

**Análisis genómico comparativo con especies y subgenotipos  
del apicomplexa *Cryptosporidium***

Laura Marcela Arias Agudelo

Trabajo de grado para optar al título de Magister en Ciencias Básicas  
Biomédicas con énfasis en Microbiología y Parasitología

Tutora:

Ana Luz Galván Díaz. MSc, PhD.

Comité asesor:

Ángela Magnet Dávila. PhD, Gisela María García Montoya. MSc, PhD y  
Juan Fernando Alzate Restrepo. MSc, PhD

Universidad de Antioquia  
Corporación de Ciencias Básicas Biomédicas  
Julio de 2020

## RESUMEN

*Cryptosporidium* es un apicomplexa asociado con diarrea, principalmente en niños y pacientes inmunocromprometidos. En la actualidad se describen 38 especies que difieren en cuanto a la especificidad de hospedero, hábitat, distribución geográfica y virulencia. Además, se han descrito subgenotipos que representan variaciones intraespecie basadas en el polimorfismo del gen *gp60*. Aunque se tienen disponibles genomas de diferentes especies y subgenotipos, son escasos los análisis comparativos que estudien las bases genéticas que expliquen las diferencias interespecie e intragenotípicas descritas para este protozoo. Teniendo en cuenta lo anterior, y con el propósito de contribuir al conocimiento de la genómica y taxonomía en este género, se realizó un estudio comparativo y filogenómico con 23 genomas de las especies más frecuentes en humanos: *C. hominis*, *C. parvum* y *C. meleagridis*.

Los genomas analizados presentaron un tamaño aproximado de 9,0Mb. Se observaron porcentajes de identidad frente al genoma de referencia *C. parvum* Iowa II del 96,8% en los aislados de *C. hominis* y 91,5% en *C. meleagridis*. En *C. parvum* se identificaron porcentajes de identidad entre un 99,7 y 99,9% para aislados zoonóticos y del 99,5% para antroponóticos. Además, en esta especie se identificó una mayor acumulación de variantes de un solo nucleótido (*SNVs*) en aislamientos antroponóticos en comparación con los zoonóticos.

La mayoría de los *SNVs* e inserciones y deleciones (*indels*) se encontraron en regiones codificantes, observándose una distribución homogénea en los cromosomas de *C. parvum*, mientras que en *C. hominis* y *C. meleagridis* predominaron en los cromosomas 1 y 3. Los genes con cambios deletéreos e *indels* anotados, están asociados con el procesamiento de la información genética, procesos enzimáticos y metabólicos; no obstante, alrededor del 50% de los genes

codifican proteínas hipotéticas.

El análisis filogenómico se realizó a partir de una matriz generada con 800.861 *SNVs*. El árbol obtenido mostró tres cladas monofiléticas para todos los aislados, observándose la presencia de dos ramas separadas para los aislados antroponóticos y zoonóticos de *C. parvum*. Todos los genomas de *C. hominis* y *C. parvum* se agruparon de acuerdo con la familia de subgenotipos basada en el gen *gp60*.

Hasta el momento, este estudio es el análisis filogenómico y comparativo más robusto realizado en el género *Cryptosporidium*, que incluye genomas completos de diferentes aislados, familias alélicas y subgenotipos de las tres principales especies intestinales de *Cryptosporidium* asociadas con infección en el hombre.

**Palabras clave:** *Cryptosporidium*, biología molecular, biología computacional, secuenciación de nucleótidos de alto rendimiento, secuenciación completa del genoma, genómica, variación genética.

## SUMMARY

*Cryptosporidium* is an apicomplexan associated with diarrhea mainly in children and immunocompromised patients. Currently, there are 38 species that differ in host specificity, habitat, geographic distribution, and virulence. Additionally, subtype families that show intraspecies variations based on the polymorphism of the gp60 gene are described. Although genomes of several species and subtypes are available, a comparative analysis that explores the genetic bases of the phenotypic differences within the genus are scarce. According to the above, and to contribute to the knowledge of genomics and taxonomy in *Cryptosporidium*, a comparative and phylogenomic study was conducted with 23 genomes of the most frequent species in humans: *C. hominis*, *C. parvum*, and *C. meleagridis*.

*Cryptosporidium* genomes have a size of 9.0Mb approximately. Comparative analysis was done using *C. parvum* Iowa II as a reference genome. Identity percentages of 96.8% and 91.5% were observed in *C. hominis* and *C. meleagridis* respectively. Intra-genotypic analysis in *C. parvum* showed an identity percentage between 99.7 and 99.9% for zoonotic isolates, and 99.5% for anthroponotic isolates. Additionally, a higher accumulation of SNVs was identified in anthroponotic isolates compared to the zoonotic ones.

Most of the SNVs and indels detected in the genomes were located in coding regions, showing a homogeneous distribution on the chromosomes in *C. parvum*; while a predominance of these variants was observed on the chromosomes 1 and 3 in both *C. hominis* and *C. meleagridis*. Genes with deleterious changes and indels that were successfully annotated were associated with the processing of genetic information and enzymatic and metabolic processes; however, about 50% of genes encode hypothetical proteins.

The phylogenomic analysis was performed from a matrix generated with 800,861 SNVs. The tree obtained showed three monophyletic clades for all isolates. In *C. parvum*, the presence of two separate branches was observed for anthroponotic and zoonotic isolates. All *C. hominis* and *C. parvum* genomes were grouped according to the gp60 subtype family.

This study is the most robust phylogenomic and genomic comparative analysis performed in the genus *Cryptosporidium*, which includes complete genomes from different isolates, allelic families, and subtypes of the three most frequent intestinal species of *Cryptosporidium* in humans.

**Keywords:** *Cryptosporidium*, molecular biology, computational biology, high-throughput nucleotide sequencing, whole-genome sequencing, genomics, genetic variation.

## ABREVIATURAS

**ICZN:** Código Internacional para la Nomenclatura Zoológica

**Pb:** Pares de bases

**Mb:** Megabases

**kDa:** KiloDalton

**ADN:** Ácido desoxirribonucleico

**ARN:** Ácido ribonucleico

**Indels:** Inserciones y deleciones

**gp60:** Gen que codifica para la glicoproteína de 60 kDa

**Proteínas MEDLE:** Proteínas secretadas específicas de *Cryptosporidium*

**tp14:** Gen que codifica para una proteína hipotética

**cp47:** Gen que codifica para una proteína de 47 kDa

**ms6-7:** Gen que codifica para el antígeno de serina

**Hsp70:** Proteína de shock térmico de 70 kDa

**CDSs:** Secuencias codificantes

**NGS:** Secuenciación de segunda generación

**WGS:** Secuenciación de genoma completo

**SNV:** Variante de un solo nucleótido

**sSNV:** Variante de un solo nucleótido sinónima

**nsSNVs:** Variante de un solo nucleótido no sinónima

**TVM:** Modelo de transversiones con frecuencia de bases desigual

**BIC:** Criterio de información bayesiano

## ÍNDICE

<b>RESUMEN</b> .....	<b>2</b>
<b>SUMMARY</b> .....	<b>4</b>
<b>ÍNDICE DE TABLAS</b> .....	<b>9</b>
<b>ÍNDICE DE FIGURAS</b> .....	<b>10</b>
<b>1. INTRODUCCIÓN</b> .....	<b>11</b>
1.1. HISTORIA .....	12
1.2. TAXONOMÍA .....	13
1.2.1. ESPECIES.....	14
1.3. CICLO DE VIDA.....	17
1.4. CRIPTOSPORIDIOSIS.....	21
1.5. GENOTIPIFICACIÓN MOLECULAR Y DIVERSIDAD GENÉTICA.....	23
1.6. EPIDEMIOLOGÍA Y DISTRIBUCIÓN GEOGRÁFICA.....	26
1.7. ESTUDIOS GENÓMICOS EN EL GÉNERO <i>CRYPTOSPORIDIUM</i> .....	29
1.7.1. Cariotipo.....	29
1.7.2. Genoma.....	30
1.7.3. Genómica Comparativa.....	36
1.8. JUSTIFICACIÓN DEL TRABAJO .....	40
<b>2. HIPÓTESIS</b> .....	<b>41</b>
<b>3. PREGUNTA DE INVESTIGACIÓN</b> .....	<b>43</b>
<b>4. OBJETIVOS</b> .....	<b>45</b>
4.1. OBJETIVO GENERAL .....	46
4.2. OBJETIVOS ESPECÍFICOS.....	46
<b>5. METODOLOGÍA</b> .....	<b>47</b>
5.1. OBTENCIÓN DE DATOS NGS, ENSAMBLAJE Y FILTRADO DE CONTIGS.....	48
5.1.1. Obtención de los reads .....	48
5.1.2. Fusión de los reads.....	48
5.1.3. Ensamblaje de novo software SPAdes.....	48
5.1.4. Identificación de contigs pertenecientes al género <i>Cryptosporidium</i> ....	48
5.1.5. Extracción de contigs y estadísticas de los ensamblajes .....	49
5.2. ANÁLISIS ESTRUCTURAL.....	49
5.2.1. Identificación de variantes de un solo nucleótido.....	49
5.2.3. Identificación de inserciones y deleciones .....	49
5.3. ANÁLISIS FUNCIONAL .....	50
5.3.1. Identificación de <i>SNVs</i> en regiones codificantes .....	50
5.3.2. Anotación funcional de los <i>SNVs</i> detectados en CDSs .....	50

5.3.3 Identificación y anotación de los genes con cambios no sinónimos deletéreos .....	50
5.4. ANÁLISIS FILOGENÓMICO DE LAS PRINCIPALES ESPECIES DE <i>CRYPTOSPORIDIUM</i> .....	51
<b>6. RESULTADOS .....</b>	<b>53</b>
6.1. AISLADOS SELECCIONADOS.....	54
6.2. ENSAMBLAJE DE LOS GENOMAS .....	56
6.3. IDENTIFICACIÓN DE VARIANTES DE UN SOLO NUCLEÓTIDO ( <i>SNVs</i> ).....	59
6.3.1. Acumulación de <i>SNVs</i> inter e intraespecie.....	59
6.3.2. Acumulación de <i>SNVs</i> en los cromosomas.....	63
6.3.3. Identificación de <i>SNVs</i> en regiones codificantes .....	65
6.4. IDENTIFICACIÓN Y ANOTACIÓN DE GENES CON MUTACIONES DELETÉREAS COMPARTIDOS Y EXCLUSIVOS DE ESPECIE .....	69
6.4.1. Acumulación de genes con mutaciones deletéreas intraespecie y por cromosoma .....	73
6.5 INSERCIONES Y DELECCIONES – INDELS.....	76
6.5.1 Acumulación de Inserciones inter e intraespecie.....	76
6.5.2. Identificación de inserciones en regiones codificantes y no codificantes .....	77
6.5.3. Acumulación de inserciones por cromosoma .....	78
6.5.4. Acumulación de deleciones inter e intraespecie .....	80
6.5.5. Identificación de deleciones en regiones codificantes y no codificantes	81
6.5.6. Acumulación de deleciones por cromosoma .....	82
6.5.7. Identificación de genes con eventos de inserción y deleción inter e intraespecie.....	84
6.6. ANÁLISIS FILOGENÓMICO CON LAS TRES PRINCIPALES ESPECIES DE <i>CRYPTOSPORIDIUM</i> ASOCIADAS CON ENFERMEDAD EN EL HUMANO .....	92
<b>7. DISCUSIÓN .....</b>	<b>94</b>
<b>8. CONCLUSIONES .....</b>	<b>112</b>
<b>9. MATERIAL SUPLEMENTARIO Y ANEXOS .....</b>	<b>115</b>
<b>10. BIBLIOGRAFÍA .....</b>	<b>119</b>

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Resumen cronológico de la nueva clasificación taxonómica del género <i>Cryptosporidium</i> .....	14
<b>Tabla 2.</b> Especies válidas en el género <i>Cryptosporidium</i> .....	16
<b>Tabla 3.</b> Familias de subgenotipos gp60..	25
<b>Tabla 4.</b> Cromosomas de <i>Cryptosporidium parvum</i> .....	30
<b>Tabla 5.</b> Proyectos de Secuenciación de diferentes especies y aislados de <i>Cryptosporidium</i> .....	33
<b>Tabla 6.</b> Características genómicas de diferentes especies en el género <i>Cryptosporidium</i> .....	39
<b>Tabla 7.</b> Descripción de los proyectos genómicos de <i>Cryptosporidium</i> incluidos en los análisis comparativos.....	55
<b>Tabla 8.</b> Estadísticas de los ensamblajes en los aislados de <i>Cryptosporidium</i> .....	58
<b>Tabla 9.</b> Resultado de los alineamientos entre los aislados seleccionados y el genoma de referencia <i>C. parvum</i> Iowa II.....	60
<b>Tabla 10.</b> Acumulación de <i>SNVs</i> en regiones codificantes .....	66
<b>Tabla 11.</b> Anotación de los genes con mutaciones deletéreas exclusivos de especie .....	71
<b>Tabla 12.</b> Número de genes con eventos de inserción y deleción respecto al genoma de <i>C. parvum</i> Iowa II .....	85
<b>Tabla 13.</b> Anotación de los genes afectados por eventos de <i>indels</i> .....	90

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Ciclo de vida de <i>Cryptosporidium</i> .....	20
<b>Figura 2.</b> Acumulación de <i>SNVs</i> interespecie.....	61
<b>Figura 3.</b> Acumulación de <i>SNVs</i> intraespecie.....	62
<b>Figura 4.</b> Acumulación de <i>SNVs/Kb</i> en cada cromosoma.....	64
<b>Figura 5.</b> Acumulación intraespecie de <i>sSNVs</i> y <i>nsSNV</i> en regiones codificantes.	68
<b>Figura 6.</b> Genes con mutaciones deletéreas en las especies evaluadas. ....	69
<b>Figura 7.</b> Análisis de ortología KEGG en los genes con mutaciones deletéreas compartidos por las tres especies de <i>Cryptosporidium</i> . ....	70
<b>Figura 8.</b> Análisis de ortología KEGG en los genes con mutaciones deletéreas exclusivas de especie.....	72
<b>Figura 9.</b> Acumulación de genes con mutaciones deletéreas intraespecie. ....	74
<b>Figura 10.</b> Acumulación de genes con <i>nsSNVs</i> deletéreos/Kb en cada cromosoma. ....	75
<b>Figura 11.</b> Acumulación de inserciones intraespecie.....	76
<b>Figura 12.</b> Acumulación de inserciones intraespecie en regiones codificantes y no codificantes.....	77
<b>Figura 13.</b> Acumulación de inserciones por cromosoma.....	79
<b>Figura 14.</b> Acumulación de deleciones intraespecie.....	80
<b>Figura 15.</b> Acumulación de deleciones intraespecie en regiones codificantes y no codificantes.....	81
<b>Figura 16.</b> Acumulación de deleciones por cromosoma.....	83
<b>Figura 17.</b> Número de genes afectados por eventos indels en las tres especies de <i>Cryptosporidium</i> . ....	87
<b>Figura 18.</b> Genes con eventos de inserción y deleción en las especies evaluadas. ....	88
<b>Figura 19.</b> Árbol filogenómico de las tres principales especies de <i>Cryptosporidium</i> ....	93

## **1. INTRODUCCIÓN**

## 1.1. Historia

El género *Cryptosporidium* está conformado por parásitos apicomplexa, los cuales infectan las células epiteliales del tracto gastrointestinal de un amplio rango de vertebrados (1). *Cryptosporidium* fue descrito en 1907 por el parasitólogo Ernest Tyzzer, quien lo descubrió mientras estudiaba las glándulas gástricas de ratones (2), y para el año 1910 propuso el nombre *Cryptosporidium* como género y *C. muris* para la especie identificada. Tyzzer también postuló que los esporozoítos de los ooquistes podrían ser los responsables de invadir las células. En 1912, este mismo autor describió una nueva especie, *C. parvum* y demostró que los ooquistes eran más pequeños que los de *C. muris* (3).

