



**Centro de Investigación en Alimentación y  
Desarrollo, A.C.**

**ANÁLISIS PANGENÓMICO DEL PROTOZOARIO *Cyclospora  
cayetanensis* PARA LA IDENTIFICACIÓN DE MARCADORES  
NUCLEARES**

---

Por:

**LBG. Guadalupe Estefanía Alarcón Fabela**

TESIS APROBADA POR LA

COORDINACIÓN EN CIENCIA Y TECNOLOGÍA PARA PRODUCTOS AGRICOLAS DE  
ZONAS TROPICALES Y SUBTROPICALES

Como requisito parcial para obtener el grado de

**MAESTRÍA EN CIENCIAS**

## APROBACIÓN

Los miembros del comité designado para la revisión de la tesis de Guadalupe Estefanía Alarcón Fabela, la han encontrado satisfactoria y recomiendan que sea aceptada como requisito parcial para obtener el grado de Maestro en Ciencias.



---

Dra. Nohelia Castro del Campo  
Directora de Tesis



---

Dr. Luis Fernando Lozano Aguirre Beltrán  
Co-director de tesis



---

Dr. Cristóbal Chaidez Quiroz  
Integrante del comité de tesis



---

Dr. José Andrés Medrano Félix  
Integrante del comité de tesis

## DECLARACIÓN INSTITUCIONAL

La información generada en la tesis “Análisis Pangenómico del Protozooario *Cyclospora cayetanensis* para la Identificación de Marcadores Nucleares” es propiedad intelectual del Centro de Investigación en Alimentación y Desarrollo, A.C. (CIAD). Se permiten y agradecen las citas breves del material contenido en esta tesis sin permiso especial de la autora Guadalupe Estefanía Alarcón Fabela, siempre y cuando se dé crédito correspondiente. Para la reproducción parcial o total de la tesis con fines académicos, se deberá contar con la autorización escrita de quien ocupe la titularidad de la Dirección General del CIAD.

La publicación en comunicaciones científicas o de divulgación popular de los datos contenidos en esta tesis, deberá dar los créditos al CIAD, previa autorización escrita del manuscrito en cuestión del director(a) de tesis.



## AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACyT), por el apoyo económico otorgado para el estudio del posgrado.

Al Centro de Investigación en Alimentación y Desarrollo, A.C., Coordinación Culiacán por aceptarme como estudiante de su programa de posgrado y confiarme el uso de sus recursos para la realización de este trabajo.

Al Laboratorio Nacional para la Investigación en Inocuidad Alimentaria (LANIIA), por el apoyo técnico-científico brindado para llevar a cabo el cumplimiento del proyecto de investigación.

A mi directora de tesis, la Dra. Nohelia Castro del Campo, a quien agradezco todo su apoyo, confianza y conocimientos que fungieron como guía y me permitieron concluir con este proyecto. Pero más que eso, debo decir gracias por su comprensión, cariño y su amistad.

A mi co-director de tesis, el Dr. Luis Fernando Lozano Aguirre Beltrán que agradezco su amabilidad infinita y su calidez tanto como persona como maestro.

Al Dr. Cristóbal Chaidez Quiroz, le agradezco por su apoyo crítico y sus aportaciones, así como fomentar el trabajo y crecimiento en equipo, tanto de manera académica como en la convivencia social entre el grupo de laboratorio.

Al Dr. José Andrés Medrano Félix, gracias por su disponibilidad, paciencia y consejos que me permitieron enriquecer mi desarrollo tanto en la realización de este proyecto como en mi persona.

Al Dr. Juan Ramón Ibarra Rodríguez y a la Dra. Nohemí Castro del Campo por brindar su apoyo tanto con su experiencia técnica que permitieron fortalecer este trabajo.

Al equipo de trabajo del laboratorio de microbiología: a la QFB. Célida Isabel Martínez Rodríguez y a Miriam Belén Vega Rodríguez por su amabilidad, disponibilidad y constante apoyo técnico.

A los maestros que con sus conocimientos apoyaron en mi formación durante la maestría: Dr. Raymundo Estrada, Dr. Benigno Valdez, Dra. Nohelia Castro, Dr. Andrés Medrano, Dr. Osvaldo López, Dr. Cristóbal Chaidez, Dra. Adriana Sañudo, Dr. Juan Ramón Ibarra y MC. Armando Carrillo. Y, en general, a todo el personal de CIAD Culiacán, ya que cada uno de ellos contribuyó de diferente manera en mi formación profesional.

A mis amigos Edén, Jean Pierre, Berenice, Lennin y Rosalía quienes compartí gratas experiencias en el proceso y agradezco infinitamente su amistad.

A mis compañeros de laboratorio Grecia, Irvin, Alondra, Vianey, Roberto, Mónica y Belia.

A mis queridas amigas de infancia Anae y Dayanara, que a pesar de las distancias y a los caminos que cada una ha tomado, siguen compartiendo y apoyando estas experiencias conmigo a lo largo de los años.

A mi amiga Brenda, por ser tan linda como siempre, estar al pendiente de mí y seguir siendo parte de mi vida.

A mis amigas Cecilia, Carolina, Kenia, Sarai y mi amigo Geovanni, a las cuales agradezco sus conocimientos y experiencias, que me han apoyado a crecer tanto profesionalmente como emocionalmente, son un ejemplo a seguir para mí. Muchas gracias, los adoro.

A mi amigo y prácticamente hermano Manuel, le agradezco por ser mi support todos estos años. Valoro mucho culminar de nuevo otra etapa con su compañía.

A mis gorditos Hilary y Ángel, a quienes agradezco su apoyo, cariño y compartir de nuevo otra etapa de crecimiento. Son excelentes amigos.

A mi compañera del team chiclospora, querida amiga y hermana Yazmín, no pude tener mejor compañía de desvelo y aventura en todo este proceso. Agradezco quedarme con tan bonitos recuerdos.

A toda mi familia, a mis queridos tíos Olimpia, Joel, Alemzue y Keysi; a mis primos María Fernanda, Everardo, Monserrat y Matías; a mis hermanos Jorge y Haziél; a mis abuelos Everardo y María Silvia; mi madre Avigail, mi padre Edgardo, a mis hermanos Alejandra y Andrés; a mi esposo César y familia; les agradezco a todos por tanto amor, por su comprensión, por su apoyo y fe en mí en cada paso, son el pilar de mi vida.

## DEDICATORIA

*A mi hermosa madre Avigail Fabela, por su amor y apoyo incondicional,  
quién me ha dado todo y es todo lo que aspiro a ser.*

*A mi papá Edgardo Salazar, por su cariño y lo que  
me ha dado con tanto esfuerzo.*

*A mis queridos hermanos, Alejandra y Andrés, por creer en mí,  
impulsarme y brindarme todo su cariño.*

*A mi amado esposo César Gastelúm, quién siempre ha estado para mí,  
que cree en mis capacidades y me ha animado  
a crecer más en todo momento.*

*A toda mi familia, que me han apoyado hoy y siempre con todo su amor,  
pero en especial a mis abuelos María Silvia y Everardo.*

*A mis amigos, algunos viejos otros más nuevos, les agradezco  
infinitamente sus consejos, conocimiento y,  
sobre todo, su comprensión y cariño.*

*Este nuevo logro es posible gracias a ustedes,  
son mi inspiración, voluntad y amor,  
gracias a todos.*

## CONTENIDO

<b>APROBACIÓN</b> .....	2
<b>DECLARACIÓN INSTITUCIONAL</b> .....	3
<b>AGRADECIMIENTOS</b> .....	4
<b>DEDICATORIA</b> .....	6
<b>CONTENIDO</b> .....	7
<b>LISTA DE FIGURAS</b> .....	9
<b>LISTA DE CUADROS</b> .....	10
<b>RESUMEN</b> .....	11
<b>ABSTRACT</b> .....	13
<b>1. INTRODUCCIÓN</b> .....	15
<b>2. ANTECEDENTES</b> .....	17
2.1. Ciclosporiasis.....	17
2.2. <i>Cyclospora cayetanensis</i> .....	17
2.2.1 Taxonomía y Filogenia.....	17
2.2.2 Morfología.....	18
2.2.3 Ciclo de Vida.....	19
2.2.4 Epidemiología.....	21
2.2.5 Ruta de Transmisión.....	22
2.2.5.1 Transmisión por Alimentos.....	23
2.2.5.2 Transmisión por Agua.....	25
2.2.5.3 Transmisión por Vía Zoonótica.....	26
2.2.6 Identificación.....	26
2.2.7 Genoma Nuclear .....	27
2.2.7.1 Genotipificación del Genoma Nuclear de <i>Cyclospora cayetanensis</i> .....	28
2.3 Genómica Comparativa.....	31
2.3.1 Análisis de Pangenoma.....	31
2.3.2 Filogenómica.....	33
2.3.3 Herramientas de Análisis.....	35
<b>3. HIPÓTESIS</b> .....	37
<b>4. OBJETIVOS</b> .....	38
4.1 Objetivo General.....	38
4.2 Objetivos Específicos .....	38

## CONTENIDO (Continuación)

<b>5. MATERIALES Y MÉTODOS.....</b>	<b>39</b>
5.1 Bases de Datos con los Genomas de <i>Cyclospora Cayetanensis</i> Reportados en el NCBI.....	39
5.2 Análisis Bioinformáticos de las Secuencias Genómicas de <i>Cyclospora Cayetanensis</i> .....	40
5.2.1 Determinación del Genoma Core y Filogenia.....	40
5.2.2 Determinación del Pangenoma y Análisis del Genoma Accesorio.....	41
<b>6. RESULTADOS Y DISCUSIÓN.....</b>	<b>43</b>
6.1 Bases de Datos con los Genomas de <i>Cyclospora cayetanensis</i> Reportados en el NCBI.....	43
6.2 Análisis de las Relaciones Filogenéticas de las Secuencias Genómicas de <i>Cyclospora cayetanensis</i> con otros Microorganismos Apicomplexa.....	47
6.3 Marcadores Óptimos de <i>Cyclospora cayetanensis</i> y Análisis de las Relaciones Filogenéticas entre los Aislados de la misma especie.....	51
6.4 Análisis del Pangenoma de <i>Cyclospora cayetanensis</i> Comparado con Otros Microorganismos Apicomplexa.....	56
6.5 Análisis del Pangenoma de <i>Cyclospora cayetanensis</i> e Identificación de Genes Presentes Diferencialmente en las Secuencias Genómicas Reportadas como Mexicanas.....	58
<b>7. CONCLUSIONES.....</b>	<b>64</b>
<b>8. RECOMENDACIONES.....</b>	<b>65</b>
<b>9. REFERENCIAS.....</b>	<b>66</b>

## LISTA DE FIGURAS

<b>Figura</b>		<b>Página</b>
1	Morfología de los ooquistes de <i>Cyclospora cayetanensis</i> .....	19
2	Ciclo de vida de <i>Cyclospora cayetanensis</i> .....	20
3	Distribución y prevalencia de <i>Cyclospora cayetanensis</i> en el mundo.....	22
4	Pangenoma de tres genomas denominados A, B y C.....	32
5	Diagrama de flujo general de los materiales y métodos.....	42
6	Heatmap con valores de cobertura de Average Nucleotide Identity (ANI) de los 37 genomas reportados de <i>Cyclospora cayetanensis</i> .....	46
7	Heatmap con valores de identidad de Average Nucleotide Identity (ANI) de los 37 genomas reportados de <i>Cyclospora cayetanensis</i> .....	47
8	Diagrama de Venn de la intersección de los métodos OMCL y BDBH que determinan el genoma core de los microorganismos Apicomplexa.....	48
9	Árbol filogenético de especies del subfilum Apicomplexa utilizando 2 marcadores.....	51
10	Diagrama de Venn de la intersección de los métodos OMCL y BDBH que determinan el genoma core de los aislados de <i>Cyclospora cayetanensis</i> .....	52
11	Árbol filogenético de los 36 aislados de <i>Cyclospora cayetanensis</i> basado en máxima verosimilitud de secuencias de 47 marcadores obtenidos por el pipeline GET_PHYLOMARKERS.....	56
12	Gráfica de la estimación estadística del tamaño del genoma core (a) y pangenoma (b) de <i>Cyclospora cayetanensis</i> .....	61

## LISTA DE CUADROS

<b>Cuadro</b>		<b>Página</b>
1	Investigaciones de los brotes de <i>C. cayetanensis</i> por la CDC.....	25
2	Base de datos de microorganismos Apicomplexa descargada de la base de datos del NCBI.....	43
3	Base de datos de <i>Cyclospora cayetanensis</i> descargada de la base de datos del NCBI .....	45
4	Lista de genes del genoma core obtenido para la base de datos de microorganismos Apicomplexa utilizando los métodos de BDBH y OMCL...	49
5	Marcadores de la base de datos de microorganismos Apicomplexa obtenido por el pipeline de GET_HOMOLOGUES y GET_PHYLOMARKERS.....	50
6	Marcadores de la base de datos de <i>Cyclospora cayetanensis</i> obtenido por el pipeline de GET_HOMOLOGUES-EST y GET_PHYLOMARKERS.....	54
7	Genes únicos de <i>Cyclospora cayetanensis</i> en comparación a otros Apicomplexa.....	58
8	Información de las secuencias genómicas de <i>Cyclospora cayetanensis</i> asociadas a México.....	62
9	Comparación de <i>Cyclospora cayetanensis</i> del grupo relacionado a México contra otras secuencias de otras regiones.....	63

## RESUMEN

El protozooario *Cyclospora cayetanensis* es reconocido mundialmente como agente causal de diversos brotes diarreicos, vinculados al consumo de alimentos contaminados. En países como Estados Unidos, el rastreo de la fuente de contaminación microbiológica durante las investigaciones epidemiológicas ha llegado a vincular principalmente alimentos importados de México. Sin embargo, se carece de un método de tipificación contundente para *C. cayetanensis* que complemente las investigaciones epidemiológicas. Esto se debe especialmente a la falta de marcadores moleculares con un alto poder de resolución que permita discriminar entre cepas durante la aparición de los brotes. Por este motivo, el objetivo del presente trabajo fue analizar el pangenoma de *C. cayetanensis* de diferentes regiones geográficas, así como aislados asociados con México, partiendo del análisis de secuencias genómicas reportadas en bases de datos en búsqueda de potenciales marcadores nucleares que permitan la subtipificación de la especie. Para ello, se generaron dos bases de datos de secuencias genómicas descargadas del NCBI: una de microorganismos Apicomplexa y otra de *C. cayetanensis*. Se utilizaron los softwares GET\_HOMOLOGUES/-EST y GET\_PHYLOMARKERS para el análisis del pangenoma, genoma core y establecer la filogenia. Como resultados se obtuvieron dos marcadores (una Proteína de clasificación de proteínas vacuolares y una proteína de translocación) como óptimos para establecer la filogenia, observando que el género *Cyclospora* forma su propio clado; mientras que en el análisis de su pangenoma mostró 656 genes únicos. En cuanto a la base de datos de *C. cayetanensis*, se seleccionaron 36 secuencias que compartían 2,297 genes como genoma core, optando por 47 de estos genes para establecer el árbol filogenético. La visualización de la filogenia expuso tres agrupaciones por origen geográfico; en particular, siete secuencias relacionadas a México se encontraron dispersas en el árbol sin un perfil único de agrupación. En cuanto a la estimación del pangenoma de *C. cayetanensis*, este revela que el pangenoma es abierto, por lo que es necesario aumentar las secuencias de la especie para una caracterización representativa. Respecto a la identificación de genes únicos, se realizaron comparaciones del contenido de una secuencia de México con diferentes agrupaciones según el origen geográfico, obteniendo de 53 a 6 genes únicos. En conclusión, se estableció un flujo de trabajo con el que se obtuvo un repertorio de 47 marcadores para *C. cayetanensis* que permitieron visualizar claras agrupaciones según el origen geográfico de

algunas secuencias. Al mismo tiempo, a causa de la pobre caracterización de los aislados asociados con México, se sustenta la necesidad de añadir secuencias con certeza del origen de aislamiento.

**Palabras clave:** *Cyclospora cayetanensis*, Pangenoma, Genoma core, Filogenia, Marcadores

## ABSTRACT

The protozoan *Cyclospora cayetanensis* is recognized worldwide as a causal agent of various diarrheal outbreaks, linked to the consumption of contaminated food. In countries like the United States, tracking the source of microbiological contamination during epidemiological investigations has come to link mainly foods imported from Mexico. However, a robust typing method for *C. cayetanensis* to complement epidemiological investigations is lacking. This is especially due to the lack of molecular markers with a high resolving power that allows to discriminate between strains during the appearance of outbreaks. For this reason, the objective of this work was to analyze the *C. cayetanensis* pangenome from different geographic regions, as well as isolates associated with Mexico, based on the analysis of genomic sequences reported in databases in search of potential nuclear markers that allow subtyping of the specie. For this, two databases of genomic sequences downloaded from the NCBI were generated: one of Apicomplexa microorganisms and the other of *C. cayetanensis*. The software GET\_HOMOLOGUES /-EST and GET\_PHYLOMARKERS were used for the analysis of the pangenome, core genome and establishing the phylogeny. As results, two markers were obtained (a vacuolar protein classification protein and a translocation protein) as optimal to establish the phylogeny, observing that the genus *Cyclospora* forms its own clade; while in the analysis of its pangenome showed 656 unique genes. Regarding the *C. cayetanensis* database, 36 sequences that shared 2,297 genes were selected as the core genome, choosing for 47 of these genes to establish the phylogenetic tree. The visualization of the phylogeny exposed three clusters by geographic origin; in particular, seven sequences related to Mexico were found scattered in the tree without a unique grouping profile. Regarding the estimation of the *C. cayetanensis* pangenome, this reveals that the pangenome is open, so it's necessary to increase the sequences of the specie for a representative characterization. Regarding the identification of unique genes, comparisons of the content of a sequence from Mexico were made with different groups according to geographical origin, obtaining 53 to 6 unique genes. In conclusion, a workflow was established from which a repertoire of 47 markers for *C. cayetanensis* was obtained that allowed the visualization of clear clusters according to the geographical origin of some sequences. At the same time, due to the poor characterization of the isolates associated with Mexico, the need to add sequences with certainty of the origin of the

isolation is supported.

**Keywords:** *Cyclospora cayetanensis*, Pangenome, Core genome, Phylogeny, Markers

## 1. INTRODUCCIÓN

*Cyclospora cayetanensis* es un protozooario intracelular emergente causante de la infección gastrointestinal en humanos llamada ciclosporiasis. Dicha enfermedad incluye signos clínicos como diarrea acuosa, náuseas, vómito, fiebre y pérdida de peso (Connor *et al.*, 1999). En los últimos años, este microorganismo ha sido reconocido mundialmente como agente causal de diversos brotes diarreicos, teniendo como vehículo de transmisión alimentos frescos como frutas y hortalizas frescas, así como agua contaminadas (CDC, 2020a). En Estados Unidos, el Centro de Control y Prevención de Enfermedades (CDC, por sus siglas en inglés) ha registrado una incidencia anual de brotes de ciclosporiasis desde el 2013, en donde sus investigaciones con base en los datos epidemiológicos han llegado a vincular principalmente alimentos frescos contaminados importados de México (CDC, 2020a). Sin embargo, no todos los casos han logrado ser vinculados con un vehículo alimentario y con la procedencia específica de dicho vehículo, dado que se carece de un método de tipificación confiable para *C. cayetanensis* que complemente a las investigaciones epidemiológicas. El estudio a profundidad de este microorganismo representa un reto debido a diversos factores tales como: la dificultad de su propagación en cultivo celular y/o modelo animal (Eberhard *et al.*, 2000); e incluso el crecimiento en seres humanos ha sido infructuoso para utilizarse como modelo de estudio (Alfano-Sobsey *et al.*, 2004). En consecuencia, hay una limitada fuente de muestra para su estudio, ya que se restringe a obtenerla a partir de muestras de heces de personas infectadas, lo que conlleva un proceso laborioso de purificación para el aislamiento de los ooquistes (Qvarnstrom *et al.*, 2018). A pesar de las limitaciones experimentales, han surgido alternativas novedosas para el estudio de este tipo de microorganismos fastidiosos; por ejemplo, los nuevos métodos basados en secuenciación y biología computacional, han permitido analizar grupos microbianos que no pueden cultivarse en laboratorio como es el caso de *Cyclospora*, facilitando disponer de secuencias en bases de datos como la de NCBI (National Center for Biotechnology Information). Aunque la secuenciación del genoma completo para este microorganismo eucariota ha sido inviable para el genotipado de rutina, se ha optado por desarrollar un panel de marcadores moleculares para la secuenciación dirigida a amplicones para su análisis por MLST (Multilocus Sequence Typing). No obstante, hasta la fecha, no se ha logrado determinar algunos de estos marcadores MLST como rutinarios, por lo que la tendencia de estos

estudios se enfoca en la búsqueda de nuevos marcadores moleculares (Guo *et al.*, 2016; Li *et al.*, 2017; Barratt *et al.*, 2019; Houghton *et al.*, 2020). Uno de los enfoques bioinformáticos comúnmente empleado para el análisis en conjunto de varios genomas, es el estudio del pangenoma, que permite explorar las similitudes y diferencias en el contenido genético de varios genomas de manera global y/o específica (Tettelin *et al.*, 2008; Contreras-Moreira y Vinuesa, 2013). Si bien, estos estudios se han realizado principalmente en organismos procariotas, poco a poco se han establecido algoritmos que permiten el análisis de secuencias de mayor tamaño como las secuencias de microorganismos eucariotas. Tal es el caso de los softwares, GET\_HOMOLOGUES y GET\_PHYLOMARKERS recientemente desarrollados (Contreras-Moreira y Vinuesa, 2013; Contreras-Moreira *et al.*, 2017; Vinuesa *et al.*, 2018), los cuales brindan la posibilidad de explorar secuencias de considerable tamaño para la búsqueda de nuevos marcadores. El presente proyecto propone abordar un análisis del pangenoma de las secuencias de *C. cayetanensis* de diferentes regiones geográficas, para estudiar la diversidad genómica del protozooario y determinar si las cepas de distintos orígenes contienen diferencias en cuanto contenido genético. Esto con el objetivo de identificar marcadores nucleares prometedores que permitan la subtipificación del parásito y coadyuven a dar una mayor resolución al genotipo de *C. cayetanensis*.

