_14677		Aciltransferasa predicha
grupo_14682	<i>terD3</i>	Proteína de resistencia al telurio terD3
grupo_14686		Sistema de transporte tipo-ABC, Proteína de unión a soluto extracelular de familia de aminoácido
grupo_14693	<i>cwp13</i>	Proteína de unión a pared celular predicha cwp13
grupo_14729		Lipoproteína predicha
grupo_14730	<i>feoA1</i>	Proteína transportadora de hierro ferroso
grupo_14782		Proteína hipotética
grupo_14794	<i>rnhB</i>	Ribonucleasa HII (ARNasa HII)
grupo_14795	<i>rimM</i>	Proteína procesadora de ARN 16S RimM
grupo_14797		Proteína de superfamilia de radical SAM predicha
grupo_14803	<i>spollP</i>	Proteína de esporulación de etapa II P
grupo_14821		Peptidasa dependiente de metal predicha
grupo_14834		Proteína predicha
grupo_14855		Proteína hipotética
grupo_14858		Proteína predicha
grupo_14859		Proteína dominio cupina
grupo_14876		Amidohidrolasa predicha, familia M20D
grupo_14922		Proteína hipotética
grupo_14923		Fosfozúcar isomerasa predicha
grupo_14950		Proteína de membrana predicha
grupo_14951	<i>cwp8</i>	Proteína de unión a pared celular cwp8
grupo_14954		Proteína de biosíntesis de polisacárido predicha
grupo_14959		Regulador transcripcional, familia LysR
grupo_14986	<i>flgG1</i>	Proteína de complejo de cuerpo gancho-basal flagelar FlgG1

<b>grupo_14987</b>	<i>fliN1</i>	Fosfatasa de switch motor flagelar FliN1
<b>grupo_14990</b>	<i>effD</i>	Antiportador de droga/sodio predicho, familia MATE
<b>grupo_14991</b>		Regulador transcripcional, similar a tipo HTH, CTn2-Orf22
<b>grupo_14996</b>	<i>radC</i>	Proteína de reparación de ADN RadC
<b>grupo_14997</b>	<i>gloA</i>	Lactoglutation liasa
<b>grupo_14999</b>		Regulador transcripcional, tipo HTH
<b>grupo_15004</b>		Proteína hipotética
<b>grupo_15013</b>		Regulador transcripcional, familia MarR
<b>grupo_15016</b>		Metiltransferasa predicha
<b>grupo_15018</b>		Regulador de respuesta de dos componentes
<b>grupo_15056</b>	<i>atpG</i>	Cadena gamma de ATP sintasa (Subunidad gama de sector F1 de ATP sintasa)
<b>grupo_15062</b>	<i>ksgA</i>	Dimetiladenosina transferasa (S-adenosilmetionina-6-N', N'-adenosil(ARNr) dimetiltransferasa)
<b>grupo_15069</b>		Proteína de resistencia al telurio predicha
<b>grupo_15073</b>	<i>pcp</i>	Pirrolidona-carboxilato peptidasa
<b>grupo_15079</b>	<i>trxA</i>	Tiorredoxina
<b>grupo_15090</b>		Ferredoxina
<b>grupo_15100</b>		NADH-azoreductasa dependiente de FMN predicha
<b>grupo_15116</b>	<i>tuf1_1</i>	Factor de elongación EFTu/EF1A
<b>grupo_15117</b>	<i>tuf1_2</i>	Factor de elongación EFTu/EF1A
<b>grupo_15124</b>		Proteína predicha
<b>grupo_15172</b>		Proteína hipotética
<b>grupo_15176</b>	<i>luxS</i>	S-ribosilhomocisteína liasa (Proteína de producción de autoinductor-2 luxS) (proteína de síntesis AI-2)
<b>grupo_1603</b>		Regulador transcripcional, familia hxlR
<b>grupo_1605</b>		Proteína de fago predicha
<b>grupo_2089</b>		Proteína hipotética
<b>grupo_2092</b>		Acetiltransferasa predicha
<b>grupo_2093</b>		Quinasa histidina censora de dos componentes
<b>grupo_2096</b>		Proteína de biosíntesis flagelar predicha
<b>grupo_2097</b>	<i>cwp16_1</i>	N-acetilmuramoil-L-alanine aminidasa predicha, autolisina cwp16
<b>grupo_2100</b>		peptidasa predicha, familia S8 o S52
<b>grupo_2103</b>		Proteína hipotética
<b>grupo_2324</b>		Proteína hipotética
<b>grupo_2749</b>	<i>bltD</i>	Diamina N-acetiltransferasa
<b>grupo_2764</b>		Quinasa histidina censora de dos componentes
<b>grupo_2781</b>		Regulador transcripcional, familia MerR
<b>grupo_2782</b>		Proteína hipotética
<b>grupo_2788</b>		N-carbamoil-L-aminoácido hidrolasa
<b>grupo_2791</b>		Proteína hipotética
<b>grupo_2794</b>		Hidrolasa predicha, superfamilia metalo-beta-lactamasa
<b>grupo_2799</b>		Proteína hipotética
<b>grupo_2800</b>		Transferasa dependiente de piridoxal fosfato predicha
<b>grupo_2804</b>		Proteína de señalización diguanilato ciclasa predicha
<b>grupo_2805</b>		Sistema de transporte tipo-ABC, permeasa similar a molibdeno

<b>grupo_2814</b>		Proteína de biosíntesis de fenazina, familia phzf
<b>grupo_2815</b>		Proteína de adhesión de unión a calcio predicha
<b>grupo_2816</b>		Proteína de señalización predicha
<b>grupo_2818</b>	<i>coaE</i>	Defosfo-CoA quinasa (Defosfocoenzima A quinasa)
<b>grupo_2819</b>	<i>cwp16_2</i>	N-acetilmuramoil-L-alanine aminidasa predicha, autolisina cwp16
<b>grupo_2820</b>		Regulador transcripcional, familia MerR
<b>grupo_2823</b>		Proteasa/amidasa predicha
<b>grupo_2825</b>	<i>rsmG</i>	Metiltransferasa G de subunidad pequeña de ARN ribosomal
<b>grupo_2828</b>	<i>cheC</i>	Fosfatasa de Proteína de quimiotaxis CheY-P CheC
<b>grupo_2829</b>		Proteína hipotética
<b>grupo_3015</b>		Regulador transcripcional, familia Rrf2
<b>grupo_3664</b>		Proteína de señalización predicha
<b>grupo_3670</b>		Sistema PTS, componente IIC de familia lactosa/celobiosa
<b>grupo_3695</b>		Sistema de transporte tipo-ABC, permeasa de familia de cisteína/aminoácido
<b>grupo_3697</b>		Antiportador de droga/sodio predicho, familia MATE
<b>grupo_3699</b>	<i>modC</i>	Sistema de transporte tipo-ABC, Proteína de unión a ATP molibdeno específica
<b>grupo_3700</b>		Proteína hipotética
<b>grupo_3701</b>		Amidohidrolasa predicha, familia peptidasa M20D
<b>grupo_3702</b>		Canal de protones OmpA/MotB predicho
<b>grupo_3704</b>		Proteína de superfamilia de radical SAM predicha
<b>grupo_3705</b>		Proteína hipotética
<b>grupo_3706</b>		Proteína con dominio de unión a 4Fe-4S
<b>grupo_3707</b>		peptidasa predicha, familia U32
<b>grupo_3708</b>		Metaloendopeptidasa exportada predicha
<b>grupo_3714</b>		Proteína hipotética
<b>grupo_3715</b>		Metiltransferasa de ARN predicha
<b>grupo_3717</b>	<i>cheB</i>	Proteína glutamato metilesterasa predicha
<b>grupo_3718</b>	<i>cheR</i>	Metiltransferasa de proteínas de quimiotaxis aceptoras de metilo, familia MCP
<b>grupo_3719</b>		Proteína hipotética
<b>grupo_3720</b>		Proteína hipotética
<b>grupo_3722</b>	<i>cwp28</i>	Proteína de unión a pared celular predicha cwp28
<b>grupo_3723</b>		Quinasa histidina censora de dos componentes
<b>grupo_3724</b>		Proteína hipotética
<b>grupo_3730</b>		Metiltransferasa predicha, familia tipo 11
<b>grupo_3731</b>		Sistema de transporte tipo-ABC, Proteína de unión a ATP de familia multidroga
<b>grupo_3732</b>	<i>aroE1</i>	Shikimato dehidrogenasa 1
<b>grupo_3733</b>		Proteína hipotética
<b>grupo_3734</b>	<i>terD4</i>	Proteína de resistencia al telurio terD4
<b>grupo_3735</b>		Coenzima A disulfuro reductasa (CoA-disulfuro reductasa) (CoADR)
<b>grupo_3736</b>		Proteína hipotética
<b>grupo_3737</b>		Proteína hipotética
<b>grupo_3738</b>	<i>ogt2</i>	ADN-metilado-[Proteína]-cisteína S-metiltransferasa 2
<b>grupo_3740</b>		Proteína hipotética