Durante los años siguientes a la primera publicación de Tyzzer, la importancia económica, y en salud humana y animal de *Cryptosporidium* no se había reconocido (1). En 1955, una nueva especie encontrada en pavos fue denominada *C. meleagridis* (3). Para el año 1976, dos grupos informaron por separado los primeros casos de criptosporidiosis en humanos (4,5), correspondientes a un niño de 3 años y un paciente inmunosuprimido de 39 años, quienes compartían antecedentes clínico-epidemiológicos que incluían entre otros vivir en una granja y presentar diarrea acuosa.

Finalmente, en la década de los 80s, *Cryptosporidium* fue reconocido como un agente importante de diarrea en pacientes VIH/SIDA, lo que aumentó el interés en el estudio de este protozoo. En 1982, el Centro para el Control y Prevención de Enfermedades (CDC) de Estados Unidos describió 21 casos de criptosporidiosis crónica en portadores del VIH, y en los siguientes años este número aumentó considerablemente, confirmando la infección por *Cryptosporidium* como una parasitosis frecuente en esta población (3). En 1993, a raíz de un brote masivo en Milwaukee (Wisconsin, USA), que afectó aproximadamente a 403.000 personas y

que fue asociado a *Cryptosporidium*, se identificó el gran potencial de transmisión hídrica de este parásito, siendo en la actualidad el responsable de la mayoría de los brotes de diarrea por transmisión hídrica relacionados con protozoos (6).

## 1.2. Taxonomía

Según la clasificación taxonómica clásica para protozoos unicelulares eucariotas desarrollada en 1970 por Levine (7), *Cryptosporidium* se clasifica en el *phylum* apicomplexa (presencia de complejo apical), clase conoidasida (presencia de conoide), subclase coccidiasina (parásitos intracelulares obligados), orden eucoccidiorida (parásitos formadores de esporas que afectan humanos y animales). Sin embargo, por décadas, la ubicación taxonómica de *Cryptosporidium* ha sido fuente de debate. En 1998, Bull y col. (8) describieron que anticuerpos monoclonales frente a *Cryptosporidium* presentaban reactividad cruzada contra esporozoítos de la gregarina *Monocystis* sp., sugiriendo una relación cercana entre ambos protozoos y aumentando el escepticismo en cuanto a la ubicación de *Cryptosporidium* dentro de las coccidias. Adicionalmente, *Cryptosporidium* y las gregarinas compartían otras características como una localización intracelular pero extra citoplasmática, la producción de ooquistes de pared delgada, resistencia a los fármacos anticoccidiales, ausencia de apicoplasto y la presencia de un orgánulo especializado para la alimentación (9).

En 1999, Carreño y col. (10), mediante el análisis de la secuencia del ADN de la subunidad pequeña ribosomal - SSU rRNA de *Cryptosporidium* y varias especies de gregarinas, establecieron que estos protozoos compartían un ancestro común. Estudios realizados en años posteriores confirmaron la cercanía del género *Cryptosporidium* con las gregarinas (**Tabla 1**). Adicionalmente, se han reportado estadios de vida libre similares a gamontes realizando sizigia, una asociación descrita en etapas extracelulares de diferentes especies de gregarinas (11,12).

En 2014, Cavalier-Smith (13) propuso una nueva clasificación taxonómica para el género *Cryptosporidium*, ubicándolo en un grupo taxonómico separado de las coccidias y cercano a las gregarinas, en la clase Gregarinomorpha y estableciendo una nueva subclase denominada Cryptogregaria en la cual se agrupan parásitos epicelulares de vertebrados, que carecen de apicoplasto y presentan un orgánulo alimenticio.

**Tabla 1. Resumen cronológico de la nueva clasificación taxonómica del género *Cryptosporidium*.** Adaptada de Thompson y col. (14)

<b>Descubrimiento</b>	<b>Año</b>
<b>Reactividad cruzada serológica con la gregarina <i>Monocystis</i> sp.</b>	1998
<b>Análisis filogenético basado en el estudio de la secuencia del ADN de la subunidad pequeña ribosomal - SSU rRNA, que demostró una relación filogenética más cercana con las gregarinas que con las coccidias.</b>	1999
<b>Descripción de nuevos estadios extracelulares de <i>C. andersoni</i> en cultivos celulares <i>in vitro</i>.</b>	2002
<b>Descripción de estadios de desarrollo extracelular en <i>C. parvum</i> en ambiente libre de células <i>in vitro</i></b>	2004
<b>Confirmación <i>in vitro</i> de asociación de estadios parasitarios mediante sizigia</b>	2004/2005
<b>Localización intracelular, pero extracitoplasmática similar a las gregarinas <i>in vivo</i></b>	2007
<b>Desenquistamiento y multiplicación de <i>Cryptosporidium</i> en biopelículas (confirmación de estadios extracelulares)</b>	2013
<b>Reclasificación de <i>Cryptosporidium</i> en una nueva clase y subclase dentro de las gregarinas, mediante análisis filogenéticos basados en el ADN del ssu rRNA de más de 500 organismos, incluyendo coccidias y gregarinas.</b>	2014
<b>Confirmación del desarrollo de merozoítos con regiones análogas al epimerito y protomerito de las gregarinas.</b>	2015

### 1.2.1. Especies

Para que una especie sea admitida en el género *Cryptosporidium* debe cumplir las normas y parámetros establecidos por el ICZN, que incluyen: descripción del tamaño

y morfología de los ooquistes; caracterización molecular basada en el gen 18S rRNA, demostración natural y experimental de la especificidad de hospedero; y finalmente la disponibilidad de las secuencias depositadas en GenBank (base de datos de secuencias genéticas del NIH-National Institutes of Health de Estados Unidos) (15). En la actualidad se aceptan 38 especies (16), de las cuales más de veinte están asociadas con enfermedad en el humano, siendo *C. parvum*, *C. hominis* y *C. meleagridis* las más frecuentes (16,17). Las especies en el género *Cryptosporidium* difieren en características fenotípicas como la especificidad de hospedero, pudiendo infectar mamíferos, reptiles y aves (15); también difieren en su hábitat, ya que se han adaptado a diversos ambientes (intestinal y gástrico) siendo el intestino delgado y el colon donde la mayoría se multiplican. En la Tabla 2, se presentan las especies válidas en el género *Cryptosporidium*.

**Tabla 2. Especies válidas en el género *Cryptosporidium***

<b>Especie</b>	<b>Ooquistes (µm)</b>	<b>Principal hospedero</b>	<b>Habitat</b>	<b>Reporte en humanos</b>
<i>C. hominis</i> *	4,9 x 5,2	Humanos	ID	+
<i>C. parvum</i>	5,0 x 4,5	Ovinos, bovinos	ID	+
<i>C. meleagridis</i>	5,2 x 4,6	Aves	ID	+
<i>C. felis</i>	4,6 x 4,0	Gatos	ID	a
<i>C. canis</i>	5,0 x 4,7	Perros	ID	a
<i>C. muris</i>	7,0 x 5,0	Roedores	Estómago	ψ
<i>C. suis</i>	4,6 x 4,2	Cerdos	ID e IG	ψ
<i>C. ubiquitum</i>	5,6 x 5,4	Ovejas	ID	ψ
<i>C. andersoni</i>	7,4 x 5,5	Bovinos	Abomaso o cuajar	ψ
<i>C. cuniculus</i>	4,6 x 4,2	Conejos	ID	a
<i>C. viatorum</i>	4,4 x 3,9	Humanos	ID	a
<i>C. bovis</i>	4,9 x 4,6	Bovinos	ID	ψ
<i>C. scrofarum</i>	4,0 x 4,7	Cerdos	ID	ψ
<i>C. erinacei</i>	4,9 x 4,4	Erizos	Intestino	ψ
<i>C. fayeri</i>	4,9 x 4,3	Marsupiales	ID	ψ
<i>C. tyzzeri</i>	5,4 x 4,7	Ratones	ID	ψ
<i>C. galli</i>	8,3 x 6,3	Gallinas	Proventrículo	~
<i>C. fragile</i>	6,2 x 5,5	Sapos	Estómago	~
<i>C. serpentis</i>	6,2 x 5,3	Serpientes	Estómago	~
<i>C. varanii</i>	4,8 x 4,7	Serpientes	Intestino y cloaca	~
<i>C. baileyi</i>	6,2 x 4,6	Gallinas	Cloaca y bursa	~
<i>C. macropodum</i>	5,4 x 4,9	Marsupiales	ID	~
<i>C. ryanae</i>	3,7 x 3,2	Bovinos	ID	~
<i>C. cichlidis</i>	4,6 x 4,4	Peces	Estómago	~
<i>C. molnari</i>	4,7 x 4,5	Peces	Estómago	~
<i>C. ducismarci</i>	5,0 x 4,8	Tortugas	Intestino	~
<i>C. huwi</i>	4,6 x 4,4	Peces	Estómago	~
<i>C. proliferans</i>	7,7 x 5,3	Roedores	Estómago	~
<i>C. reichenbachklinkei</i>	3,6 x 4,0	Peces	Estómago	~
<i>C. rubeyi</i>	4,7 x 4,3	Ardillas	Intestino y colon	~
<i>C. wrairi</i>	5,4 x 4,6	Cerdos de guinea	ID	~
<i>C. xiaoi</i>	3,9 x 3,4	Ovejas	Intestino	~
<i>C. homai</i>	< 5 µm**	Cerdos de guinea	Intestino	~
<i>C. avium</i>	6.26 x 4.86	Periquitos	Ileon y ciego	~
<i>C. apodeme</i>	4.2 x 4.0	Ratones	Intestino	~
<i>C. ditrichi</i>	4.7 x 4.2	Ratones	ID	~
<i>C. testudinis</i>	6.4 x 5.9	Tortugas	Intestino	~
<i>C. occultus</i>	5.2 x 4.9	Ratas salvajes	Colon	~

ID: Intestino delgado; IG: Intestino grueso; +: Comúnmente, <sup>a</sup>: Ocasionalmente,  $\psi$ : Raramente, ~: Nunca. \* *C. hominis* recientemente se ha aislado en primates no humanos, equinos, rumiantes, entre otros; \*\* Tamaño de trofozoítos y merontes. (15,18,19).

Además de las especies aceptadas, en el género *Cryptosporidium* se han reportado más de 60 genotipos, que si bien difieren en la secuencia del gen 18S rRNA, no alcanzan a tener suficientes datos adicionales que les confieran el estatus de especie. Hasta el momento siete genotipos se han asociado con enfermedad en humanos (genotipo del ciervo, mono, cerdo, zorrillo, caballo, conejo y ratón) (16,20,21).

### **1.3. Ciclo de vida**

*Cryptosporidium* puede comportarse como parásito de una gran variedad de hospederos vertebrados y algunos autores han descrito su presencia en moluscos bivalvos, los que podrían comportarse como vectores mecánicos para este protozoo (22,23). Adicionalmente, en los últimos años se ha demostrado la capacidad de completar su ciclo de vida en un ambiente libre de células (24).

Las especies del género *Cryptosporidium* presentan un ciclo de vida monoxeno, desarrollando todas sus fases en el tracto gastrointestinal (principalmente estómago e intestino delgado) de un único hospedero. La infección se produce luego de la ingestión de los ooquistes esporulados de pared gruesa (con cuatro esporozoítos en su interior) a partir de agua o alimentos contaminados. Los ooquistes pasan por el esófago hasta alcanzar el estómago donde son expuestos a enzimas gástricas y a condiciones reductoras, las cuales sumadas a moléculas endógenas del parásito (serina y cisteína proteasas, aminopeptidasas, fosfolipasas A2, proteínas de choque térmico, entre otras) establecen las condiciones favorables para que se dilate una pequeña sutura presente en uno de los extremos del ooquiste y los esporozoítos puedan ser liberados en el intestino delgado (25) (**Figura 1**). Una vez en la luz intestinal, los esporozoítos realizan movimientos circulares y helicoidales, que les

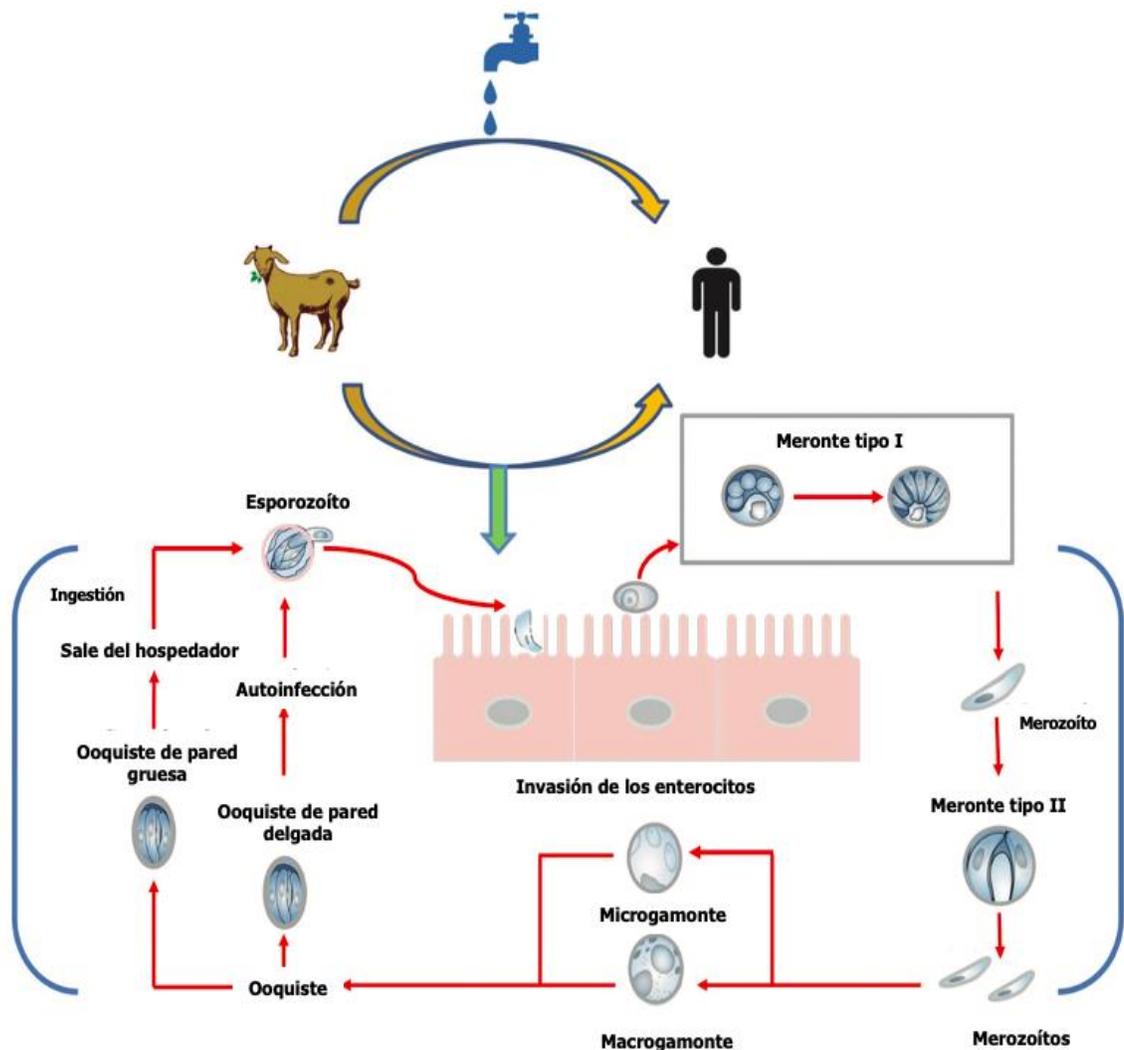
permiten acercarse y establecer contacto con la membrana de la célula hospedera (26). Posteriormente la única roptra de este estadio parasitario se extiende al sitio de unión, los micronemas y los gránulos densos se mueven a la región apical y se expresan una serie de glucoproteínas que contienen ligandos involucrados en la unión a las células epiteliales intestinales, tales como el antígeno tipo circumsporozoito (CSL), glicoproteínas tipo mucina (GP60, GP900), y proteínas adhesivas de la familia de las trombospondinas (TRAPs) (27).