## 2. ANTECEDENTES

### 2.1 Ciclosporiasis

En humanos, la ciclosporiasis es una enfermedad intestinal causada por el protozooario *Cyclospora cayetanensis*, el cual es catalogado en los últimos años como patógeno emergente. Posiblemente el primer reporte de la infección se registró en el año 1979, en el cual se describió la presencia de coccidias en materia fecal en dos niños y una mujer en Papúa, Nueva Guinea (Ashford, 1979). Sin embargo, fue hasta 1993, que este parásito fue clasificado como *C. cayetanensis*, debido a la caracterización de la morfología y esporulación *in vitro* de los ooquistes del parásito, realizada en la universidad de Cayetano Heredia en Perú por Ortega *et al.*, (1993). La infección se manifiesta con signos clínicos como náusea, calambre abdominal y diarrea acuosa. Sin un tratamiento, la infección puede durar aproximadamente de 10 a 12 semanas. Los grupos de la población principalmente afectados son niños, ancianos y personas inmunocomprometidas (CDC, 2020b). La adquisición del agente causal es mediante el consumo de alimentos o agua contaminados. El riesgo de infección aumenta con las personas que viajan o son originarios de países donde la enfermedad es considerada endémica, tales como: Bangladesh, Brasil, Chile, Cuba, República Dominicana, Egipto, Guatemala, Haití, India, Indonesia, México, Nepal, Nigeria, Perú, Tanzania, Tailandia, Venezuela y Zimbabue (Ortega y Sherchand, 2015).

### 2.2 *Cyclospora cayetanensis*

#### 2.2.1 Taxonomía y Filogenia

*C. cayetanensis* se encuentra descrita en el subfilum Apicomplexa, subclase Coccidiasina y familia *Eimeriidae* (Ortega y Sanchez, 2010). La mayor parte de los microorganismos Apicomplexa son

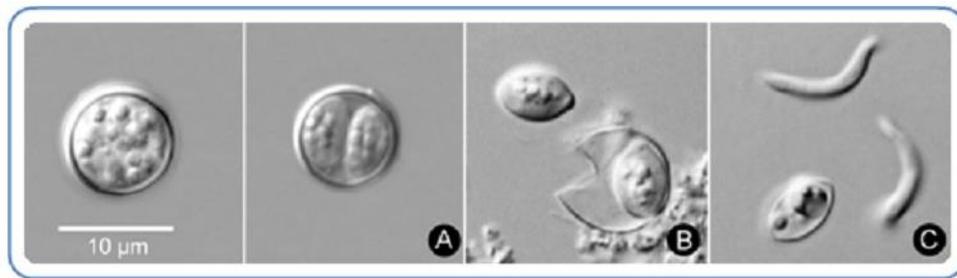
parásitos obligados y algunos son agentes causales de enfermedades importantes en humanos y animales, por ejemplo: malaria (*Plasmodium* spp.), toxoplasmosis (*Toxoplasma gondii*), criptosporidiosis (*Cryptosporidium* spp.), babesiosis (*Babesia* spp.) y coccidiosis (*Eimeria* spp.) (Sato, 2011). El género *Cyclospora* cuenta con 20 especies (Marangi *et al.*, 2015), las cuales se pueden dividir en dos grupos según la morfología: 1) especies que infectan a diplópodos, mamíferos insectívoros, roedores y reptiles, se caracterizan por su morfología de ooquistes subesféricos y grandes; y 2) especies que infectan a primates no humanos y a los humanos, producen ooquistes esféricos y pequeños (Eberhard *et al.*, 1999; Lainson, 2005; Chacín-bonilla y Barrios, 2011; Marangi *et al.*, 2015). Entre estas 20 especies, cinco infectan a primates (*C. papionis*, *C. colobi*, *C. cercopithecii* y *C. macacae*), entre las que se encuentra *C. cayetanensis*, la única descrita en la literatura que causa una infección en humanos (Marangi *et al.*, 2015). Aún no se han reportado investigaciones que determinen si todos los aislados de *Cyclospora* encontrados en humanos pertenecen a la misma especie, si las especies que afectan a primates pueden infectar al hombre así como la existencia de subtipos de la especie (Chacín-Bonilla, 2010).

En los estudios de *Cyclospora* sobre la identificación, diferenciación entre especies y la determinación de la filogenia del parásito, se han utilizado diferentes genes como el ITS-1, el ITS-2, el HSP70 y el ARNr 18s como marcadores moleculares (Adam *et al.*, 2001; Riner *et al.*, 2010; Sulaiman *et al.*, 2013; Ortega *et al.*, 2014). En general, todas estas secuencias han permitido la detección de *Cyclospora*, siendo el gen del ARNr 18S el más utilizado de estos marcadores, ya que ha permitido la diferenciación de *C. cayetanensis* de las demás especies. No obstante, la aplicación de estos marcadores presenta limitaciones debido a la baja variabilidad genética que contienen en sus secuencias, lo cual afecta la capacidad de resolución para discriminar si existen subtipos de cepas en la especie (Zhou *et al.*, 2011; Sulaiman *et al.*, 2013, 2014).

En cuanto a los análisis filogenéticos, se ha confirmado que existe un alto grado de homología entre las especies de *Cyclospora* con el género *Eimeria* (Reiman *et al.*, 1996). En concreto, el análisis filogenético del genoma del apicoplasto y la mitocondria han sugerido que *C. cayetanensis* se ubica más cercanamente relacionado con las especies de *Eimeria* que infectan el ciego de aves, como *E. tenella* y *E. necatrix* (Tang *et al.*, 2015).

### 2.2.2 Morfología

Los estadios morfológicos usualmente observados en *Cyclospora* son el ooquiste inmaduro o no esporulado y el ooquiste maduro o esporulado (Figura 1). El ooquiste no esporulado posee una doble pared gruesa de 63 nm y 50 nm de grosor. La parte interior consta de una masa celular redondeada de aspecto morular y granulosa, que deja un espacio transparente entre esta y la pared del ooquiste. El ooquiste esporulado de *C. cayetanensis* mide aproximadamente de 8 a 10  $\mu\text{m}$  de diámetro y también posee una doble pared de 63 y 50 nm de grosor. En su interior se encuentran dos esporoquistes ovoides de 4,0 x 6,3  $\mu\text{m}$  aproximadamente, los cuales, cada uno contiene dos esporozoítos (Ortega *et al.*, 1993; Vazquez-Tsuji *et al.*, 2000).

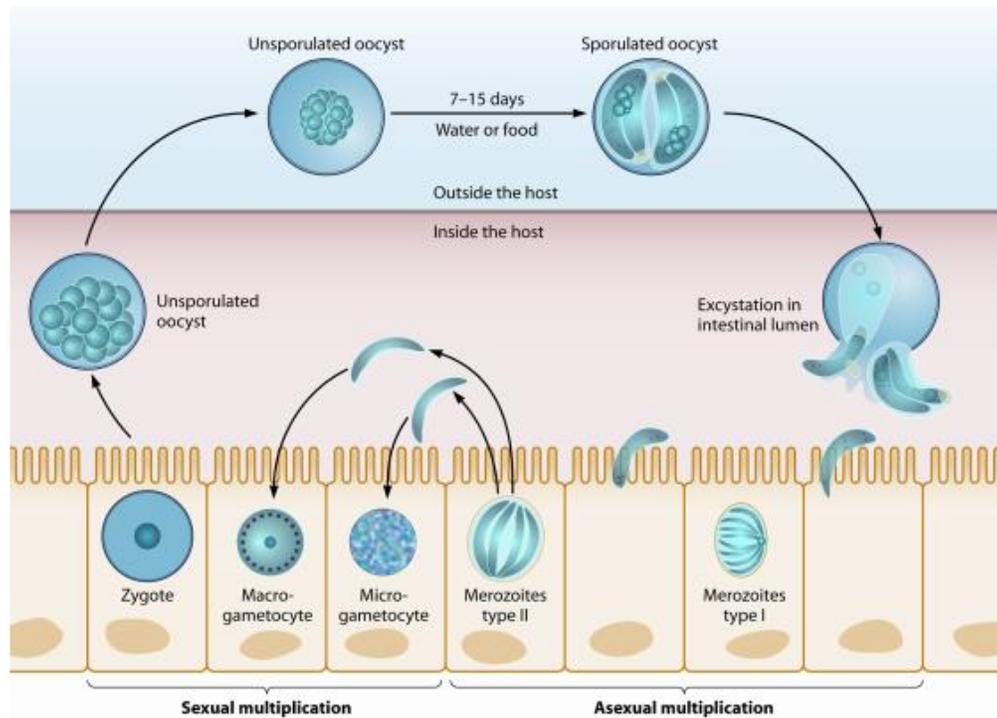


**Figura 1.** Morfología de los ooquistes de *Cyclospora cayetanensis*. De izquierda a derecha se observan diferentes estadios morfológicos de *C. cayetanensis*. El ooquiste no esporulado (izquierda), se observa con citoplasma indiferenciado y un tamaño aproximado de 10  $\mu\text{m}$ ; mientras que en los otros paneles se observa (A) el ooquiste esporulado que contiene dos esporoquistes inmaduros, (B) el ooquiste liberando uno de sus dos esporoquistes y (C) un esporoquiste y dos esporozoítos libres (Tomado de CDC, 2020b).

### 2.2.3 Ciclo de Vida

*Cyclospora cayetanensis* es un parásito intracelular obligado del epitelio del tracto gastrointestinal, cuyo ciclo de vida no está completamente caracterizado (Figura 2). El ciclo parece ser monoxeno, requiriendo un único hospedero (en este caso el ser humano). La infección se inicia cuando la persona ingiere ooquistes esporulados del parásito. En presencia de la bilis, tripsina y otros factores

del tracto gastrointestinal, los ooquistes liberan los esporozoítos, los cuales invaden las células epiteliales del intestino delgado. Estos esporozoítos empiezan un ciclo de multiplicación asexual o merogonia, en la que se generan merozoítos (Tipo I y tipo II). Finalmente, cada merozoíto hijo entra en una célula epitelial y se diferencia en microgametocito masculino o macrogametocito femenino. Estas células sexuales se unen y forman un cigoto, que madura en un ooquiste de pared gruesa. Esta estructura rompe la célula hospedadora y se excreta en la materia fecal. Los ooquistes no son infectantes cuando son eliminados, estos requieren por lo menos una semana para esporular en el ambiente bajo condiciones favorables; como a una temperatura entre 22 a 30 °C. Durante la esporulación, su contenido (esporonte) se divide en dos esporoquistes, cada uno con dos esporozoítos en forma de banana (Ortega *et al.*, 1993; Dubey *et al.*, 2020). La presencia de dos esporozoítos en cada esporoquiste es el criterio fundamental del género *Cyclospora* (Ortega *et al.*, 1993).



**Figura 2.** Ciclo de vida de *Cyclospora cayetanensis*. Los ooquistes no esporulados se diferencian en ooquistes esporulados, que se someten al proceso de excistación. Los esporozoítos infectan las células del epitelio gastrointestinal para formar merozoítos de tipo I, y estos forman merozoítos de tipo II. El microgametocito de etapa sexual fertiliza al macrogametocito para convertirse en un cigoto y, por lo tanto, para diferenciarse como un ooquiste no esporulado (Tomado de Ortega y Sanchez, 2010).

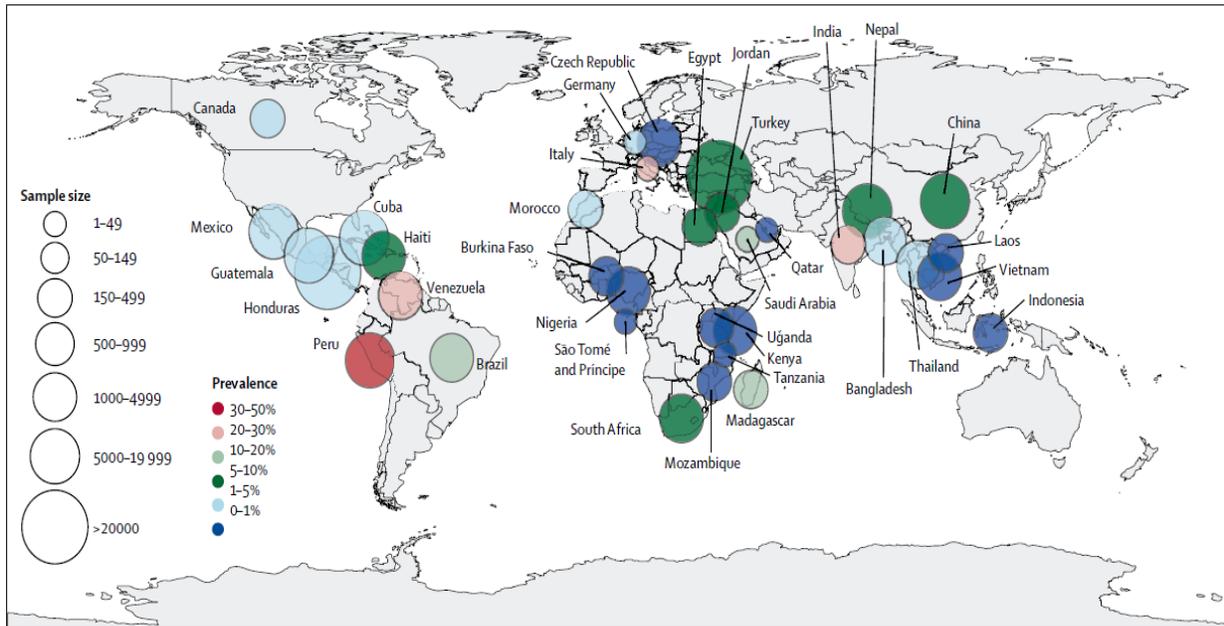
## 2.2.4 Epidemiología

*Cyclospora cayetanensis* se describió por primera vez aproximadamente en los años 90's y su importancia ha ido en aumento debido a que el protozooario ha sido vinculado con varios brotes por consumo de alimentos frescos en Estados Unidos de América (Ortega *et al.*, 1993; Shields y Olson, 2003; CDC, 2020a). A nivel mundial, se ha descrito la enfermedad tanto en países desarrollados como en vía de desarrollo, presentándose con mayor frecuencia y mayor prevalencia en las regiones tropicales y subtropicales (con tasas de infección mayor al 20 %, que pueden llegar hasta el 41%). Debido a esta tendencia, ciertos países han sido considerado endémicos. Entre los países que cuentan con esta característica están: Bangladesh, Brasil, Chile, Cuba, República Dominicana, Egipto, Guatemala, Haití, India, Indonesia, México, Nepal, Nigeria, Perú, Tanzania, Tailandia, Venezuela y Zimbabue (Shields y Olson, 2003; Ortega y Sherchand, 2015). Además, se han hecho reportes de personas extranjeras que al visitar estos países adquieren la infección; mientras que, en algunos países sin climas tropicales, como Estados Unidos de América, la aparición de brotes se ha adjudicado al consumo de alimentos procedentes de países tropicales (CDC, 2020a).

La prevalencia de *C. cayetanensis* varía alrededor del mundo, según los estudios reportados hasta la fecha (Figura 3). Por mencionar algunos, se han registrado prevalencias de 5.6% en China, 9.2% en Nepal, 17.4% en Turquía y hasta 22% en India. En los países africanos, las prevalencias más altas registradas hasta la fecha se encuentran en Egipto (10%) y Sudáfrica (7.2%). Por otra parte, en América Latina las prevalencias oscilan desde 7.9% en Haití, 10.8% en Brasil, 24.2% en Venezuela y hasta 41.6% en Perú (Giangaspero y Gasser, 2019).

En México, son pocos los estudios que se han realizado con relación a la prevalencia del parásito. En 2014, se publicó un estudio clínico realizado en el hospital pediátrico de Morelia, Michoacán en el que se examinaron 8,877 muestras de heces de niños. Las muestras fueron recolectadas desde el año 2000 al 2009 y de las muestras, sesenta niños (0.67%) resultaron positivos para la presencia de *C. cayetanensis*, de los cuales 90% de los niños eran inmunocomprometidos. Además, la mayoría de los casos sucedieron durante los meses de junio a agosto, meses que coinciden con la época de lluvia en Morelia (Orozco-Mosqueda *et al.*, 2014). Otro de estos estudios fue desarrollado en el estado de Jalisco, en 2019, donde se determinó que la prevalencia de 2.9% (3/104 muestras) de *C. cayetanensis* es baja en comparación con otros parásitos entéricos (Galván-Ramírez *et al.*,

2019). Sin embargo, las tasas de prevalencia se pueden subestimar debido a la baja sensibilidad de los métodos de detección convencionales, la falta de monitoreo de posibles fuentes de contaminación y carencia de reportes de casos por los sistemas de salud de los países en donde posiblemente son endémicos (Giangaspero y Gasser, 2019).



**Figura 3.** Distribución y prevalencia de *Cyclospora cayentanensis* en el mundo (tomado de Giangaspero y Gasser, 2019).

### 2.2.5 Ruta de Transmisión

La transmisión principal de los ooquistes de *C. cayentanensis* al ser humano ocurre por la vía fecal-oral. Aunque la transmisión de persona a persona es poco probable, la transmisión puede ocurrir de manera indirecta: a través de la exposición a agua o alimentos contaminados con heces conteniendo los ooquistes (Ortega *et al.*, 1998; Chacín-Bonilla, 2008). Dicha contaminación posiblemente se deba a la baja infraestructura de saneamiento en ciertas regiones, permitiendo que el parásito llegue a fuentes de agua, y estas faciliten la contaminación de alimentos al utilizarlas como agua para irrigar los cultivos en la práctica agrícola (Onstad *et al.*, 2019). Por otra parte, puede darse el caso

de uso de desechos de animales deficientemente tratados como fertilizantes y, aunado a ello, el ambiente o vehículo (alimento) puede poseer las características óptimas para la persistencia del parásito antes de la ingesta por el hospedero (Tram *et al.*, 2008). Asimismo, se ha reportado que *C. cayetanensis* es resistente a los tratamientos de saneamiento del agua como la cloración (Rabold *et al.*, 1994).

Las características biológicas y epidemiológicas de *C. cayetanensis* que fomentan la transmisión a los humanos aún se desconocen. En los países en desarrollo, donde se ha descrito la presencia de manera endémica del microorganismo, se han realizado pocos estudios para abordar las rutas de diseminación del parásito (Chacín-Bonilla, 2010).