<b>grupo_3742</b>	<i>grdB_3</i>	Subunidad gama de componente B de complejo glicina reductasa
<b>grupo_3743</b>		Acetiltransferasa predicha
<b>grupo_3746</b>		Quinasa histidina censora de dos componentes
<b>grupo_3747</b>		Proteína de transporte de triptófano predicha
<b>grupo_3748</b>		Polisacárido deacetilasa exportada predicha
<b>grupo_3749</b>	<i>cotE</i>	Proteína de cubierta de espora: peroxirredoxina/quitinasa
<b>grupo_3750</b>		Helicasa predicha, familia DinG
<b>grupo_3751</b>		Proteína hipotética
<b>grupo_3759</b>		Regulador transcripcional, familia MerR
<b>grupo_3760</b>		Muroamioiltetrapeptido carboxipeptidasa predicha, familia S66
<b>grupo_3761</b>		Proteína similar a XkdN de fago predicha
<b>grupo_3762</b>		Proteína hipotética
<b>grupo_3763</b>		Galactosa-1-fosfato uridiltransferasa predicha
<b>grupo_3764</b>		Proteína hipotética
<b>grupo_3765</b>		Magnesio quelatasa predicha
<b>grupo_3766</b>	<i>trmD</i>	ARNr (guanina-N(1)-)metiltransferasa (M1G-metiltransferasa) (ARNt [GM37] metiltransferasa)
<b>grupo_3769</b>		Aminotransferasa predicha
<b>grupo_3770</b>		Amino peptidasa Xaa-Pro predicha, familia M24
<b>grupo_3771</b>		Sistema PTS, componente IIA similar a fructosa
<b>grupo_3772</b>		Coenzima A disulfuro reductasa
<b>grupo_3773</b>		Proteína hipotética
<b>grupo_3774</b>		Regulador de respuesta de dos componentes
<b>grupo_3775</b>		Proteína hipotética
<b>grupo_3776</b>		Antiportador Na(+)/H(+) predicho
<b>grupo_3777</b>		Proteína hipotética
<b>grupo_3778</b>	<i>xdhB</i>	Subunidad XdhB de xantina deshidrogenasa, unión a FAD
<b>grupo_3779</b>		Oxidorreductasa predicha, Subunidad Fe-S
<b>grupo_3780</b>	<i>sat</i>	Estreptogramina A acetiltransferasa (virginamicina acetiltransferasa D) (Vat(D))
<b>grupo_3781</b>		Proteína hipotética
<b>grupo_3782</b>		Proteína de señalización diguanilato quinasa predicha
<b>grupo_3783</b>		Proteína hipotética
<b>grupo_3784</b>		Transportador, Superfamilia de facilitador mayor (MFS)
<b>grupo_3785</b>		Regulador transcripcional, familia MerR
<b>grupo_3786</b>		Proteína predicha
<b>grupo_3787</b>		Relacionada a oligoendopeptidasa, clado 2, familia M3B
<b>grupo_3788</b>	<i>mtlR</i>	Antiterminador de transcripción, regulador de operón PTS
<b>grupo_3789</b>	<i>mtlD</i>	Manitol-1-fosfato 5-dehidrogenasa
<b>grupo_3790</b>		Sistema de transporte tipo-ABC, Proteína de unión a soluto extracelular similar a molibdeno
<b>grupo_3793</b>		Oxidorreductasa predicha, dependiente de FAD
<b>grupo_3795</b>		Proteína hipotética
<b>grupo_3796</b>		Proteína de ensamblaje de citocromo C predicha
<b>grupo_3797</b>		Transportador predicho
<b>grupo_3798</b>		Aminotransferasa predicha

<b>grupo_3799</b>		Regulador de respuesta de dos componentes
<b>grupo_3801</b>	<i>xynA</i>	Sistema PTS, componente IIA xilosido-específico
<b>grupo_3802</b>	<i>tagABC</i>	Sistema PTS, componente IIABC tagatosa específico
<b>grupo_3804</b>		Proteína predicha
<b>grupo_3807</b>		Regulador transcripcional, familia DeoR
<b>grupo_3808</b>		Regulador transcripcional, familia MarR
<b>grupo_3810</b>		Proteína de eflujo de catión predicha
<b>grupo_3812</b>		Glutamato carboxipeptidasa predicha, familia M20
<b>grupo_3813</b>		Racemasa de aminoácido predicha
<b>grupo_3814</b>	<i>garR</i>	Tartronato semialdehído reductasa
<b>grupo_3815</b>		Proteína de membrana predicha
<b>grupo_3818</b>	<i>cwp5</i>	Proteína de superficie celular predicha <i>cwp5</i>
<b>grupo_3820</b>	<i>cwp14</i>	Proteína de unión a pared celular predicha <i>cwp14</i>
<b>grupo_3821</b>		Quinasa histidina censora de dos componentes
<b>grupo_3822</b>		Antiterminador de transcripción, regulador de operón PTS
<b>grupo_3823</b>		peptidasa predicha, familia M22
<b>grupo_3824</b>		Maltosa O-acetiltransferasa predicha
<b>grupo_3826</b>	<i>flgM</i>	Regulador negativo de la síntesis de flagelina (factor anti-sigma-d)
<b>grupo_3827</b>	<i>fliG</i>	Proteína de switch motor flagelar FliG
<b>grupo_3829</b>	<i>fliM</i>	Proteína de switch motor flagelar FliM
<b>grupo_3830</b>		Regulador transcripcional, familia hxlR
<b>grupo_3831</b>		Proteína predicha
<b>grupo_3832</b>	<i>pdp</i>	Pirimidina-nucleósido fosforilasa
<b>grupo_3833</b>		Lipoproteína de unión a soluto predicha
<b>grupo_3835</b>		Proteína predicha
<b>grupo_3836</b>		Polisacárido deacetilasa
<b>grupo_3837</b>	<i>folC</i>	Folilpoliglutamato sintasa
<b>grupo_3840</b>		Regulador transcripcional, familia MarR
<b>grupo_3841</b>		Sistema de transporte tipo-ABC, permeasa
<b>grupo_3844</b>		Antiportador de droga/sodio predicho, familia MATE
<b>grupo_3845</b>	<i>cobA + hemD</i>	Uroporfirinógeno-III metiltransferasa bifuncional/uroporfirinógeno-III sintasa
<b>grupo_3846</b>	<i>cbiJ</i>	Cobalto-precocina-6A reductasa
<b>grupo_3847</b>	<i>cbiT</i>	Cobalto-precocina-6Y C(15) metiltransferasa
<b>grupo_3848</b>	<i>tmk</i>	Timidilato quinasa dTMT quinasa
<b>grupo_3849</b>		Transportador, Superfamilia de facilitador mayor (MFS)
<b>grupo_3850</b>	<i>lplA</i>	Lipoato-proteína ligasa predicha
<b>grupo_3851</b>		Regulador de respuesta de dos componentes
<b>grupo_3852</b>		Regulador transcripcional, familia TetR
<b>grupo_3853</b>	<i>mobA</i>	Proteína A de biosíntesis de dinucleótido molibdopterina-guanina
<b>grupo_3854</b>		Regulador transcripcional predicho
<b>grupo_3855</b>		Sistema de transporte tipo-ABC, permeasa de familia multidroga
<b>grupo_3856</b>	<i>hisB</i>	Imidazolglicerol-fosfato deshidratasa
<b>grupo_3857</b>	<i>hisC_2</i>	Histidinol-fosfato aminotransferasa (imidazol acetol-fosfato transaminasa)

<b>grupo_3858</b>	<i>ggt</i>	Gama glutamil transpeptidasa
<b>grupo_3859</b>		Proteína de hierro-sulfuro predicha
<b>grupo_3860</b>	<i>truA1</i>	ARNt pseudouridina sintasa A 1
<b>grupo_3861</b>		Proteína de membrana predicha
<b>grupo_3866</b>		FMN reductasa dependiente de NADPH predicha
<b>grupo_3867</b>		Proteína hipotética
<b>grupo_3873</b>		Sistema de transporte tipo-ABC, permeasa de familia multidroga
<b>grupo_3874</b>		Sistema de transporte tipo-ABC, Proteína de unión a ATP de familia multidroga
<b>grupo_3875</b>		Proteína hipotética
<b>grupo_3876</b>	<i>birA</i>	Proteína bifuncional BirA (incluye: represor de operón de biotina y biotina--[acetil-CoA-carboxilasa] sintetasa)
<b>grupo_4049</b>		Oxidorreductasa predicha
<b>grupo_4284</b>	<i>fliJ</i>	Proteína flagelar FliJ
<b>grupo_4362</b>		Proteína hipotética
<b>grupo_4686</b>		Proteína contenedora de repeticiones MORN predicha
<b>grupo_4736</b>		Antiterminador de transcripción, regulador de operón PTS
<b>grupo_477</b>		Proteína de señalización predicha
<b>grupo_492</b>		Proteína de unión a nucleótido cíclico predicha
<b>grupo_4986</b>		Regulador transcripcional, dependiente de sigma-54
<b>grupo_5018</b>		Proteína hipotética
<b>grupo_5036</b>		Glutamato carboxipeptidasa predicha
<b>grupo_5059</b>		Regulador transcripcional, familia hxlR
<b>grupo_5139</b>		Proteína hipotética
<b>grupo_5156</b>		Proteína de membrana predicha
<b>grupo_5157</b>		Sistema de transporte tipo-ABC, Proteína/permeasa de unión a ATP de familia multidroga
<b>grupo_5159</b>	<i>speA</i>	Arginina decarboxilasa
<b>grupo_5162</b>		Sistema de transporte tipo-ABC, Proteína de unión a soluto extracelular de familia de carbohidrato
<b>grupo_5163</b>	<i>maa_1</i>	Maltosa O-acetiltransferasa
<b>grupo_5165</b>		Regulador transcripcional, familia MarR
<b>grupo_5168</b>	<i>rbo</i>	Rubredoxina oxidorreductasa (desulfoferredoxina)
<b>grupo_5170</b>		Flavodoxina/óxido nítrico sintasa predicha
<b>grupo_5172</b>	<i>iscS1</i>	Cisteína desulfurasa
<b>grupo_5174</b>		Proteína hipotética
<b>grupo_5181</b>		peptidasa predicha, familia S9
<b>grupo_5182</b>		peptidasa predicha, familia S9
<b>grupo_5185</b>		Ferredoxina predicha
<b>grupo_5187</b>		Proteína similar a DnaB helicasa replicativa
<b>grupo_5188</b>		Regulador transcripcional, familia LytR
<b>grupo_5189</b>		Sistema de transporte tipo-ABC, Proteína de unión a ATP
<b>grupo_5194</b>		Proteasa dependiente de ATP predicha
<b>grupo_5195</b>		Proteína de unión a conglomerado de hierro-sulfuro predicha
<b>grupo_5197</b>		Proteína hipotética
<b>grupo_5198</b>		D-hidatoinasa (dihidropirimidinasa)