Una vez se da la adhesión de los esporozoítos a la célula hospedera, se inicia el proceso de invasión, el cual involucra la reorganización de la actina del citoesqueleto del enterocito, generando una prolongación de las microvellosidades, las cuales abrazan el parásito y lo rodean formando la vacuola pre-parasitófora; simultáneamente las proteínas secretadas por los micronemas y la roptra se fusionan con esta membrana, formando la vacuola parasitófora. Aunque los esporozoítos se sitúan en el interior de la vacuola, no están directamente en contacto con el citoplasma celular, de manera que permanecen intracelulares pero extra citoplasmáticos (18,28).

Los esporozoítos internalizados se transforman en estadios parasitarios redondeados denominados trofozoítos, los cuales inician el ciclo de reproducción asexual (merogonia I y II), luego del cual se produce un ciclo de reproducción sexual (25). En la primera merogonia el núcleo del trofozoíto se divide desarrollando merontes tipo I (4 a 8 merozoítos I en su interior), estos merozoítos al ser liberados (ruptura del meronte), infectarán otras células para desarrollar nuevamente ya sea merontes tipo I ó merontes tipo II (4 merozoítos II en su interior) (1).

Los merozoítos producidos en los merontes tipo II infectan nuevas células para diferenciarse en microgamontes (masculino) o macrogamonte (femenino). Cada

microgamonte se multiplica por fisión múltiple, generando hasta 16 gametos móviles (microgametocitos), los cuales al liberarse buscan el macrogamonte para fertilizarlo (1,25). Una vez se da la fusión de los dos gametos se produce un cigoto, que se rodea de una pared transformándose en un ooquiste, en el cual se dan procesos de división meiótica seguidos de una esporogonia, con la producción final de 4 esporozoítos. En este protozoo se pueden generar ooquistes de pared delgada, asociados con el desarrollo de un ciclo de autoinfección interna, y ooquistes de pared gruesa, que se eliminan con las heces y representan la forma de resistencia e infectante para un nuevo hospedero (1,29).



**Figura 1. Ciclo de vida de *Cryptosporidium*.** Adaptado de Gerace y col. (30).

*Cryptosporidium* no es un parásito intracelular obligado, análisis *in vitro* describen un ciclo de vida completo en ambientes libres de células, con los trofozoítos como el estadio dominante (31–33). En estudios realizados por Hijjawi y col. (34), en los que se infectaron cultivos celulares con oocistos de *C. andersoni* y cultivos libres de células con *C. parvum* y *C. muris*, se pudieron observar estadios extracelulares similares a trofozoítos y gamontes de las gregarinas. Estos estadios fueron observados en grandes cantidades y algunos de ellos realizando sizigia, una

asociación en la que dos gametos de diferente sexo se fusionan y se rodean por una cubierta protectora transformándose en gametoquistes, los cuales se han descrito también en el ciclo de vida de las gregarinas (24). Todavía se desconoce el origen de las etapas extracelulares similares a gamontes, pero se ha sugerido que podrían originarse a partir de esporozoítos libres que extracelularmente se transformaron en trofozoítos móviles (24,35). Otros autores sugieren que estos estadios derivan de los trofozoítos, y que tendrían un posible papel en la generación de más trofozoítos y merozoítos para la producción de nuevos ooquistes infectantes en el ambiente (18).

#### **1.4. Criptosporidiosis**

La criptosporidiosis es el síndrome clínico asociado con la infección por *Cryptosporidium*, en el que se presenta una diarrea osmótica/secretora, voluminosa y de mal olor, acompañada con dolor abdominal, náuseas, vómitos, anorexia, calambres abdominales, pérdida de peso, fiebre y fatiga, entre otros (36). La infección puede presentarse de forma auto limitada en individuos inmunocompetentes, produciendo manifestaciones clínicas agudas (6 a 14 días), o por el contrario evolucionar a una infección crónica en personas inmunocomprometidas con pérdida de electrolitos y un síndrome de desgaste en los pacientes en fase SIDA (1,37). En cuanto a la criptosporidiosis infantil, la diarrea se asocia con retraso en el crecimiento y déficit cognitivo, e incluso alteración de la microbiota, favoreciendo el ciclo de malnutrición en esta población (19,38).

La sintomatología de la criptosporidiosis se presenta en respuesta al daño en el epitelio gastrointestinal con la subsecuente atrofia de las microvellosidades e hiperplasia en las criptas de Lieberkühn, también se produce un aumento en la secreción de cloruro y disminución en la absorción de vitamina B12 y glucosa (1,37).

La duración media de la enfermedad es de 9 días con un máximo de 12 deposiciones diarias (6).

La clínica y gravedad de la infección dependen de factores del hospedero, como la edad y el estado inmune-nutricional, así como de características biológicas del parásito (37,39). Aunque *Cryptosporidium* suele infectar principalmente el intestino delgado en hospederos inmunocompetentes, también se ha descrito en todo el tracto gastrointestinal y respiratorio de pacientes inmunocomprometidos, quienes desarrollan cuadros clínicos atípicos extraintestinales en el tracto hepatobiliar, páncreas y vejiga (1). Con relación a los primeros pacientes con VIH / SIDA que desarrollaron una criptosporidiosis diseminada, la infección fue presumiblemente originada en el tracto intestinal, pero los casos posteriores que carecían de evidencia de compromiso gastrointestinal insinuaron la posibilidad de transmisión respiratoria a través de aerosoles y saliva (40). Se ha descrito que la criptosporidiosis de las vías respiratorias superiores en humanos es poco común y produce la misma sintomatología que la mayoría de las enfermedades respiratorias clásicas (41). Las especies *C. hominis* y *C. parvum* se han descrito ocasionando inflamación de la mucosa nasal, senos paranasales, laringe y tráquea, acompañados de secreción nasal y cambios en la voz (1). En contraste, la criptosporidiosis del tracto respiratorio inferior produce tos productiva, disnea, fiebre e hipoxemia (40).

Además de las condiciones propias del paciente, también se han observado diferencias en las manifestaciones clínicas y severidad de la sintomatología dependiendo de la especie de *Cryptosporidium* implicada. Algunos autores han descrito que la infección por *C. hominis* se asocia con una mayor eliminación de ooquistes (42) y duración de la diarrea (43) en comparación con la infección por *C. parvum*. También se ha demostrado que hay variaciones intragenotípicas al interior de la especie *C. hominis* que pueden tener una mayor virulencia, asociándose

con una mayor severidad de la sintomatología, mayor excreción de ooquistes y potencial para generar brotes (44).

### **1.5. Genotipificación molecular y diversidad genética**

Avances en las tecnologías y los estudios moleculares han permitido mejoras significativas en la caracterización genética, clasificación taxonómica de especies e identificación de subgenotipos de este protozoo (45). En la actualidad, diferentes métodos y marcadores se utilizan para este propósito mediante protocolos que combinan la PCR con el análisis de las secuencias (1). Los marcadores blanco utilizados incluyen los genes que codifican para el 18S rRNA, las proteínas de la pared del ooquiste (COWP), B-tubulina, adhesinas relacionadas con la trombospondina TRAP-C1 y TRAP-C2, proteína de choque térmico (HSP70), glicoproteína GP60 y dihidrofolato reductasa (DHFR), entre otros (17).

El análisis del gen que codifica para la glicoproteína de membrana *gp60* (también conocida como *gp40/15*) se ha utilizado con frecuencia en el estudio de la dinámica de transmisión de *Cryptosporidium* en zonas endémicas y en la identificación de las fuentes de contaminación en brotes epidémicos (45). Este gen es hasta el momento el marcador más utilizado en la genotipificación molecular de *Cryptosporidium*. El polimorfismo se debe a las altas tasas de mutación con regiones microsatélites de repeticiones en tándem para los trinucleótidos que codifican la serina (TCA / TCG / TCT), y adicionalmente variaciones en la secuencia de las regiones no repetidas, lo que aumenta su poder discriminatorio (46). A partir del análisis de secuencia de *gp60* se clasifican los aislados en distintas familias alélicas y dentro de cada familia, los subgenotipos se diferencian principalmente por el número de repeticiones de trinucleótidos (1,39). Las principales familias de subgenotipos, así como los hospederos y otras fuentes descritas se presentan en la Tabla 3.

La evaluación molecular de *gp60* ha permitido determinar que *C. parvum* presenta la mayor riqueza y diversidad en el género, teniendo 20 familias alélicas (IIa - IIi) y siendo considerada la especie zoonótica más importante. A la fecha, la familia IIa tiene la mayor diversidad genética con 112 subgenotipos, seguida de la familia IIi con 51. Adicionalmente las dos familias presentan el rango de hospederos más amplio, infectando diversas especies de mamíferos y siendo reportadas en muestras ambientales. En la actualidad, la familia familia IIc es la responsable de la mayoría de los casos de criptosporidiosis de índole antroponótico (47).

*Cryptosporidium hominis* tiene trece familias alélicas (Ia-Io) siendo la familia Ib la más frecuente a nivel mundial (1,39). Aunque *C. hominis* fue considerada una especie restringida a humanos, en la actualidad es reportada con frecuencia en ganado vacuno, equinos y primates no humanos (48). En 2010, Xiao y col. propusieron que la menor diversidad de *C. hominis* en comparación con *C. parvum* podría deberse a su rango de hospedero más estrecho y menor intensidad en la transmisión (39). Con respecto a *Cryptosporidium meleagridis*, se han reportado diez familias alélicas (IIIa – IIIj), de las cuales la familia IIIb es la más frecuente, posee el mayor número de subgenotipos, y es encontrada por igual en pavos y humanos (49).

**Tabla 3. Familias de subgenotipos *gp60*.** Información obtenida de las secuencias depositadas en GenBank (base de datos NIH-National Institutes of Health de Estados Unidos) a enero 25 de 2020.

<b>Especie</b>	<b>Familia</b>	<b>Subgenotipos (n)</b>	<b>Hospederos y otras fuentes descritas</b>
<b><i>C. hominis</i></b>	Ia	57	Humanos y primates no humanos
	Ib	22	Humanos, rumiantes, caprinos y primates no humanos
	Id	23	Humanos, primates no humanos y animales salvajes
	Ie	6	Humanos
	If	12	Humanos y primates no humanos
	Ig	6	Humanos
	Ih	1	Humanos
	Ii	1	Humanos, equinos y macacos
	Ij	1	babuinos
	Ik	7	Humanos, equinos, camellos y monos
	Im	1	Macacos
	In	3	Macacos
	Io	2	Macacos
	<b><i>C. parvum</i></b>	IIa	112
IIb		1	Humanos
IIc		17	Humanos, erizos, roedores, cabras y rumiantes
IId		51	Caprinos, ovinos, humanos, bovinos, corderos, equinos, roedores, erizos, cerdos, ciervos, primates no humanos y muestras ambientales
IIe		4	Humanos
IIf		1	Humanos
IIg		3	Roedores y humanos
IIh		1	Humanos
IIi		2	Humanos
IIj		4	Bovinos
IIk		1	Humanos
IIl		5	Bovinos, humanos
IIm		2	Humanos
IIn		2	Humanos
IIo		2	Macacos, humanos
IIp		1	Roedores
IIq		1	Humanos
IIr		1	Muestras ambientales
IIs		1	Humanos
IIt		1	Humanos
<b><i>C. meleagridis</i></b>	IIIa	3	Pavos y humanos
	IIIb	23	Pavos y humanos
	IIIc	1	Humanos
	IIId	2	Humanos
	IIIe	8	Cacatúas, gallinas, perdices, faisanes y humanos
	IIIf	1	Humanos
	IIIg	12	Pájaros, gallinas, pollos y humanos
	IIIh	1	Humanos
	IIIi	1	Humanos
	IIIj	1	Humanos

En los últimos años la genotipificación molecular y los estudios de diversidad genética inter e intraespecie mediante el análisis de diferentes locus concatenados ha ganado importancia, mejorando el poder discriminatorio y resolución taxonómica en comparación con el uso de un solo marcador (50). Los estudios de múltiples marcadores suelen realizarse mediante la detección de diferencias en la longitud de los fragmentos amplificados - MLFT (del inglés Multilocus Fragment Typing), o por secuenciación - MLST (del inglés Multilocus Sequence Typing) (1). Los análisis multilocus proporcionan información útil en el estudio de brotes, el seguimiento de las fuentes de infección y la evaluación de la dinámica de transmisión de las especies en regiones endémicas (transmisión antroponótica vs zoonótica) (51,52). Para realizar análisis multilocus en el género *Cryptosporidium* (48,53–55) se han utilizado alrededor de 55 marcadores en diversas combinaciones, siendo los más utilizados CP47, *gp60*, ML1, ML2, MS5, MS9, MSB, MSC 6-7 y TP14 (50,56). Actualmente no se tiene disponible un consenso de los marcadores que deben utilizarse, lo que dificulta la comparación de los resultados obtenidos en los diferentes estudios.

### **1.6. Epidemiología y distribución geográfica**

La infección con *Cryptosporidium* se ha reportado en la mayoría de países alrededor del mundo, con una prevalencia que varía en función de las características sociodemográficas de las poblaciones, encontrándose en las heces del 1% al 3% de los habitantes de países desarrollados (Europa y América del Norte), 10% en países africanos y hasta 40% en suramérica (57). Se ha descrito que los países en desarrollo, y las zonas rurales presentan las tasas más altas de infección, como resultado de las deficiencias en el acceso al agua potable y saneamiento básico y un mayor contacto con animales (1,30,58).

Esta parasitosis suele presentarse principalmente en población infantil y pacientes inmunocomprometidos y VIH/SIDA, especialmente aquellos con un recuento bajo de

linfocitos T CD4+ (1). En relación con la población infantil, se han reportado prevalencias entre 8,1% y 49,2%. (59,60). Estudios multicéntricos realizados en países en desarrollo describen a *Cryptosporidium* como una de las cinco primeras causas de diarrea en niños menores de 24 meses (59, 60). En adultos de países en desarrollo, se han descrito prevalencias que oscilan entre 25% y 74,3% en pacientes inmunocomprometidos, y de 14% a 88,9% en inmunocompetentes (57).

Las infecciones en los seres humanos se asocian principalmente con *C. parvum* y *C. hominis*, las cuales causan la mayoría de los casos de criptosporidiosis en el mundo (63). La tercera especie en frecuencia es *C. meleagridis*, a la fecha la única especie que infecta aves de corral y mamíferos, siendo identificada en el 1% de los pacientes con diarrea en Inglaterra, y cercano al 10% en Perú, en donde la prevalencia es tan alta como la de *C. parvum* (44). Otras especies y genotipos se han descrito con una menor frecuencia en infecciones en humanos, incluyendo *C. felis*, *C. canis*, *C. viatorum*, *C. ubiquitum* y *C. muris* (64–66).

Algunos autores han reportado diferencias geográficas en los patrones de transmisión de las especies y subgenotipos de *Cryptosporidium*, siendo *C. hominis* dominante en países en desarrollo, principalmente el subgenotipo IbA9G3 (16, 65). Por el contrario, el subgenotipo epidémico IbA10G2 se distribuye por igual en países en desarrollo e industrializados, asociándose con la mayoría de casos esporádicos y brotes relacionados con transmisión de origen hídrico (44, 66). En cuanto a *C. parvum*, se ha evidenciado que la familia IIa tiene una mayor prevalencia en Europa, particularmente el subgenotipo IIaA15G2R1 (16, 45), el cual se asocia con el contacto estrecho con ganado (69). En contraste, en países en desarrollo, la familia IIa rara vez se identifica en humanos, y las infecciones son ocasionadas principalmente por la familia IIc (16, 68). En la actualidad no se ha establecido una

asociación entre las familias de subgenotipos de *C. meleagridis* con diferencias en la distribución geográfica.

En cuanto a las condiciones climáticas, las cargas de la enfermedad difieren en países industrializados que poseen variaciones estacionales; siendo *C. hominis* más frecuente en el Reino Unido y Nueva Zelanda durante el otoño, mientras que *C. parvum* suele aumentar su incidencia en la primavera en los Países Bajos, Canadá e Irlanda (71). En países en desarrollo que no poseen estaciones, la infección es más común en climas tropicales, en los cuales se describe un aumento de la incidencia durante la época lluviosa (57,72).