2.2.5.1 Transmisión por Alimentos. En países desarrollados donde se han presentado brotes epidemiológicos, éstos han sido asociados a la ingesta de alimentos de consumo en fresco importados, tales como: frambuesas, zarzamoras, albahaca, cilantro, chícharos y lechugas. La mayoría de estos registros son compilados por el Centro de Control y Prevención de Enfermedades (CDC, por sus siglas en inglés) de Estados Unidos. Desde el 2013 hasta la fecha se observa que hay una incidencia anual de brotes por *C. cayetanensis* en el país antes mencionado (Cuadro 1). En algunas de las investigaciones realizadas por la CDC de estos casos han logrado rastrear y asociar el vehículo de transporte (alimento) del protozooario mediante su identificación; siendo de esta manera que rastrean el origen del alimento para vincular el origen de la cepa. En dichas investigaciones se ha llegado a vincular principalmente alimentos importados de México, como en los casos de los años 2013, 2014 y 2019 (CDC, 2020a).

En 2013 se reportaron 631 personas afectadas en 25 estados, siendo este un brote multiestatal. Las investigaciones realizadas en Iowa y Nebraska asociaron los casos de ciclosporiasis con la mezcla de ensalada empaquetada producida por Taylor Farms de México. Mientras que las investigaciones de trazabilidad realizadas en Texas indicaron que el vehículo alimentario fue cilantro fresco proveniente del estado de Puebla, México. Para el primer caso, las investigaciones se dirigieron a una evaluación ambiental de Taylor Farms en Guanajuato, México, en donde participaron tanto la FDA (Food and Drugs Administration, por sus siglas en inglés) y la CDC de Estados Unidos, como otras organizaciones mexicanas tales como la COFEPRIS (Comisión Federal para la Protección Contra Riesgos Sanitarios), SENASICA (Servicio Nacional de Sanidad, Inocuidad y Calidad

Agroalimentaria), SAGARPA (Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación) y CESAPEG (Comité Estatal de Sanidad Vegetal de Guanajuato). Durante la evaluación se verificó que el lugar cumplía con las condiciones sanitarias y los empleados cumplían con las buenas prácticas de manipulación de los alimentos. Al mismo tiempo, se generó un extenso muestreo, incluyendo muestras de productos (tejido vegetal antes de la cosecha, procesado y producto terminado), muestras de agua (de proceso, de riego, ambientales), muestra de suelo, muestras de heces humanas de instalaciones sanitarias, entre otras. Finalmente, el patógeno del brote no se detectó en ninguna de estas muestras. En el segundo caso del mismo año, las investigaciones concluyeron que los enfermos consumieron cilantro: la información del rastreo indicó que durante el período en que los casos de ciclosporiasis iniciaron, cilantro fresco proveniente de Puebla, México, fue distribuido en restaurantes y supermercados. No obstante, no se obtuvo mucha más información ni se reportaron más investigaciones para identificar el patógeno en territorio mexicano. En consecuencia, se incrementaron los esfuerzos en la vigilancia de los productos de cilantro exportados desde México a Estados Unidos (CDC, 2020a).

Para el año 2014, los brotes de ciclosporiasis fueron multiestatales, pero hay que resaltar que al menos 133 personas eran residentes de Texas, de las cuales aproximadamente 57% consumió cilantro fresco días antes de presentar la enfermedad. Cuatro restaurantes diferentes, donde varias personas enfermas informaron haber comido, fueron investigados, y la información de rastreo reveló que Puebla, México, era una fuente de cilantro que se servía en los cuatro restaurantes (CDC, 2020a). Pero, similar al caso en Texas del 2013, no se reportó una investigación más exhaustiva o evaluación ambiental para la identificación de patógeno en México.

Por otro lado, en los brotes multiestatales de 2019, se reportó que la albahaca fresca exportada por Siga Logistics de RL de CV de Morelos, México, era una fuente probable para la infección. Tales alusiones se generaron a partir de las entrevistas a las personas enfermas en las que varios de ellos consumieron el alimento contaminado en los mismos restaurantes, esto sin que las personas estuvieran relacionadas. Asimismo, la información de la distribución del producto apoyo a considerar que esta era la fuente.

En resumen, los casos que se resaltan en este apartado comparten que no hay suficiente información o resultados contundentes para aseverar que las cepas de *C. cayetanensis* de estos brotes son procedentes de México.

**Cuadro 1.** Investigaciones de los brotes de *C. cayetanensis* por la CDC.

Lugar	Año	Alimento asociado	Origen del alimento
EU: Multiestatal	2013	Mezcla de ensalada empacada	Guanajuato, México
		Cilantro	Puebla, México
EU: Multiestatal	2014	Cilantro	Puebla, México
		Desconocido	Desconocido
EU: Multiestatal	2015	Desconocido	Desconocido
EU: Multiestatal	2016	Desconocido	Desconocido
EU: Multiestatal	2017	Desconocido	Desconocido
EU: Multiestatal	2018	Vegetales frescos Del Monte	Desconocido
EU: Multiestatal	2018	Ensalada mixta fresca vendida por McDonald's	Desconocido
EU: Multiestatal	2019	Albahaca	Morelos, México
EU: Multiestatal	2020	Mezcla de ensalada empacada	Desconocido.

EU: Estados Unidos.

2.2.5.2 Transmisión por Agua. En general, los brotes han sido relacionados principalmente a los alimentos, pero posiblemente el vehículo principal de contaminación de estos sea el agua (Marshall *et al.*, 1997). En este sentido, se han reportado casos en los que las fuentes de agua para beber están contaminadas con el protozoo, debido a que los tratamientos utilizados para su saneamiento, no fueron lo suficientemente eficaces para eliminarlo; por ejemplo, la desinfección por cloración. Tal es el caso del brote diarreico en soldados británicos en 1994 en Nepal. Las investigaciones revelaron que el tanque de almacenamiento de agua, la cual estaba clorada y filtrada, contenía ooquistes de *Cyclospora*. Esto no sólo sugiere que el protozoo es resistente a las concentraciones estándares de saneamiento, sino que, el uso de sistemas de filtración no siempre posee la capacidad de retención de partículas tan pequeñas como las de un ooquiste; por tanto, una filtración inadecuada del agua potable, puede provocar una infección directa (Rabold *et al.*, 1994). También se ha detectado tanto en agua para beber como en agua residual en Nepal y Perú (Sherchand *et al.*, 1999; Cross *et al.* 1997). Por otra parte, en otro estudio se ha demostrado que este parásito prevalece en aguas residuales y que la eficiencia de los procesos de tratamiento de estas aguas es limitada contra el mismo (Kitajima *et al.*, 2014). A estos reportes se suma que se ha logrado la identificación de *C. cayetanensis* tanto en agua tratada, agua de pozo y agua de uso para irrigación de cultivos (Giangaspero *et al.*, 2015). Estos hallazgos sugieren que *C. cayetanensis* tiene potencial de propagarse en diferentes cuerpos de agua y la ineficiencia en algunos sistemas de saneamiento del agua.

2.2.5.3 Transmisión por Vía Zoonótica. La participación de los animales como reservorios de *C. cayetanensis* aún no ha sido esclarecida. Existen estudios donde se han realizado esfuerzos para identificar reservorios animales del parásito; como el de Eberhard *et al.* (2000), quienes intentaron establecer la infección en una amplia variedad de modelos animales: perros, pollos, patos, conejos, puercos, borregos, cabras, hámsters, ratas, diferentes cepas de ratones, monos e inclusive monos inmunosuprimidos. Pese a los esfuerzos, no se logró establecer una infección exitosa y se ha sugerido que *C. cayetanensis* tiene una alta especificidad por su hospedero, el ser humano.

## 2.2.6 Identificación

Dentro de los métodos para identificación de *Cyclospora* se encuentran los tradicionales y los moleculares. El método de detección tradicional ha sido la microscopía, en donde las muestras son teñidas. En general, el método modificado de Ziehl-Neelsen (que tiñe los ooquistes de rosa a rojo) parece lograr la mejor sensibilidad, en comparación con las tinciones con safranina. Asimismo, se cuenta con la microscopía basada en fluorescencia, en donde los ooquistes son detectados por su capacidad de autofluorescencia utilizando un filtro de excitación de longitud de onda de 330–380 nm (fluorescencia azul) o de longitud de onda de 450–490 nm (fluorescencia verde). No obstante, la microscopía tiene como limitante que en ocasiones no son específicos y no permiten la identificación por especie (Orlandi *et al.*, 2004; Ortega y Sanchez, 2010; Qvarnstrom *et al.*, 2017). Por otro lado, los métodos moleculares permiten la detección e identificación específica de *C. cayetanensis*. La FDA cuenta con dos métodos en su Manual Analítico Bacteriológico (BAM, en inglés por Bacteriological Analytical Manual) para este fin. Los métodos se encuentran en el capítulo 19a y 19b del BAM. Estos consisten en una PCR (Polymerase Chain Reaction) anidada y en tiempo real, basados en la amplificación dirigida a la región codificante del ARN ribosomal 18S que permite identificar de forma específica a *C. cayetanensis* (FDA, 2004a; FDA, 2018b; Murphy *et al.*, 2018). Aunque este método fue desarrollado y validado usando muestras de alimentos, también puede ser usado para muestras de otras matrices.

En lo que respecta a los ensayos de la viabilidad, debido a la carencia de métodos de cultivo y modelo animal, el único indicador fácil de realizar en el laboratorio es la esporulación de los

ooquistes con dicromato de potasio al 2.5%. Después de 7 a 13 días a una temperatura de 25 °C a 35 °C, se puede visualizar la morfología característica de los ooquistes maduros (dos esporoquistes con dos esporozoítos) (Eberhard *et al.*, 1997; Shields y Olson, 2003). Asimismo, se puede realizar el desenquistamiento del ooquiste mediante la exposición a agentes como tripsina (0-5%) y taurocolato de sodio (1-5%) en solución salina tamponada con fosfato, seguida de una disrupción mecánica (Ortega *et al.*, 1993; Smith *et al.*, 1997).

### **2.2.7 Genoma Nuclear**

En 2015 fue reportado el primer genoma secuenciado de *C. cayetanensis* (Qvarnstrom *et al.*, 2015). Hasta la fecha, en el “National Center of Biotechnology Information” (NCBI) se cuenta con 37 ensamblajes en “draft genome” del genoma nuclear y tres reportes de anotaciones de organelos (dos del genoma mitocondrial y uno del genoma del apicoplasto). El genoma nuclear tiene una longitud promedio de 44.2483 Mb, un conteo promedio de 5793 proteínas y un contenido de G-C del 51.9%. El genoma de referencia hasta el momento es el *Cyclospora cayetanensis CcayRef3* publicado por el “Center for Food Safety and Applied Nutrition” (CFSAN) en 2018. La representación del genoma de referencia es completa, con un nivel de ensamblaje de contig; utilizaron el método de ensamblaje SPAdes v3.10.1 con la tecnología de secuenciación Illumina Miseq y obtuvieron una cobertura del genoma de 20.0x.

En general, la naturaleza del origen de las muestras de los 37 genomas publicados hasta el momento es a partir de heces de humanos de diferentes áreas geográficas (Nepal 6 genomas, China 2 genomas, Indonesia 3 genomas, EUA 20 genomas, Guatemala 4 genomas, México 1 genoma y un genoma sin especificar la fuente). Sin embargo, el origen geográfico real de varias de las cepas de estos genomas puede no concordar con el lugar de obtención de la muestra. Algunas investigaciones realizadas para los brotes en Estados Unidos han vinculado a diversos alimentos procedentes de otros países como la fuente de contaminación. Así mismo, algunos de estos ensamblajes son a partir de muestras de personas que visitaron otros países y se infectaron. En este sentido, podemos ubicar que de los genomas publicados, siete de ellos se encuentran relacionados a México; dos genomas de viajeros de Estados Unidos a México (aislados CDC:HCTX208:15 y

CDC:HCTX227:15), dos genomas de los brotes del 2013 al 2014 en Texas, los cuales fueron asociados a consumo de cilantro proveniente del estado de Puebla (aislados CDC:HCTX204:15, CDC:HCTX205:15, CDC:HCTX69:14 y CDC:HCTX48:14 ) y el genoma que es referido a México (el ensamblaje del aislado CDC HCMX010\_16) (Qvarnstrom *et al.*, 2018).

Respecto a su contenido genético, poco se sabe aún de los genomas de *C. cayetanensis*; sólo dos de los 37 ensamblajes publicados en NCBI cuentan con anotación del genoma y la mayoría de los genes están no caracterizadas o como hipotéticos.

2.2.7.1 Genotipificación del Genoma Nuclear de *Cyclospora cayetanensis*. Para llevar a cabo estudios epidemiológicos y poder vincular el origen de la cepa, así como dilucidar su posible ruta de contaminación, es necesaria la caracterización genética de aislados de *C. cayetanensis*. Hasta el momento, son contados los estudios del genotipo de *Cyclospora cayetanensis*. En un inicio, en los intentos de búsqueda de marcadores moleculares para la tipificación de *Cyclospora*, se usaron los genes más utilizados como marcadores en otros microorganismos. En las bases de datos de GenBank, la mayoría de las secuencias de *Cyclospora* hasta el 2014 eran genes como del RNAr (SSU rRNA, 5.8S rRNA), la proteína de choque térmico de 70 kDA (HSP70) y del ITS-1 e ITS-2 (Adam *et al.*, 2001; Riner *et al.*, 2010; Sulaiman *et al.*, 2013; Ortega *et al.*, 2014). En general, todas estas secuencias han permitido la detección de *Cyclospora*, siendo el gen del ARNr 18S el más utilizado de estos marcadores, ya que ha permitido la diferenciación de *C. cayetanensis* de las otras especies. No obstante, estos marcadores se ven limitados sólo a esta aplicación debido a la baja variabilidad genética que contienen en sus secuencias como para discriminar si existen subtipos de *C. cayetanensis* (Zhou *et al.*, 2011; Sulaiman *et al.*, 2013, 2014). Por tanto, estos marcadores por sí solos no fueron útiles para la vinculación de casos e investigaciones de trazabilidad durante la ocurrencia de brotes.

A partir de la secuenciación del genoma de *Cyclospora cayetanensis*, nuevas perspectivas se desarrollaron para su tipificación. Qvarnstrom *et al.*, (2015) reportan la primera secuencia genómica en “draft genome”, lo cual se refiere a que es una secuencia de ADN genómico en “borrador”, que aún no es totalmente precisa. En general, para este genoma obtuvieron un total de 147,760,794 reads (lecturas), las cuales fueron ensambladas *de novo* a contigs (lecturas superpuestas para representar una región consenso de ADN). En sí un ensamblaje *de novo* es un

alineamiento y mezcla de múltiples fragmentos de una secuencia de ADN mucho mayor para reconstruir la secuencia original, en este caso, al no contar con una secuencia de referencia, la secuencia se hace *de novo*. El resultado fue de 865 contigs en total (incluyendo un contig que poseía el genoma de la mitocondria y dos contigs con el genoma del apicoplasto) con una longitud total de 44,563,857 bases.

Guo *et al.*, (2016) secuenciaron el aislado CHEN\_HEN01 de Henan, China para la búsqueda de regiones microsatélite y minisatélite en los contigs ensamblados utilizando un software de búsqueda de secuencias repetidas en tándem. Cinco loci microsatélites fueron elegidos como marcadores para tipificación MLST (Multilocus Sequence Typing): CYC3, CYC13, CYC15, CYC21 y CYC22. La resolución de estos marcadores permitió ubicar 25 genotipos MLST en 34 secuencias de 64 de diferentes regiones como EUA, Perú, Nepal, China y España. En su caso, su desventaja fue que no lograron obtener secuencias informativas de todos los amplicones en su evaluación de los marcadores. Ellos lo atribuyeron a la presencia de productos de PCR de diferente longitud. Pero, a pesar de ello, observaron la formación de agrupaciones geográficas en algunos loci, como en el caso de las secuencias de China que se agrupaban en 1 o 2 grupos. Con esto se resalta que el uso de un solo marcador no llega al alcance de lograr una trazabilidad del origen geográfico; en cambio, el uso de dos o más marcadores polimórficos parece ser más prometedor. Posteriormente, Li *et al.*, (2017), continuó con un estudio similar en donde ellos determinaron otros 29 genotipos MLST a partir de 45 muestras, utilizando los mismos cinco marcadores publicados por Guo *et al.*, (2016). El análisis fue realizado concatenando las cinco secuencias de los marcadores. Con estos nuevos genotipos se sumaron 54 genotipos MLST en total y obtienen resultados similares al observar la formación de agrupaciones según el origen geográfico.

Hofstetter *et al.*, (2019) realizan esfuerzos por mejorar la técnica de tipificación MLST reportada por Guo *et al.*, (2016) donde omiten el paso de la PCR anidada y se enfocan en el uso del loci CYC21 y CYC22. Sin embargo, no lograron diferenciar los aislados no relacionados, por lo que concluyen que es difícil mejorar la técnica debido a las características de los marcadores utilizados junto con la ocurrencia de genotipos mixtos en las infecciones por *Cyclospora*.

La búsqueda de nuevos marcadores continuó con Barratt *et al.*, (2019). La complejidad de las infecciones por *C. cayetanensis* por su heterogeneidad ha establecido que el genotipado sea todo un reto. En este caso, seleccionaron tres marcadores con SNP de alta confianza (aquellos SNP dentro de las regiones genómicas de mayor cobertura) a partir de 11 genomas secuenciados a partir

de muestras de heces: dos loci nucleares, Nu indefinida (360i2) y Nu familia de proteína Sec14 (378); y un locus mitocondrial, Mt rRNA (MSR). Una vez determinados los marcadores, realizaron una examinación de los “draft genomes” mediante BLAST y “discontiguous” BLAST utilizando como query las secuencias de los loci MLST. Esto para descartar la existencia de parálogos y confirmar que al encontrar más de una copia de estos loci sería por diferentes haplotipos; es decir, una heterocigidad real. En este caso, hacen una descripción más detallada del flujo de trabajo que siguieron para la selección de dichos marcadores.

Sin embargo, no se ha logrado determinar uno de estos métodos como rutinario para el genotipado de *C. cayetanensis* dada su limitada capacidad para resolver los brotes epidemiológicos y dilucidar la naturaleza de parásito. Por consiguiente, la investigación y análisis de nuevos marcadores moleculares adicionales parece ser la corriente más acertada para la genotipificación del parásito.

Uno de los estudios más recientes, abordó el desarrollo de un flujo de trabajo que permite la identificación de nuevos marcadores genotípicos nucleares a partir de los genomas secuenciados ya reportados. El flujo de trabajo o pipeline, consiste en un control de calidad de las secuencias ya reportadas desde sus lecturas crudas (FASTQC v0.11.7, AdaptorRemoval v2.2.2, BLASTN, SPAdes v3.11.1), la anotación de los genomas ensamblados (GeneMark-EP+, Blast2Go) y un llamado de SNPs utilizando Mauve y STAR. Con su flujo de trabajo lograron detectar 485 genes candidatos y llevan este estudio hasta la evaluación de cuatro marcadores prometedores. Dicha evaluación constó de amplificación de los cuatro marcadores por PCR y su capacidad de resolución de los aislados en genotipos. Lograron amplificar exitosamente los cuatro marcadores en 57 de 84 aislados de heces de casos en Estados Unidos, colectados entre los años de 2013 a 2017. Estos marcadores fueron capaces de discriminar estos 57 aislados en 19 genotipos (Houghton *et al.*, 2020). Este enfoque puede ser útil para la búsqueda de nuevos marcadores.

Por lo anteriormente mencionado, se concluye que es necesario identificar y estudiar más marcadores para seleccionar aquellos que puedan mejorar estos métodos de genotipificación, así como implementar diferentes flujos de trabajo que permitan ir acercándose a cumplir con este objetivo.