<b>grupo_5205</b>	<i>argC</i>	N-acetil-gamma-glutamyl-fosfato reductasa
<b>grupo_5206</b>		Lipoproteína predicha
<b>grupo_5207</b>		Proteína de membrana predicha
<b>grupo_5208</b>	<i>miaA</i>	ARNt delta(2)-isopentenilpirofosfato transferasa
<b>grupo_5209</b>		Permeasa predicha
<b>grupo_5210</b>		Sistema de transporte tipo-ABC, Proteína de unión a ATP
<b>grupo_5211</b>		Regulador de respuesta de dos componentes
<b>grupo_5213</b>		Proteína predicha
<b>grupo_5216</b>	<i>eutE</i>	Etanolamina acetaldehído oxidorreductasa
<b>grupo_5217</b>	<i>eutG</i>	Etanolamina hierro dependiente alcohol deshidrogenasa
<b>grupo_5218</b>		Proteína de membrana predicha
<b>grupo_5222</b>	<i>tyrC</i>	Prefenato deshidrogenasa
<b>grupo_5223</b>	<i>aroB</i>	3-dehidroquinato sintasa
<b>grupo_5224</b>	<i>kdpD</i>	Quinasa histidina censora de dos componentes KdpD
<b>grupo_5226</b>	<i>metA</i>	Homoserina O-succiniltransferasa
<b>grupo_5228</b>	<i>ade</i>	Adenina deaminasa
<b>grupo_5229</b>	<i>ispH</i>	4-hidroxi-3-metilbut-2-enil difosfo reductasa
<b>grupo_5230</b>	<i>cmk</i>	Citidilato quinasa (CK) (Citidina monofosfato quinasa) (CMP quinasa)
<b>grupo_5231</b>		Subunidad de nitrato y sulfato reductasa predicha
<b>grupo_5232</b>	<i>truA2</i>	Pseudouridilato sintasa (ARNt pseudouridina sintasa A)
<b>grupo_5233</b>		Proteína de dominio metiltransferasa
<b>grupo_5234</b>		Regulador transcripcional, tipo HTH
<b>grupo_5235</b>		Proteína hipotética
<b>grupo_5237</b>		Regulador de respuesta de dos componentes
<b>grupo_5238</b>		Simportador de ion divalente predicho
<b>grupo_5244</b>	<i>cotB</i>	Proteína de cubierta de espora
<b>grupo_5246</b>		Antiportador de droga/sodio predicho, familia MATE
<b>grupo_5247</b>		3-metiladenina ADN glicosilasa predicha
<b>grupo_5248</b>		Quinasa histidina censora de dos componentes, asociado a la sporulacion spo0A
<b>grupo_5249</b>	<i>ssuB</i>	Sistema de transporte tipo-ABC, Proteína de unión a ATP de familia de sulfonatos
<b>grupo_5250</b>		Sistema de transporte tipo-ABC, permeasa de unión a ATP de familia multidroga
<b>grupo_5251</b>		Sistema de transporte tipo-ABC, Proteína/permeasa de unión a ATP de familia multidroga
<b>grupo_5253</b>		Proteína hipotética
<b>grupo_5254</b>	<i>aroF</i>	Fosfo-2-dehidro-3-deoxiheptonato aldolasa
<b>grupo_5255</b>	<i>folK</i>	2-amino-4-hidroxi-6- hidroximetildihidropteridinapirofosfoquinasa
<b>grupo_5256</b>	<i>pabA</i>	Componente II de para-aminobenzoato/antranilato sintasa glutamina amidotransferasa [incluye: Componente II de para-aminobenzoato/antranilato sintasa glutamina amidotransferasa; componente II de antranilato sintasa]
<b>grupo_5257</b>		Proteína hipotética
<b>grupo_5266</b>		Proteína hipotética
<b>grupo_5267</b>		Hidrolasa predicha, familia CocE/NonD
<b>grupo_5268</b>		Regulador transcripcional predicho
<b>grupo_5269</b>		Proteína hipotética
<b>grupo_5270</b>		Regulador transcripcional, familia TetR

grupo_5271		Regulador transcripcional, tipo fago
grupo_5275		ARNt 2-selenouridina sintasa predicha, selenofosfato dependiente; Proteína con dominio similar a rodanasa
grupo_5276		Proteína de unión a ARN predicha
grupo_5277		Regulador transcripcional, familia AraC
grupo_5278		Acetiltransferasa predicha
grupo_5279		Quinasa histidina censora de dos componentes
grupo_5281		Recombinasa sitio específica predicha
grupo_5282		Quinasa histidina censora de dos componentes, asociado a la esporulación spo0A
grupo_5283	<i>dacF1</i>	D-alanil-D-alanina carboxipeptidasa (Proteína de unión a penicilina específica de esporulación predicha)
grupo_5284		Hidrolasa predicha, superfamilia HAD, subfamilia IIB
grupo_5285	<i>LinCd</i>	Transportador, Superfamilia de facilitador mayor (MFS)
grupo_5287	<i>glgB</i>	enzima ramificadora de 1,4-alfa-glucano
grupo_5289		Lipoproteína predicha
grupo_5290	<i>rmz</i>	Ribonucleasa Z (ARNasa Z) (ARNtasa Z) (ARNt 3 endonucleasa)
grupo_5291		Sistema de transporte tipo-ABC, permeasa de familia de carbohidrato
grupo_5293	<i>rsmB</i>	Metiltransferasa B de subunidad pequeña de ARN ribosomal (ARNr (citosina-C(5)-)-metiltransferasa)
grupo_5294		Enzima dependiente de piridoxal 5'-fosfato predicha
grupo_5295		Oxidoreductasa cianobacterial-específica contenedora de FeS predicha
grupo_5296		Proteína hipotética
grupo_5297		Proteína hipotética
grupo_5299		Proteína de señalización predicha
grupo_5300		Regulador transcripcional, familia AraC
grupo_5301		Proteína relacionada a flavodocina predicha
grupo_5303		Fosfatasa asociada a membrana predicha
grupo_5304	<i>cwp22</i>	Proteína de pared celular predicha cwp22
grupo_5305		Proteína hipotética
grupo_5306		peptidasa predicha C60B, sortasa B
grupo_5307		Helicasa de ARN dependiente de ATP predicha
grupo_5308		Proteína de reparación de mal emparejamientos de ADN
grupo_5309	<i>ldh</i>	L-lactato deshidrogenasa (L-LDH)
grupo_5310		Proteína de unión a hierro-sulfuro predicha
grupo_5311		peptidasa predicha, familia M20D
grupo_5315		Deshidrogenasa anaeróbica predicha
grupo_5316		Fragmento de N-acetilmuramoil-L-alanina amidasa (hidrolasa de pared celular)
grupo_5317		Hidrolasa de pared celular relacionada a fago predicha
grupo_5319		Proteína predicha
grupo_5320		Regulador transcripcional, familia LysR
grupo_5321		Acetiltransferasa de familia gnat
grupo_5322		Regulador transcripcional, familia Crp
grupo_5323		Glucocinasa predicha, familia ROK
grupo_5324		Proteína de membrana predicha
grupo_5326	<i>prsA</i>	Peptidil-prolil cis-trans isomerasa, tipo PpiC

grupo_5331		ARNt-nucleotidiltransferasa/miembro de familia Poli(A) polimerasa predicha
grupo_5332		Metiltransferasa predicha
grupo_5334		Antiterminador de transcripción, regulador de operón PTS
grupo_5336		Hidrolasa de pared celular predicha; Proteína asociada a fosfatasa
grupo_5337		Exonucleasa predicha
grupo_5338		Hidrolasa predicha, familia NUDIX
grupo_5339		Proteína hipotética
grupo_5343		Proteína de señalización predicha
grupo_5344		Proteína exportada predicha
grupo_5345		Transportador predicho
grupo_5347		Proteína tirosina fosfatasa predicha
grupo_5348	<i>fruABC</i>	Sistema PTS, componente IIABC fructosa-específico
grupo_5353		Proteína hipotética
grupo_5355		Tiolasa predicha
grupo_5360	<i>uppP2</i>	Undecaprenil-difosfatasa 2 (undecaprenil pirofosfato fosfatasa 2) (Proteína de resistencia a bacitracina 2)
grupo_5361		Quinasa histidina censora de dos componentes
grupo_5362	<i>ssuA2</i>	Sistema de transporte tipo-ABC, proteína de unión a soluto extracelular de familia de sulfonato
grupo_5364		Proteína hipotética
grupo_5368		Sistema PTS, componente IIA manosa específico
grupo_5369		Proteína de captación de glucosa predicha
grupo_5370		Oxidoreductasa
grupo_5371		Proteína hipotética
grupo_5374		ATPasa predicha
grupo_5377	<i>xylB</i>	Xilulosa quinasa
grupo_5384	<i>bgIF4</i>	Sistema PTS, componente específico de beta-glucósido IIABC, bgIF4
grupo_5388		Sistema de transporte tipo-ABC, Proteína de unión a ATP de familia multidroga
grupo_5389		Regulador transcripcional, dependiente de sigma-54
grupo_5390	<i>dlitD</i>	D-alanina transferasa predicha DItD
grupo_5391		Deshidrogenasa de unión a zinc predicha
grupo_5393		Regulador de respuesta de dos componentes
grupo_5395		Proteína predicha
grupo_5397		Proteína hipotética
grupo_5398		Proteína de membrana predicha
grupo_5400		Metiltransferasa de ADN N6 adenina específica predicha, clase N12
grupo_5403	<i>hemZ</i>	Coproporfirinógeno III oxidasa
grupo_5404		Polisacárido deacetilasa predicha
grupo_5405		Proteína exportada predicha
grupo_5406	<i>spolID</i>	Proteína de esporulación de etapa II D
grupo_5408		Glicina/sarcosina/betaina reductasa b
grupo_5411		Cetopantoato reductasa predicha
grupo_5412		Proteína hipotética
grupo_5414	<i>cysL</i>	Regulador transcripcional de tipo HTH CysL
grupo_5416	<i>fliH</i>	Proteína de ensamblaje flagelar FliH