Otro aspecto importante en la epidemiología de *Cryptosporidium* es su potencial de transmisión hídrica, el cual se relaciona con características biológicas del parásito que facilitan su supervivencia y diseminación por largos periodos en fuentes de agua (49). En consecuencia, desde el 2006, el género *Cryptosporidium* es considerado por la Organización Mundial de la Salud como un patógeno de referencia para la evaluación de la calidad del agua potable (73). Entre enero de 2011 y diciembre de 2016 se presentaron más de 381 brotes de diarrea transmitidos por el agua, de los cuales *Cryptosporidium* fue el principal agente etiológico, produciendo el 63% de los brotes a nivel mundial (74). Revisiones recientes de los protozoos presentes en el agua potable establecen a *Cryptosporidium* como el parásito más importante relacionado con la transmisión de origen hídrico (19,75).

En Colombia los estudios epidemiológicos en este género son escasos y la mayoría se han basado en técnicas convencionales de tinción, las cuales presentan una baja sensibilidad y solo permiten la identificación de género. Además, la criptosporidiosis no está incluida dentro del Sistema de Vigilancia en Salud Pública (SIVIGILA), por lo tanto no es de notificación obligatoria y no se realiza su seguimiento epidemiológico

(57). Estudios realizados en la década del 2000, estimaron prevalencias en adultos inmunocomprometidos que oscilaban entre 2,5% y 23,7%, y en inmunocompetentes de 1,9% a 9,9% (76–78). En cuanto a la población infantil, se han descrito prevalencias ente 7% y 46,8% en niños inmunocompetentes, y en un 9,42% a 51,4% en niños inmunocomprometidos (59,79–81).

En una revisión sistemática de la criptosporidiosis en Colombia, se analizaron un total de 42 estudios, donde se encontró una prevalencia promedio de 9,9% en adultos inmunocompetentes y 19,7% en niños. Adicionalmente también se describe la presencia del parásito en animales como bovinos con un 21% y perros con un 17,4%, así como fuentes hídricas con un 38,9%. Las prevalencias descritas en esta revisión demuestran el gran potencial que tiene la transmisión zoonótica e hídrica de este parásito en Colombia (82).

## **1.7. Estudios genómicos en el género *Cryptosporidium***

### **1.7.1. Cariotipo**

El cariotipo de *C. parvum* se estudió por primera vez en 1988, utilizando electroforesis en gel de campo invertido – FIGE (del inglés Field Inversión Gel Electrophoresis), donde se revelaron cinco bandas cromosómicas que variaban en tamaño de 3,3 a 1,4 Mb (83). Estudios adicionales realizados en un periodo comprendido entre 1993 y 1998, mediante técnicas de electroforesis en gel, hibridación y densitometría, indicaron que *C. parvum* tenía de siete a nueve cromosomas, cuyos tamaños oscilaban entre 0,945 y 2,2 Mb (84–86). El cariotipo definitivo de *C. parvum* se estableció en 1998 utilizando enzimas de restricción y electroforesis de campo pulsado (87). En este estudio se demostró la presencia de 8 cromosomas, estimándose un tamaño de genoma cercano a 9,0 Mb (**Tabla 4**). Además, se construyó un mapa óptico del genoma, el cual sirvió como punto de partida para que en 2004 se ensamblara el genoma de *C. parvum* (88).

**Tabla 4. Cromosomas de *Cryptosporidium parvum*.** Adaptada de Abrahamsen y col. (89) y actualizada con información de CryptoDB v.43 (90).

Organismo	Tamaño estimado del genoma (Mb)	Cromosoma	Código	Tamaño estimado (pb)	Número de genes
<i>C. parvum</i>	9.0	1	CM000429	875.659	404
		2	CM000430	985.969	455
		3	CM000431	1.099.352	456
		4	CM000432	1.104.417	478
		5	CM000433	1.080.900	483
		6	CM000434	1.332.857	577
		7	CM000435	1.278.458	593
		8	CM000436	1.344.712	574

### 1.7.2. Genoma

El estudio del genoma completo de *Cryptosporidium* comenzó a finales de 1990, periodo en el que se recopiló y publicó un set de datos correspondiente al 2,5% del genoma de *C. parvum*, utilizando la secuenciación tipo Sanger (91); en este estudio se realizó la primera descripción de los genes involucrados en la biosíntesis macromolecular, señalización celular y producción de energía, asimismo, se logró determinar que el 18,9% de los genes detectados se clasificaban como proteínas hipotéticas (91). Posteriormente, en el año 2004 fueron publicados los dos primeros genomas de este género correspondientes a las dos especies intestinales *C. parvum* y *C. hominis* (88,92).

Abrahamsen y col. (88) mediante la purificación de ooquistes a partir de materia fecal de terneros infectados y la estrategia de secuenciación de genoma completo - WGS (del inglés Whole-Genome Shotgun), ensamblaron el genoma de *C. parvum* Iowa. Estos autores obtuvieron un genoma de 9,1 Mb, con un porcentaje de Guanina – Citosina (GC) del 30%, y una cobertura global de 13X. Además, se determinó que el genoma contenía genes que codificaban aproximadamente 3.805 proteínas. A

pesar de que el genoma de referencia de *C. parvum* Iowa estaba completo con sus ocho cromosomas, estudios posteriores evidenciaron incongruencias en la anotación en cuanto a predicción del codón de inicio y finalización de la traducción, así como los sitios de empalme durante el procesamiento del ARN (93).

Con respecto a *C. hominis*, Buck y col. (92) a finales de 2004 publicaron el primer genoma de referencia (aislado TU502). Los ooquistes, que se obtuvieron de un niño infectado con el parásito, fueron inoculados en cerdos gnotobióticos y posteriormente el ADN obtenido se secuenció siguiendo una estrategia WGS con secuenciación Sanger. Se obtuvo un genoma de 8,74 Mb, con un porcentaje de GC de 31,7% y 3.994 genes codificantes. Este ensamblaje se consideró como un borrador del genoma, debido a su alto nivel de fragmentación (1422 contigs) (94). En 2012, Ifeonu y col. (94) realizaron una nueva secuenciación de este aislado con el fin de mejorar el ensamblaje. Para ello utilizaron ooquistes del aislado inicial y secuenciaron el ADN con la plataforma HiSeq 2000 de Illumina. El nuevo genoma denominado *C. hominis* TU502-2012 consta de solo 119 contigs (9,1 Mb), alcanzando una reducción de 10 veces en el número de contigs con respecto al ensamblaje de 2004. Adicionalmente, los genes fueron en promedio 500 pb más largos que sus contrapartes en la anotación original, lo que resultó en un aumento del 17% en la fracción del genoma codificante. En 2015, Isaza y col. secuenciaron el genoma de un aislado de *C. hominis* (UdeA01) obtenido a partir de una paciente infectada con el virus de la inmunodeficiencia humana, de Medellín (Colombia), utilizando una estrategia WGS con las plataformas 454 (Roche) y MiSeq (Illumina). El tamaño estimado fue de 9,04 Mb, con una cobertura de 68X y 0,09% de bases ambiguas. En la actualidad este es el único genoma de *C. hominis* ensamblado en cromosomas, que además presenta la mayor curación de la anotación, no solo con respecto a esta especie sino para todas las especies de este género, para las que se encuentran a la fecha genomas disponibles (93).

El advenimiento de nuevas técnicas para la purificación de ooquistes y extracción de ADN, y la secuenciación de genomas de *Cryptosporidium* de alta calidad a partir de muestras clínicas, han brindado la oportunidad de incrementar la cantidad de genomas disponibles en el género (95). A noviembre de 2019 se habían secuenciado los genomas de 8 especies diferentes, cuyas secuencias se encuentran en repositorios importantes como el Instituto Sanger (Wellcome Trust Sanger Institute), el Centro Nacional de Información Biotecnológica (NCBI) y CryptoDB (90). En la Tabla 5 se presentan algunos de los proyectos genómicos en este género, los cuales están organizados en orden descendente de acuerdo al estado del ensamblaje. Actualmente, además de *C. hominis* UdeA01, solo *C. parvum* Iowa II y *C. tyzzeri* UGA55 se encuentran ensamblados en cromosomas.

**Tabla 5. Proyectos de Secuenciación de diferentes especies y aislados de *Cryptosporidium*.** Adaptado de Isaza JP (93).

Estado del ensamblaje	Especie	Aislado	Año de publicación
<b>Cromosomas</b>	<i>C. tyzzeri</i>	UGA55	2017
	<i>C. hominis</i>	UdeA01	2016
	<i>C. parvum</i>	Iowa II	2004
<b>Scaffolds</b>	<i>C. parvum</i>	UKP2	2015
	<i>C. parvum</i>	UKP3	2015
	<i>C. parvum</i>	UKP4	2015
	<i>C. parvum</i>	UKP5	2015
	<i>C. hominis</i>	UKH3	2015
	<i>C. hominis</i>	UKH4	2015
	<i>C. hominis</i>	UKH5	2015
	<i>C. hominis</i>	37999	2014
	<i>C. hominis</i>	30976	2014
	<i>C. muris</i>	RN66	2008
	<i>C. andersoni</i>	30847	2016
	<i>C. hominis</i>	TU502 - 2012	2014
	<b>Contigs</b>	<i>C. hominis</i>	UKH1
<i>C. hominis</i>		30974	2014
<i>C. hominis</i>		33537	2014
<i>C. baileyi</i>		TAMU09-Q1	2013
<i>C. ubiquitum</i>		39726	2014
<i>C. meleagridis</i>		UKMEL1	2014

A continuación, se presenta un resumen de las estrategias y plataformas utilizadas para secuenciar las especies de *Cryptosporidium* disponibles a la fecha en CryptoDB, principal repositorio digital para las secuencias e información genómica en este género.

### ***C. baileyi* aislado TAMU09-Q1**

Secuenciado en 2013 en el Instituto de Ciencias Genómicas de la Universidad de Maryland (USA). Los ooquistes se inocularon en huevos embrionados de gallina y posteriormente, el ADN obtenido fue secuenciado en Illumina MiSeq, obteniendo un tamaño de 8,05 Mb, en 153 contigs con %GC de 24,3 y una cobertura de 70X (94).

### ***C. meleagridis* aislado UKMEL1**

Los ooquistes fueron purificados de la materia fecal de ratones de laboratorio inmunosuprimidos, que fueron inoculados con ooquistes provenientes de un paciente infectado de la India. Con el ADN obtenido, se realizó WGS en la plataforma Illumina MiSeq con una cobertura de 6,026X, representados en 57 contigs, %GC 30,1 y tamaño estimado de 8,97 Mb (94).

### ***C. ubiquitum* aislado 39726**

Fue secuenciado en 2014 en el Centro para el Control y Prevención de Enfermedades – CDC. Se utilizó la plataforma Illumina GAIIx a partir del ADN obtenido de ooquistes de materia fecal de humanos infectados naturalmente. Se obtuvo un tamaño estimado del genoma de 8,97 Mb representados en 39 contigs con %GC de 30,8; 0,02% de Ns y 760X de cobertura (90).

### ***C. tyzzeri* aislado UGA55**

La especie fue secuenciada en la Universidad de Georgia (USA) en 2017, para ello utilizaron WGS en la plataforma illumina MiSeq, a partir de parásitos obtenidos de ratones infectados y propagados en cultivos celulares. El tamaño del genoma es de 9,01 Mb, con un %GC de 29,99 y 0,87% de Ns distribuidos en ocho cromosomas con 3.979 genes codificantes (90).

### ***C. muris* aislado RN66**

Los ooquistes fueron obtenidos de ratones inmunosuprimidos y posteriormente se realizó experimento WGS en la plataforma 454 de Roche. El genoma depositado en CryptoDB presenta un tamaño de 9,25 Mb distribuidos en 84 scaffolds, GC de 28,5%, 0,07% Ns, cobertura de 13X y 3.934 genes codificantes (58).

### ***C. andersoni* aislado 30847**

Secuenciado en 2016 en Alberta (Canadá), se utilizó tecnología WGS a partir de ooquistes obtenidos de una vaca infectada; ensamblando un genoma de 9,09 Mb, porcentaje GC de 28,46% y 0,05% de Ns distribuidos en 135 scaffolds con 3.876 genes codificantes (90).

Los diferentes estudios genómicos han permitido determinar que *Cryptosporidium* presenta un genoma estimado de 9.0 Mb, distribuido en 8 cromosomas y una cantidad media de GC de 30%. De acuerdo con información actualizada en la base de datos CryptoDB se estima que *Cryptosporidium* tiene 4.200 genes de los que aproximadamente el 61% son anotados como codificantes de proteínas hipotéticas, y entre el 5% y 20% poseen al menos un intrón (15, 92, 94, 96–101). Adicionalmente, diferentes estrategias de anotación han permitido determinar que al menos 31 genes de *Cryptosporidium* implicados en el metabolismo de carbohidratos, nucleótidos y proteínas fueron adquiridos por transferencia a partir de eubacterias, proteobacterias, algas, plantas y posiblemente arqueas (102). Debido a que *Cryptosporidium* presenta vías metabólicas reducidas, depende del hospedero para la síntesis *de novo* de ácidos nucleicos, carbohidratos, lípidos y proteínas. Por tanto, éste debe adquirir estos compuestos del hospedero para compensar la actividad biosintética limitada (58). Los análisis genómicos también han permitido evidenciar que el genoma codifica un amplio repertorio de transportadores de azúcares, nucleótidos y aminoácidos, aunque su especificidad y cinética de sustrato no se han determinado (58,88,92). Adicionalmente, se han identificado siete ATPasas tipo P que típicamente están involucradas en el transporte de cationes (17).

### 1.7.3. Genómica Comparativa

Los análisis comparativos inter e intraespecie con genomas de *Cryptosporidium* se han realizado con el propósito principal de aumentar el conocimiento de la biología (58) y de las bases genéticas que expliquen la adaptación de las diferentes especies y subgenotipos a su hospedero (103–106).

En la actualidad, la mayoría de los estudios de genómica comparativa se han enfocado principalmente en análisis estructurales del ADN en cuanto al ordenamiento de los genes y la comparación de regiones codificantes y no codificantes entre *C. hominis* y *C. parvum* (45,107–110). Con relación a la organización espacial de los genes, análisis comparativos han permitido establecer que la sintenia entre *C. parvum* y *C. hominis* se conserva exhibiendo una similitud en términos de nucleótidos de 95-97% sin evidencia de rearrreglos genómicos, inserciones o deleciones significativas (103). Las comparaciones interespecie entre *C. parvum* Iowa y *C. hominis* TU502 han proporcionado una visión general de los cambios de un solo nucleótido - *SNVs* (del inglés Single Nucleotide Variant), identificándose una diferencia entre ambas especies de 37.500 *SNVs* en regiones codificantes y ~8.000 en zonas no codificantes (104). A partir del año 2010 los análisis comparativos intraespecie han ido en aumento (106,111–114). En diferentes estudios se han identificado variantes estructurales que incluyen el número y localización de *SNVs* y eventos de inserción y deleción (94,96).

En cuanto al estudio de las secuencias codificantes CDS (del inglés Coding Sequence), en 2017 Khan y col. (15) al comparar aislados individuales de *C. hominis* y *C. parvum* (Iowa II), encontraron 11 CDS en el genoma de *C. hominis* que no están en *C. parvum*, de los cuales nueve corresponden a proteínas hipotéticas y los dos restantes proteínas ribosomales. En 2014, Caccio y col. (17) realizaron un estudio comparando la especie gástrica *C. muris* RN66 con las principales especies

intestinales, *C. hominis* (TU502) y *C. parvum* (Iowa). Estos autores identificaron 3.511 y 3.359 genes con ortólogos en *C. parvum* y *C. hominis* respectivamente. Adicionalmente, se identificaron 3.167 genes compartidos por las tres especies. Posteriormente, en 2015 Hijjawi y Col. (115) determinaron que a diferencia de los genomas de *C. hominis* y *C. parvum*, la especie *C. muris* presentaba genes que codificaban el conjunto completo de enzimas pertenecientes al ciclo del ácido cítrico.