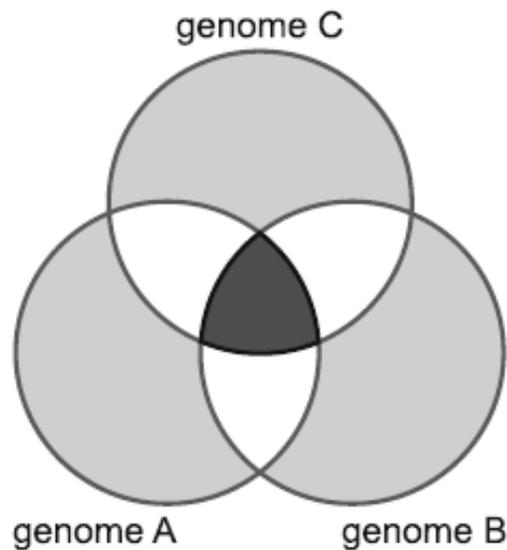
## 2.3 Genómica Comparativa

Con el advenimiento de nuevas tecnologías de secuenciación como las NGS (Next Generation Sequencing), hubo un rápido incremento en el número de secuencias genómicas disponibles en bases de datos, tales como GenBank (Qvarnstrom *et al.*, 2015; Qvarnstrom, *et al.*, 2018; Cinar *et al.*, 2020). En su mayoría, las secuencias depositadas corresponden a microorganismos bacterianos como *Escherichia coli*, las cuales llegan a poseer un gran número de genomas completos (NCBI, 2020). Mientras que organismos de un orden superior como los protozoarios, se observa que hay un limitado número de genomas disponible, como en el caso de *Cyclospora*, los cuales además se encuentran no terminados o en “draft genome” (NCBI, 2020). No obstante, con la generación masiva de secuencias genómicas, se impulsó el análisis de secuencias múltiples y con ello la genómica comparativa. La genómica comparativa permite el estudio de la variabilidad genética entre varios objetos de estudio, ya sean de diferente género o de la misma especie, comparando su repertorio genético (Guimarães *et al.*, 2015). Un enfoque concurrido es el análisis de pangenoma, que se basa en la identificación de genes ortólogos; estos genes se derivan de un ancestro en común y divergen a partir de un evento de especiación, pero conservan su función entre organismos (Contreras-Moreira y Vinuesa, 2013). Por esta razón, este tipo de estudios proporcionan información valiosa para entender la función y procesos evolutivos que intervienen sobre los genomas, e inferir sus estilos de vida, repertorio de genes y tamaño mínimo del genoma (Tettelin *et al.*, 2005).

### 2.3.1 Análisis de Pangenoma

Los análisis de pangenoma se han desarrollado para el estudio de múltiples genomas de manera comparativa estableciendo nuevos conceptos entorno a su diversidad genética (Tettelin *et al.*, 2005; Tettelin *et al.*, 2008). Estos estudios se han realizado principalmente para el estudio de organismos procariontes. Sin embargo, a pesar del grado de dificultad de análisis de secuencias eucariotas, poco a poco se han establecido algoritmos que permiten su estudio (Contreras-Moreira *et al.*, 2017).

El término pangenoma (Figura 4) se refiere al contenido genético total de una especie. Este se conforma por tres subconjuntos principales: genoma core, el genoma accesorio y los genes únicos (Tettelin *et al.*, 2005; Rouli *et al.*, 2015). El genoma core se refiere al conjunto de genes compartidos en todos los organismos que se estén comparando, que en general son genes esenciales para la biología básica de cualquier organismo. Por otra parte, el genoma accesorio se refiere a aquellos genes prescindibles o parcialmente compartidos; es decir, puede estar presente en dos o más cepas, pero no en todas. Y, por último, los genes únicos que se encuentran presentes solo en un aislado, y son genes que contribuyen a la diversidad genética del organismo. Tanto genoma accesorio como los genes únicos, probablemente proporciona funciones que no son esenciales, sino más bien, funciones específicas que están relacionadas con la supervivencia, adaptabilidad, resistencia e invasión (Tettelin *et al.*, 2005; Tettelin *et al.*, 2008; Rouli *et al.*, 2015; McInerney *et al.*, 2017).



**Figura 4.** Pangenoma de tres genomas denominados A, B y C. En el diagrama el genoma core se muestra como la parte negra; el genoma prescindible es la intersección entre los genomas coloreado en blanco; y la sección coloreada en gris representa el genoma único (Tomado de Zekic *et al.*, 2018).

Dado el limitado número de secuencias de algunas especies, ha surgido el particular interés en identificar el número de genomas secuenciados necesarios para lograr caracterizar una especie.

Esto se puede lograr realizando estimaciones del tamaño del pangenoma, con lo que se puede inferir si este es abierto o cerrado. En el pangenoma cerrado de una especie, se visualiza que en la suma de genes se agregan muy pocos de estos por genoma secuenciado y gráficamente se observa que se llega a estabilizar su contenido genético en una asíntota; por tanto, el tamaño del pangenoma puede predecirse teóricamente y caracterizarse el contenido genético de manera representativa al contar con un número suficiente de genomas. En caso contrario, en las especies con un pangenoma abierto, la acumulación de genes de la especie tras la secuenciación de una nueva cepa va en un aumento tal, que el repertorio de genes del pangenoma se amplía considerablemente, y en consecuencia es imposible predecir el tamaño que este tiene (Tettelin *et al.*, 2005, 2008). En otras palabras, es favorable llegar a establecer que un pangenoma es cerrado para poseer una visión del contenido genómico real del objeto de estudio. Además, se ha sugerido que la naturaleza del pangenoma está ligada al estilo de vida de las especies bacterianas. Las especies que suelen vivir en nichos aislados tienen un pangenoma cerrado y un genoma pequeño; mientras que especies con pangenoma abierto, tienden a tener genomas más grandes, vivir en comunidad y tener altas tasas de transferencia horizontal de genes (Rouli *et al.*, 2015).

Este tipo de análisis ya se ha realizado en esfuerzos por determinar el contenido genético de *C. cayetanensis*. En 2016, se generó un estudio de genómica comparativa del ensamblaje GCA\_000769155.2 de la cepa China CHN\_HEN01 con otros microorganismos Apicomplexa (*Cryptosporidium parvum*, *Toxoplasma gondii*, *Plasmodium falciparum* y *Eimeria tenella*), en donde generaron una estimación del proteoma de *C. cayetanensis* para detectar ortólogos de proteínas. Este estudio determinó que *C. cayetanensis* comparte rutas metabólicas clásicas de los coccidios; pero presenta la ruta de síntesis de manitol a partir de fructosa de manera única, junto con *E. tenella*. Además, probablemente posee un sistema de invasión celular (anclaje por motores de actina, interacción con el hospedero por antígenos de superficie, formación de una unión móvil con una serie de proteínas secretadas para iniciar la invasión, etc.) semejante al de *Eimeria* spp. y *T. gondii*. Esto último lo discuten, ya que se detectaron cuatro genes putativos codificantes para antígenos de superficie tipo-TA4 como los de *Eimeria* spp. y una secuencia paróloga con alta similaridad a la proteína 1 de la membrana del eritrocito de *P. falciparum*. Adicionalmente, se encontraron ortólogos de proteínas para el antígeno apical de membrana 1 (AMA1) y proteínas RON, las cuales participan en la invasión celular en otros microorganismos como *T. gondii* (Liu *et al.*, 2016).

Sin embargo, se observó que comparte un número reducido de dominios de antígenos de superficie con otros coccidios. Por tanto, se asume que *C. cayetanensis* podría poseer antígenos de superficie únicos, entre otras proteínas, que lo hacen ser hospedero específico (Liu *et al.*, 2016).

Dicho lo anterior, por los pocos estudios que se han generado respecto al genoma nuclear de *C. cayetanensis*, la necesidad de adicionar nuevos marcadores moleculares y a que se cuenta con algunas secuencias del genoma nuclear en bases de datos, esto da pauta al estudio del pangenoma de *C. cayetanensis* con otros microorganismos del mismo subfilum, así como con la misma especie.

### 2.3.2 Filogenómica

La filogenómica es definida como la reconstrucción de la historia evolutiva de organismos tomando en cuenta regiones representativas de todo el genoma, permitiendo su análisis. El genoma core suele utilizarse como input para este tipo de análisis, proporcionando la base genómica de la filogenia. A partir de ello, podemos encontrar el sentido del cómo y por qué las similitudes o diferencias entre genomas, o hasta entender las características del metabolismo, la fisiología o patogenicidad. Igualmente, pueden contribuir a resolver las relaciones que poseen organismos entre y dentro de la misma especie (Eisen y Fraser, 2003).

En el estudio antes mencionado de Liu *et al.*, (2016) se generó un análisis de pangenoma utilizando el proteoma de *C. cayetanensis*. En sus resultados muestran que obtienen un genoma core de 1,020 proteínas, que están compartidas entre los microorganismos como *Cryptosporidium parvum*, *Toxoplasma gondii*, *Plasmodium falciparum*, *Eimeria tenella* y *C. cayetanensis*. A partir de este resultado, generaron un análisis de filogenia basado en neighbor-joining, concatenando las secuencias de proteínas de 100 de estos ortólogos. En su estimación de árbol filogenético confirman la estrecha relación que tiene *C. cayetanensis* con *E. tenella*.

Cabe mencionar, que se han realizado otros análisis de las relaciones filogenéticas de *C. cayetanensis* en intentos de apoyar a vincular los casos de ciclosporiasis. En los estudios mencionados en el apartado de genotipificación de *C. cayetanensis*, se generaron árboles a partir de concatenar los marcadores evaluados: los marcadores MLST basado en microsatélite (Guo *et al.*, 2016b; Li *et al.*, 2017) y los marcadores basados en SNP de alta confianza (Barratt *et*

*al.*, 2019; Houghton *et al.*, 2020). En el caso de los marcadores MLST, en los árboles filogenéticos obtenidos se visualizaba una tendencia a formar agrupaciones dependiendo del origen geográfico, particularmente siendo más notorio para las cepas de China en el estudio de Li *et al.* (2017). Por otro lado, en el enfoque de marcadores de SNP de alta confianza, dividen los grupos por genotipos y generan dendrogramas para representar las agrupaciones; pero, en estos estudios, utilizan muestras de casos de los brotes en Estados Unidos. La concordancia de agrupamiento genético con datos epidemiológicos es más detallada en el caso del estudio generado por Barrat *et al.*, (2019) donde mencionan que resolvieron cuatro de ocho grupos de brotes relacionados epidemiológicamente.

En el caso del presente proyecto, visualizar las relaciones filogenéticas basadas en nuevos marcadores permitirá estimar si estos tienden a formar agrupaciones según los datos de origen y discriminar si estos marcadores pueden ser potencialmente utilizados para resolver investigaciones epidemiológicas.

### **2.3.3 Herramientas de Análisis**

El flujo de trabajo habitual de las herramientas para análisis de pangenoma se puede mencionar en grandes rasgos en los siguientes pasos: 1) asignación de genes ortólogos, 2) identificación de pangenoma y 3) los análisis con los que se desee continuar, por ejemplo, estimación de árboles filogenéticos (Zekic *et al.*, 2018).

El pangenoma de un grupo de genomas puede ser determinado basado en el contenido de genes. Para ello, primeramente, se deben tener los genomas anotados o generar una predicción de los genes como preprocesamiento. Para este último objetivo, existen diferentes herramientas de predicción, como Prodigal, Glimmer o AUGUSTUS. Posteriormente, se puede continuar con la asignación de genes ortólogos (Zekic *et al.*, 2018).

Una herramienta que se ha desarrollado recientemente para agrupar genes/transcritos de organismos eucariotas, es el software GET\_HOMOLOGUES-EST (Contreras-Moreira *et al.*, 2017). Este software está construido con base a una versión previa llamada GET\_HOMOLOGUES (Contreras-Moreira y Vinuesa, 2013), la cual es un paquete de software abierto que se basa en tres

algoritmos para generar una base de datos de ortología para el análisis de pangenomas microbianos de manera accesible. Utiliza tres métodos o algoritmos para la agrupación de genes homólogos: OrthoMCL, COGtriangles y BDHM (“Bidirectional Best Hits”). Los criterios para generar los grupos de genes homólogos resultantes de estos métodos pueden ajustarse con el control de los parámetros de corrida. A partir de la construcción de grupos de homólogos, se pueden generar matrices pangenómicas y luego utilizarlas para crear árboles filogenéticos. Estimar el tamaño del genoma core y pangenoma es una opción que se puede computar bajo el modelo exponencial o binomial. Además, se puede realizar genómica comparativa básica para identificar genes específicos de linaje o expansiones de familias de genes (Contreras-Moreira y Vinuesa, 2013; Vinuesa y Contreras-Moreira, 2015). Sin embargo, el desarrollo de la bifurcación de GET\_HOMOLOGUES, GET\_HOMOLOGUES-EST, fue para adaptarlo al análisis de datos de gran tamaño y agregar nuevas características para manejar adecuadamente secuencias de transcripción fragmentadas o redundantes (Contreras-Moreira *et al.*, 2017).

Al obtener archivos gemelos de ADN (.fna) y CDS de péptidos (.faa), la salida de GET\_HOMOLOGUES-EST se puede utilizar para conducir los análisis de filogenómica y genética de poblaciones con el pipeline de GET\_PHYLOMARKERS. Este último es un paquete de software diseñado para seleccionar marcadores óptimos para estimar un árbol de especies por máxima verosimilitud (ML). Los marcadores se identifican a través de una serie de filtros secuenciales (detección de recombinantes, detección de árboles atípicos y detección de señal filogenética) que funcionan en grupos de genes/ transcripciones/ proteínas ortólogos, dando la posibilidad al usuario de controlar la robustez de estos filtros (Vinuesa *et al.*, 2018).

En el presente proyecto se plantea abordar un análisis pangenómico de las secuencias disponibles de *Cyclospora cayetanensis* utilizando como herramientas principales los softwares de GET\_HOMOLOGUES/GET\_HOMOLOGUES-EST y GET\_PHYLOMARKERS, ya que estas brindan la posibilidad de explorar en la búsqueda de nuevos marcadores en secuencias genómicas de considerable tamaño, dado que el microorganismo de estudio es de naturaleza eucariota y la necesidad del momento es ubicar nuevos marcadores que permitan complementar y mejorar la resolución en la tipificación del parásito.

### 3. HIPÓTESIS

1. El conjunto de marcadores nucleares de *Cyclospora cayetanensis* determinados con los pipelines de GET\_HOMOLOGUES/GET\_HOMOLOGUES-EST y GET\_PHYLOMARKERS permite establecer agrupaciones de acuerdo al origen geográfico en el árbol filogenético.
2. *Cyclospora cayetanensis* posee genes únicos para la especie respecto a otros protozoarios relacionados.
3. El aislado de *Cyclospora cayetanensis* relacionado a México tiene al menos un gen único respecto a los aislados de otras regiones geográficas.

## 4. OBJETIVOS

### 4.1 Objetivo General

Analizar la diversidad genética del genoma nuclear de *Cyclospora cayetanensis* de secuencias de diferentes regiones geográficas reportadas en bases de datos para la identificación de marcadores moleculares candidatos para subtipificación.

### 4.2 Objetivos Específicos

1. Obtener las secuencias del genoma nuclear de *Cyclospora cayetanensis* reportados en bases de datos.
2. Determinar los marcadores óptimos entre microorganismos Apicomplexa y analizar las relaciones filogenéticas de las secuencias de *Cyclospora cayetanensis* con otros Apicomplexa.
3. Determinar los marcadores óptimos de *Cyclospora cayetanensis* y analizar las relaciones filogenéticas entre los aislados de la misma especie.
4. Analizar el pangenoma de *Cyclospora cayetanensis* comparado con otros Apicomplexa.
5. Analizar el pangenoma de *Cyclospora cayetanensis* con aislados de la misma especie para identificar genes presentes diferencialmente en las secuencias reportadas como mexicanas.

## 5. MATERIALES Y MÉTODOS

### 5.1 Bases de Datos con los Genomas de *Cyclospora cayetanensis* Reportados en el NCBI

Para los análisis de genómica comparativa se generaron dos bases de datos de genomas: la primera de Apicomplexa y la segunda de *Cyclospora cayetanensis* (Figura 5). La base de datos de microorganismos Apicomplexa constó de genomas de los géneros *Babesia*, *Cryptosporidium*, *Eimeria*, *Plasmodium*, *Toxoplasma* y *Cyclospora*. Por lo que se refiera a la base de datos de *C. cayetanensis*, constó de todas las secuencias del genoma del patógeno que se encuentran reportadas hasta el momento de diferentes regiones geográficas. El propósito de generar dos bases de datos es para explorar características de *C. cayetanensis* tanto de manera interespecífica como intraespecífica dado que existe poca información genética del parásito; aunque el enfoque principal va hacia ubicar marcadores nucleares que permitan la diferenciación intraespecífica. En este sentido, la primera base de datos se generó para observar cuales son los marcadores óptimos determinados con los softwares de GET\_HOMOLOGUES/-EST y GET\_PHYLOMARKERS, ya que en la identificación del parásito se ha utilizado principalmente como marcador el ARN ribosomal 18S; y, con este mismo, se ha observado la estrecha relación del género de *Eimeria* con *Cyclospora*. Por otro lado, también para analizar el pangenoma de estos microorganismos dirigido a identificar los genes únicos que posee el género de *Cyclospora*, puesto que se ha hipotetizado que el género podría poseer antígenos de proteínas únicos, por ejemplo. Por lo que se refiere a la base de datos de *C. cayetanensis*, como se mencionó antes, fue para determinar cuáles genes pueden ser utilizados como marcadores para la subtipificación del parásito.

Las secuencias genómicas fueron obtenidas de las bases de datos del NCBI para genomas (<ftp://ftp.ncbi.nlm.nih.gov/genomes/>) de especies del subfilum Apicomplexa y de la especie de *C. cayetanensis*. Los criterios de selección de las secuencias de la base de datos de Apicomplexa fueron: que se tuviera una representación completa del genoma, que fueran del subfilum Apicomplexa y las secuencias se encontrarán anotadas. Para el caso de las secuencias de *C. cayetanensis*, la selección fue realizada con base a su identidad y cobertura generando Heatmaps con el programa ANI, seleccionando aquellas con un porcentaje mayor al 96% y una cobertura

similar. También se realizó una predicción de genes para aquellas secuencias no anotadas utilizando el programa AUGUSTUS (Hoff y Stanke, 2013), ya que sólo dos de los genomas reportados en NCBI para la especie se encuentran anotados. Una de las dos secuencias genómicas anotadas, es la secuencia de referencia de *C. cayetanensis*, la cual permitió entrenar al programa para obtener una mejor predicción de los genes de las secuencias no anotadas.