<b>grupo_5417</b>	<i>motA</i>	Proteína de rotación motora flagelar MotA
<b>grupo_5418</b>	<i>fliZ</i>	Proteína flagelar FliZ
<b>grupo_5419</b>	<i>fliA</i>	Factor sigma 28 de ARN polimerasa para operón flagelar
<b>grupo_5420</b>		Alcohol deshidrogenasa contenedora de hierro predicha
<b>grupo_5423</b>		Sistema de transporte tipo-ABC, permeasa de familia de bacitracina/multidroga
<b>grupo_5424</b>	<i>rbsK</i>	Riboquinasa, familia pfkB
<b>grupo_5425</b>	<i>argE</i>	Acetilornitina deacetilasa ArgE
<b>grupo_5426</b>		Proteína hipotética
<b>grupo_5427</b>		Peptidoglicano glicosiltransferasa predicha
<b>grupo_5428</b>	<i>pupG</i>	Purina nucleósido fosforilasa
<b>grupo_5429</b>		Proteína hipotética
<b>grupo_5430</b>	<i>spolIAB</i>	Proteína de esporulación de etapa III AB
<b>grupo_5431</b>	<i>fabD</i>	Transacilasa de proteína acarreadora de malonil CoA-acilo (MCT)
<b>grupo_5432</b>		Proteína hipotética
<b>grupo_5433</b>		Regulador transcripcional, familia GntR
<b>grupo_5434</b>	<i>rng</i>	Ribonucleasa G (ARNasaG) (Proteína de filamento axial citoplasmático)
<b>grupo_5435</b>		Proteína con dominio SH3 predicha
<b>grupo_5438</b>		Sistema PTS, componente IIA manosa/fructosa/sorbosa
<b>grupo_5439</b>		Proteína predicha
<b>grupo_5440</b>		Pseudouridilato sintasa predicha
<b>grupo_5441</b>	<i>sbcC</i>	Subunidad C de nucleasa SbcCD
<b>grupo_5442</b>	<i>sbcD</i>	Subunidad D de nucleasa SbcCD
<b>grupo_5443</b>	<i>addA</i>	Subunidad A de nucleasa dependiente de ATP (helicasa dependiente de ATP addA)
<b>grupo_5444</b>	<i>addB</i>	Subunidad B de nucleasa dependiente de ATP (helicasa dependiente de ATP addB), superfamilia 1 de familia UvrD
<b>grupo_5445</b>		Proteína hipotética
<b>grupo_5446</b>		Proteína hipotética
<b>grupo_5447</b>		Proteína de extrusión multiantimicrobial predicha, familia MatE
<b>grupo_5449</b>		Sistema de transporte tipo-ABC, Proteína de unión a ATP glicina betaina/carnitina/colina
<b>grupo_5452</b>	<i>proC1</i>	Pirrolina-5-carboxilato reductasa
<b>grupo_5453</b>		Oxidorreductasa predicha, dependiente de FAD
<b>grupo_5454</b>		Proteína hipotética
<b>grupo_5455</b>		Proteína de unión a ATP/GTP predicha
<b>grupo_5456</b>		Deshidrogenasa de 2-hidroxiácido D-isomero específica predicha
<b>grupo_5457</b>	<i>cprR</i>	Regulador de respuesta de dos componentes
<b>grupo_5459</b>		Proteína hipotética
<b>grupo_5461</b>		Sistema PTS, componente IIA manosa/fructosa/sorbosa
<b>grupo_5463</b>		Regulador de respuesta de dos componentes
<b>grupo_5464</b>		Sistema de transporte tipo-ABC, Proteína de unión a ATP
<b>grupo_5467</b>		Proteína hipotética
<b>grupo_5474</b>	<i>cbiK</i>	Sirohidroclorina cobaltoquelatasa
<b>grupo_5477</b>	<i>cobD</i>	Treonina-fosfato decarboxilasa (L-treonina-O-3-fosfato decarboxilasa), familia CobD/CbiB
<b>grupo_5478</b>	<i>cbiB</i>	Proteína de biosíntesis de cobalamina CbiB, familia CobD/CbiB

grupo_5479	<i>cobQ</i>	Ácido cobárico sintasa CobQ
grupo_5481		5-metiltetrahidrofolato ciclo-ligasa
grupo_5482		Proteína hipotética
grupo_5483	<i>gtaB1</i>	UTP--glucosa-1-fosfato-uridililtransferasa (Proteína de estrés general 33)
grupo_5484	<i>prsA1</i>	Proteína similar a PrsA peptidilprolil isomerasa
grupo_5485		Prepilina peptidasa tipo IV predicha, familia A24A
grupo_5486		Pilina tipo IV predicha
grupo_5487		Proteína de membrana predicha
grupo_5488	<i>phnM</i>	Subunidad de complejo de carbón-fósforo liasa predicha
grupo_5490	<i>cotA</i>	Proteína de ensamblaje de cubierta de espora
grupo_5491		Sistema de transporte tipo-ABC, permeasa
grupo_5492		Proteína de señalización diguanilato quinasa predicha
grupo_5494		Proteína hipotética
grupo_5495		Proteína hipotética
grupo_5496		peptidasa predicha, familia M19
grupo_5500		Regulador transcripcional, tipo RmlC
grupo_5501		Proteína hipotética
grupo_5502		Sistema de transporte tipo-ABC, Proteína de unión a ATP de familia multidroga
grupo_5505	<i>thiM</i>	Hidroxietiltiazol quinasa
grupo_5506	<i>kdpC</i>	Cadena C de ATPasa transportadora de potasio
grupo_5507		Proteína hipotética
grupo_5508		Ribonucleasa asociada a membrana predicha similar a BN (Rbn)
grupo_5509	<i>ilvB</i>	Subunidad grande de acetolactato sintasa
grupo_5510		Factor accesorio de xantina deshidrogenasa predicho
grupo_5511		Polisacárido deacetilasa predicha
grupo_5512	<i>hisA</i>	1-(5-fosforibosil)-5-[(5-fosforibosilamino)metilideneamino]imidazol-4-carboxamida isomerasa
grupo_5513	<i>hisZ</i>	Subunidad reguladora ATP fosforibosiltransferasa
grupo_5516		Aciltransferasa predicha
grupo_5519		Antiterminador de transcripción, regulador de operón PTS
grupo_5522		4-hidroxi-2-oxoglutarato aldolasa predicha
grupo_5523		Fosfozúcar isomerasa predicha
grupo_5541		Fosfoesterasa predicha
grupo_5543		Nucleótido pirofosfatasa predicha
grupo_5544		Regulador transcripcional, familia PadR
grupo_5545		Sistema de transporte tipo-ABC, Proteína de unión a ATP
grupo_5546	<i>ispE</i>	4-difosfocitidil-2-C-metil-D-eritrol quinasa (4-(citidina-5'-difosfo)-2-C-metil-D-eritriol quinasa)
grupo_5547		Hidrolasa de corteza-lítica de espora predicha
grupo_5548		Peptidasa predicha
grupo_5550		Proteína quimiocensora similar a CheY predicha
grupo_5551	<i>cheY</i>	Proteína de quimiotaxis CheY
grupo_5553		Regulador transcripcional, familia DeoR
grupo_5555		Metionina sintasa dependiente de S-adenosil-L-metionina predicha
grupo_5556	<i>aspC</i>	Aspartato aminotransferasa

grupo_5734	<i>thil</i>	Proteína de biosíntesis de tiamina Thil
grupo_7784		Sistema de transporte tipo-ABC, Proteína/permeasa de unión a ATP de familia multidroga
grupo_7789		Fosfatasa/hidrolasa predicha
grupo_7792		Regulador transcripcional, familia GntR
grupo_7794		Dominio de unión a ADN, familia de escicionasa
grupo_7796		Proteína de membrana predicha
grupo_7798	<i>cwp25</i>	Proteína de unión a pared celular predicha cwp25
grupo_7799		Proteína hipotética
grupo_7803		Regulador de captación férrica predicho
grupo_7804		Proteína hipotética
grupo_7805		N-acetilmuramoil-L-alanine aminidasa predicha
grupo_7806	<i>srlR</i>	Antiterminador de transcripción, regulador de operón PTS
grupo_7808		Canal de protón MotA/TolQ/ExbB predicho
grupo_7824		Permeasa predicha
grupo_7833		Proteína hipotética
grupo_7851		Proteína hipotética
grupo_7854		Proteína hipotética
grupo_7858		Proteína hipotética
grupo_7862		Proteína de membrana predicha
grupo_7863		Sistema de transporte tipo-ABC, Proteína de unión a ATP de familia multidroga
grupo_7870		Sistema de transporte tipo-ABC, permeasa
grupo_7871		Sistema de transporte tipo-ABC, Proteína de unión a ATP
grupo_7876	<i>eutW</i>	Quinasa histidina censora de dos componentes, etanolamina específica
grupo_7877		Regulador de unión a oligonucleótido predicho
grupo_7882	<i>aroA</i>	3-fosfoshikimato 1-carboxiviniltransferasa
grupo_7885		Proteína hipotética
grupo_7888		Oxidorreductasa predicha
grupo_7891		Lipoproteína predicha
grupo_7892		Arsenato reductasa predicha
grupo_7895	<i>grdB_4</i>	Subunidad gama de componente B de complejo glicina reductasa
grupo_7896		Regulador transcripcional, tipo HTH
grupo_7908	<i>cwp20</i>	Proteína de superficie celular predicha; Proteína de unión a penicilina predicha cwp20
grupo_7910		Sistema de transporte tipo-ABC, Proteína de unión a ATP
grupo_7911		Fosfohidrolasa predicha, superfamilia HD
grupo_7928		Proteína de unión a ATP predicha
grupo_7929		Proteína de membrana predicha
grupo_7931		Proteína hipotética
grupo_7932		Sintetasa predicha
grupo_7940		Proteína de fago predicha
grupo_7941		Proteína de membrana predicha
grupo_7942	<i>cprA</i>	Sistema de transporte tipo-ABC, Proteína de unión a ATP de familia de lantibiotico/multidroga
grupo_7949		Lipoproteína predicha
grupo_7950		Regulador transcripcional, familia TetR