Por otro lado, conclusiones de diferentes análisis comparativos realizados entre los años 2012 y 2016 sugieren que los miembros de varias familias de proteínas secretadas, como las proteínas MEDLE, las proteasas tipo insulina y las glicoproteínas tipo mucina, pueden ser determinantes de la especificidad de hospedero en las especies y subgenotipos de *Cryptosporidium* (116,117). Estos estudios han permitido observar que hay una reducción en el número de genes que codifican para la familia de proteínas secretadas conocidas como MEDLE y las proteasas similares a la insulina ubicadas en las regiones 3' subteloméricas de los cromosomas 5 y 6 de *C. hominis* frente a los mismos cromosomas en *C. parvum* (118). También se ha observado diferencias en el número de genes que codifican para las proteínas MEDLE entre las familias alélicas de *C. parvum* IIa y IIc, lo que podría asociarse con la especificidad de hospedero (118). Se ha descrito que las regiones ortólogas que codifican las proteasas tipo insulina subteloméricas y las proteínas MEDLE están totalmente ausentes en los genomas de *C. ubiquitum* y la especie gástrica *C. andersoni* (119).

Estudios comparativos con el propósito de identificar genes específicos de especie también han sido realizados por diferentes autores. Bouzid y col. (120) identificaron un gen de ubicación telomérica que codifica proteínas ricas en serina aparentemente específico para cada especie, denominados Cops-1 para *C. parvum* y Chos-1 para *C. hominis*. Es importante destacar que la secuencia Cops-1 contiene *SNVs*

específicos con un buen poder discriminatorio y que podrían ser de utilidad en estudios epidemiológicos moleculares (120). Aunque el trabajo inicial sugirió que ambos genes eran específicos de especie, en un estudio posterior se identificó un ortólogo de Cops-1 en *C. hominis*, denominado ChCops-1. Estos genes ortólogos exhiben 78,8% de identidad de secuencia, con un tamaño de 1434 pb CpCops-1 y de 1263 pb para ChCops-1. Estas diferencias en tamaño corresponden a una región truncada de aproximadamente 170 pb en *C. hominis* (121).

En 2019, Xu y col. (119) publicaron el primer estudio comparativo donde se incluyeron genomas individuales de la mayoría de las especies de *Cryptosporidium* secuenciadas a la fecha: *C. parvum*, *C. hominis*, *C. meleagridis*, *C. ubiquitum*, *C. baileyi*, *C. andersoni* y *C. muris*. La investigación estuvo centrada en el análisis genómico estructural entre las especies intestinales y la divergencia de secuencia en los posibles determinantes de la patogénesis que son secretados por el parásito. Esta investigación determinó que los genes altamente polimórficos codifican en su mayoría proteínas de secreción subtelomérica que están bajo una fuerte selección positiva.

En la Tabla 6 se resumen diferentes características de los genomas de la mayoría de las especies de *Cryptosporidium* secuenciadas a la fecha. Se identifican diferencias en cuanto a los genes que presentan intrones en las principales especies intestinales, *C. parvum* Iowa, *C. hominis* UdeA01 y *C. meleagridis* UKMEL1 con 163, 417 y 506 genes respectivamente. En las demás especies >20% del total de genes identificados poseen al menos un intrón.

**Tabla 6. Características genómicas de diferentes especies en el género *Cryptosporidium*.** Adaptada de Xu y col. (119).

<b>Característica</b>	<b><i>C. parvum</i> Iowa II*</b>	<b><i>C. hominis</i> UdeA01*</b>	<b><i>C. meleagridis</i> UKMEL1*</b>	<b><i>C. ubiquitum</i> 39726</b>	<b><i>C. baileyi</i> TAMU09_Q1</b>	<b><i>C. andersoni</i> 30847</b>	<b><i>C. muris</i> RN66</b>
<b>Hábitat</b>			Intestinal			Gástrico	
<b>Número de genes</b>	3.805	3.819	3.782	3.767	3.728	3.905	3.937
<b>Longitud total de CDS (Mb)</b>	6,83	6,81	6,91	6,94	6,69	6,86	6,93
<b>Densidad génica (genes/Mb)</b>	418,1	421,5	421,6	420	438,6	429,6	427,5
<b>Porcentaje codificante (%)</b>	75	75,2	77	77,4	78,7	75,5	75,2
<b>No. de genes con intrones</b>	163	417	506	758	763	832	798
<b>Genes con intrones (%)</b>	4,2	10,9	13,4	20,1	20,5	21,3	20,3
<b>No. de tRNAs</b>	45	45	45	45	46	44	45
<b>Proteínas con péptidos señal</b>	397	391	397	399	344	309	323
<b>Proteínas con dominios transmembrana</b>	832	817	805	772	813	839	836

\*Principales especies intestinales que infectan al humano.

## 1.8. Justificación del trabajo

En la actualidad un número reducido de estudios han explorado las bases genéticas que expliquen las particularidades fenotípicas de las especies y subgenotipos del género *Cryptosporidium*, incluyendo el hábitat, especificidad de hospedero, distribución geográfica y virulencia. Adicionalmente, son escasos los estudios genómicos que identifiquen marcadores para el estudio de la variabilidad genética del parásito y mejoren la resolución taxonómica al interior del género.

El uso de técnicas moleculares y el acceso a plataformas de secuenciación de primera y segunda generación han permitido análisis genómicos en este género. Sin embargo, la mayoría se han enfocado en el estudio de aislados únicos de *C. hominis* y *C. parvum* (45,106,108–110,112,113,118), con pocos datos de genómica comparativa de diferentes subgenotipos de ambas especies, y de otras especies que también se han asociado con infecciones en el hombre, como *C. meleagridis*.

Teniendo en cuenta lo anterior, y con el objetivo de contribuir al conocimiento de la genómica estructural y funcional en el género *Cryptosporidium*, se realizó un estudio comparativo con genomas de las principales especies asociadas con infección el humano (*C. hominis*, *C. parvum* y *C. meleagridis*), incluyendo diferentes familias alélicas y subgenotipos. Los datos obtenidos, además, pueden servir de base para el desarrollo de otras líneas de investigación que requieren de atención en el manejo y control de esta parasitosis, como lo son la identificación de blancos terapéuticos específicos y de antígenos con capacidad inmunogénica útiles en el desarrollo de vacunas.

## **2. HIPÓTESIS**

El estudio de genómica comparativa de aislados de diferentes especies de *Cryptosporidium* permitirá identificar variaciones estructurales y/o funcionales que podrán relacionarse con características fenotípicas particulares de cada especie (especificidad de hospedero, distribución geográfica, virulencia, entre otros).

### **3. PREGUNTA DE INVESTIGACIÓN**

¿Cuáles son los principales cambios genómicos estructurales y funcionales que se presentan en diferentes aislados de las especies de *Cryptosporidium*: *C. hominis*, *C. meleagridis* y *C. parvum*?

## **4. OBJETIVOS**

#### **4.1. OBJETIVO GENERAL**

Analizar estructural y funcionalmente diferentes genomas de especies y subgenotipos de *Cryptosporidium* asociados con infección en humanos.

#### **4.2. OBJETIVOS ESPECÍFICOS**

- Ensamblar *de novo* los genomas de diferentes especies y subgenotipos de *Cryptosporidium* asociados con infección en humanos a partir de *reads* depositados en bases de datos públicas.
- Identificar variantes de un solo nucleótido (*SNVs*) y eventos de inserción y deleción (*indels*) en los genomas ensamblados.
- Comparar las variantes estructurales encontradas según especie, aislado y cromosoma.
- Identificar y anotar funcionalmente los genes con mutaciones deletéreas compartidos y exclusivos de cada especie.
- Realizar un análisis filogenómico a partir de los *SNVs* detectados entre los genomas de diferentes especies y subgenotipos de *Cryptosporidium* asociados con infección en humanos.

## **5. METODOLOGÍA**

## **5.1. Obtención de datos NGS, ensamblaje y filtrado de contigs**

Para la realización de los análisis de genómica comparativa, inicialmente se seleccionaron por conveniencia 23 proyectos genómicos de las tres principales especies de *Cryptosporidium* asociadas con infección en humanos, luego se descargaron los *reads* y posteriormente fueron ensamblados *de novo*.

### **5.1.1. Obtención de los *reads***

Se descargaron desde el *SRA* (del inglés Sequence Read Archive) del NCBI (122) los *reads* binarios de los diferentes proyectos genómicos de 23 aislados de las especies *C. hominis*, *C. parvum* y *C. meleagridis*.

### **5.1.2. Fusión de los *reads***

Las parejas de *reads* se fusionaron con el programa *FLASH* (del inglés *Fast Length Adjustment of Short reads*) versión 1.2.11 (123).

### **5.1.3. Ensamblaje *de novo* software *SPAdes***

A partir de los *reads*, para obtener los contigs se realizó un ensamblaje *de novo* con los parámetros por defecto en el programa *SPAdes* versión 3.11.1 (124), se seleccionaron los kmers 33, 55, 77, 99, 111 y la opción *-careful*.

### **5.1.4. Identificación de contigs pertenecientes al género *Cryptosporidium***

Para determinar si los *reads* depositados en el *SRA* correspondían exclusivamente a *Cryptosporidium* y descartar la contaminación con otros microorganismos, se realizó un *BLASTn* (125) usando como base datos los genomas de referencia con una mayor depuración y ensamblados en cromosomas (*C. hominis* UdeA01, *C. parvum* Iowa II y *C. tyzzeri* UGA55), los cuales se descargaron de *CryptoDB* versión 43 (126). Se filtraron y seleccionaron los contigs con un valor de Bit score  $\geq 300$ .

### **5.1.5. Extracción de contigs y estadísticas de los ensamblajes**

Los contigs que fueron filtrados con *BLASTn* fueron extraídos utilizando el script *SeqSelect.py*; posteriormente con el script *Count\_fasta* de Perl se verificaron estadísticas tales como: longitud total de las secuencias, N50, %GC y %Ns, entre otros. Se realizó un mapeo con el alineador *BWA* (del inglés *Burrows-Wheeler Aligner*) versión 0.7.17 (127) de los *reads* de cada uno de los aislados contra sus *contigs* ensamblados *de novo* respectivamente, y finalmente se usó la herramienta *SAMtools Depth* para calcular la profundidad de lectura en todas las posiciones (profundidad de cobertura máxima de 1000000). Se estimó la cobertura media con una fórmula de *awk*.

## **5.2. Análisis estructural**

Para la realización de los análisis comparativos se utilizó como referencia el genoma de *Cryptosporidium parvum* aislado Iowa II versión 43 depositado en *CryptoDB* (126). Inicialmente se identificaron las variantes de un solo nucleótido y luego los eventos de inserción y deleción.

### **5.2.1. Identificación de variantes de un solo nucleótido**

Las variantes de un sólo nucleótido - *SNVs* (del inglés Single Nucleotide Variants) se detectaron usando el alineador MUMmer (del inglés *Maximal Unique Matches*) versión 3.0 (128) con las herramientas *show-snps* y *dnaDIFF*.

### **5.2.3. Identificación de inserciones y deleciones**

Para la identificación de inserciones y deleciones - *indels* se utilizó el alineador MUMmer (del inglés *Maximal Unique Matches*) versión 3.0 (128) según lo descrito previamente por Isaza y colaboradores (93), excluyendo aquellos con una longitud < 50 nucleótidos con el software *online Assemblytics* (129).

### **5.3. Análisis funcional**

Para la realización de los análisis funcionales en los 23 proyectos genómicos se utilizó como referencia el genoma de *Cryptosporidium parvum* aislado Iowa II versión 43 depositado en *CryptoDB* (126). Inicialmente se detectaron las variantes en regiones codificantes y luego se identificaron y anotaron los genes con mutaciones deletéreas compartidos y exclusivos de especie.

#### **5.3.1. Identificación de *SNVs* en regiones codificantes**

Para la detección de variantes de un solo nucleótido en los CDSs, los *reads* de los 23 proyectos genómicos fueron mapeados al genoma de *C. parvum* aislado Iowa II con el alineador *BWA* (del inglés *Burrows-Wheeler Aligner*) versión 0.7.17 (127). Posteriormente se generaron los archivos bam y VCF con *SAMtools* y *BCFtools* respectivamente (130, 131). Solo fueron incluidas las variantes con una puntuación de calidad Phred (del inglés *Phred quality score*)  $\geq 50$  (132). Finalmente se usó la herramienta *SAMtools Depth* para calcular la profundidad de lectura en todas las posiciones, con una profundidad de cobertura máxima de 1000000. Posteriormente, la cobertura media se estimó con una fórmula de awk.

#### **5.3.2. Anotación funcional de los *SNVs* detectados en CDSs**

Las variantes de un solo nucleótido en CDSs fueron anotadas con el predictor de efecto *SIFT4G* (del inglés *sorting intolerant from tolerant*) versión 3.0 (133).

#### **5.3.3 Identificación y anotación de los genes con cambios no sinónimos deletéreas**

A partir de la predicción obtenida con *SIFT4G*, se filtraron los genes con mutaciones deletéreas en los 8 cromosomas para cada uno de los aislados. Estos genes fueron extraídos con el script *SeqSelect.py*. y se realizó análisis de enriquecimiento en términos de Ontología Génica (del inglés *Gene Ontology*, abreviado como *GO*) con

el programa EggNOG mapper versión 1.0 (134). Posteriormente se visualizó con WEGO (del inglés *Web Gene Ontology Annotation Plotting*) versión 2.0 (135). Seguidamente, se realizó análisis de ortología funcional con KEGG en el servidor KAAS (136), seleccionando los mejores hits mediante una estrategia bidireccional-BBH (del inglés *bi-directional best hit*). Para la detección de genes con dominios transmembrana se utilizó el servidor TMHMM versión 2.0 (137). Se realizó además, predicción de los genes que codifican para proteínas con secreción por vía clásica y no clásica con los servidores online SIGNALP versión 5.0 (138) y SECRETOME-P versión 2.0 (139) respectivamente.

#### **5.4. Análisis filogenómico de las principales especies de *Cryptosporidium***

Se realizó un análisis filogenómico basado en los *SNVs* detectados en los 24 genomas de *Cryptosporidium* seleccionados (23 proyectos genómicos y el genoma de referencia). Se construyó una matriz con todos los *SNVs* encontrados y se cargó en el programa IQ-TREE versión 1.6.12 (140). Se construyó un árbol de máxima verosimilitud ML (del inglés *maximum likelihood*) utilizando este software de inferencia filogenómica con valores predeterminados. Se seleccionó el modelo de transversiones con frecuencia de bases desigual TVM (del inglés *Transversion model, and unequal base freq*) de acuerdo con el criterio de información bayesiano BIC (del inglés *Bayesian information criterion*). Se realizaron 200 iteraciones y 1.000 réplicas de *Bootstrap* ultra rápido y de SH- aLRT (prueba de razón de probabilidad aproximada) respectivamente. El árbol fue editado usando FigTree versión 1.4.4. (141).

Los análisis estadísticos se realizaron mediante pruebas de Kruskal-Wallis para comparar los *SNVs* inter e intraespecie, la acumulación de transiciones/transversiones, los genes con mutaciones deletéreas inter e

intraespecie, así como las inserciones y deleciones mediante software R studio versión 3.3.3 (142). Se acepto como un valor significativo una  $p \leq 0.050$ .

## **6. RESULTADOS**

### 6.1. Aislados seleccionados

Para realizar el análisis comparativo se escogieron 23 proyectos genómicos NGS de las tres especies de *Cryptosporidium* de mayor importancia clínica en el humano, distribuidas así: diez aislados de *C. hominis*, nueve aislados de *C. parvum* y cuatro aislados de *C. meleagridis*. Con respecto a *C. meleagridis* se incluyeron todos los genomas publicados a 2019, evidenciándose la escasez de proyectos genómicos en esta especie. Los genomas están clasificados según la subgenotipificación con *gp60* en diferentes familias alélicas y subgenotipos. En *C. hominis* se incluyeron aislados de cinco familias (Ia, Ib, Ie, Id y If) con seis subgenotipos diferentes en total, de las cuales la Ia, Ib, Id y If han sido asociadas con transmisión en animales. En *C. parvum* tres familias alélicas (IIa, IId y IIc) con seis subgenotipos, dos de los cuales pertenecen a la familia alélica IIc, asociada con transmisión antroponótica; y en *C. meleagridis* cuatro familias alélicas (IIIb, IIIg, IIIh y IIIe), cada una con un subgenotipo distinto.