## 5.2 Análisis Bioinformáticos de las Secuencias Genómicas de *Cyclospora cayetanensis*

### 5.2.1 Determinación del Genoma Core y Filogenia

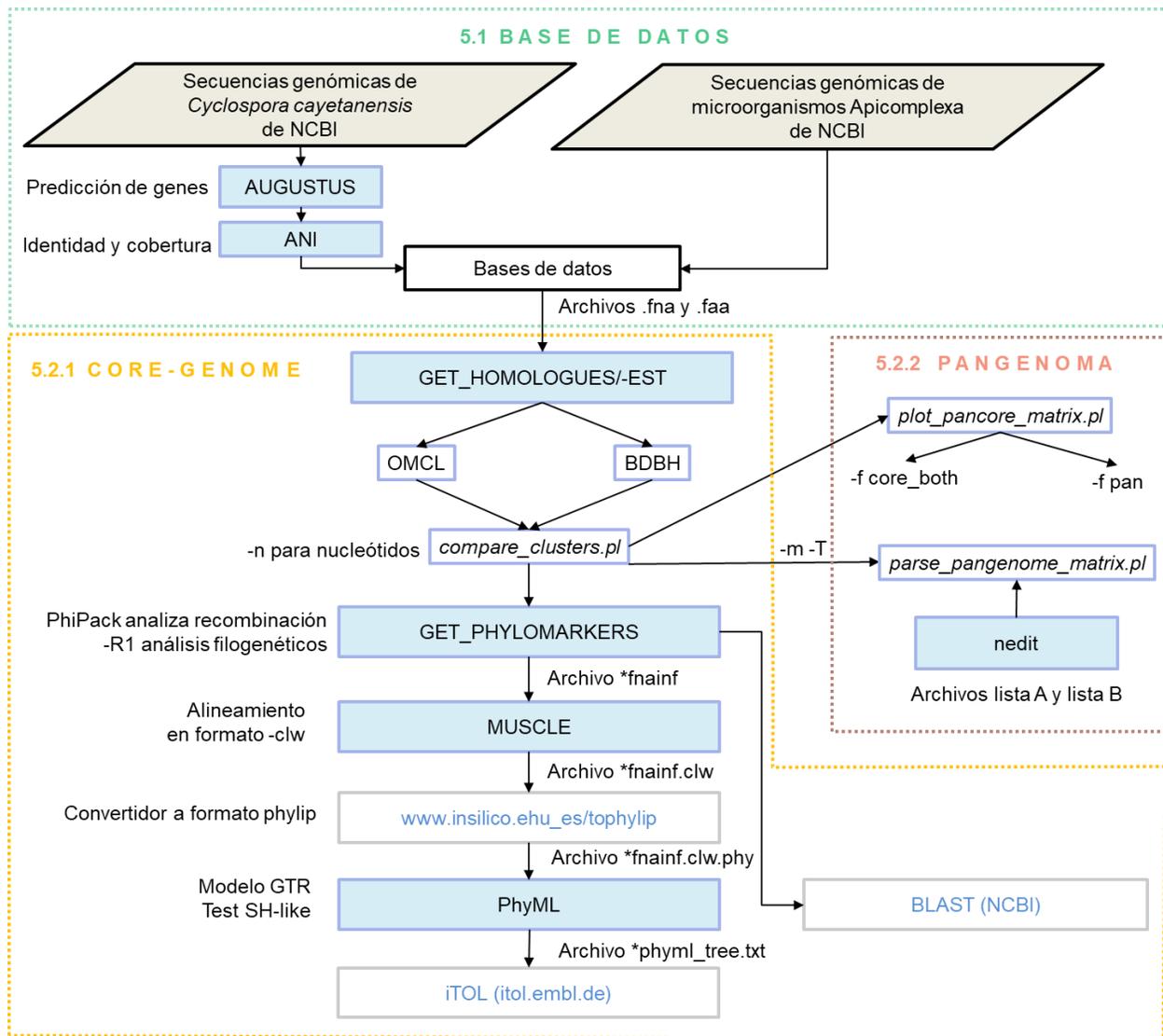
Para determinar el genoma core (Figura 5), los archivos gemelos de ADN (.fna) y proteínas (.faa) de los genomas, en ambas bases de datos, fueron sometidos a la selección de genes homólogos por razones de similitud de secuencia con el software GET\_HOMOLOGUES (Contreras-Moreira y Vinuesa, 2013) y su nueva versión, GET\_HOMOLOGUES-EST (Contreras-Moreira *et al.*, 2017), utilizando dos métodos en la corrida: OMCL (OrthoMCL) y BDBH (Bidirectional Best Hits). Seguido, se realizó una selección de los genes unicopia e intersección de los clusters con el script *compare\_clusters.pl* usando como input los directorios con los clusters (uno por cada método) generados como output con GET\_HOMOLOGUES/GET\_HOMOLOGUES-EST, indicándole el número de secuencias a analizar y el formato de input (dos corridas `-n` para el formato de nucleótidos y sin `-n` para el formato de proteínas). Posteriormente se estableció la filogenia mediante una serie de pasos: se examinó la recombinación, la topología de árbol y la señal filogenética de los genes del genoma core para analizar cuáles de estos tienen potencial como biomarcador. Esto último fue realizado con el software de GET\_PHYLOMARKERS (Vinuesa *et al.*, 2018), que se corre dentro del directorio con la intersección de los clusters de los métodos BDBH y OMCL con el modo de ejecución `-R1` para filogenética. Este software cuenta con PhiPack para analizar la recombinación y algoritmos de búsqueda rápida de árboles filogenéticos para estimar la filogenia (kdetrees y FastTree). Al finalizar la corrida, se generó un directorio con los genes clasificados como mejores marcadores, los cuales aparecen con nombre si los genomas están

anotados o, de lo contrario, con un número de identificación; por tanto, para el segundo caso, se determinó la identidad de los marcadores realizando BLASTn y BLASTp en NCBI. Para generar el árbol filogenético se tomó el archivo con los marcadores concatenados (concat\_cdnAIns.fnainf) generado por GET\_PHYLOMARKERS y, posteriormente, se estimó la filogenia con PhyML v3.3.20190321. Para ello, primero el archivo \*.fnainf fue alineado con el programa MUSCLE v3.8.1551, indicando que el archivo se encontraba en formato CLUSTALW con el parámetro -clw (.clw); para después convertirlo a formato PHYLIP (.phy) en la página en línea “www.insilico.ehu.es/tophylip”. Una vez obtenido el archivo en formato PHYLIP, este fue utilizado como input en el programa PhyML v3.3.20190321 para generar un árbol filogenético por máxima verosimilitud. Como opciones de corrida, se eligió el modelo de sustitución el modelo GTR, y como test para el soporte de las ramas, SH-like. El árbol resultante (archivo phym\_l\_tree.txt) fue visualizado en la herramienta online iTOL v5 (Letunic y Bork, 2019).

### **5.2.2 Determinación del Pangenoma y Análisis de Genoma Accesorio**

Utilizando los mismos resultados de GET\_HOMOLOGUES/GET\_HOMOLOGUES-EST (Contreras-Moreira y Vinuesa, 2013; Contreras-Moreira *et al.*, 2017) mencionados en el apartado anterior, se obtuvieron las gráficas de tamaño tanto del genoma core como del pangenoma utilizando el script *plot\_pancore\_matrix.pl*. En el primer caso se eligió el parámetro -f core\_Tettelin utilizando como input el archivo del genoma core con terminación \*.tab, el cual se crea en el directorio de la corrida de GET\_HOMOLOGUES. En el segundo caso, para la gráfica de pangenoma se utiliza el parámetro -f pan utilizando como input el archivo \*.tab del pangenoma. Además, se generó un nuevo directorio para el análisis del genoma accesorio, en particular para la búsqueda de genes únicos según el linaje, con el script *compare\_clusters.pl* con los parámetros -m y -T. Este nuevo directorio contenía los genes del pangenoma, en donde se corrió el script *parse\_pangenome\_matrix.pl* para analizar conjuntos de pangenoma, a fin de encontrar genes presentes en un subconjunto A de cepas que están ausentes en el subconjunto B. Para indicarle al programa que debe comparar y determinar los genes únicos, se establecieron los subconjuntos A y B generando una lista para cada conjunto con los nombres de los archivos de los genomas que

comprendería cada uno utilizando el editor de texto nedit (Figura 5).



**Figura 5.** Diagrama de flujo general de los materiales y métodos.

## 6. RESULTADOS Y DISCUSIÓN

### 6.1 Bases de Datos con los Genomas de *Cyclospora cayetanensis* Reportados en el NCBI

Como parte de los resultados, se generaron dos bases de datos de genomas: la primera de Apicomplexa y la segunda de *Cyclospora cayetanensis*. Para la base de datos de Apicomplexa se descargaron 16 secuencias genómicas de NCBI de diferentes géneros: 2 de *Babesia*, 4 de *Cryptosporidium*, 5 de *Eimeria*, 2 de *Plasmodium*, 1 de *Toxoplasma* y las 2 de *Cyclospora cayetanensis* (Cuadro 2). Todas las secuencias genómicas descargadas contaban con anotación del genoma.

**Cuadro 2.** Base de datos de microorganismos Apicomplexa descargada de la base de datos del NCBI.

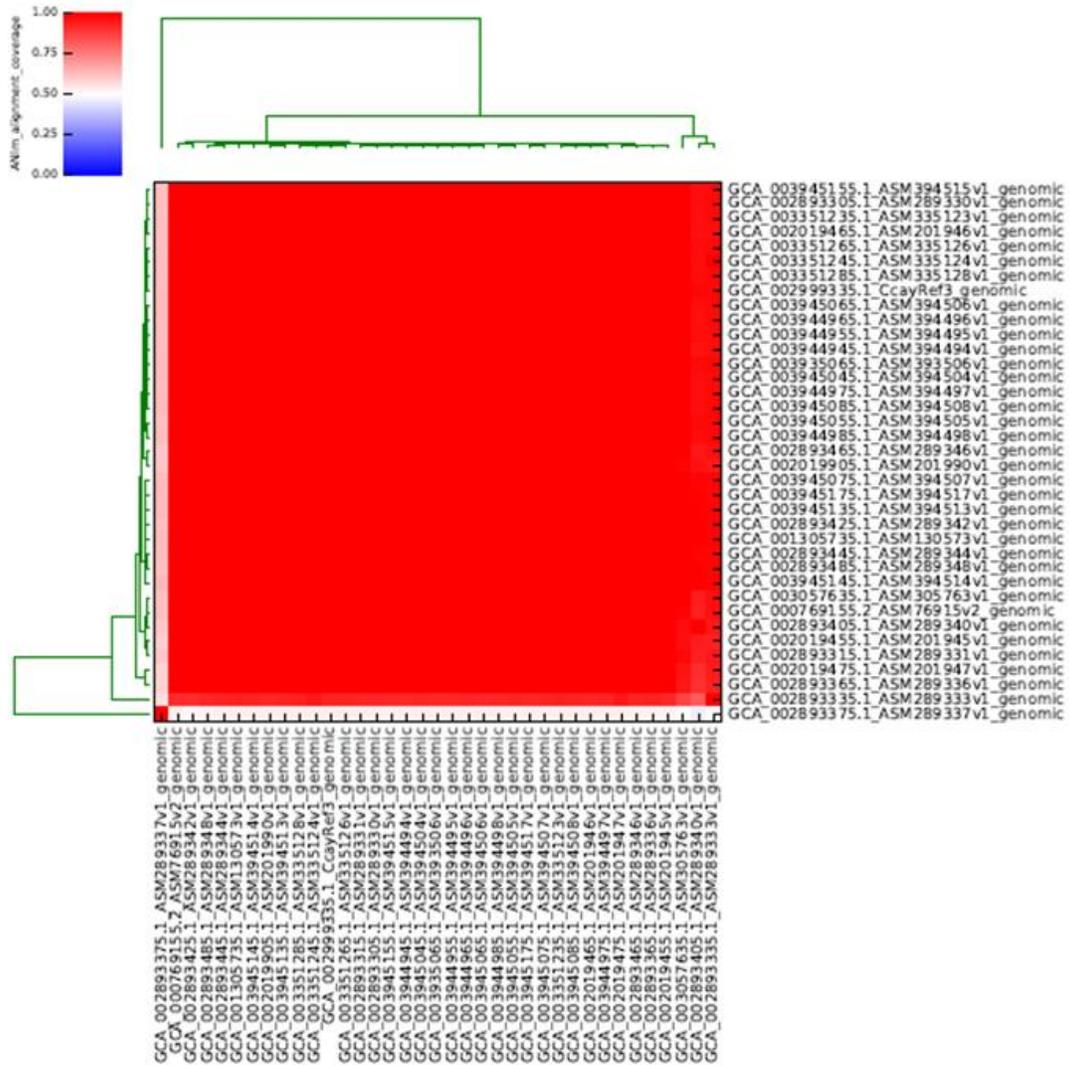
Base de datos Apicomplexa		
Número	Nombre en la base de datos	Ensamblaje en NCBI
1	<i>Cyclospora cayetanensis</i>	GCA_000769155.2
2	<i>Cyclospora cayetanensis</i>	GCA_002999335.1
3	<i>Babesia bovis</i>	GCA_000165395.1
4	<i>Babesia microti</i>	GCA_000691945.2
5	<i>Cryptosporidium hominis</i>	GCA_000006425.1
6	<i>Cryptosporidium muris</i>	GCA_000006515.1
7	<i>Cryptosporidium parvum</i>	GCA_000165345.1
8	<i>Cryptosporidium ubiquitum</i>	GCA_001865345.1
9	<i>Eimeria necatrix</i>	GCA_000499385.1
10	<i>Eimeria acervulina</i>	GCA_000499425.1
11	<i>Eimeria tenella</i>	GCA_000499545.2
12	<i>Eimeria maxima</i>	GCA_000499605.1
13	<i>Eimeria mitis</i>	GCA_000499745.1
14	<i>Plasmodium vivax</i>	GCA_000002415.2
15	<i>Plasmodium falciparum</i>	GCA_000002765.4
16	<i>Toxoplasma gondii</i>	GCA_000006565.2

En el caso de la base de datos de *C. cayetanensis*, se seleccionaron las secuencias genómicas con una cobertura e identidad similar. Según lo observado con los resultados de ANI, se seleccionaron 36 de los 37 genomas (Cuadro 3). La secuencia genómica GCA\_0026893375.1 poseía una cobertura menor en comparación a las otras secuencias, por lo que fue descartada (Figura 6). En cuanto a la identidad de las secuencias genómicas en general es mayor al 98% (Figura 7).

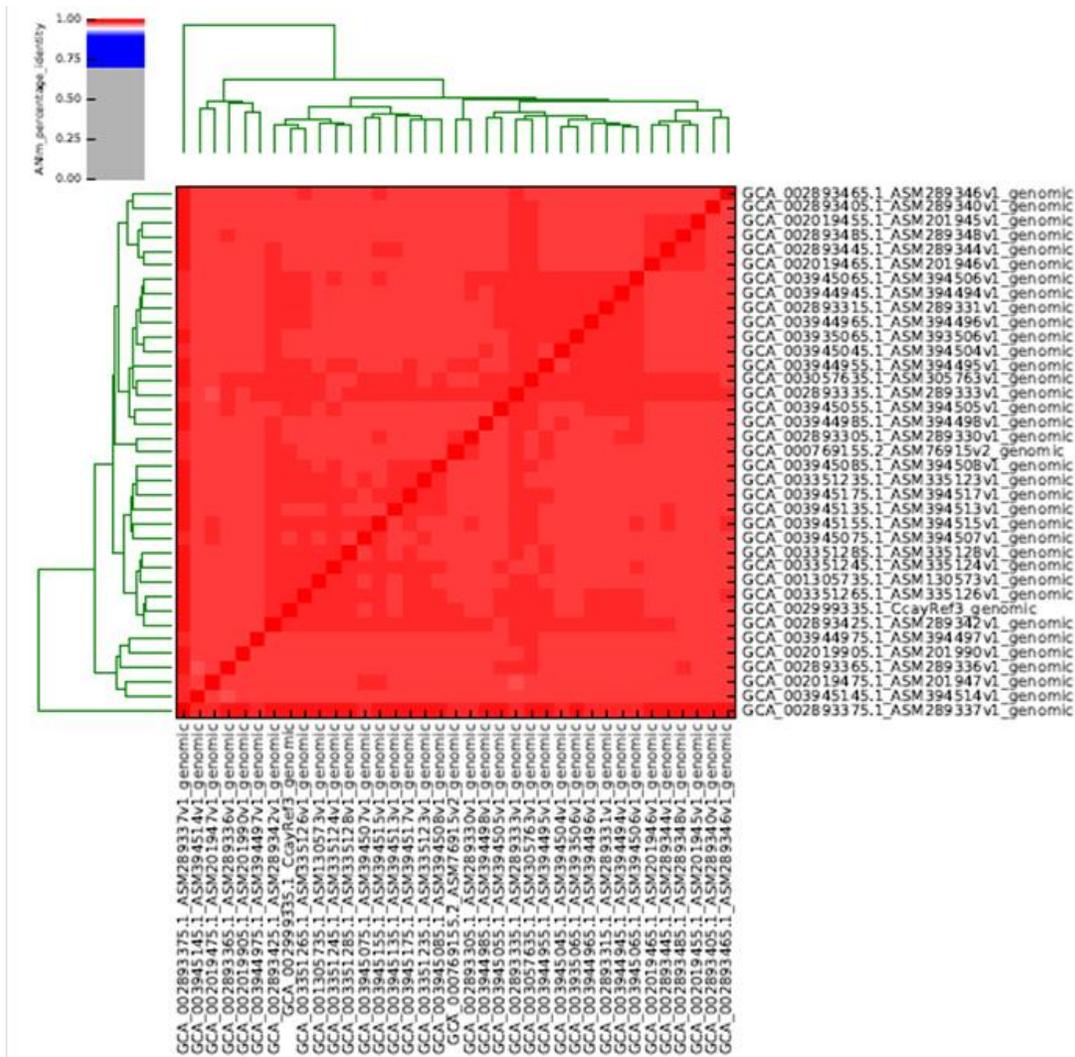
Por otra parte, de las 36 secuencias genómicas seleccionadas, sólo dos se encuentran anotadas en el NCBI (GCA\_002999335.1 CcayRef3 y GCA\_000769155.2), por lo que a las 34 secuencias genómicas restantes se les realizó una predicción de genes con AUGUSTUS, tomando una de las secuencias anotadas (secuencia genómica de referencia de *C. cayetanensis*, GCA\_002999335.1 CcayRef3) para entrenar el programa y así obtener una mejor predicción de los genes. De esta manera, se obtuvieron los archivos gemelos de nucleótidos y proteínas (.fna y .faa) para los posteriores análisis.

**Cuadro 3.** Base de datos de *Cyclospora cayetanensis* descargada de la base de datos del NCBI.

<b>Base de datos de <i>Cyclospora cayetanensis</i></b>				
#	Nombre en la base de datos	Ensamblaje	Procedencia	Fecha de colecta
1	Cc_1305735_EUA:NY	GCA_001305735.1	EUA:New York	2001
2	Cc_2019455_EUA:Tx2014	GCA_002019455.1	EUA: Texas	2014
3	Cc_2019465_Guatemala	GCA_002019465.1	Guatemala	1997
4	Cc_2019475_Indonesia	GCA_002019475.1	Indonesia	2014
5	Cc_2019905_EUA:Rhode	GCA_002019905.1	EUA:Rhode Island	1997
6	Cc_2893305_China	GCA_002893305.1	China	2001
7	Cc_2893315_EUA:Tx:VCR	GCA_002893315.1	EUA:Texas (Viajero a Costa Rica)	2015
8	Cc_2893335_EUA:Tx:Vmex	GCA_002893335.1	EUA:Texas (Viajero a México)	2015
9	Cc_2893365_EUA:Tx:Cilantro	GCA_002893365.1	EUA:Texas (Consumo de cilantro)	2013
10	Cc_2893405_EUA:FL	GCA_002893405.1	EUA:Florida	2015
11	Cc_2893425_EUA:NY	GCA_002893425.1	EUA:New York	2001
12	Cc_2893445_Guatemala	GCA_002893445.1	Guatemala	1997
13	Cc_2893465_EUA:Tx:Vmex	GCA_002893465.1	EUA:Texas (Viajero a México)	2015
14	Cc_2893485_EUA:Tx2014	GCA_002893485.1	EUA: Texas	2014
15	CcayRef3_2999335_Nepal	GCA_002999335.1	Nepal	-
16	Cc_3057635_EUA:Tx:Cilantro	GCA_003057635.1	EUA:Texas (Consumo de cilantro)	2013
17	Cc_3351235_Nepal	GCA_003351235.1	Nepal	-
18	Cc_3351245_Nepal	GCA_003351245.1	Nepal	-
19	Cc_3351265_Nepal	GCA_003351265.1	Nepal	-
20	Cc_3351285_Nepal	GCA_003351285.1	Nepal	-
21	Cc_3935065_EUA	GCA_003935065.1	EUA	2015
22	Cc_3944945_EUA	GCA_003944945.1	EUA	2013
23	Cc_3944955_EUA	GCA_003944955.1	EUA	2015
24	Cc_3944965_EUA	GCA_003944965.1	EUA	2014
25	Cc_3944975_EUA	GCA_003944975.1	EUA	2016
26	Cc_3944985_EUA	GCA_003944985.1	EUA	2016
27	Cc_3945045_EUA	GCA_003945045.1	EUA	1997
28	Cc_3945055_EUA	GCA_003945055.1	EUA	2016
29	Cc_3945065_Mexico	GCA_003945065.1	México	2016
30	Cc_3945075_Indonesia	GCA_003945075.1	Indonesia	2015
31	Cc_3945085_Guatemala	GCA_003945085.1	Guatemala	1997
32	Cc_3945135_Guatemala	GCA_003945135.1	Guatemala	1998
33	Cc_3945145_Nepal	GCA_003945145.1	Nepal	2017
34	Cc_3945155_Indonesia	GCA_003945155.1	Indonesia	2015
35	Cc_3945175_EUA	GCA_003945175.1	EUA	1996
36	Cc_769155_China	GCA_000769155.2	China	2011



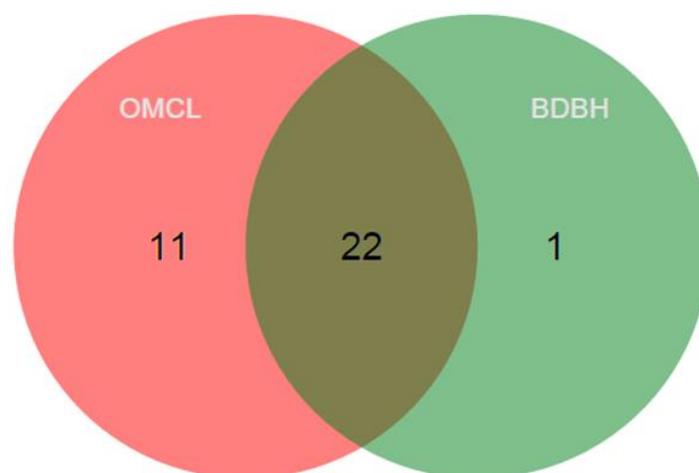
**Figura 6.** Heatmap con valores de cobertura de Average Nucleotide Identity (ANI) de los 37 genomas reportados de *Cyclospora cayetanensis*. La secuencia GCA\_0026893375.1 muestra menor cobertura que el resto de las secuencias genómicas.