<b>grupo_7954</b>		Hidrolasa predicha, superfamilia HAD, subfamilia IA
<b>grupo_7956</b>	<i>fmt</i>	Metionil-ARNt formiltransferasa
<b>grupo_7957</b>	<i>fbpA</i>	Proteína A de unión a fibronectina
<b>grupo_7958</b>		Regulador transcripcional predicho
<b>grupo_7961</b>		Proteína hipotética
<b>grupo_7969</b>		Proteína hipotética
<b>grupo_7970</b>		Sistema de transporte tipo-ABC, permeasa
<b>grupo_7973</b>		Proteína hipotética
<b>grupo_7974</b>		Transportador de malato predicho
<b>grupo_7975</b>	<i>cdsA</i>	Fosfatidato citidiltransferasa CdsA
<b>grupo_7977</b>		Proteína de resistencia a vancomicina predicha, familia vanW
<b>grupo_7978</b>		Proteína predicha
<b>grupo_7983</b>		Acil transferasa/acil hidrolasa/lisofosfolipasa
<b>grupo_7998</b>	<i>cynT</i>	Anhidrasa carbónica
<b>grupo_8000</b>		Regulador transcripcional, tipo HTH
<b>grupo_8012</b>	<i>grpE</i>	Proteína grpE (cofactor HSP-70)
<b>grupo_8015</b>		Acarreador de C4-dicarboxilato anaeróbico predicho, familia DcuC
<b>grupo_8017</b>		Regulador transcripcional, represor de beta-lactamas
<b>grupo_8019</b>		Proteína de membrana predicha
<b>grupo_8020</b>		Proteína de membrana predicha
<b>grupo_8021</b>		Proteína hipotética
<b>grupo_8028</b>		Regulador de respuesta de dos componentes
<b>grupo_8036</b>		Proteína de membrana integral de esporulación predicha
<b>grupo_8044</b>		Regulador transcripcional, familia GntR
<b>grupo_8045</b>		Antiportador de droga/sodio predicho, familia MATE
<b>grupo_8048</b>	<i>atpI</i>	Subunidad I de ATP sintasa sodio tipo V
<b>grupo_8050</b>		Regulador de respuesta de dos componentes
<b>grupo_8051</b>	<i>nrdE</i>	Subunidad alfa de Ribonucleósido-difosfato reductasa (subunidad grande de ribonucleótido reductasa)
<b>grupo_8054</b>		Sistema PTS, componente IIA similar a glucosa
<b>grupo_8061</b>		Glicosil hidrolasa
<b>grupo_8063</b>		Proteína de membrana predicha
<b>grupo_8064</b>	<i>tagK</i>	Tagatosa-6-fosfato quinasa
<b>grupo_8072</b>		Sistema PTS, componente IIABC similar a fructosa
<b>grupo_8073</b>		Fructosa-1-6-bisfosfato aldolasa predicha
<b>grupo_8075</b>		Transportador de malato predicho
<b>grupo_8076</b>		Proteína hipotética
<b>grupo_8080</b>		Proteína hipotética
<b>grupo_8087</b>	<i>cwp9</i>	Proteína de unión a pared celular cwp9
<b>grupo_8090</b>		Lipoproteína predicha
<b>grupo_8092</b>		Fosfolipasa predicha
<b>grupo_8094</b>		Monogalactosildiacilglicerol sintasa predicha
<b>grupo_8097</b>		4-hidroxifenilacetato decarboxilasa, subunidad activadora HpdA predicha; proteína activadora de radical glicil predicha
<b>grupo_8100</b>	<i>flhG</i>	Regulador de número flagelar FlhG

grupo_8101	<i>flgG</i>	Proteína de bastón de cuerpo basal flagelar FlgG
grupo_8103		Proteína hipotética
grupo_8104		Sistema de transporte tipo-ABC, Proteína de unión a ATP de familia de bacitracina/multidroga
grupo_8106		Proteína de membrana predicha
grupo_8109	<i>spoIIIAF</i>	Proteína de esporulación de etapa III AF
grupo_8111		Proteína hipotética
grupo_8113		Proteína de familia nitroreductasa
grupo_8115		Fosfoesterasa predicha
grupo_8116		Sistema PTS, componente de familia IIB fructosa/manitol
grupo_8119		Amino peptidasa I predicha, familia M18
grupo_8122		Sistema de transporte tipo-ABC, Proteína de unión a ATP de familia multidroga
grupo_8138		Proteína de membrana predicha
grupo_8145		Transportador, Superfamilia de facilitador mayor (MFS)
grupo_8153		Deshidrogenasa azúcar-fosfato predicha
grupo_8171		Regulador transcripcional, familia GntR
grupo_8173	<i>cobU</i>	Proteína de biosíntesis de adenosilcobalamina bifuncional CobU [Adenosilcobinamida quinasa; adenosilcobinamida-fosfato guanililtransferasa]
grupo_8178	<i>tkt2</i>	Transcetolasa, subunidad de unión a piridina
grupo_8182		Proteína hipotética
grupo_8188		Proteína hipotética
grupo_8194		FMN reductasa NADPH.dependiente predicha
grupo_8200	<i>thiD</i>	Fosfometilpirimidina quinasa
grupo_8201		Proteína hipotética
grupo_8205		Regulador transcripcional
grupo_8207	<i>pgm</i>	Alfa-fosfoglucomutasa
grupo_8212	<i>hisH</i>	Imidazol glicerol fosfato sintasa de subunidad HisH
grupo_8214		Proteína hipotética
grupo_8216		Regulador de respuesta de dos componentes
grupo_8224		Proteína hipotética
grupo_8226		Antiterminador de transcripción, regulador de operón PTS
grupo_8232		Proteína hipotética
grupo_8233		Proteína de metabolismo de selenio yedf
grupo_8250		Proteína hipotética
grupo_8252		Receptor de quimiotaxis metilo aceptor predicho, familia MCP
grupo_8253	<i>cheD</i>	Quimiorreceptor glutamina deaminidasa predicha CheD
grupo_8255		Proteína hipotética
grupo_8256		Proteína hipotética
grupo_8266		Oligosacárido deacetilasa predicha
grupo_8267		Transportador de captación de ADN
grupo_8771		Dioxigenasa predicha
grupo_8843		Regulador transcripcional, familia CarD
hcp_2		Hidroxilamina reductasa
hdfR		Regulador transcripcional de tipo HTH HdfR
hydR_2		Regulador transcripcional, familia TetR

<b>iorA_2</b>		Indol piruvato ferredoxina/ flavodoxina oxidoreductasa
<b>iorB_2</b>		Indol piruvato ferredoxina/ flavodoxina oxidoreductasa
<b>IsrF</b>		3-hidroxi-5-fosfonooxipentano-2,4-diona tiolasa
<b>maa_3</b>		Maltosa O-acetiltransferasa
<b>prdF_1</b>		Prolina racemasa
<b>rtcB</b>		Proteína hipotética
<b>SinR_1</b>		Regulador transcripcional, tipo HTH
<b>xyIR_2</b>		Regulador transcripcional, familia TetR

**Anexo V. Conglomerados de genes identificados del bin exclusivo de las cepas del Clado C-I identificado por Anvi'o.**

Identificador de conglomerado de genes	Índice de homogeneidad funcional	Índice de homogeneidad geométrica	Categoría COG	Accesión de función COG	Función COG
GC_00004213	0.8303303	0.9044067	E	COG0765	Sistema de transporte de aminoácidos tipo ABC, componente permeasa
GC_00004217	0.9985348	1			
GC_00004236	0.9728395	1			
GC_00004238	0.9854906	1	R	COG2350	Proteína conservada no caracterizada Ycil, contiene sitio activo fosfohistidina predicho
GC_00004239	0.978858	0.9918194	K	COG1309	Regulador transcripcional de unión a ADN, familia AcrR
GC_00004241	0.9960695	1			
GC_00004242	0.9863346	1			
GC_00004244	0.9852571	1			
GC_00004246	0.7737465	0.8558289	K	COG1396	Regulador transcripcional, contiene dominio HTH de familia XRE
GC_00004247	0.9963971	1	G	COG1830	Fructosa bisfosfato aldolasa clase Ia, familia Dhna
GC_00004250	0.9738506	0.9705329	H R	COG1335	Amidasa relacionada a nicotinamidasa
GC_00004251	0.9529818	1	K	COG1695	Regulador transcripcional de unión a ADN, familia PadR
GC_00004253	0.9773937	0.9889393	K	COG0789	Regulador transcripcional de unión a ADN, familia MerR
GC_00004255	0.9788986	1			
GC_00004258	0.984083	1	P	COG2221	Sulfito reductasa disimilatoria, subunidades alfa y beta
GC_00004259	0.9666179	1	K X	COG0789  COG2452	Regulador transcripcional de unión a ADN, familia MerR   Integrasa resolvasa sitio específica predicha
GC_00004260	0.9686067	1	K	COG1309	Regulador transcripcional de unión a ADN, familia AcrR
GC_00004262	0.9780167	0.9983871	D P	COG0239	Exportador de ión fluoruro CrcB/FEX
GC_00004264	0.9786834	0.9969349	E P	COG0601	Sistema de transporte dipéptido/oligopéptido/niquel tipo ABC, componente permeasa
GC_00004267	0.9876358	1			
GC_00004270	0.948429	0.9706525	T	COG2199  COG2202  COG2202	Dominio GGDEF, diguanilato ciclasa (c-di-GMP sintetasa) o sus variantes enzimáticamente inactivas, dominio PAS
GC_00004271	0.9692998	1	T	COG2199  COG2203	Dominio GGDEF, diguanilato ciclasa (c-di-GMP sintetasa) o sus variantes enzimáticamente inactivas, dominio GAF

<b>GC_00004273</b>	0.9918166	1	K	COG1309	Regulador transcripcional de unión a ADN, familia AcrR
<b>GC_00004274</b>	0.9884117	1	S	COG5423	Proteína de unión a metal predicha
<b>GC_00004276</b>	0.9549308	1	J O	COG1670	Proteína N-acetiltransferasa, familia RimJ/RimL
<b>GC_00004278</b>	0.9735237	1	J	COG0456	Proteína ribosomal S18 acetilasa RimL y acetiltransferasas relacionadas
<b>GC_00004280</b>	0.9871605	1			
<b>GC_00004282</b>	0.9743538	1	K T	COG3706	Regulador de respuesta de dos componentes, familia PleD, Dominio GGDEF, diguanilato ciclasa (c-di-GMP sintetasa) o sus variantes enzimáticamente inactivas
<b>GC_00004283</b>	0.9750725	1	G	COG0662	Manosa-6-fosfato isomerasa, familia cupina
<b>GC_00004286</b>	0.8482018	0.8241404	R	COG1149	ATPasa de haza P superfamilia MinD, contiene dominio ferredoxina incertado
<b>GC_00004287</b>	0.8688453	0.9193178	F	COG0563	Adenilato quinasa o quinasa relacionada
<b>GC_00004288</b>	0.9548868	0.9787334	E T	COG0834  COG0834  COG2199  COG2200	Sistema de transporte/transducción de señal de aminoácido tipo ABC, componente/dominio periplásmico, dominio EAL   Dominio GGDEF, diguanilato ciclasa (c-di-GMP sintetasa) o sus variantes enzimáticamente inactivas, fosfodiesterasa específica de c-di-GMP clase I, o su variante enzimáticamente inactiva
<b>GC_00004290</b>	0.9699604	1	K T	COG3947	Regulador de respuesta de dos componentes, familia SAPR, consiste en dominios REC, wHTH y BTAD
<b>GC_00004291</b>	0.9645606	0.9956427			
<b>GC_00004293</b>	0.9792616	1	F G V	COG0737  COG1653	2', 3'-nucleótido cíclico 2' fosfodiesterasa/5' o 3' nucleotidasa, familia 5' nucleotidasa, sistema de transporte glicerol-3-fosfato tipo ABC, componente periplásmico
<b>GC_00004294</b>	0.9899976	0.9987261	K	COG1846	Regulador transcripcional de unión a ADN, familia MarR
<b>GC_00004296</b>	0.9668562	0.9975194	K	COG1309	Regulador transcripcional de unión a ADN, familia AcrR
<b>GC_00004298</b>	0.9593671	1			
<b>GC_00004299</b>	0.9896769	1	R	COG0110	Acetiltransferasa (superfamilia de parche de isoleucina)
<b>GC_00004300</b>	0.9608644	1	K R T	COG0457  COG3706	Repetición de tetratricopéptido (TPR), regulador de respuesta de dos componentes, familia PleD
<b>GC_00004302</b>	0.9824831	1			
<b>GC_00004303</b>	0.9762564	1	K R	COG1917  COG2207	Proteína con dominio cupina relacionado a quecetina dioxigenasa, dominio de unión a ADN tipo AraC y proteínas contenedoras de AraC
<b>GC_00004305</b>	0.9712746	1			
<b>GC_00004306</b>	0.9910921	1	K	COG2390	Regulador transcripcional de unión a ADN LsrR, familia DeoR