Los proyectos genómicos de *Cryptosporidium* seleccionados para los análisis de genómica comparativa provenían de diferentes ubicaciones geográficas, siendo el Reino Unido (UK) el origen más frecuente, y en menor número, aislados de países en desarrollo como Colombia (n=1) y países africanos (n=3). El material genético se obtuvo de ooquistes de pacientes con infección natural, y además cuatro de los aislados se propagaron en modelos *in vivo* (lechones, ratones o gallinas). La mayoría fueron secuenciados en las plataformas HiSeq y MiSeq de Illumina. Únicamente dos fueron secuenciados en la plataforma Ion Torrent, los cuales correspondieron a los genomas de *C. hominis* provenientes de Kenia (SWEH2) y Dinamarca (SWEH5) (**Tabla 7**).

**Tabla 7. Descripción de los proyectos genómicos de *Cryptosporidium* incluidos en los análisis comparativos**

Especies	Aislado	Subgenotipo	Lugar	Propagación en animales	Técnica de secuenciación	Proyecto genómico	Referencias
<i>C. hominis</i>	UKH1	IbA10G2	UK: Gales	Cerdos	Illumina MiSeq	PRJNA222837	(94)
	UKH3	IbA10G2	UK: Inglaterra	N/A	Illumina MiSeq	PRJNA253834	(143)
	UKH4	IaA14R3	UK: Inglaterra	N/A	Illumina HiSeq	PRJNA253838	(143)
	UKH5	IbA10G2	UK: Inglaterra	N/A	Illumina HiSeq	PRJNA253839	(143)
	37999	IbA10G2	USA: Twin Fall, Idaho	N/A	Illumina GAIIx	PRJNA252787	(103)
	TU502_2012	IbA10G2	Uganda	Cerdos	Illumina MiSeq	PRJNA222836	(94)
	30976	IaA28R4	USA: St. Louis	N/A	Illumina GAIIx	PRJNA252787	(103)
	UdeA01	IeA11G3T3	Colombia: Medellín	N/A	Illumina MiSeq	PRJEB10000	(93)
	SWEH2	IdA14	Kenia	N/A	Ion Torrent	PRJNA307563	(51)
	SWEH5	IfA12G1	Dinamarca	N/A	Ion Torrent	PRJNA307563	(51)
<i>C. parvum</i>	UKP2	IiA19G1R2	UK: Inglaterra	N/A	Illumina MiSeq	PRJNA253836	(143)
	UKP3	IiA18G2R1	UK: Gales	N/A	Illumina HiSeq	PRJNA253840	(143)
	UKP4	IiA15G2R1	UK: Inglaterra	N/A	Illumina HiSeq	PRJNA253843	(143)
	UKP5	IiA15G2R1	UK +	N/A	Illumina HiSeq	PRJNA253845	(143)
	UKP6	IiA15G2R1	UK +	N/A	Illumina HiSeq	PRJNA253846	(143)
	UKP7	IiA17G1R1	UK: Inglaterra	N/A	Illumina HiSeq	PRJNA253847	(143)
	UKP8	IIdA22G1	UK: Inglaterra	N/A	Illumina HiSeq	PRJNA253848	(143)
	UKP14*	IiCA5G3a	UK: Inglaterra	N/A	Illumina HiSeq	PRJNA315506	(70)
	UKP15*	IiCA5G3a	UK: Gales	N/A	Illumina HiSeq	PRJNA315507	(70)
	<i>C. meleagridis</i>	UKMEL1	IIIbA22G1R1	India	Ratones	Illumina MiSeq	PRJNA222838
UKMEL3		IIIgA23G3	UK+	N/A	Illumina HiSeq	PRJNA315502	(70)
UKMEL4		IIIhA7	UK +	N/A	Illumina HiSeq	PRJNA315503	(70)
TU1867		IIIeA21G2R1	Uganda	Cerdos, ratones y gallinas	Illumina MiSeq	PRJNA192428	(144)

\*: Aislados de *C. parvum* antroponóticos.

+: Se desconoce el país exacto de procedencia.

N/A: No aplica.

## 6.2. Ensamblaje de los genomas

Los *reads* de los 23 proyectos genómicos seleccionados se descargaron del Sequence Read Archive - SRA de la base de datos pública de NCBI y se ensamblaron siguiendo una estrategia *de novo* con los parámetros por defecto del software *SPAdes*.

Teniendo en cuenta el tamaño del genoma esperado para este género (aproximadamente 9,0 Mb), se observó una contaminación en los ensamblajes de un aislado de *C. hominis* (UKH4: 11 Mb) y varios aislados de *C. parvum* (UKP3: 12 Mb; UKP4: 11 Mb y UKP6: 9,8 Mb). Debido a esto se asumió una aproximación metagenómica en el ensamblaje, y para extraer únicamente los contigs pertenecientes al género *Cryptosporidium*, se realizó una selección con el algoritmo *BLASTn* utilizando como base de datos los tres genomas ensamblados en cromosomas: *C. hominis* UdeA01, *C. parvum* Iowa II y *C. tyzzeri* UGA55.

A partir de los resultados del *BLASTn* se realizó una depuración de los *contigs* por un *bitscore*  $\geq 300$ , recuperándose con esta estrategia, genomas cercanos a los 9,0 Mb en la mayoría de los aislados. En tres genomas se obtuvieron tamaños de ensamblaje inferiores a 9,0 Mb: *C. parvum* UKP14 (8,2 Mb), y *C. hominis* SWEH2 y SWEH5 (8,2 Mb) (**Tabla 8**). De todos los genomas ensamblados, solo cuatro tuvieron una cobertura media de mapeo inferior a 50X contra el genoma de referencia de *C. parvum* Iowa II, tres de los cuales pertenecen a la especie *C. hominis* (UKH3, SWEH2 y SWEH5) y un aislado de *C. parvum* (UKP5).

Los ensamblajes más fragmentados correspondieron a genomas de la especie *C. parvum*, puntualmente UKP3 y el antroponótico UKP14 con 2.971 y 2.787 *contigs* respectivamente. Sin embargo, no se evidenció disminución de la cobertura media del mapeo de lectura para estos aislados. Todos los genomas ensamblados *de novo*

presentaron un contenido medio de GC del 30% y bases ambiguas que no superaron el 0,26%.

Para los aislamientos que presentaban ensamblajes disponibles en bases de datos, se compararon las métricas y se mantuvieron los ensamblajes *de novo* filtrados por bitscore para los análisis posteriores. En el caso de los aislados de *C. hominis* SWEH2, SWEH5 y de *C. meleagridis* TU1867 no se contaba con ensamblajes previos disponibles.

Teniendo en cuenta que para algunos aislados no se presentaron diferencias en cuanto al número de contigs y/o aumento en la estadística N50 entre los ensamblajes *de novo* y los descargados de CryptoDB, para el estudio se seleccionaron seis ensamblajes de esta base de datos (versión 43): *C. hominis* UKH1, TU502\_2012, 30976, 37999, UdeA01 y *C. meleagridis* UKMEL1. Los 17 genomas restantes se ensamblaron *de novo* (**Tabla 8**).

**Tabla 8. Estadísticas de los ensamblajes en los aislados de *Cryptosporidium***

Especies	Aislados	No. de reads	Genoma ensamblado (pb)	No. de contigs	Contig más largo (pb)	Valor N <sub>50</sub> (pb)	Contenido GC (%)	No. de Ns	Cobertura media del mapeo (X) contra <i>C. parvum</i> Iowa II	Cobertura media del mapeo (X) contra contigs ensamblados <i>de novo</i>	
<b><i>C. meleagridis</i></b>	UKMEL1 +	11.431.022	8.973.200	57	732.862	322.908	30,97	0	101,6	110,4	
	UKMEL3 *	3.097.090	9.028.762	421	214.804	62.491	31,01	6.697	75,2	93,1	
	UKMEL4 *	5.648.885	9.150.400	167	296.518	110.331	30,94	693	148,7	182,4	
	TU1867 *	16.007.104	8.997.078	46	1.076.730	672.302	30,94	763	344,7	378,5	
<b><i>C. hominis</i></b>	UKH1 +	3.798.205	9.141.398	156	542.781	179.408	30,13	0	191,7	197,3	
	UKH3 *	1.238.762	9.071.292	179	552.915	167.737	30,14	256	35,2	35,8	
	UKH4 *	11.895.367	9.390.791	2164	239.037	48.766	30,66	12.842	326,2	321,5	
	UKH5 *	12.649.912	9.068.487	526	730.267	81.885	30,2	2.681	354,8	362,2	
	30976 +	35.360.353	9.059.225	53	1.279.890	470.636	30,12	1.699	502,4	511	
	37999 +	16.569.871	9.054.010	78	1.029.232	406.678	30,13	3,168	349,9	367,4	
	TU502-2012+	1.810.060	9.107.739	119	1.270.815	238.509	30,13	0	94,8	96	
	UdeA01 +	1.080.448	9.043.938	8	1.322.947	1.103.974	30,15	8.500	52,6	53,4	
	SWEH2 *	1.791.829	8.818.287	1629	62.020	9.465	30,43	0	33,9	35,2	
	SWEH5 *	2.058.197	8.821.738	1342	94.133	14.514	30,47	96	40,8	42,4	
	<b><i>C. parvum</i></b>	UKP2 *	2.015.213	9.113.527	118	859.882	216.383	30,17	485	52,3	52,4
		UKP3*	8.227.141	9.636.677	2971	364.354	51.320	30,87	25.249	242,4	229,3
UKP4 *		9.342.641	9.429.553	1933	418.794	62.049	30,59	18.349	279,5	270,2	
UKP5 *		1.321.400	9.017.288	473	376.426	68.305	30,24	254	40,1	40,4	
UKP6 *		5.598.532	9.129.374	64	1.336.351	383.563	30,18	1.484	160,3	160,4	
UKP7 *		3.746.781	9.007.490	555	340.451	67.504	30,32	861	112,3	113,8	
UKP8 *		11.161.344	9.108.067	451	787.834	95.411	30,24	2.482	300,4	301,5	
UKP14 *		2.378.388	8.221.521	2787	43.566	5.120	31,95	441	69,4	85,4	
UKP15 *		6.627.377	8.904.901	1830	94.121	10.293	31,09	693	147,6	209,8	

+ Ensamblajes tomados de CryptoDB Versión 43. \* Ensamblajes *de novo* filtrados por bitscore  $\geq 300$ .

**Nota:** Todos los *reads* se mapearon contra el genoma de referencia *C. parvum* Iowa II descargado de la base de datos CryptoDB versión 43 y contra los *contigs* ensamblados *de novo* de cada uno de los aislados.

### **6.3. Identificación de variantes de un solo nucleótido (*SNVs*)**

#### **6.3.1. Acumulación de *SNVs* inter e intraespecie**

Con el propósito de identificar los *SNVs* se realizaron comparaciones intra e interespecie alineando > 96% del total de bases de todos los aislados. A partir de los alineamientos realizados se observó que los porcentajes de identidad no cambiaron entre los genomas de los diferentes aislados de *C. hominis* (96,8%) y *C. meleagridis* (91,5%), contra el genoma de referencia Iowa II (**Tabla 9**).

El porcentaje de *SNVs* en el genoma también fue constante entre los aislados de estas dos especies con 6,9% y 2,4% respectivamente. Por el contrario, en *C. parvum* se encontraron diferencias intragenotípicas, con porcentajes de identidad entre 99,7% y 99,9% para aislados zoonóticos y de 99,5% para los antroponóticos (UKP14 y UKP15). Adicionalmente se encontraron diferencias intraespecie en el porcentaje total de *SNVs* en los genomas evaluados de esta especie. Este porcentaje osciló entre 0,02% y 0,06% para aislados zoonóticos y fue igual para los antroponóticos (0,22%).

Al evaluar la acumulación de *SNVs* por especie, se encontró el mayor número en *C. meleagridis*, con una mediana de 624.606, seguido por *C. hominis* (223.640) y finalmente *C. parvum*, en el que se observó un número menor, con una mediana de 2.116 *SNVs* (**Figura 2**). Estas diferencias fueron estadísticamente significativas (p-value 8.299e-05\*).

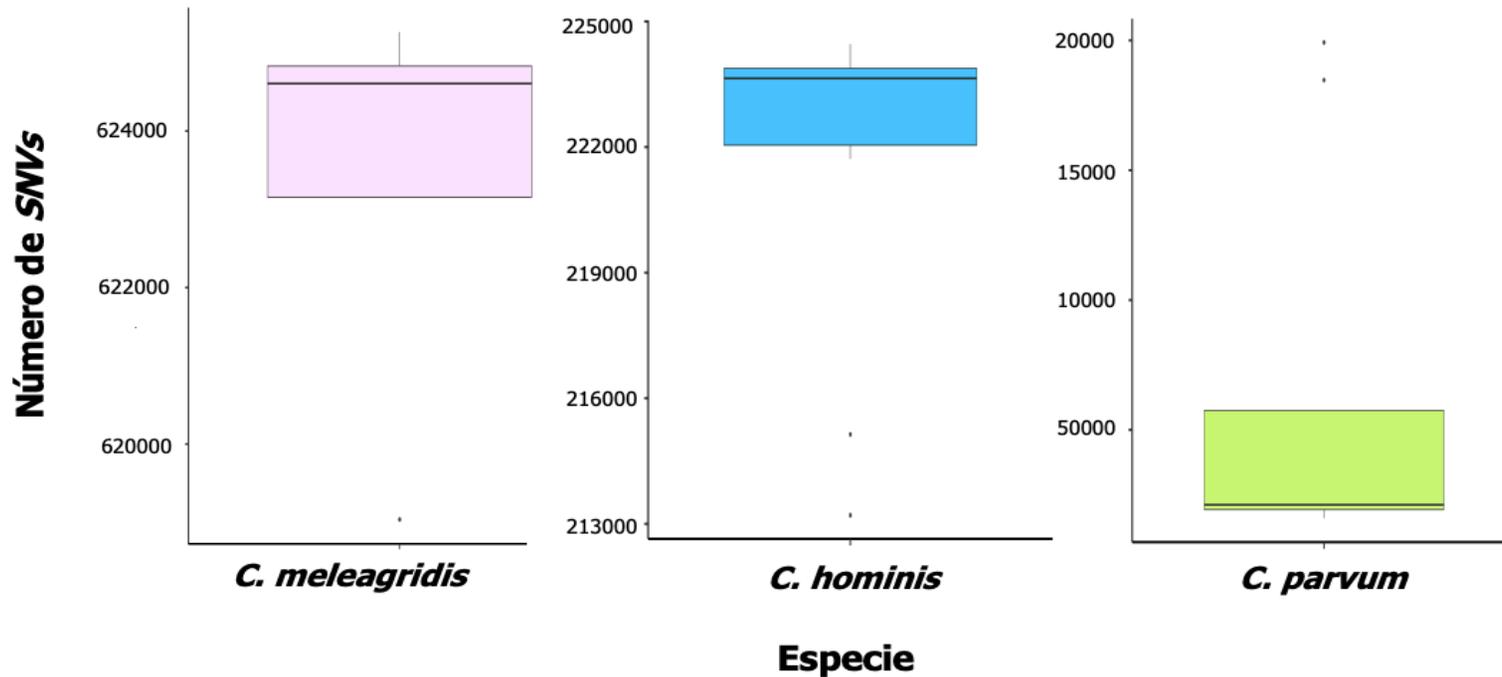
En cuanto a la acumulación de *SNVs* en cada uno de los aislados de las especies evaluadas, se observó un comportamiento homogéneo en *C. hominis* (p-value 0,4373) y *C. meleagridis* (p-value 0,3916). Aunque se observó un número menor de *SNVs* en los aislados SWEH2 y SWEH5 de *C. hominis*, no se encontraron diferencias estadísticamente significativas (**Figura 3**). Con relación a la acumulación de *SNVs*

en *C. parvum*, los datos sugieren un comportamiento heterogéneo, con un mayor número de cambios en los aislados antroponóticos (>18.000), en comparación con los zoonóticos (< 6.000). Sin embargo, no se encontraron diferencias estadísticamente significativas (p-value 0,4335) (**Figura 3**).