**Figura 7.** Heatmap con valores de identidad de Average Nucleotide Identity (ANI) de los 37 genomas reportados de *Cyclospora cayetanensis*.

## 6.2 Análisis de las Relaciones Filogenéticas de las Secuencias Genómicas de *Cyclospora cayetanensis* con Otros Microorganismos Apicomplexa

En la selección del grupo de genes homólogos y unicopia para determinar el genoma core con el pipeline de GET\_HOMOLOGUES, se obtuvieron 33 genes con OMCL y 23 con BDBH. Sin embargo, 22 genes (Cuadro 4) eran compartidos entre los dos métodos (Figura 8) y son los seleccionados para proseguir con el análisis del pipeline de GET\_PHYLOMARKERS.



**Figura 8.** Diagrama de Venn de la intersección de los métodos OMCL y BDBH que determinan el genoma core de los microorganismos Apicomplexa.

**Cuadro 4.** Lista de genes del genoma core obtenido para la base de datos de microorganismos Apicomplexa utilizando los métodos de BDBH y OMCL.

Marcador	Loci	Nombre
1	81534	Subunidad 20s del proteosoma
2	82501	Proteína 1 del complejo T
3	82274	ARN helicasa dependiente de ATP
4	80604	Proteína fosfatasa
5	81490	Piruvato quinasa
6	80810	Antígeno nuclear de células proliferantes
7	80615	Subunidad ribosomal larga
8	81556	Proteína de clasificación de proteínas vacuolares
9	81078	Proteína de translocación
10	81226	Lisil - tRNA sintetasa
11	81858	Proteína de unión al nucleótido guanina
12	80892	Subunidad ribosomal larga
13	81122	Peroxiredoxina
14	82542	Aminopeptidasa
15	81864	Subunidad ribosomal larga
16	80844	Clatrina polipéptido pesado
17	82393	Subunidad del complejo AP-1
18	82437	Diftina sintasa
19	80900	Subunidad ribosomal larga
20	83365	Proteína tumoral controlada traslacionalmente
21	81876	GAP40
22	83115	Factor de elongación

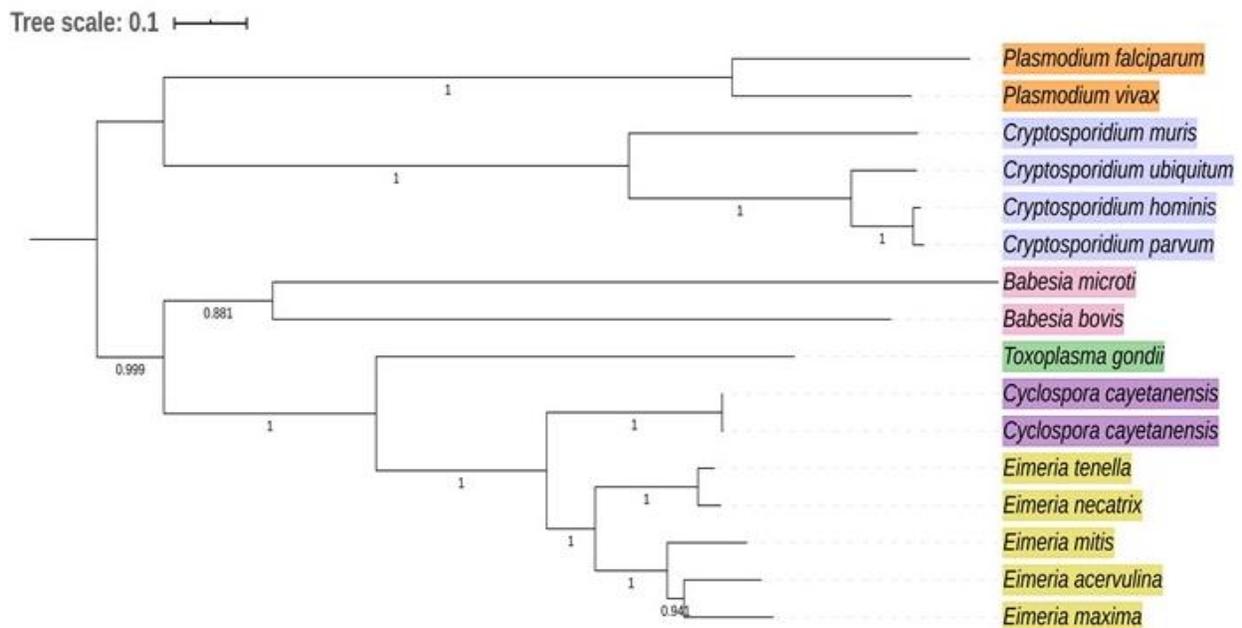
Una vez analizada la recombinación, estimada la filogenia y señales filogenéticas, se generó un directorio con los genes que entraron como los mejores marcadores. Como resultado, se detectaron genes de los cuales dos genes fueron los óptimos (Cuadro 5) y son los que se utilizaron para estimar el árbol filogenético con PhyML. Para este caso, la identidad de estos marcadores venía en los archivos generados, ya que esta base de datos cuenta con secuencias anotadas desde el NCBI.

**Cuadro 5.** Marcadores de la base de datos de microorganismos Apicomplexa obtenido por el pipeline de GET\_HOMOLOGUES y GET\_PHYLOMARKERS.

Marcador	Loci	Nombre	Función
1	81556	Proteína de clasificación de proteínas vacuolares	Clasificación intracelular y suministro de proteínas vacuolares solubles
2	81078	Proteína de translocación	Media el transporte cotraduccional y postraduccional de ciertos polipéptidos precursores a través del retículo endoplásmico

Con ambos marcadores, se procedió a generar un árbol filogenético con máxima verosimilitud utilizando PhyML. El árbol visualizado con la herramienta online iTOL, muestra que *Cyclospora cayetanensis* forma su propio clado (agrupación en color morado) al igual que los otros géneros de microorganismos Apicomplexa (Figura 9). Esto coincide con lo observado por Sulaiman *et al.*, (2013), quienes realizaron un análisis filogenético utilizando como marcador la proteína de choque térmico 70 (HSP70, por sus siglas en inglés), para determinar la relación evolutiva entre las especies de *Cyclospora* y otros microorganismos Apicomplexa (*Babesia*, *Cryptosporidium*, *Eimeria*, *Plasmodium*, *Theileria* y *Toxoplasma*). Como resultado revelaron que en el árbol generado por neighbor-joining se forman clados aparentemente género específicos. Sin embargo, la relación entre *Eimeria* y *Cyclospora* no se observa tan estrecha como en el árbol filogenético resultante de los dos marcadores establecidos en el presente flujo de trabajo, la proteína de clasificación de proteínas vacuolares y la proteína de translocación, donde el soporte de las ramas entre sus agrupaciones es de 1. Este resultado puede ser comparable por los reportado por Reiman *et al.*, (1996) con la filogenia establecida con el ARNr 18S, ya que establece la cercanía de *Cyclospora* con *Eimeria*. Igualmente, esto coincide con la estimación filogenética reportada por Eberhard *et al.*, (1999) basada en la secuencia de la subunidad pequeña del ARN ribosomal (SSU

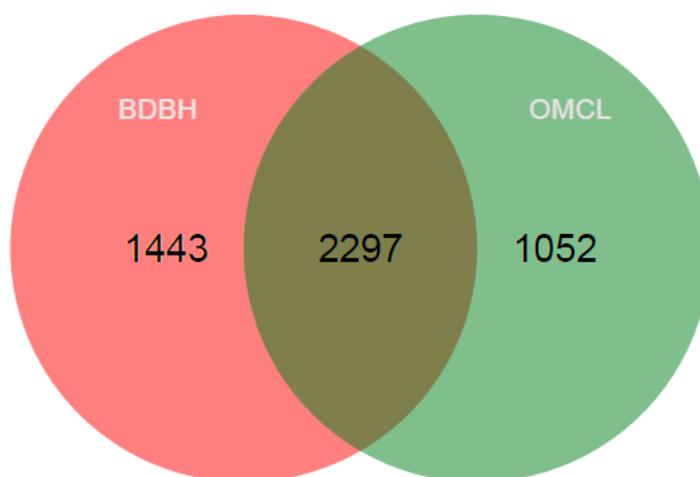
rRNA, por sus siglas en inglés) en donde las especies de *Cyclospora* formaron su propio clado en una relación cercana al género de *Eimeria*. Por otro lado, el árbol similar fue reportado por Liu *et al.* (2016), el cual construyeron con base en cien ortólogos de proteínas identificados por genómica comparativa. En dicho árbol, también *Cyclospora cayetanensis* se posiciona en su propio clado estrechamente relacionado con *Eimeria tenella*. El contraste de los resultados obtenidos en el presente estudio está en que la topología del árbol construido con sólo dos marcadores (la proteína de clasificación de proteínas vacuolares y la proteína de translocación) permite evidenciar la estrecha relación entre *Cyclospora cayetanensis* y las especies de *Eimeria* de forma clara y posicionando en cada género un clado monofilético.



**Figura 9.** Árbol filogenético de especies del subfilum Apicomplexa basado en máxima verosimilitud utilizando dos marcadores, una proteína de clasificación de proteínas vacuolares y una proteína de traslocación, obtenidos por el pipeline de GET\_PHYLOMARKERS. Los números en las ramas indican el nivel de soporte de acuerdo con el estadístico SH-Like supports y la raíz se determinó utilizando el método de punto medio.

### 6.3 Marcadores Óptimos de *Cyclospora cayetanensis* y Análisis de las Relaciones Filogenéticas entre los Aislados de la Misma Especie

En el caso de la construcción del genoma core de *Cyclospora* se obtuvieron 3,349 genes con el método OMCL y 3,740 genes con BDBH, siendo 2,297 genes los compartidos entre ambos métodos (Figura 10).



**Figura 10.** Diagrama de Venn de la intersección de los métodos OMCL y BDBH que determinan el genoma core de los aislados de *Cyclospora cayetanensis*.

Posteriormente, en el análisis de la recombinación y estimación de la filogenia con el pipeline de GET\_PHYLOMARKERS, se generó un directorio con los mejores marcadores, obteniendo como resultado 547 marcadores. De estos últimos, 47 fueron los óptimos para estimar la filogenia (Cuadro 6). Para confirmar la identidad de los marcadores, se tomaron las secuencias de cada uno de los genes para someterlos a BLAST, tanto de nucleótidos como de aminoácidos. En el Cuadro 6 se menciona cada uno de estos marcadores, el loci y la identidad con su número de acceso en la base de datos del NCBI. Consecutivamente, estos 47 marcadores se utilizaron como input en el programa PhyML para generar el árbol filogenético por máxima verosimilitud.

**Cuadro 6.** Marcadores de la base de datos de *Cyclospora cayetanensis* obtenido con el pipeline de GET\_HOMOLOGUES-EST y GET\_PHYLOMARKERS.

<b>Marcador</b>	<b>Loci</b>	<b>Nombre</b>	<b>Núm. de acceso</b>
1	4800	Proteína con dominio para dedo de zinc	OEH76891.1
2	3084	Proteína hipotética cyc_03188	OEH80123.1
3	2846	Proteína no caracterizada LOC34618073	XP_022589823.2
4	5955	Proteína no caracterizada LOC113146566	XP_026190111.1
5	2976	Proteína no caracterizada LOC34621809	XP_026191972.1
6	4539	Proteína hipotética cyc_06096	OEH79049.1
7	3757	Proteína de procesamiento de ARN ribosomal 8	XP_022591841.2
8	5544	Subunidad del citocromo c oxidasa	OEH79646.1
9	5134	Proteína no caracterizada LOC34617310	XP_026192690.1
10	1179	Proteína no caracterizada LOC34621648	XP_026191075.1
11	5260	Proteína 1 que contiene el dominio del citocromo b5	XP_022586600.2
12	3894	Proteína RAE1	XP_026189620.1
13	1069	Proteína hipotética cyc_00082	OEH77281.1
14	3805	Proteína no caracterizada LOC113146571	XP_026190138.1
15	4694	Proteína no caracterizada LOC34623958	XP_026193307.1
16	4491	Proteína hipotética cyc_00152	OEH76892.1
17	4516	Caseína quinasa uno	OEH80498.1
18	5842	Putativo citocromo b5	OEH75136.1
19	5680	Proteína hipotética cyc_01913	OEH74244.1
20	3616	Proteína relacionada a ARNt de guanosina-2-o-metiltransferasa tipo trm11	OEH76651.1
21	3546	Proteína no caracterizada LOC34618467	XP_022586438.2
22	5461	Proteína hipotética cyc_08588	OEH76288.1
23	5406	Proteína no caracterizada LOC34617534	XP_026191431.1
24	2325	Factor clp1 de escisión de ARNm y poliadenilación	XP_026193039.1
25	5110	Proteína no caracterizada LOC34623089	XP_022589451.1
26	3085	Proteína no caracterizada LOC34623090	XP_022589452.2
27	2764	Proteína con dominio del pliegue alfa beta hidrolasa	OEH74248.1
28	4046	Proteína no caracterizada LOC34621722	XP_022592440.2
29	3624	Glutamato deshidrogenasa específica de NADP	XP_022592842.2
30	4411	Proteína relacionada con aspartato-amoniaco ligasa	OEH76662.1
31	2834	Canal 1 activado por nucleótidos cíclicos	XP_026189641.1
32	1395	Proteína relacionada con proteofosfoglicanos	OEH75887.1
33	3743	Proteína hipotética cyc_06088	OEH77794.1
34	5323	Proteína no caracterizada LOC34620711	XP_026191746.1
35	3135	Proteína no caracterizada LOC34620710	XP_022590316.1

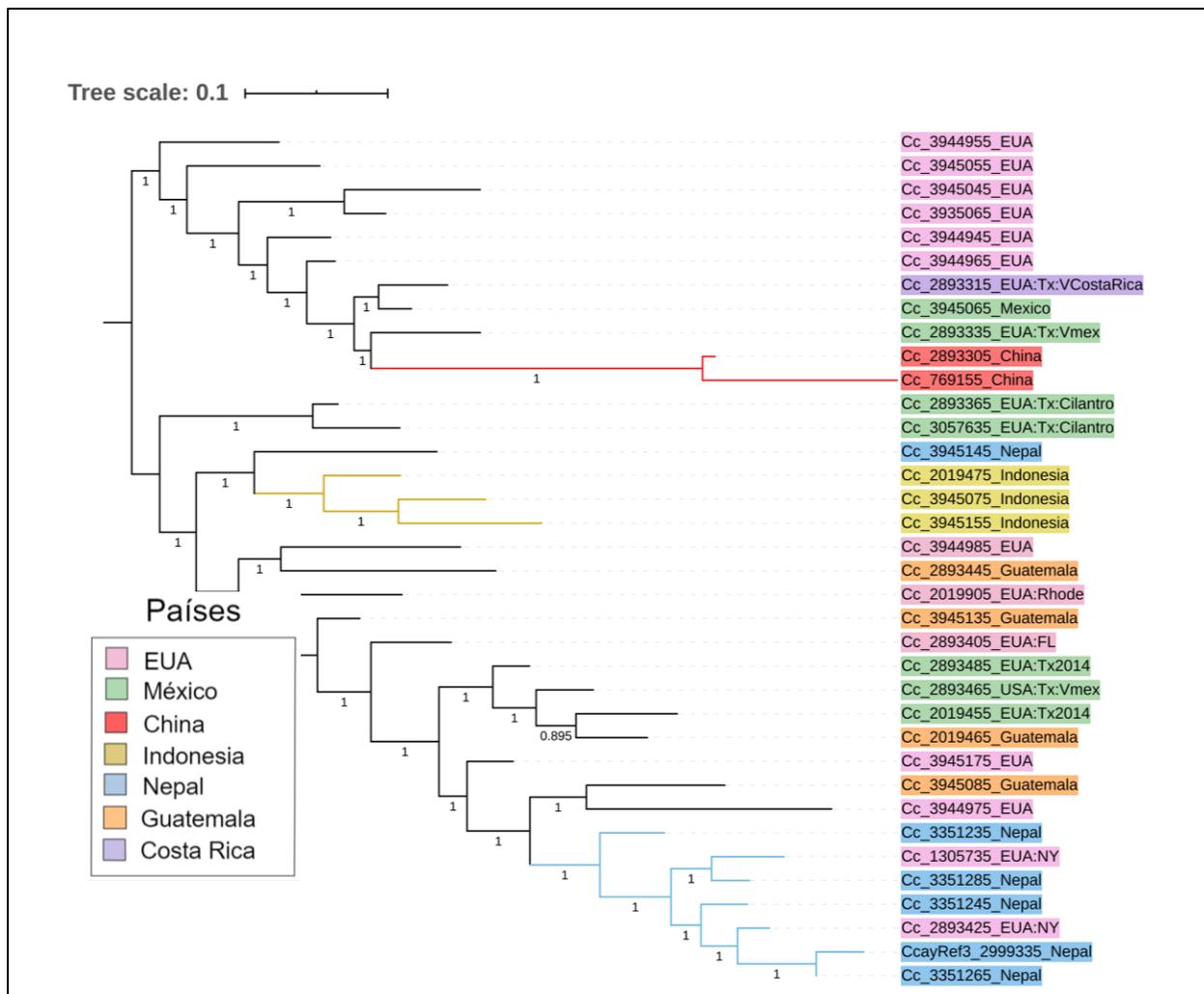
**Cuadro 6.** Marcadores de la base de datos de *Cyclospora cayetanensis* obtenido con el pipeline de GET\_HOMOLOGUES-EST y GET\_PHYLOMARKERS (continuación).

Marcador	Loci	Nombre	Núm. de acceso
36	1516	Proteína con repetición similar a beta-catenina armadillo	OEH77972.1
37	377	Proteína no caracterizada LOC113147069	XP_026192073.1
38	1061	Proteína no caracterizada LOC34620739	XP_026190438.1
39	780	Proteína no caracterizada LOC34621284	XP_026193789.1
40	2638	Rootletina	XP_026192874.1
41	2665	Proteína hipotética cyc_04389	OEH74177.1
42	2706	Factor de ensamblaje de dineína 1	XP_026191153.1
43	1084	Proteína hipotética cyc_06285	OEH73796.1
44	3486	Proteína quinasa dependiente del cAMP	XP_022589857.2
45	3489	Probable proteína de unión a histonas lin-53	XP_026191357.1
46	3826	Proteína no caracterizada LOC34622134	XP_022589743.1
47	3914	Proteína no caracterizada LOC34622471	XP_026192464.1

En la figura (Figura 11) se puede observar que algunas de las secuencias formaron clusters según el origen geográfico, como en los casos de las secuencias genómicas de China, Nepal e Indonesia. El resultado de las secuencias de China es similar a lo reportado por Li *et al.* (2017); estos autores mencionan que los aislados de China se agruparon estrechamente en comparación de otros países (EUA, Nepal y Perú) en su análisis evolutivo utilizando cinco loci microsatélite como marcadores MLST. Por otra parte, se puede visualizar que dos secuencias de brotes en Nueva York, EUA (las secuencias Cc\_2893425\_EUA:NY y Cc\_2019455\_EUA:Tx), se encuentran entre la agrupación de las secuencias de Nepal. Por tanto, la relación de estas secuencias es cercana y se puede sugerir que los ooquistes de *Cyclospora* causante de este brote provienen de ese país.