<b>GC_00004307</b>	0.9773463	0.9799093	R	COG4552	Acetiltransferasa predicha
<b>GC_00004308</b>	0.9843493	1	S	COG0730	Proteína de membrana no caracterizada YfcA
<b>GC_00004309</b>	0.9682905	0.9975917			
<b>GC_00004310</b>	0.9698026	0.9965931	K	COG0789	Regulador transcripcional de unión a ADN, familia MerR
<b>GC_00004312</b>	0.9695556	1	K	COG1309	Regulador transcripcional de unión a ADN, familia AcrR
<b>GC_00004313</b>	0.9664609	0.9750687	G K	COG1940	Quinasa de carbohidrato de la familia NBD/HSP70, puede contener dominio N terminal HTH
<b>GC_00004314</b>	0.9801717	1	L	COG3973	ADN helicasa IV
<b>GC_00004315</b>	0.9143519	0.9194711			
<b>GC_00004316</b>	0.8552681	0.9594417	K	COG1309	Regulador transcripcional de unión a ADN, familia AcrR
<b>GC_00004317</b>	0.9766097	1	E	COG0703  COG1605	Shikimato quinasa, corismato mutasa
<b>GC_00004318</b>	0.9687444	1	K R	COG0454	N-acetiltransferasa, superfamilia GNAT (incluye histona acetiltransferasa HPA2)
<b>GC_00004320</b>	0.965713	1	E P Q	COG0444  COG4172	Sistema de transporte dipéptido/oligopéptido/niquel tipo ABC, componente permeasa   Sistema de transporte de microcina C tipo ABC, componente ATPasa YejF duplicado
<b>GC_00004321</b>	0.9767247	0.9904243	G	COG2814	Permeasa de eflujo de arabinosa predicha, familia MFS
<b>GC_00004322</b>	0.8949555	0.9197223	O J R	COG1670  COG2320	Proteína N-acetiltransferasa, familia RimJ/RimL   Dominio GrpB, nucleotidiltransferasa predicha, familia UPF0157
<b>GC_00004324</b>	0.8895792	0.8358105	K T	COG0789  COG4978	Regulador transcripcional de unión a ADN, familia MerR, dominio de unión a efector bacteriano
<b>GC_00004325</b>	0.9556212	0.9610485	R	COG4188	Dienelactona hidrolasa predicha
<b>GC_00004326</b>	0.9297469	0.9465858	E	COG0765	Sistema de transporte de aminoácidos tipo ABC, componente permeasa
<b>GC_00004330</b>	0.9752102	0.9917654			
<b>GC_00004331</b>	0.9940385	1	G	COG2376	Dihidroxiacetona quinasa
<b>GC_00004332</b>	0.9541209	0.9708169	K	COG1846	Regulador transcripcional de unión a ADN, familia MarR
<b>GC_00004334</b>	0.9914609	1	K	COG1476	Regulador transcripcional de unión a ADN, dominio HTH familia XRE
<b>GC_00004335</b>	0.9782267	0.9956349	T	COG2199	Dominio GGDEF, diguanilato ciclasa (c-di-GMP sintetasa) o sus variantes enzimáticamente inactivas
<b>GC_00004336</b>	0.9561009	0.9793219	E	COG4608	Sistema de transporte de oligopéptido tipo ABC, componente ATPasa
<b>GC_00004337</b>	0.9797555	1	R	COG1461	Quinasa predicha relacionada a dihidroxiacetona kinasa
<b>GC_00004338</b>	0.9773191	1	R	COG0673	Deshidrogenasa predicha

<b>GC_00004340</b>	0.9519569	0.9933038	K T	COG3706	Regulador de respuesta de dos componentes, familia PleD
<b>GC_00004341</b>	0.9656273	1	M	COG1136	Sistema de exporte de lipoproteína tipo ABC, componente ATPasa
<b>GC_00004342</b>	0.9814121	1	E	COG0646  COG0685	Metionina sintasa I (dependiente de cobalamina), dominio metiltransferasa, 5, 10-metilentetrahidrofolato reductasa
<b>GC_00004343</b>	0.9708937	0.9949275	C	COG1853	NADH-FMN oxidorreductasa RutF, flavina reductasa familia (DIM6/NTAB)
<b>GC_00004344</b>	0.866837	0.8759979	C	COG0716	Flavodoxina
<b>GC_00004345</b>	0.9842593	1	V	COG0534	Bomba de eflujo multidroga empujada por Na+
<b>GC_00004346</b>	0.9641975	1	K T	COG0789  COG4978	Regulador transcripcional de unión a ADN, familia MerR, dominio de unión a efector bacteriano
<b>GC_00004347</b>	0.9581802	0.9972045	E	COG2957	Agmatina/peptidilarginina deiminasa
<b>GC_00004349</b>	0.8434067	0.9151963			
<b>GC_00004350</b>	0.9936322	1	V	COG0534	Bomba de eflujo multidroga empujada por Na+
<b>GC_00004351</b>	0.9710598	0.997302	X	COG5361	Proteína conservada no caracterizada
<b>GC_00004352</b>	0.9507792	0.9331772	R	COG3153	N-acetiltransferasa predicha YhbS
<b>GC_00004353</b>	0.9732992	1	E T	COG0834  COG0834  COG2199  COG2200	Sistema de transporte/transducción de señal de aminoácido tipo ABC, componente/dominio periplásmico, diguanilato ciclasa (c-di-GMP sintetasa) o sus variantes enzimáticamente inactivas, dominio EAL fosfodiesterasa específica de c-di-GMP clase I (o sus variantes enzimáticamente inactivas)
<b>GC_00004354</b>	0.7403859	0.8040424			
<b>GC_00004355</b>	0.9480851	0.9954719			
<b>GC_00004356</b>	0.9658692	1	E	COG4166	Sistema de transporte ABC de oligopéptido, componente periplásmico
<b>GC_00004357</b>	0.9186626	0.9648483	S	COG4859	Proteína no caracterizada
<b>GC_00004358</b>	0.9792394	1	K	COG2378	Regulador transcripcional de unión a ADN predicho YafY, contiene dominios WYL y HTH
<b>GC_00004359</b>	0.9522528	1	V	COG0577  COG0577	Sistema de transporte de péptido antimicrobiano tipo ABC, componente permeasa.

**Anexo VI. Características de las secuencias firma identificadas por dbGWAS de más de 100pb presentes en los primeros 60 nodos analizados según valor de  $q$ , y sus resultados al alinearlas con la colección de nucleótidos del NCBI por BLASTn.**

dbGWAS					BLAST						
Locus tag/sp de la anotación	Anotación	Identificador de la proteína	E del nodo	Largo de la secuencia	Descripción	Puntaje máximo	Puntaje total	Covertura	Valor E	Porcentaje de identidad	Accesión
CD630_12890	proteína de familia ribonucleasa J	CAJ68145.1	6.07E-52	134	<i>Clostridioides difficile</i> strain TW11-RT078 chromosome	204	204	100%	1.00E-48	94.03%	<a href="#">CP035499.1</a>
<i>Methylibium petroleiphilum</i>	Ribulosa bisfosfato carboxilasa de cadena larga	A2SFV1	4	104	<i>Clostridioides difficile</i> strain 020709 chromosome	176	176	100%	2.00E-40	97.12%	<a href="#">CP028529.1</a>
CD630_16820	Hidrolasa nucleosidasa predicha, familia IUNH	CAJ68549.1	6.68E-51	117	<i>Clostridioides difficile</i> strain CBA7204 chromosome, complete genome	195	195	100%	5.00E-46	96.58%	<a href="#">CP029566.1</a>
CD630_16730	Proteína trax	CAJ68539	4.46E-37	101	<i>Clostridioides difficile</i> strain DSM 29627 chromosome, complete genome	152	152	99%	3.00E-33	94.00%	<a href="#">CP016102.1</a>
CD630_24630	Regulador transcripcional, represor inducible por calor HrcA	CAJ69350	3.69E-48	109	<i>Clostridioides difficile</i> strain CBA7204 chromosome, complete genome	185	185	100%	3.00E-43	97.25%	<a href="#">CP029566.1</a>
CD630_10632	Proteína hipotética conservada	CAJ67906.1	4.87E-17	106	<i>Clostridioides difficile</i> strain TW11-RT078 chromosome	161	161	100%	5.00E-36	94.39%	<a href="#">CP035499.1</a>
CD630_03940	D-lactato dehidrogenasa	CAJ67215	1.25E-42	102	<i>Clostridioides difficile</i> strain TW11-RT078 chromosome	167	167	100%	1.00E-37	96.08%	<a href="#">CP035499.1</a>
CD630_05430	Proteína hipotética conservada	CAJ67376	4.46E-37	101	<i>Clostridioides difficile</i> strain CBA7204	154	154	100%	7.00E-34	94.06%	<a href="#">CP029566.1</a>