**Tabla 9. Resultado de los alineamientos entre los aislados seleccionados y el genoma de referencia *C. parvum* Iowa II**

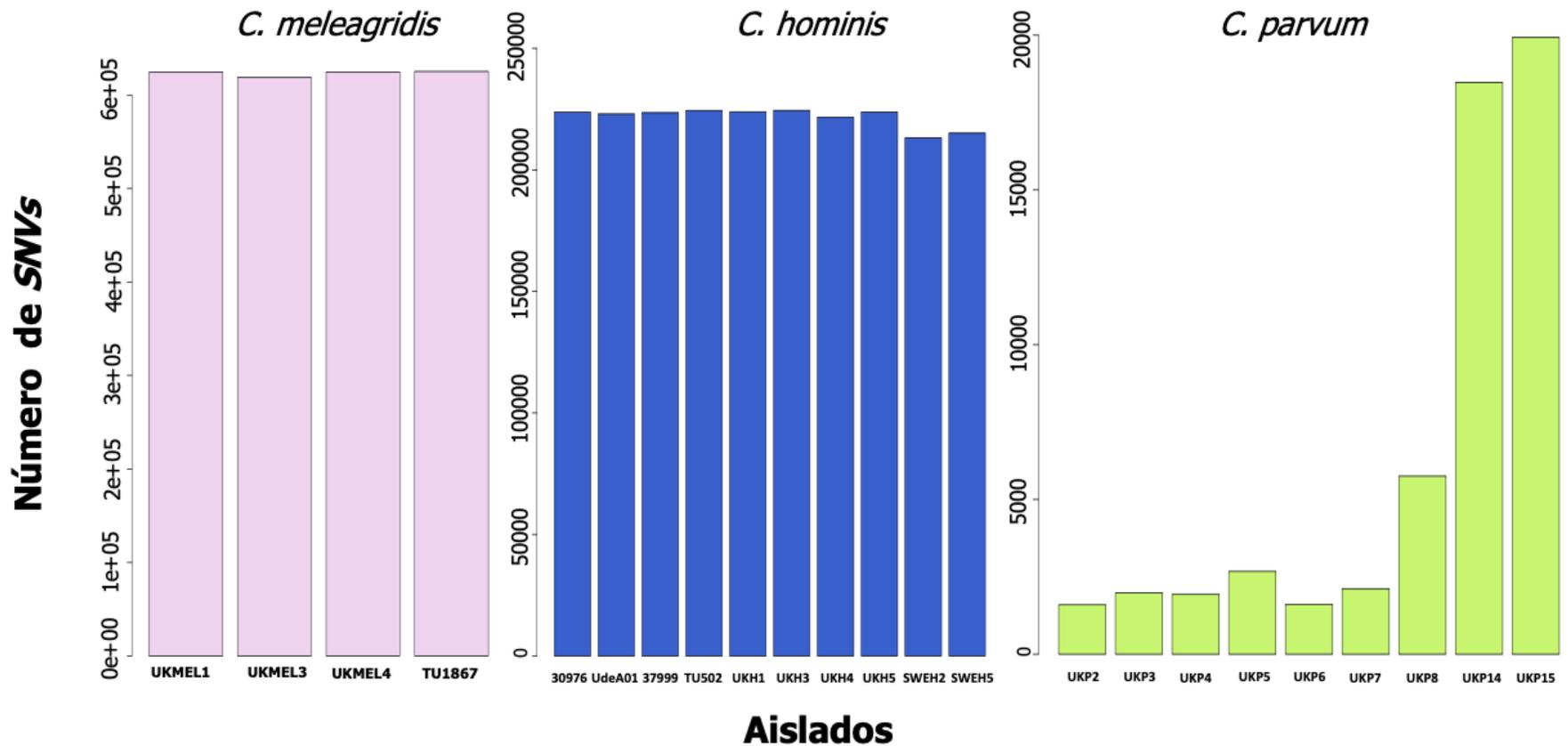
<b>Especie</b>	<b>Aislado</b>	<b>(%) identidad</b>	<b>Bases alineadas (%)</b>	<b>SNVs totales</b>	<b>(%) SNVs en el genoma</b>
<b><i>C. meleagridis</i></b>	UKMEL1	91,55	97,4	624.528	6,96
	UKMEL3	91,59	96,96	619.036	6,86
	UKMEL4	91,54	97,27	624.685	6,83
	TU1867	91,53	97,34	625.265	6,95
<b><i>C. hominis</i></b>	30976	96,82	99,00	223.712	2,47
	UdeA01	96,84	99,29	223.031	2,47
	37999	96,81	99,23	223.568	2,47
	TU502	96,81	99,22	224.430	2,46
	UKH1	96,81	99,22	223.901	2,45
	UKH3	96,81	99,26	224.460	2,47
	UKH4	96,87	98,02	221.714	2,36
	UKH5	96,84	99,2	223.819	2,47
	SWEH2	96,88	99,33	213.206	2,42
	SWEH5	96,85	99,38	215.136	2,44
<b><i>C. parvum</i></b>	UKP2	99,93	99,2	1.595	0,02
	UKP3	99,91	99,25	1.978	0,02
	UKP4	99,92	99,32	1.930	0,02
	UKP5	99,91	99,51	2.674	0,03
	UKP6	99,93	99,39	1.612	0,02
	UKP7	99,86	99,61	2.116	0,02
	UKP8	99,77	99,47	5.752	0,06
	UKP14	99,58	96,91	18.471	0,22
	UKP15	99,51	97,38	19.923	0,22

## Acumulación de *SNVs* interespecie



**Figura 2. Acumulación de *SNVs* interespecie.** Diagramas de cajas y bigotes de las especies de *Cryptosporidium* analizadas y el número de *SNVs* totales detectados. Se observan las medianas de los *SNVs* encontrados en cada especie: *C. hominis* (223.640), *C. meleagridis* (624.607) y *C. parvum* (2.116). Se encontró una diferencia estadísticamente significativa en la acumulación de *SNVs* interespecie p-value = 8,299e-05\*. Los valores atípicos en *C. meleagridis* corresponden a UKMEL3 y *C. hominis* a SWEH2 y SWEH5 respectivamente. En el caso de *C. parvum* son los aislados antroponóticos UKP14 y UKP15. Figura realizada con R-estudio versión 3.3.3.

## Acumulación de *SNVs* intraespecie



**Figura 3. Acumulación de *SNVs* intraespecie.** Gráficas de barras que muestran la acumulación de *SNVs* en los cada uno de los aislados evaluados por especie. No se encontraron diferencias estadísticamente significativas. Figuras realizadas con R-studio Versión 3.3.3.

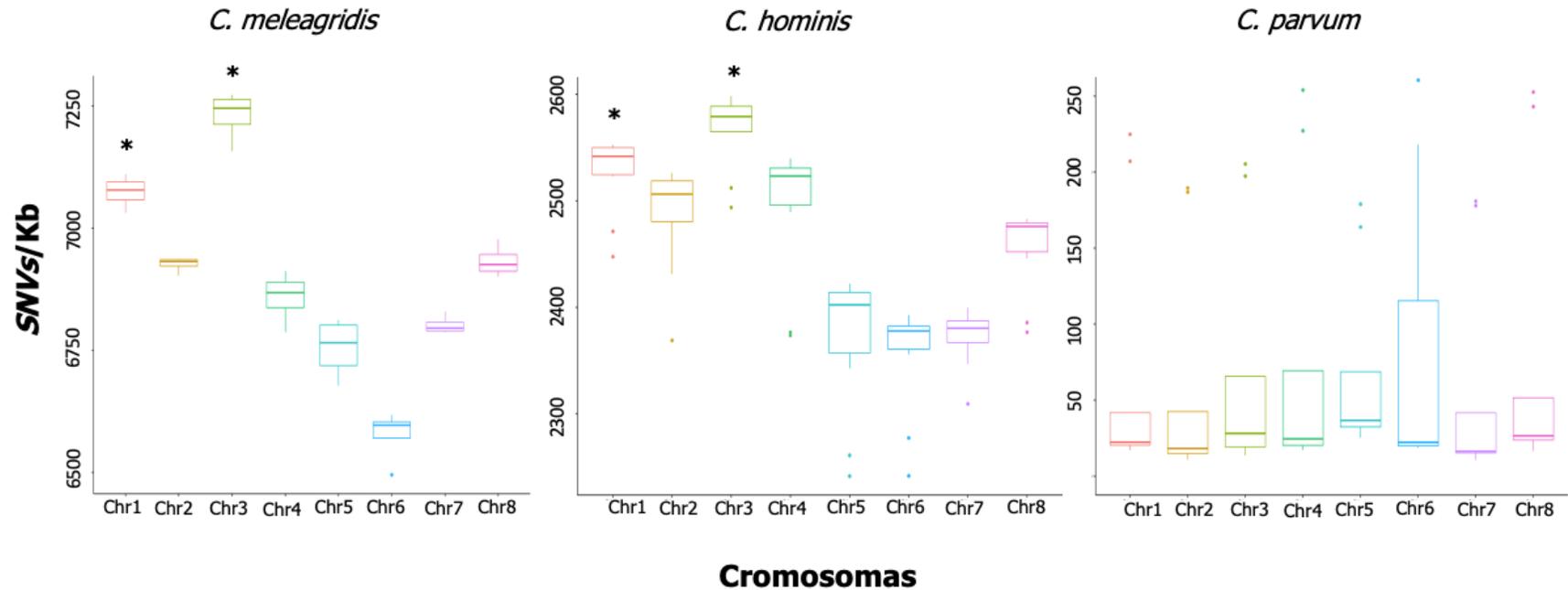
En todos los genomas evaluados, los cambios de un solo nucleótido más frecuentes fueron las transiciones, en *C. hominis* la mayoría de los cambios fueron G>A – T>C y A>G – T>C representando un 18% y 19% de todos los cambios identificados en el genoma respectivamente. Por el contrario, en *C. meleagridis* se identificó una acumulación de C>T y G>A superior al 20% y para A>G y T>C de solo 15%. En *C. parvum* las variantes G>A – T>C y A>G – T>C tuvieron una mayor fluctuación, oscilando entre 8% y 19% respecto al genoma de referencia (**Tabla suplementaria 1**).

Las transversiones también fueron detectadas en todos los genomas, representando menos del 7% de todos los *SNVs* en *C. meleagridis* y *C. hominis*. En contraste, la especie *C. parvum* presentó el mayor número en comparación con las otras especies, puntualmente en los cambios de A>T y T>G (**Tabla suplementaria 2**). No se encontraron diferencias estadísticamente significativas en la acumulación de transiciones y transversiones inter e intraespecie.

### **6.3.2. Acumulación de *SNVs* en los cromosomas**

En relación con la acumulación de *SNVs* por cromosoma, en *C. parvum* se observó una distribución homogénea de estas variantes en todos los cromosomas; sin embargo, aunque se identificaron valores diferentes en los aislados antroponóticos, no se encontraron diferencias estadísticamente significativas (p-value 0,2982). Con respecto a *C. hominis* y *C. meleagridis* se observó un patrón diferente de distribución de *SNVs*, encontrándose una mayor acumulación en los cromosomas 1 y 3, hallazgo que presentó diferencias estadísticamente significativas (*C. hominis* p-value 7,678e-10 y *C. meleagridis* p-value 0,0001496) (**Figura 4**).

## SNVs/Kb en los cromosomas



**Figura 4. Acumulación de *SNVs/Kb* en cada cromosoma.** Diagramas de cajas y bigotes que representan la acumulación de *SNVs/Kb* en los ocho cromosomas de las tres especies evaluadas. El número de los *SNVs* fue normalizado con las longitudes de los cromosomas en miles de pares de bases. Los valores atípicos corresponden a los aislados que presentaron un mayor o menor número de *SNVs* en ese cromosoma. Se encontraron diferencias estadísticamente significativas en *C. meleagris* (p-value 0,0001496\*) y *C. hominis* (p-value 7,678e-10\*). Figuras realizadas con R-studio versión 3.3.3.

### 6.3.3. Identificación de *SNVs* en regiones codificantes

A partir del mapeo de los *reads* contra los CDSs de *C. parvum* Iowa II versión 43 de CryptoDB, se identificaron los *SNVs* ubicados en regiones codificantes para cada uno de los aislados evaluados. Los genomas de *C. hominis* y *C. meleagridis* presentaron el mayor número de estas variantes con más de 150.000 y 400.000 respectivamente, de las cuales cerca del 60% correspondían a cambios sinónimos (*sSNVs*). No se evidenciaron diferencias intraespecie en la acumulación de variantes en regiones codificantes ni en los cambios sinónimos y no sinónimos. Por el contrario, en *C. parvum* el número de *SNVs* en CDSs fue menor de 20.000 en todos los genomas; de los cuales más del 50% correspondieron a mutaciones no sinónimas. Los resultados en *C. parvum* sugieren una mayor heterogeneidad intraespecie con valores que oscilan entre 840 y 3.476 *SNVs* en CDSs en aislados zoonóticos y valores mayores a 14.000 en los antroponóticos. Estas diferencias no fueron estadísticamente significativas (**Tabla 10**).

**Tabla 10. Acumulación de *SNVs* en regiones codificantes**

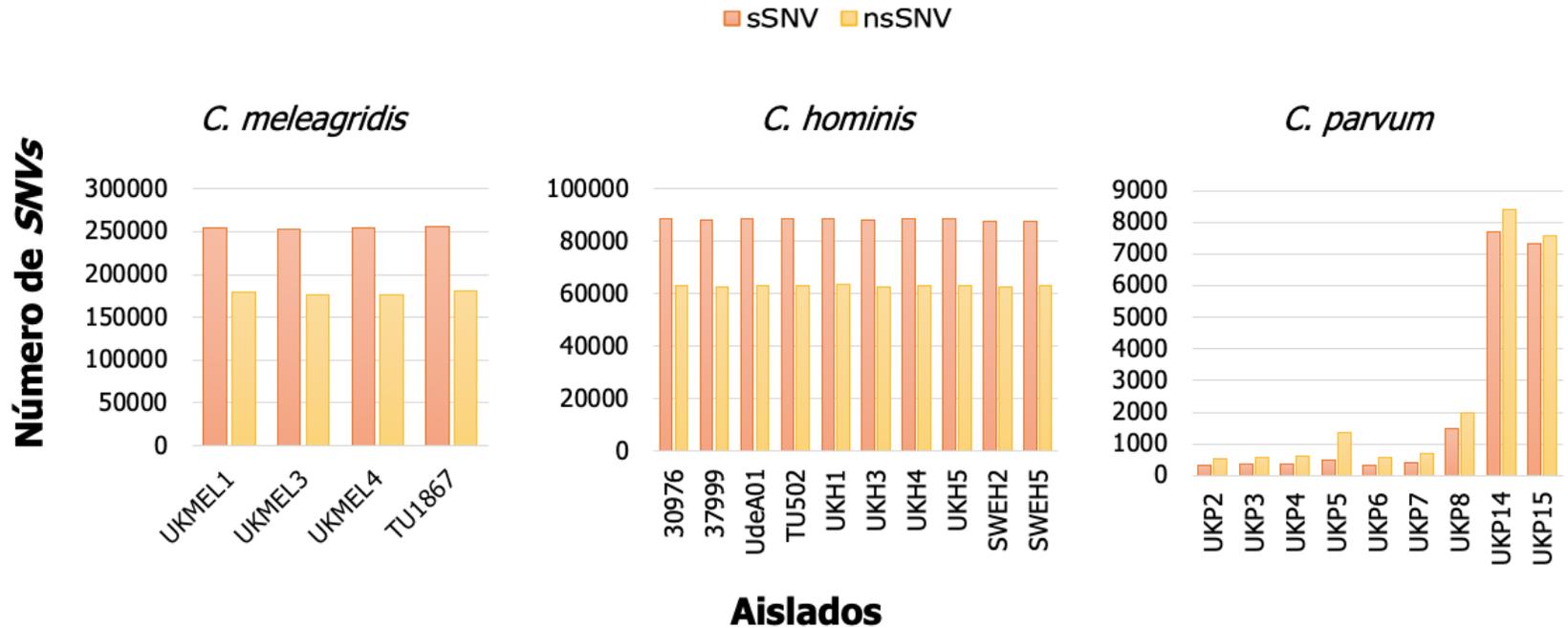
<b>Especie</b>	<b>Aislado</b>	<b><i>SNVs</i> en CDSs</b>	<b><i>sSNVs</i></b>	<b><i>nsSNVs</i></b>	<b><i>nsSNVs</i> (%)</b>
<b><i>C. meleagris</i></b>	UKMEL1	433.966	254.689	179.277	41,3
	UKMEL3	429.483	253.329	176.154	41,0
	UKMEL4	430.124	254.126	175.998	40,9
	TU1867	436.196	255.433	180.763	41,4
<b><i>C. hominis</i></b>	30976	152.048	88.668	63.380	41,7
	37999	150.752	88.008	62.744	41,6
	UdeA01	151.844	88.569	63.275	41,7
	TU502	152.107	88.693	63.414	41,7
	UKH1	152.497	88.825	63.672	41,8
	UKH3	150.993	88.825	62.742	41,6
	UKH4	152.043	88.708	63.335	41,7
	UKH5	151.979	88.678	63.301	41,7
	SWEH2	150.667	87.755	62.912	41,8
	SWEH5	151.045	87.865	63.180	41,8
	<b><i>C. parvum</i></b>	UKP2	840	314	526
UKP3		919	358	561	61,0
UKP4		949	351	598	63,0
UKP5		1.858	503	1.355	72,9
UKP6		891	340	551	61,8
UKP7		1.081	390	691	63,9
UKP8		3.476	1.494	1.982	57,0
UKP14		16.076	7.684	8.392	52,2
UKP15		14.930	7.337	7.593	50,9

Respecto a la acumulación de cambios sinónimos y no sinónimos intraespecie, en todos los aislados de *C. meleagris* observó un comportamiento similar, con aproximadamente 250.000 variantes sinónimas y 160.000 no sinónimas. Los aislados de *C. hominis* también presentaron un comportamiento homogéneo, conservando un número de variantes sinónimas cercanas a 90.000 y no sinónimas de 60.000; en ambas especies, se evidencia que el 40% de los genes tienen por lo menos una mutación no sinónima y que la mayoría de las variantes detectadas en regiones codificantes para proteínas corresponden a cambios sinónimos.