De la misma manera, se observaron las relaciones de las secuencias que se encuentran asociadas a México: las secuencias genómicas de los brotes en Texas en 2013 y 2014 relacionado con el consumo de cilantro proveniente del estado de Puebla, México (Cc\_3057635\_EUA\_Tx\_Cilantro, Cc\_2893365\_EUA\_Tx\_Cilantro, Cc\_2019455\_EUA:Tx2014 y Cc\_2893485\_EUA:Tx2014); así como otras dos de casos concernientes a viajeros estadounidenses que visitaron México y adquirieron la infección (Cc\_2893335\_Tx\_Vmex y Cc\_2893465\_Tx\_Vmex); y otra más que la procedencia se menciona es México (Cc\_3945065\_Mexico) según Qvarnstrom *et al.* (2018). Estas siete secuencias se encontraron dispersas en el árbol al no verse un único perfil o agrupamiento en concreto entre ellas. No obstante, es interesante observar que las secuencias genómicas del brote

de Texas relacionado a cilantro de Puebla de 2013 están estrechamente relacionadas entre sí, pero se encuentran separadas a las secuencias del brote de Texas del 2014, las cuales también se aluden que son de cilantro contaminado procedente del mismo sitio geográfico. Al mismo tiempo, se visualizó que las secuencias del brote de Texas del 2014 están estrechamente relacionadas con uno de los casos de viajeros estadounidenses, así como con una secuencia genómica de *Cyclospora* procedente de Guatemala. La causa de dichas inconsistencias para estos hechos puede deberse a que estas secuencias no pertenecen al mismo origen geográfico, ya que las investigaciones sobre estos casos se ha basado en especulaciones de acuerdo a declaraciones de los individuos afectados y no meramente al rastreo de la fuente de contaminación del patógeno como tal, ya que se carece de un método de subtipificación (la persona pudo adquirir el patógeno de otra fuente o tal vez en otro lugar, o hasta pudo existir alguna contaminación durante la exportación del alimento, por ejemplo); o cabe la posibilidad que las secuencias de *C. cayetanensis* procedentes de México son diversas entre ellas, dando la posibilidad de la existencia de más de una cepa en México; pero es difícil asegurar esto último contando con sólo un genoma con certidumbre de su procedencia y sin marcadores de subtipificación. Esto enfatiza la importancia de sumar secuencias con certeza del origen de aislamiento, en particular, procedentes de México en general, ya que da pauta a optimizar el perfil de marcadores nucleares y a hacer inferencias certeras respecto a las relaciones filogenéticas de aislados de *Cyclospora cayetanensis*.



**Figura 11.** Árbol filogenético de los 36 aislados de *Cyclospora cayetanensis* basado en máxima verosimilitud utilizando 47 marcadores obtenidos por el pipeline de GET\_PHYLOMARKERS. Los números en las ramas indican el nivel de soporte de acuerdo con el estadístico SH-Like supports y la raíz se determinó utilizando el método de punto medio.

En el sentido del flujo de trabajo utilizado para seleccionar los 47 marcadores resultantes, el pipeline de GET\_PHYLOMARKERS en conjunto con GET\_HOMOLOGUES, mostraron que son excelentes herramientas para este propósito debido a su fácil uso y resultados prometedores: marcadores de alta calidad con los que estimar filogenias genómicas robustas. Un estudio reciente por Houghton *et al.*, (2020) proponen un flujo de trabajo para identificar marcadores moleculares del genoma nuclear de *C. cayetanensis*, donde identificaron 485 marcadores candidatos. Estos resultados, son similares a los obtenidos en la presente investigación, en la cual se identificaron 547 marcadores preliminares con el flujo de trabajo propuesto; no obstante, el pipeline de

GET\_PHYLOMARKERS genera una selección de marcadores óptimos mediante una serie de filtros: analizar la recombinación de las secuencias, ya que las secuencias recombinantes impactan negativamente en la inferencia filogenética; verifica la topología mediante estimaciones de árboles, descartando aquellos atípicos y calcula su señal filogenética a partir de valores de soporte de rama, descartando aquellas debajo del valor de corte. Así, de los 547 marcadores, 47 fueron los mejores para estimar la filogenia. En el árbol filogenético resultante, fue evidente que se logró observar patrones de agrupamiento por origen geográfico, por lo que agregar más secuencias genómicas a este mismo flujo de trabajo podría ofrecer resultados similares (agrupamientos por origen geográfico) y más precisos. Además, cabe mencionar que los 47 marcadores óptimos seleccionados de los 547 marcadores, podrían ser prometedores para ser utilizados en tipificación multilocus y complementar el panel de marcadores que se han utilizado en investigaciones previas. Tal es el caso del estudio generado por Barrat *et al.*, (2019) en donde usaron tres marcadores para tipificación por MLST, dos marcadores derivados del genoma nuclear y otro más del genoma mitocondrial, con los cuales lograron obtener concordancia de agrupamiento con los datos epidemiológicos de la mitad de los brotes analizados. Los resultados generados en el presente estudio pueden servir como base para generar análisis a una profundidad tal que permitiría revisar los 47 marcadores y seleccionar aquellos que muestren mejores características y evaluarlos para utilizarlos para MLST. Estos marcadores podrían utilizarse en conjunto con otros ya reportados, tanto nucleares como derivados de otros genomas del parásito, para llegar al objetivo de mejorar la resolución del genotipo de *C. cayetanensis*.

#### 6.4 Análisis del Pangenoma de *Cyclospora cayetanensis* Comparado con Otros Microorganismos Apicomplexa

En el caso de la base de datos de microorganismos Apicomplexa, el enfoque fue analizar si *Cyclospora cayetanensis* posee genes únicos que permitan diferenciarlos de otro microorganismo del subfilum. Con este objetivo, se realizó la comparación tomando como lista A las secuencias de *Cyclospora cayetanensis* anotadas y como lista B las otras 14 secuencias de Apicomplexa (2 de *Babesia*, 4 de *Cryptosporidium*, 5 de *Eimeria*, 2 de *Plasmodium*, y 1 de *Toxoplasma*). El resultado

del análisis del pangenoma presentó 656 genes únicos para el género de *Cyclospora*. De los resultados más interesantes (Cuadro 7), se obtuvieron algunas proteínas involucradas en el metabolismo de manitol y relacionadas a proteínas de superficie; también concernientes a la invasión celular, transducción de señales, moduladores del sistema inmune y proteínas relacionadas como factores de virulencia. Otras más fueron clasificadas dentro de proteínas varias (275 proteínas), siendo algunas proteínas de transporte de organelos, factores de transcripción, proteínas estructurales y algunas proteínas putativas. Adicionalmente, los resultados muestran que 355 genes codificaban para proteínas cuya función está indeterminada, ya que fueron para proteínas hipotéticas o no caracterizadas.

**Cuadro 7.** Genes únicos de *Cyclospora cayetanensis* en comparación a otros Apicomplexa.

Clasificación	Genes únicos de <i>Cyclospora</i> de interés	#
Organización celular y biogénesis	Proteína de esporulación	1
Metabolismo	Manitol-1-fosfatasa	1
	Manitol-1-fosfato deshidrogenasa	1
Transducción de señal	Regulador de adhesión	1
	Proteína de unión-GTP	1
	Proteína que contiene el dominio de la familia Ras	2
Proteína de superficie	Proteína de unión-cola citoplasmática antígeno cd2	1
	Glicoproteína de superficie celular	1
	Proteína relacionada a pared del ooquiste	1
	Proteína transmembrana	5
	Miembro de la familia SAG	1
Invasión celular	Proteína del sistema de anclaje de glicina	1
	Proteína micronema	1
	Proteína de la familia roptria quinasa ROP23	2
	Roptria metaloproteasa toxolisina	1
	Proteína roptria RON1	1
	Proteína relacionada a RON10	1
	Proteína de la familia PP2C	1
	Proteasas	4
Modulador sistema inmune	Inmunomodulador de célula T	1
Factor de virulencia	Transportador de flujo de resistencia multidroga	1
	Proteína relacionada a marcador de resistencia a cloroquina	1
	Proteínas varias	270
	Proteínas hipotéticas	355

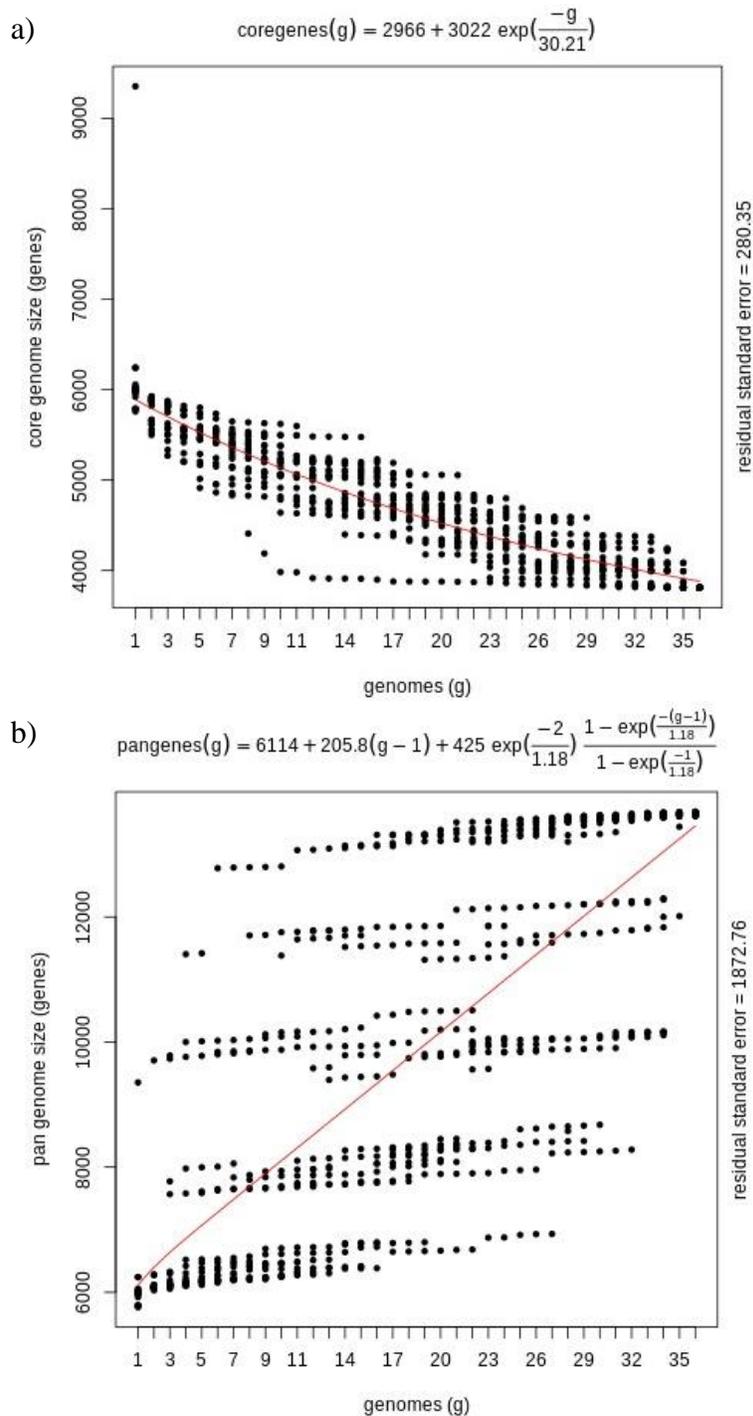
Parte de los resultados obtenidos coinciden con lo planteado por Liu *et al.*, (2016), quienes discuten que *C. cayetanensis* posee la particularidad de metabolizar de manera inversa manitol a partir de fructosa; que posee un sistema de invasión celular similar a *Eimeria tenella* y *Toxoplasma gondii*; y que podría poseer antígenos de superficie propios de la especie. Los resultados del presente estudio confirman que hay dos enzimas involucradas en la metabolización de manitol (manitol-1-fosfatasa y manitol-1-fosfato deshidrogenasa) y que son particulares para el objeto de estudio. Además, se detectaron varias proteínas que se clasifican como proteínas de superficie, confirmando lo que hipotizaron Liu *et al.*, (2016). En este caso, estas proteínas son proteínas transmembrana, una proteína relacionada con la pared del ooquiste, una glicoproteína de superficie celular, una proteína de unión a la cola citoplasmática antígeno cd2 y una proteína miembro de la familia SAG. Igualmente, se presentan otras proteínas más que están relacionadas con la invasión celular, como proteínas RON, roptrias, micronemas, etc. También es necesario considerar que más de la mitad de las proteínas que son únicas, no se encuentran caracterizadas o son hipotéticas; algunas de estas podrían estar funcionando como antígenos de superficie o como proteínas del complejo apical en la invasión celular, lo que podría apoyar a explicar la naturaleza estricta de poseer como único hospedero el ser humano. Lo anterior permite justificar de mejor manera la necesidad de caracterizar detalladamente las características de *Cyclospora*.

#### 6.5 Análisis del Pangenoma de *Cyclospora cayetanensis* e Identificación de Genes Presentes Diferencialmente en las Secuencias Genómicas Reportadas como Mexicanas

Por parte del análisis del pangenoma de *C. cayetanensis*, primero se generaron las gráficas tanto del tamaño del genoma core como el pangenoma. El objetivo de generar estas gráficas en particular fue para observar la tendencia que siguen estas y estimar si el pangenoma es abierto o cerrado. Esta información es valiosa ya que, con ella, se puede inferir si el número de genomas secuenciados son suficientes para caracterizar a la especie o si es necesario secuenciar aún más para representar con precisión todo su repertorio de genes.

Para estimar el tamaño del genoma core y pangenoma se extrapoló el número de genes compartidos ajustando los modelos exponenciales de Tettelin *et al.*, (2005) (Figura 12). Como era de esperar, el

número de genes compartidos fue disminuyendo conforme se adicionaba una nueva secuencia genómica. Según la extrapolación de la curva, el genoma core (Figura 12a) alcanza un mínimo de 2,966 genes con 36 genomas. Por otra parte, en la extrapolación del pangenoma (Figura 12b), se observa que el número de genes va en aumento conforme se añade una nueva secuencia genómica, prácticamente en aumento lineal. La tasa de crecimiento extrapolada del tamaño del pangenoma de los 36 genomas es de 205.8 en promedio, lo que significa que, por cada nuevo genoma secuenciado, se añadirán 205.8 genes específicos en promedio. Asimismo, la visualización gráfica muestra que el pangenoma supera los 12,000 genes. Según esta tendencia, podemos sugerir que el pangenoma de *C. cayetanensis* es un pangenoma abierto (en los casos contrarios, en un pangenoma cerrado se observa que se alcanza una asíntota). En consecuencia, el número de nuevos genes aumentará de manera significativa conforme se adicionen nuevas secuencias del genoma de *C. cayetanensis*; por ende, se puede considerar que aún es necesario agregar más secuencias del genoma de *C. cayetanensis* para lograr caracterizar de forma representativa el repertorio de genes de la especie.



**Figura 12.** Gráfica de la estimación estadística del tamaño del genoma core (a) y pangenoma (b) de *Cyclospora cayetanensis* basadas en el ajuste de la función exponencial de Tettelin *et al.*, 2005. Ambas curvas se construyen muestreando un número creciente de genomas al azar y contando el número de genes pertenecientes a cada categoría. En el caso de la curva de genoma core, se observa un decrecimiento que alcanza a 2,966 genes. Mientras que en la curva del pangenoma, este sigue creciendo considerablemente sin llegar a una estabilización en la medida que se agregan más genomas, evidenciando un pangenoma abierto.

Para continuar con el análisis del pangenoma, se realizó la identificación de genes específicos de linaje, considerando dos subconjuntos de comparación dentro de la misma especie de *C. cayetanensis*. En particular, se realizó la identificación de rasgos genéticos únicos de los genomas reportados como mexicanos; por ello, se seleccionó como subconjunto A las secuencias de *C. cayetanensis* que estuvieran asociadas a México. En total esta lista constó de siete secuencias (Cuadro 8): 4 secuencias relacionadas a los brotes vinculados a cilantro proveniente de Puebla, México; 2 relacionadas a viajeros que, con base a los registros epidemiológicos, adquirieron el parásito en el país; y una secuencia que se alude su procedencia es de México (Qvarnstrom *et al.*, 2018).

**Cuadro 8.** Información de las secuencias genómicas de *Cyclospora cayetanensis* asociadas a México.

Secuencia	Cepa	Localización geográfica	Fecha de colecta	Fuente de aislamiento
Cc_3057635_EUA_Tx_Cilantro	CDC:HCTX205:15	EUA: TX	2013	Heces, relacionadas a consumo de cilantro
Cc_2893365_EUA_Tx_Cilantro	CDC:HCTX204:15	EUA: TX	2013	Heces, relacionadas a consumo de cilantro
Cc_2019455_EUA:Tx2014	CDC:HCTX69:14	EUA: TX	2014	Heces
Cc_2893485_EUA:Tx2014	CDC:HCTX48:14	EUA: TX	2014	Heces
Cc_2893335_Tx_Vmex	CDC:HCTX208:15	EUA: TX	2015	Heces, viajero a México
Cc_2893465_Tx_Vmex	CDC:HCTX227:15	EUA: TX	2015	Heces, viajero a México
Cc_3945065_Mexico	CDC:HCMX010_16	México	2016	Heces

Estas siete secuencias fueron comparadas con diferentes agrupaciones de secuencias como subconjunto B. Las comparaciones y el número de genes únicos para el grupo de *Cyclospora* relacionadas a México se observa en el siguiente cuadro (Cuadro 9).

**Cuadro 9.** Comparación de *Cyclospora cayetanensis* del grupo relacionado a México contra secuencias de otras regiones.

<b>Lista A</b>	<b>Lista B</b>	<b>Genes únicos</b>
<i>(Cyclospora México)</i>	<i>(Cyclospora mundo)</i>	
7 secuencias <i>Cyclospora</i> asociadas a México	2 secuencias China	0
	6 secuencias Nepal	0
	3 secuencias Indonesia	0
	11 secuencias asiáticas	0
	4 secuencias Guatemala	0
	11 secuencias brotes en USA	0
Secuencia Cc_3945065_México	2 secuencias China	53
	6 secuencias Nepal	31
	3 secuencias Indonesia	32
	11 secuencias asiáticas	13
	11 secuencia brotes en USA	6

El resultado de genes únicos no presentó variación entre las secuencias obtenidas de cepas relacionadas a México con respecto a las secuencias aisladas en otras regiones del mundo. Tanto la comparación de las 11 secuencias genómicas asiáticas (las secuencias de China, Nepal e Indonesia), así como las comparaciones de las secuencias de manera independiente (contra las 2 secuencias de China, las 6 secuencias de Nepal y las 3 secuencias de Indonesia), arrojaron como resultado que el número de genes fue cero. De la misma manera, se hizo la comparación de las secuencias mexicanas con las secuencias de los brotes reportados en Estados Unidos que no lograron vincular el vehículo y origen del patógeno en las investigaciones epidemiológicas. Éstas sumaron un total de 11 secuencias, y el resultado fue la ausencia de genes únicos. Esto significa, en general, que por lo menos una de las secuencias genómicas de *Cyclospora* asociadas a México es muy similar por lo menos a una de las secuencias de los grupos de otras regiones geográficas. Con base en estos resultados, se tomó la secuencia genómica Cc\_3945065\_México (Qvarnstrom *et al.*, 2018), cuya procedencia es México, para generar comparaciones. En este caso, para la comparación de esta secuencia con los diferentes grupos, se obtuvo variación en el número de genes específicos para esta cepa. El número más alto de genes específicos para la cepa mexicana fue de 53 genes en la comparación con las dos secuencias de China. Esto se soporta con los mencionado por Li *et al.*, (2017), donde observaron en su análisis evolutivo que las secuencias de China se agrupan estrechamente entre ellas en comparación con las de otras regiones, por lo que

podrían ser consideradas significativamente diferentes a las de otros países. Por otro lado, el número de genes obtenido contra las secuencias de Indonesia y Nepal, de manera independiente, fue bastante similar; de 32 y 31, respectivamente. En cuanto a la comparación con las 11 secuencias genómicas asiáticas, se obtuvo un total de 13 genes únicos; mientras que, en la comparación con las 11 secuencias de Estados Unidos sin vinculación epidemiológica, el número resultante fue de 6 genes.