					chromosome, complete genome						
<b>Rhodopirellul a baltica</b>	ADN polimerasa propensa a errores	Q7UXK9	2.20E-01	110	<u>Clostridioides difficile strain CBA7204 chromosome, complete genome</u>	158	158	100%	6.00E- 35	92.73%	<u>CP029566 .1</u>
<b>NA</b>	NA	NA	NA	105	NA	NA	NA	NA	NA	NA	NA
<b>CD630_27510</b>	Proteína hipotética conservada	CAJ69639.1	4.05E-35	180	<u>Clostridioides difficile strain TW11-RT078 chromosome</u>	281	281	99%	7.00E- 72	94.97%	<u>CP035499 .1</u>
<b>CD630_31760</b>	Factor sigma-54 de ARN polimerasa	CAJ70073.1	9.54E-44	101	<u>Clostridioides difficile strain TW11-RT078 chromosome</u>	171	171	100%	7.00E- 39	97.03%	<u>CP035499 .1</u>
<b>CD630_29610</b>	Proteína no caracterizada	CAJ69854.1	3.79E-26	106	<u>Clostridioides difficile strain 020709 chromosome</u>	163	163	100%	1.00E- 36	94.34%	<u>CP028529 .1</u>
<b>CD630_07110</b>	Arginil-ARNt sintetasa	CAJ67544.1	1.05E-24	134	<u>Clostridioides difficile strain TW11-RT078 chromosome</u>	226	226	100%	2.00E- 55	97.01%	<u>CP035499 .1</u>
<b>ubiB</b>	Proteína kinasa probable UbiB	UBIB_ACTP 7	3.40E+00	116	<u>Clostridioides difficile strain CBA7204 chromosome, complete genome</u>	191	191	99%	7.00E- 45	96.52%	<u>CP029566 .1</u>
<b>CD630_34740</b>	ATP sintasa cadena A	CAJ70377.1	2.34E-62	180	<u>Clostridioides difficile strain TW11-RT078 chromosome</u>	311	311	100%	9.00E- 81	97.78%	<u>CP035499 .1</u>
<b>CD630_13440</b>	Proteína predicha de biosíntesis de pared celular	CAJ68202.1	9.53E-44	101	<u>Clostridioides difficile strain Mta-79 chromosome, complete genome</u>	176	176	100%	2.00E- 40	98.02%	<u>CP042267 .1</u>
<b>CD630_30370</b>	Regulador transcripcional, familia CarD	CAJ69930.1	5.97E-46	105	<u>Clostridioides difficile strain CBA7204 chromosome,</u>	178	178	100%	5.00E- 41	97.14%	<u>CP029566 .1</u>

					complete genome							
<b>CD630_08070</b>	Proteína no caracterizada	CAJ67641.1	9.95E-39	104	<u><i>Clostridioides difficile</i> strain TW11-RT078 chromosome</u>	165	165	100%	4.00E-37	95.19%	<u>CP035499</u> .1	
<b>NA</b>	NA	NA	NA	175	NA	NA	NA	NA	NA	NA	NA	
<b>CD630_08120</b>	Proteína A de estrés universal predicha (UspA)	CAJ67646.1	5.19E-63	148	<u><i>Clostridioides difficile</i> strain DSM 28670 chromosome, complete genome</u>	241	241	100%	9.00E-60	95.95%	<u>CP012312</u> .1	
<b>CD630_01270</b>	Proteína determinante de la forma de bastón MreB	CAJ66947.1	5.76E-68	163	<u><i>Clostridioides difficile</i> strain TW11-RT078 chromosome</u>	257	257	100%	1.00E-64	95.09%	<u>CP035499</u> .1	
<b>CD630_01260</b>	Proteína D de esporulación etapa III	CAJ66946.1	1.44E-37	115	<u><i>Clostridioides difficile</i> strain TW11-RT078 chromosome</u>	185	185	100%	3.00E-43	95.65%	<u>CP035499</u> .1	
<b>CD630_01160</b>	Subunidad alfa de ferredoxina/flavodoxina oxidoreductasa predicha	CAJ66936.1	7.25E-45	100	<u><i>Clostridioides difficile</i> strain CBA7204 chromosome, complete genome</u>	174	174	100%	6.00E-40	98.00%	<u>CP029566</u> .1	
<b>NA</b>	NA	NA	NA	120	<u><i>Clostridioides difficile</i> strain TW11-RT078 chromosome</u>	189	189	100%	3.00E-44	95.00%	<u>CP035499</u> .1	
<b>CD630_01250</b>	Endopeptidasa de pared celular predicha	CAJ66945.1	2.39E-45	116	<u><i>Clostridioides difficile</i> strain CBA7204 chromosome, complete genome</u>	176	176	100%	2.00E-40	93.97%	<u>CP029566</u> .1	
<b>CD630_01160</b>	Subunidad alfa de ferredoxina/flavodoxina oxidoreductasa predicha	CAJ66936.1	3.54E-59	132	<u><i>Clostridioides difficile</i> strain TW11-RT078 chromosome</u>	222	222	100%	3.00E-54	96.97%	<u>CP035499</u> .1	
<b>CD630_33010</b>	Proteasa La dependiente de ATP, peptidasa de familia S16	CAJ70198.2	7.81E-45	106	<u><i>Clostridioides difficile</i> strain DSM 29627 chromosome,</u>	180	180	100%	1.00E-41	97.17%	<u>CP016102</u> .1	

					<u>complete genome</u>							
<b>CD630_33050</b>	Subunidad 1 proteolítica de proteasa Clp dependiente de ATP	CAJ70202.1	9.66E-20	130	<u>Clostridioides difficile strain TW11-RT078 chromosome</u>	224	224	100%	8.00E-55	97.69%	<u>CP035499</u> .1	
<b>CD630_32890</b>	Proteína hipotética conservada	CAJ70186.1	1.04E-38	109	<u>Clostridioides difficile strain 020477 chromosome</u>	185	185	100%	3.00E-43	97.25%	<u>CP028524</u> .1	
<b>CD630_32890</b>	Proteína hipotética conservada	CAJ70186.1	9.89E-09	100	<u>Clostridioides difficile strain CBA7204 chromosome, complete genome</u>	174	174	100%	6.00E-40	98.00%	<u>CP029566</u> .1	
<b>CD630_15461</b>	Proteína hipotética conservada	CCA62825.1	5.55E-57	125	<u>Clostridioides difficile strain CBA7204 chromosome, complete genome</u>	215	215	100%	5.00E-52	97.60%	<u>CP029566</u> .1	
<b>CD630_15360</b>	Subunidad alfa de ferredoxina-NADPH(+) reductasa	CAJ68401.1	3.69E-48	109	<u>Clostridioides difficile strain FDAARGOS_723 chromosome, complete genome</u>	191	191	100%	6.00E-45	98.17%	<u>CP046327</u> .1	
<b>CD630_15360</b>	Subunidad alfa de ferredoxina-NADPH(+) reductasa	CAJ68401.1	9.53E-44	101	<u>Clostridioides difficile strain TW11 chromosome, complete genome</u>	176	176	100%	2.00E-40	98.02%	<u>CP045224</u> .1	
<b>CD630_22430</b>	Proteína de membrana predicha	CAJ69128.1	2.55E-15	118	<u>Clostridioides difficile strain CBA7204 chromosome, complete genome</u>	196	196	100%	2.00E-46	96.61%	<u>CP029566</u> .1	
<b>CD630_19270</b>	Fragmento de sistema de transporte tipo-ABC, proteína de unión a ATP	NA	2.77E-39	104	<u>Clostridioides difficile strain TW11-RT078 chromosome</u>	165	165	100%	4.00E-37	95.19%	<u>CP035499</u> .1	

<b>CD630_22580</b>	Proteína hipotética conservada	CAJ69143.1	6.03E-57	134	<a href="#">Clostridioides difficile strain DSM 28670 chromosome, complete genome</a>	220	220	100%	1.00E-53	96.27%	<a href="#">CP012312</a> <a href="#">.1</a>
<b>CD630_22550</b>	Proteína hipotética conservada	CAJ69141.1	1.14E-48	119	<a href="#">Clostridioides difficile strain DSM 102978 chromosome, complete genome</a>	198	198	100%	4.00E-47	96.64%	<a href="#">CP020380</a> <a href="#">.1</a>
<b>NA</b>	NA	NA	NA	100	<a href="#">Clostridioides difficile strain CBA7204 chromosome, complete genome</a>	161	161	100%	4.00E-36	96.00%	<a href="#">CP029566</a> <a href="#">.1</a>
<b>NA</b>	NA	NA	NA	106	<a href="#">Clostridioides difficile strain CBA7204 chromosome, complete genome</a>	178	178	98%	5.00E-41	97.17%	<a href="#">CP029566</a> <a href="#">.1</a>
<b>CD630_27640</b>	Hidrolasa predicha, superfamilia HAD, subfamilia IIB	CAJ69652.1	5.99E-36	104	<a href="#">Clostridioides difficile strain TW11-RT078 chromosome</a>	182	182	100%	4.00E-42	98.08%	<a href="#">CP035499</a> <a href="#">.1</a>
<b>Q18BH3</b>	Factor de maduración de ribosoma RimP	RIMP_PEPD 6	1.46E-16	101	<a href="#">Clostridioides difficile strain DSM 28669 chromosome, complete genome</a>	154	154	100%	7.00E-34	94.06%	<a href="#">CP012323</a> <a href="#">.1</a>
<b>CD630_32500</b>	Proteína hipotética conservada	CAJ70147.2	4.03E-53	118	<a href="#">Clostridioides difficile strain TW11-RT078 chromosome</a>	202	202	100%	3.00E-48	97.46%	<a href="#">CP035499</a> <a href="#">.1</a>
<b>CD630_32510</b>	Dehidrogenasa predicha	CAJ70148.1	1.68E-46	106	<a href="#">Clostridioides difficile strain TW11-RT078 chromosome</a>	180	180	100%	1.00E-41	97.17%	<a href="#">CP035499</a> <a href="#">.1</a>
<b>CD630_32470</b>	Proteína de transferencia de electrones predicha	CAJ70144.1	8.20E-50	112	<a href="#">Clostridioides difficile strain TW11-RT078 chromosome</a>	191	191	100%	7.00E-45	97.32%	<a href="#">CP035499</a> <a href="#">.1</a>

NA	NA	NA	NA	112	NA	NA	NA	NA	NA	NA	NA
<b>CD630_21070</b>	Proteína de familia permeasa de Xantina/uracilo/tiamina/ascorbato	CAJ68992.1	2.17E-40	104	<u><i>Clostridioides difficile</i> strain CBA7204 chromosome, complete genome</u>	159	159	100%	2.00E-35	94.23%	<u>CP029566</u> .1
<b>CD630_21020</b>	Antiportador Na(+)/H(+) predicho	CAJ68987.2	1.18E-33	120	<u><i>Clostridioides difficile</i> strain DSM 28669 chromosome, complete genome</u>	172	172	100%	3.00E-39	92.50%	<u>CP012323</u> .1
<b>CD630_21020</b>	Antiportador Na(+)/H(+) predicho	CAJ68987.2	7.33E-40	100	<u><i>Clostridioides difficile</i> strain CBA7204 chromosome, complete genome</u>	158	158	100%	6.00E-35	95.00%	<u>CP029566</u> .1
<b>A5G447</b>	Proteína de reparación de mal emparejamiento de ADN MutS	MUTS_GEOUR	9.40E+00	122	<u><i>Clostridioides difficile</i> strain CBA7204 chromosome, complete genome</u>	215	215	100%	4.00E-52	98.36%	<u>CP029566</u> .1
<b>P42458</b>	Proteína de unión a FK506 probable	FKBP_CORGL	8.20E+00	117	<u><i>Clostridioides difficile</i> strain TW11 chromosome, complete genome</u>	204	204	100%	9.00E-49	98.29%	<u>CP045224</u> .1
NA	NA	NA	NA	138	<u><i>Clostridioides difficile</i> strain CBA7204 chromosome, complete genome</u>	222	222	100%	3.00E-54	95.65%	<u>CP029566</u> .1
<b>CD630_13560</b>	Proteína con dominio CBS predicha	CAJ68214.2	1.07E-48	113	<u><i>Clostridioides difficile</i> strain FDAARGOS_723 chromosome, complete genome</u>	187	187	100%	9.00E-44	96.46%	<u>CP046327</u> .1
<b>CD630_13570</b>	Peptidil-prolil isomerasa predicha	CAJ68215.1	1.64E-36	103	<u><i>Clostridioides difficile</i> strain TW11</u>	158	158	100%	6.00E-35	94.17%	<u>CP045224</u> .1

					<a href="#">chromosome, complete genome</a>						
<b>CD630_10440</b>	Proteína de membrana predicha	CAJ67885.1	3.92E-43	114	<a href="#">Clostridioides difficile strain DSM 28669 chromosome, complete genome</a>	200	200	100%	1.00E-47	98.25%	<a href="#">CP012323.1</a>
<b>NA</b>	NA	NA	NA	123	NA	NA	NA	NA	NA	NA	NA
<b>CD630_19370</b>	subunidad beta de acetil CoA carboxilasa	CAJ68812.1	1.44E-52	117	<a href="#">Clostridioides difficile strain TW11-RT078 chromosome</a>	200	200	100%	1.00E-47	97.44%	<a href="#">CP035499.1</a>
<b>NA</b>	NA	NA	NA	119	<a href="#">Clostridioides difficile strain TW11-RT078 chromosome</a>	198	198	100%	4.00E-47	96.64%	<a href="#">CP035499.1</a>
<b>NA</b>	NA	NA	NA	133	<a href="#">Clostridioides difficile strain TW11-RT078 chromosome</a>	178	178	100%	6.00E-41	91.04%	<a href="#">CP035499.1</a>
<b>CD630_12470</b>	Proteína no caracterizada	CAJ68102.1	7.14E-11	118	<a href="#">Clostridioides difficile strain CBA7204 chromosome, complete genome</a>	200	200	100%	1.00E-47	97.46%	<a href="#">CP029566.1</a>
<b>CD630_12510</b>	Receptor de partícula de reconocimiento de señal	CAJ68106.1	5.66E-13	150	<a href="#">Clostridioides difficile strain TW11-RT078 chromosome</a>	261	261	100%	7.00E-66	98.00%	<a href="#">CP035499.1</a>
<b>CD630_12490</b>	Proteína de superfamilia radical SAM predicha	CAJ68104.1	1.29E-42	105	<a href="#">Clostridioides difficile strain TW11-RT078 chromosome</a>	183	183	100%	1.00E-42	98.10%	<a href="#">CP035499.1</a>
<b>CD630_12500</b>	Proteína de mantenimiento estructural de cromosoma SMC	CAJ68105.1	5.26E-17	114	<a href="#">Clostridioides difficile strain TW11-RT078 chromosome</a>	189	189	100%	2.00E-44	96.49%	<a href="#">CP035499.1</a>
<b>CD630_02600</b>	Proteína de biosíntesis flagelar FlIP	CAJ67082.1	1.72E-46	109	<a href="#">Clostridioides difficile strain CBA7204 chromosome, complete genome</a>	180	180	100%	1.00E-41	96.33%	<a href="#">CP029566.1</a>

<b>CD630_26820</b>	Piruvato ferredoxina oxidorreductasa	CAJ69568.1	1.61E-52	129	<u><i>Clostridioides difficile</i> strain TW11-RT078 chromosome</u>	206	206	100%	3.00E-49	95.35%	<u>CP035499</u> <u>.1</u>
<b>NA</b>	NA	NA	NA	134	<u><i>Clostridioides difficile</i> strain CBA7204 chromosome, complete genome</u>	226	226	100%	2.00E-55	97.01%	<u>CP029566</u> <u>.1</u>
<b>CD630_27160</b>	Proteína hipotética conservada	CAJ69603.1	1.23E-48	127	<u><i>Clostridioides difficile</i> strain TW11-RT078 chromosome</u>	202	202	97%	4.00E-48	95.97%	<u>CP035499</u> <u>.1</u>
<b>CD630_27140</b>	UDP-glucosa-4-epimerasa	CAJ69600.1	4.63E-37	104	<u><i>Clostridioides difficile</i> strain FDAARGOS_723 chromosome, complete genome</u>	159	159	100%	2.00E-35	94.23%	<u>CP046327</u> <u>.1</u>
<b>CD630_27130</b>	Proteína de pared celular predicha cwp22	CAJ69599.2	3.81E-48	112	<u><i>Clostridioides difficile</i> strain TW11-RT078 chromosome</u>	185	185	100%	3.00E-43	96.43%	<u>CP035499</u> <u>.1</u>

### Anexo VII. Resultados de PCR punto final para los pares de cebadores.

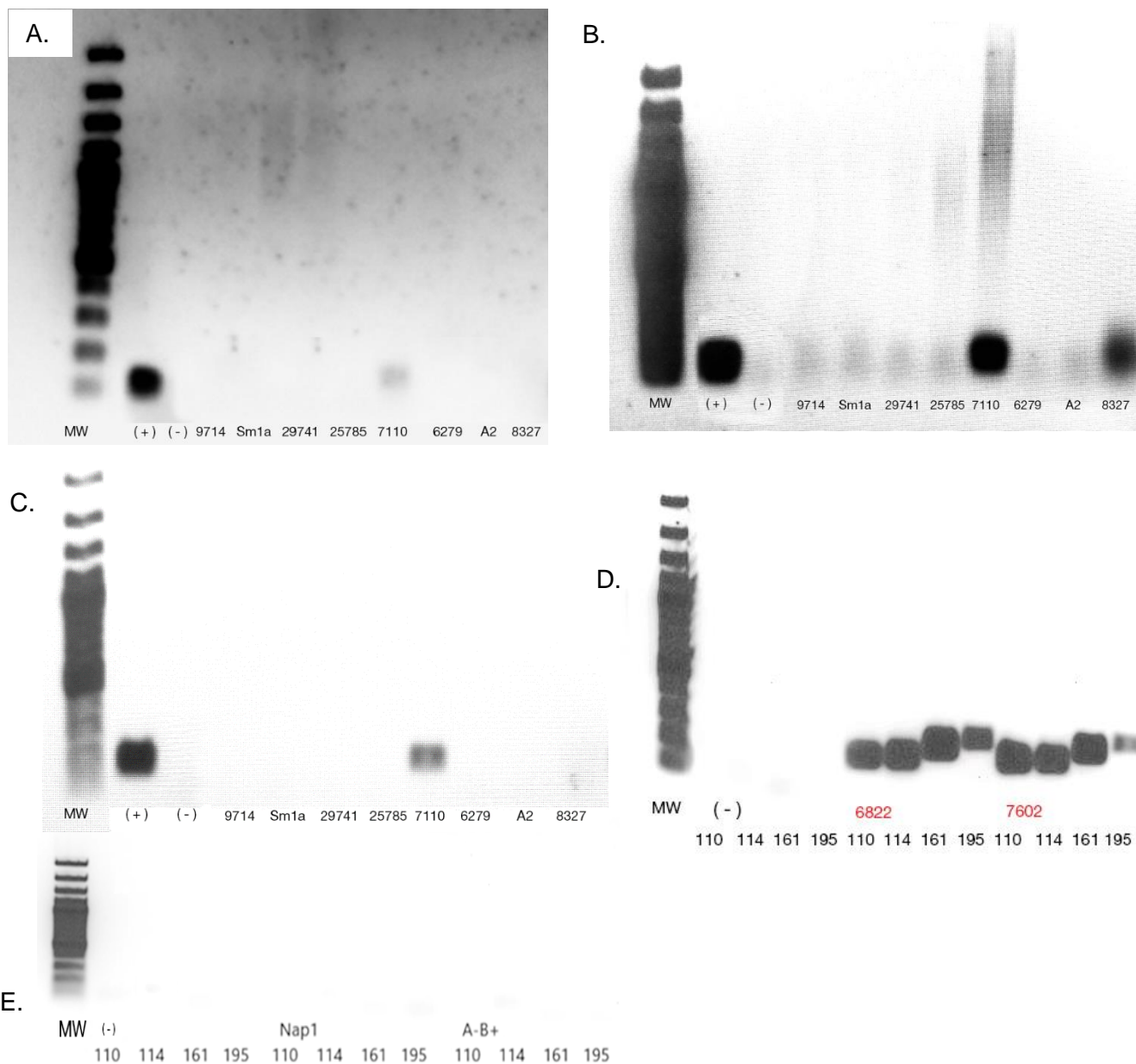


Figura 12. Amplificación de las cepas utilizadas con los cebadores 110 (panel A), 114 (panel B) y 161 (panel C), así como las cepas 6822, 7602 (panel D), NAP1 y A-B+ (panel E) con los cebadores 110, 114, 161 y 195. Los nombres de las cepas del Clado C-I se muestran en rojo. Se muestran los controles negativos de mezcla de reacción sin ADN como (-). En los paneles A, B y C el carril (+) corresponde a la cepa C-I 7602 y el (-) al control negativo sin ADN. El carril MW corresponde al marcador de peso molecular (100pb).