En resumen, estos resultados evidencian la pobre caracterización de los aislados en general y se confirma que las investigaciones epidemiológicas no han sido concluyentes, dando soporte a la necesidad de implementar marcadores que permitan la subtipificación de la cepa. Además del hecho anteriormente mencionado, que es de suma importancia agregar secuencias con certidumbre de su origen geográfico para realizar este tipo de análisis, ya que esto brindará resultados certeros para la caracterización de los aislados según su origen geográfico.

## 7. CONCLUSIONES

Se estableció un flujo de trabajo con los softwares GET\_HOMOLOGUES y GET\_PHYLOMARKERS con los cuales obtuvimos 547 marcadores, siendo 47 de estos los mejores para elaborar el árbol filogenético de *Cyclospora cayetanensis*. Con este repertorio de marcadores, se logró visualizar claras agrupaciones según el origen geográficos de algunos aislados (China, Nepal e Indonesia), pero no para los aislados relacionados a México.

Se confirmó que el género *Cyclospora* posee hasta el momento 656 genes únicos respecto a otros parásitos del subfilum Apicomplexa, los cuales se involucran en diferentes procesos celulares como el metabolismo, invasión celular, como proteínas de superficie y hasta factores de virulencia; además, la mayoría de estos (355) son genes hipotéticos o sin caracterizar. Este número podría aumentar dado que, de acuerdo con los resultados, se estima que el pangenoma de *C. cayetanensis* es un pangenoma abierto. En consecuencia, esto significa que el repertorio de genes irá en aumento considerable conforme incrementa el número de genomas reportados y añadidos al análisis.

Se evidenció la necesidad de aumentar en el número de secuencias del genoma con certeza de origen de aislamiento para poder lograr una caracterización más profunda de parásito, tanto por su pangenoma abierto como debido a la carencia en información epidemiológica de las cepas asociadas a México.

## 8. RECOMENDACIONES

En esta investigación se identificaron 47 potenciales marcadores como los mejores de 547 para *C. cayetanensis*, con los cuales se estimó un árbol filogenético con agrupaciones estrechamente relacionadas por su origen geográfico. Por lo que el estudio más profundo de estos (seleccionar aquellos que muestren mejores características, como densidad de SNPs) y su evaluación podría dirigirse a establecer más marcadores candidatos que podrían ser añadidos con otros marcadores MLST que coadyuven a dar una mayor resolución al genotipo de *C. cayetanensis*. El pipeline de GET\_HOMOLOGUES y GET\_PHYLOMARKERS puede ser utilizado para correr una nueva base de datos con secuencias con certidumbre de origen mexicano, lo cual permitirá inferir con certeza las características que podrían tener las cepas del país y establecer mejores relaciones filogenéticas. Además, los pipelines tienen más parámetros/scripts que podrían ser explorados, u obtener muchos más datos respecto al genotipo de *C. cayetanensis*. Por ejemplo, GET\_HOMOLOGUES-EST puede ser utilizado para el análisis de transcritos y GET\_PHYLOMARKERS tiene dos modos de ejecución: -R1 para análisis filogenético y -R2 para análisis de genética de poblaciones. En este trabajo sólo utilizamos el modo -R1 para análisis filogenético.

## 9. REFERENCIAS

- Adam, R. D., Ortega, Y. R., Gilman, R. H., y Sterling, C. R. (2000). Intervening transcribed spacer region 1 variability in *Cyclospora cayetanensis*. *Journal of Clinical Microbiology*, 38(6), 2339–2343.
- Barratt, J. L. N., Park, S., Nascimento, F. S., Hofstetter, J., Plucinski, M., Casillas, S., ... Talundzic, E. (2019). Genotyping Genetically Heterogeneous *Cyclospora cayetanensis* Infections to Complement Epidemiological Case Linkage. *Parasitology*, 1–33. <https://doi.org/10.1017/S0031182019000581>
- Center for disease control and prevention (CDC). (2020a). List of Selected Multistate Foodborne Outbreak Investigations. <https://www.cdc.gov/foodsafety/outbreaks/multistateoutbreaks/outbreaks-list.html>
- Center for disease control and prevention (CDC). (2020b). Parasites - Cyclosporiasis (*Cyclospora* infection). <https://www.cdc.gov/parasites/cyclosporiasis>
- Chacín-Bonilla, L. (2008). Transmission of *Cyclospora cayetanensis* infection: a review focusing on soil-borne cyclosporiasis. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 102(3), 215–216. <https://doi.org/10.1016/j.trstmh.2007.06.005>
- Chacín-Bonilla, L. (2010, September). Epidemiology of *Cyclospora cayetanensis*: A review focusing in endemic areas. *Acta Tropica*, Vol. 115, pp. 181–193. <https://doi.org/10.1016/j.actatropica.2010.04.001>
- Chacín-Bonilla, L., y Barrios, F. (2011). Chacín-Bonilla L, Barrios F REVISIÓN DE TEMA *Cyclospora cayetanensis*: biología, distribución ambiental y transferencia. *Biomédica*, 31, 132–175. Retrieved from <http://www.scielo.org.co/pdf/bio/v31n1/v31n1a16.pdf>
- Cinar, H. N., Gopinath, G., Murphy, H. R., Almeria, S., Durigan, M., Choi, D., y Da Silva, A. (2020). Molecular typing of *Cyclospora cayetanensis* in produce and clinical samples using targeted enrichment of complete mitochondrial genomes and next-generation sequencing. *Parasites and Vectors*, 13(1), 1–12. <https://doi.org/10.1186/s13071-020-3997-3>
- Contreras-Moreira, B., Cantalapiedra, C. P., García-Pereira, M. J., Gordon, S. P., Vogel, J. P., Igartua, E., y Vinuesa, P. (2017). Analysis of plant pan-genomes and transcriptomes with GET\_HOMOLOGUES-EST, a clustering solution for sequences of the same species. *Frontiers in Plant Science*, 1–16. <https://doi.org/10.3389/fpls.2017.00184>
- Contreras-Moreira, B., y Vinuesa, P. (2013). GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Applied and Environmental Microbiology*, 79(24), 7696–7701. <https://doi.org/10.1128/AEM.02411-13>
- Dubey, J. P., Almeria, S., Mowery, J., y Fortes, J. (2020). Endogenous Developmental Cycle of the Human Coccidian *Cyclospora cayetanensis*. *Journal of Parasitology*, 106(2), 295. <https://doi.org/10.1645/20-21>
- Eberhard, M. L., Da Silva, A. J., Lilley, B. G., y Pieniazek, N. J. (1999). Morphologic and molecular characterization of new *Cyclospora* species from Ethiopian monkeys: C.

*cercopithecii* sp.n., *C. colobi* sp.n., and *C. papionis* sp.n. *Emerging Infectious Diseases*, 5(5), 651–658.

- Eberhard, M. L., Pieniazek, N. J., y Arrowood, M. J. (1997). Laboratory diagnosis of *Cyclospora* infections. *Archives of Pathology and Laboratory Medicine*, Vol. 121, pp. 792–797.
- Eisen, J. A., y Fraser, C. M. (2003). Phylogenomics: Intersection of evolution and genomics. *Science*, 300(5626), 1706–1707. <https://doi.org/10.1126/science.1086292>
- Food and Drugs Administration (FDA). (2018a). BAM 19a: Detection of *Cyclospora* and *Cryptosporidium*. <https://www.fda.gov/food/laboratory-methods-food-safety/bam-19a-detectionCyclospora-and-cryptosporidium#slide1>
- Food and Drugs Administration (FDA). (2018b). BAM 19b: Molecular Detection of *Cyclospora cayetanensis* in Fresh Produce Using Real-Time PCR. <https://www.fda.gov/food/laboratorymethods-food-safety/bam-19b-molecular-detection-Cyclospora-cayetanensis-fresh-produceusing-real-time-pcr>
- Galván-Ramírez, M. de la L., Madriz-Elisondo, A. L., Ramírez Temores, C. G., Romero Rameño, J. de J., De La O Carrasco, D. A., y Cardona López, M. A. (2019). Enteroparasitism and risk factors associated with clinical manifestations in children and adults of Jalisco state in western Mexico. *Osong Public Health and Research Perspectives*, 10(1), 39–48. <https://doi.org/10.24171/j.phrp.2019.10.1.08>
- Giangaspero, A., y Gasser, R. B. (2019, March). Human cyclosporiasis. *The Lancet Infectious Diseases*, Vol. 19, pp. e226–e236. [https://doi.org/10.1016/S1473-3099\(18\)30789-8](https://doi.org/10.1016/S1473-3099(18)30789-8)
- Giangaspero, A., Marangi, M., Koehler, A. V., Papini, R., Normanno, G., Lacasella, V., ... Gasser, R. B. (2015). Molecular detection of *Cyclospora* in water, soil, vegetables and humans in southern Italy signals a need for improved monitoring by health authorities. *International Journal of Food Microbiology*, 211, 95–100. <https://doi.org/10.1016/j.ijfoodmicro.2015.07.002>
- Guimarães, L. C., Benevides De Jesus, L., Vinícius, M., Viana, C., Silva, A., Thiago, R., y Azevedo, V. (2015). Inside the Pan-genome-Methods and Software Overview. *Current Genomics*. 245–252.
- Guo, Y., Roellig, D. M., Li, N., Tang, K., Frace, M., Ortega, Y., Xiao, L. (2016). Multilocus sequence typing tool for *Cyclospora cayetanensis*. *Emerging Infectious Diseases*, 22(8), 1464–1467. <https://doi.org/10.3201/eid2208.150696>
- Hoff, K. J., y Stanke, M. (2013). WebAUGUSTUS--a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Research*, 41(Web Server issue), 123–128. <https://doi.org/10.1093/nar/gkt418>
- Houghton, K. A., Lomsadze, A., Park, S., Nascimento, F. S., Barratt, J., Arrowood, M. J., ... Qvarnstrom, Y. (2020). Development of a workflow for identification of nuclear genotyping markers for *Cyclospora cayetanensis*. *Parasite*, 27(10). <https://doi.org/10.1051/parasite/2020022>
- Kitajima, M., Haramoto, E., Iker, B. C., y Gerba, C. P. (2014). Occurrence of *Cryptosporidium*, *Giardia*, and *Cyclospora* in influent and effluent water at wastewater treatment plants in Arizona. *Science of the Total Environment*, 484(1), 129–136.

<https://doi.org/10.1016/j.scitotenv.2014.03.036>

- Letunic, I., y Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, 47(W1), 2–5. <https://doi.org/10.1093/nar/gkz239>
- Li, J., Chang, Y., Shi, K., Wang, R., Fu, K., Li, S., ... Zhang, L. (2017). Multilocus sequence typing and clonal population genetic structure of *Cyclospora cayetanensis* in humans. *Parasitology*, 144(14), 1890–1897. <https://doi.org/10.1017/S0031182017001299>
- Marangi, M., Koehler, A. V., Zanzani, S. A., Manfredi, M. T., Brianti, E., Giangaspero, A., y Gasser, R. B. (2015). Detection of *Cyclospora* in captive chimpanzees and macaques by a quantitative PCR-based mutation scanning approach. *Parasites and Vectors*, 8(1), 1–5. <https://doi.org/10.1186/s13071-015-0872-8>
- Murphy, H. R., Cinar, H. N., Gopinath, G., Noe, K. E., Chatman, L. D., Miranda, N. E., da Silva, A. J. (2018). Interlaboratory validation of an improved method for detection of *Cyclospora cayetanensis* in produce using a real-time PCR assay. *Food Microbiology*, 69, 170–178. <https://doi.org/10.1016/j.fm.2017.08.008>
- Olivier, C., Van De Pas, S., Lepp, P. W., Yoder, K., y Relman, D. A. (2001). Sequence variability in the first internal transcribed spacer region within and among *Cyclospora* species is consistent with polyparasitism. *International Journal for Parasitology*, 31(13), 1475–1487. [https://doi.org/10.1016/S0020-7519\(01\)00283-1](https://doi.org/10.1016/S0020-7519(01)00283-1)
- Onstad, N. H., Miller, M. R., Green, M. L., Witola, W. H., y Davidson, P. C. (2019). A review of *Cyclospora cayetanensis* transport in the environment. *Transactions of the ASABE*, 62(3), 795–802. <https://doi.org/10.13031/trans.13313>
- Orozco-Mosqueda, G. E., Martínez-Loya, O. A., y Ortega, Y. R. (2014). *Cyclospora cayetanensis* in a Pediatric Hospital in Morelia, México. *The American Journal of Tropical Medicine and Hygiene*, 91(3), 537–540. <https://doi.org/10.4269/ajtmh.13-0535>
- Ortega, Ynés R., y Sanchez, R. (2010). Update on *Cyclospora cayetanensis*, a Food-Borne and Waterborne Parasite. *Clinical Microbiology Reviews*, 23(1), 218–234. <https://doi.org/10.1128/CMR.00026-09>
- Ortega, Ynes R., y Sherchand, J. B. (2017). *Cyclospora cayetanensis*. *Biology of Foodborne Parasites*. <https://doi.org/10.1201/b18317>
- Ortega, Ynés R., Sterling, C., y Gilman, R. (1998). *Cyclospora cayetanensis*. *Advances in Parasitology*, 40, 399–418. [https://doi.org/10.1016/S0065-308X\(08\)60128-1](https://doi.org/10.1016/S0065-308X(08)60128-1)
- Ortega, Ynes R., Sterling, C. R., Gilman, R. H., Cama, V. A., y Diaz, F. (1993). *Cyclospora* Species -- A New Protozoan Pathogen of Humans. *New England Journal of Medicine*, 328(18), 1308–1312. <https://doi.org/10.1056/NEJM199305063281804>
- Qvarnstrom, Y., Benedict, T., Marcet, P. L., Wiegand, R. E., Herwaldt, B. L., y Da Silva, A. J. (2018). Molecular detection of *Cyclospora cayetanensis* in human stool specimens using UNEX-based DNA extraction and real-time PCR. *Parasitology*, 145(7), 865–870. <https://doi.org/10.1017/S0031182017001925>
- Qvarnstrom, Y., Wei-Pridgeon, Y., Li, W., Nascimento, F. S., Bishop, H. S., Herwaldt, B. L., y Arrowood, M. J. (2015). Draft genome sequences from *Cyclospora cayetanensis* oocysts purified from a human stool sample. *Genome Announcements*, 3(6), 10–11.

<https://doi.org/10.1128/genomeA.01324-15>

- Qvarnstrom, Y., Wei-Pridgeon, Y., Van Roey, E., Park, S., Srinivasamoorthy, G., Nascimento, F. S., y Arrowood, M. J. (2018). Purification of *Cyclospora cayetanensis* oocysts obtained from human stool specimens for whole genome sequencing. *Gut Pathogens*, *10*(1), 1–9. <https://doi.org/10.1186/s13099-018-0272-7>
- Rabold, J. G., Hoge, C., Shlim, D., Kefford, C., Rajah, R., & Echeverria, P. (1994). *Cyclospora* outbreak associated with chlorinated drinking water. *The Lancet*, *344*(8933), 1360-1361.
- Reiman, D. A., Schmidt, T. M., Gajadhar, A., Sogin, M., Cross, J., Yoder, K., y Echeverria, P. (1996). Molecular Phylogenetic Analysis of *Cyclospora*, the Human Intestinal Pathogen, Suggests that It Is Closely Related to *Eimeria* Species. *Journal of Infectious Diseases*, *173*(2), 440–445. <https://doi.org/10.1093/infdis/173.2.440>
- Riner, D. K., Nichols, T., Lucas, S. Y., Mullin, A. S., Cross, J. H., y Lindquist, H. D. A. (2010). Intragenomic Sequence Variation of the ITS-1 Region Within a Single Flow-Cytometry–Counted *Cyclospora cayetanensis* Oocysts. *Journal of Parasitology*, *96*(5), 914–919. <https://doi.org/10.1645/ge-2505.1>
- Rouli, L., Merhej, V., Fournier, P. E., y Raoult, D. (2015). The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes and New Infections*, *7*, 72–85. <https://doi.org/10.1016/j.nmni.2015.06.005>
- Sato, S. (2011). The apicomplexan plastid and its evolution. *Cellular and Molecular Life Sciences*, *68*(8), 1285–1296. <https://doi.org/10.1007/s00018-011-0646-1>
- Shields, J. M., y Olson, B. H. (2003). *Cyclospora cayetanensis*: A review of an emerging parasitic coccidian. *International Journal for Parasitology*, *33*(4), 371–391. [https://doi.org/10.1016/S0020-7519\(02\)00268-0](https://doi.org/10.1016/S0020-7519(02)00268-0)
- Smith, H. V, Paton, C. A., Mtambo, M. M. A., y Girdwood, R. W. A. (1997). Sporulation of *Cyclospora* sp. oocysts. *Applied and Environmental Microbiology*, *63*(4), 1631–1632.
- Sulaiman, I. M., Ortega, Y., Simpson, S., y Kerdahi, K. (2014). Genetic characterization of human-pathogenic *Cyclospora cayetanensis* parasites from three endemic regions at the 18S ribosomal RNA locus. *Infection, Genetics and Evolution*, *22*, 229–234. <https://doi.org/10.1016/j.meegid.2013.07.015>
- Sulaiman, I. M., Torres, P., Simpson, S., Kerdahi, K., y Ortega, Y. (2013). Sequence Characterization of Heat Shock Protein Gene of *Cyclospora cayetanensis* Isolates from Nepal, Mexico, and Peru. *Journal of Parasitology*, *99*(2), 379–382. <https://doi.org/10.1645/ge-3114.1>
- Tang, K., Guo, Y., Zhang, L., Rowe, L. A., Roellig, D. M., Frace, M. A., ... Xiao, L. (2015). Genetic similarities between *Cyclospora cayetanensis* and cecum-infecting avian *Eimeria* spp. in apicoplast and mitochondrial genomes. *Parasites and Vectors*, *8*(1), 1–11. <https://doi.org/10.1186/s13071-015-0966-3>
- Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., ... Fraser, C. M. (2005). Erratum: Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial “pan-genome”. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(45), 16530.

<https://doi.org/10.1073/pnas.0508532102>

- Tettelin, H., Riley, D., Cattuto, C., y Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology*, 11(5), 472–477. <https://doi.org/10.1016/j.mib.2008.09.006>
- Tram, N. T., Hoang, L. M. N., Cam, P. D., Chung, P. T., Fyfe, M. W., Isaac-Renton, J. L., y Ong, C. S. L. (2008). *Cyclospora* spp. in herbs and water samples collected from markets and farms in Hanoi, Vietnam. *Tropical Medicine and International Health*, 13(11), 1415–1420. <https://doi.org/10.1111/j.1365-3156.2008.02158.x>
- Vazquez-Tsuji, Ó., Jiménez-Domínguez, R., Rivera, T. C., Rojas, S. V., Romero Cabello, R., Aranda, V. G., y Martínez-Barbabosa, I. (2000). Infección por *Cyclospora cayetanensis*. Diagnóstico de Laboratorio. *Revista Latinoamericana de Microbiología*, 42, 45–52.
- Vinuesa, P., y Contreras-Moreira, B. (2015). Robust Identification of Orthologues and Paralogues for Microbial Pan-Genomics Using GET\_HOMOLOGUES: A Case Study of pIncA/C Plasmids. In *Methods in molecular biology (Clifton, N.J.)* (Vol. 1231, pp. 203–232). [https://doi.org/10.1007/978-1-4939-1720-4\\_14](https://doi.org/10.1007/978-1-4939-1720-4_14)
- Vinuesa, P., Ochoa-Sánchez, L. E., y Contreras-Moreira, B. (2018). GET\_PHYLOMARKERS, a software package to select optimal orthologous clusters for phylogenomics and inferring pan-genome phylogenies, used for a critical geno-taxonomic revision of the genus *Stenotrophomonas*. *Frontiers in Microbiology*, 9(MAY), 1–22. <https://doi.org/10.3389/fmicb.2018.00771>
- Zekic, T., Holley, G., y Stoye, J. (2018). Pan-genome storage and analysis techniques. *Methods in Molecular Biology*, 1704, 29–53. [https://doi.org/10.1007/978-1-4939-7463-4\\_2](https://doi.org/10.1007/978-1-4939-7463-4_2)
- Zhou, Y., Lv, B., Wang, Q., Wang, R., Jian, F., Zhang, L., y Xiao, L. (2011). Prevalence and molecular characterization of *Cyclospora cayetanensis*, Henan, China. *Emerging Infectious Diseases*, 17(10), 1887–1890. <https://doi.org/10.3201/eid1710.101296>