En *C. parvum*, por el contrario, se observó un comportamiento intraespecie heterogéneo con cambios sinónimos que oscilaron entre 314 y 7.337 *SNVs*, presentándose la mayor acumulación en los aislados UKP8 (1.494) y los antroponóticos UKP14 (7.684), UKP15 (7.337). Asimismo, las sustituciones nucleotídicas no sinónimas fluctuaron entre 526 y 8.392 *SNVs*, observándose una mayor acumulación en los aislados zoonóticos UKP5 (1.355) y UKP8 (1.982), y en los antroponóticos UKP14 (8.392) y UKP15 (7.593). En esta especie, más del 50% de los genes presentaron mutaciones no sinónimas, lo que sugiere un predominio de estas mutaciones a expensas de las variantes sinónimas (**Figura 5**).

Posterior a la caracterización de las mutaciones sinónimas y no sinónimas en regiones codificantes, se analizaron por separado los genes con *nsSNVs*, encontrando 2.532 presentes en por lo menos un aislado de cada especie, de los cuales 110 fueron exclusivos de *C. meleagridis*, 11 de *C. hominis* y 2 de *C. parvum*.

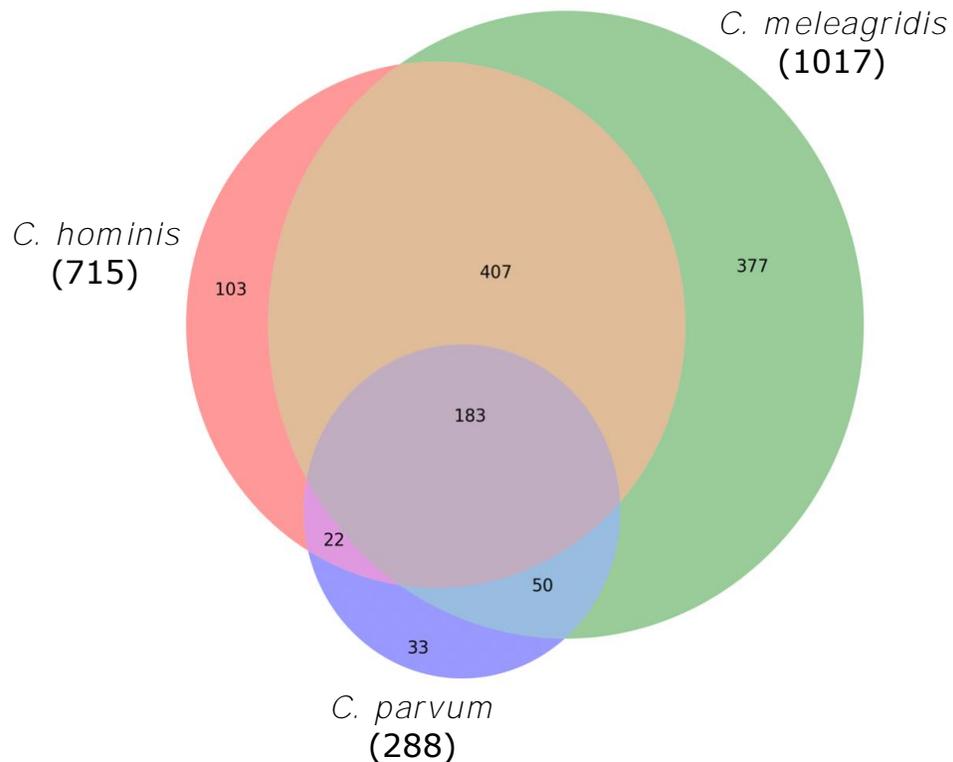
## Acumulación de *sSNV* y *nsSNV* en CDSs



**Figura 5. Acumulación intraespecie de sSNVs y nsSNV en regiones codificantes.** Gráficos de barras para representar los cambios sinónimos (*sSNVs*) y no sinónimos (*nsSNVs*) ubicados en regiones codificantes a partir de la predicción realizada con el anotador SIFT4G. No se encontraron diferencias estadísticamente significativas.

#### 6.4. Identificación y anotación de genes con mutaciones deletéreas compartidos y exclusivos de especie

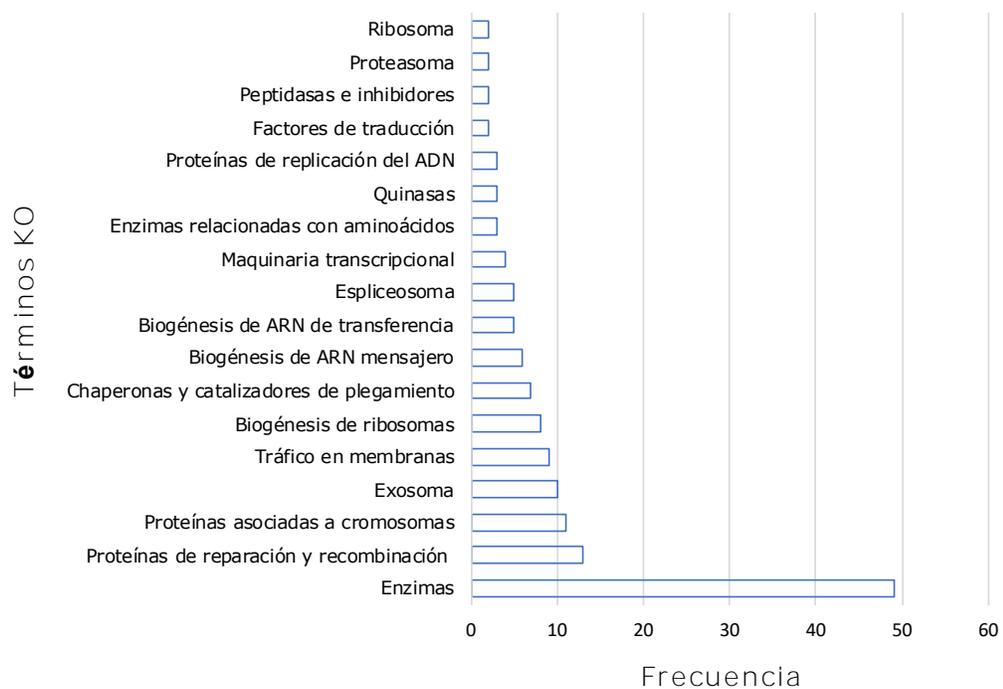
A partir de la identificación de los genes que presentaron variantes no sinónimas (*nsSNVs*), se caracterizaron aquellos que adicionalmente fueron catalogados por el predictor de efecto SIFT4G como cambios no tolerados, indicativos de posibles mutaciones deletéreas. Con esta estrategia fueron predichos un total de 1.017 genes en *C. meleagridis*, 715 en *C. hominis* y 288 en *C. parvum*. De estos, 183 son compartidos y están presentes en por lo menos un aislado de cada especie; 377 son exclusivos de *C. meleagridis*; 103 de *C. hominis* y 33 de *C. parvum* (**Figura 6**).



**Figura 6. Genes con mutaciones deletéreas en las especies evaluadas.**

Diagrama de Venn que esquematiza los genes con mutaciones deletéreas compartidos y exclusivos en las tres especies de *Cryptosporidium*.

En relación con los genes que tienen mutaciones deletéreas y son compartidos por las tres especies, tan solo el 29,5% pudieron ser anotados con *EggNOG – mapper*, cuyo análisis de enriquecimiento en términos GO indicó que se encuentran involucrados principalmente en procesos biológicos y hacen parte de componentes celulares. El análisis de ortología funcional realizado en el servidor KAAS frente a la base de datos del KEGG tuvo un mejor rendimiento en comparación con la asignación GO, anotando el 46,5% de los genes compartidos (**Figura 7**) e indicando que la mayoría de estos genes están implicados en procesos enzimáticos, de reparación y recombinación de ADN. La anotación complementaria permitió determinar que solo el 1,1% presentan secreción por vía clásica, el 10,9% se secretan por vía no clásica y el 2,3% exhiben dominios con hélices transmembrana.



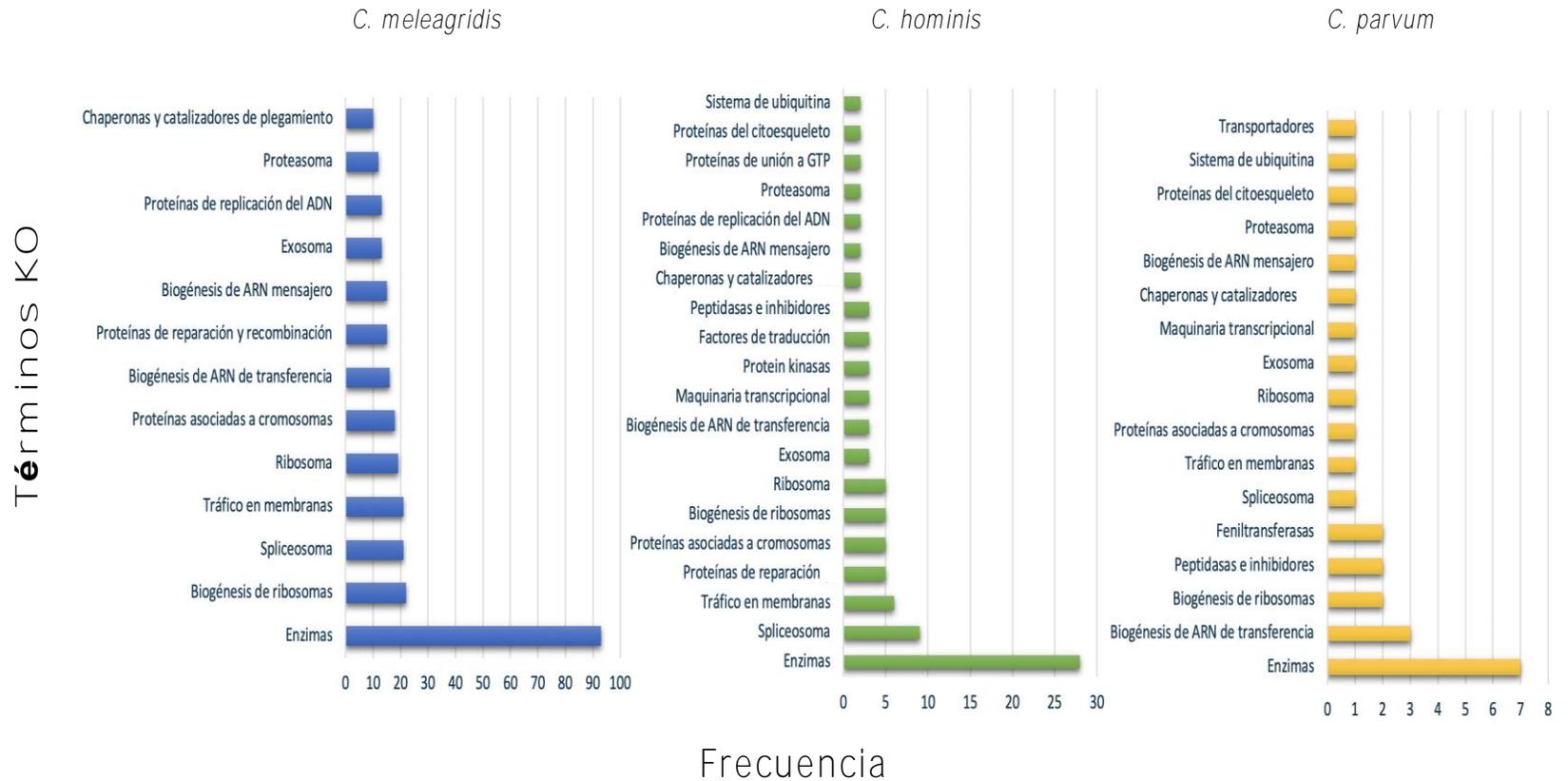
**Figura 7. Análisis de ortología KEGG en los genes con mutaciones deletéreas compartidos por las tres especies de *Cryptosporidium*.** Asignación de términos KO a los genes con mutaciones deletéreas detectados en las tres especies de *Cryptosporidium* a través del servidor KAAS en la base de datos KEGG.

Al analizar los genes con mutaciones deletéreas exclusivos de especie, se pudieron clasificar en por lo menos una categoría de GO el 29,1 y 28,9% de los genes de *C. hominis* y *C. meleagris* respectivamente. En *C. parvum* sólo se asignaron términos de ontología en 3 genes (9,09%), los cuales se clasificaron únicamente en la categoría componente celular. El análisis de ortología funcional de nuevo logró un mejor resultado en comparación con el software *EggNOG – mapper*, asignando 60,2% KO en *C. hominis*, 59,1% en *C. meleagris* y 42,4% en *C. parvum* (**Tabla 11**).

**Tabla 11. Anotación de los genes con mutaciones deletéreas exclusivos de especie**

<b>Especie</b>	<b><i>C. meleagris</i></b>	<b><i>C. hominis</i></b>	<b><i>C. parvum</i></b>
<b>Número de genes con mutaciones deletéreas</b>	377	103	33
<b>Asignación términos GO (%)</b>	28,9	29,1	9,1
<b>Asignación términos KO (%)</b>	59,1	60,2	42,4
<b>Dominios transmembrana (TMHMM) (%)</b>	15,6	13,6	18,2
<b>Péptidos de secreción (SIGNALP) (%)</b>	7,4	4,9	3,0
<b>Secreción por vía no clásica (SECRETOME) (%)</b>	28,1	28,0	42,4

En relación con la asignación de los ortólogos funcionales en vías metabólicas, se estableció que para las tres especies la mayoría de estos genes están implicados en procesos enzimáticos. La segunda vía asignada con mayor frecuencia en los genes con mutaciones deletéreas exclusivos de especie fue la biogénesis de ribosomas en *C. meleagris*, el complejo del espliceosoma en *C. hominis* y la biogénesis de ARN de transferencia en *C. parvum* (**Figura 8**).



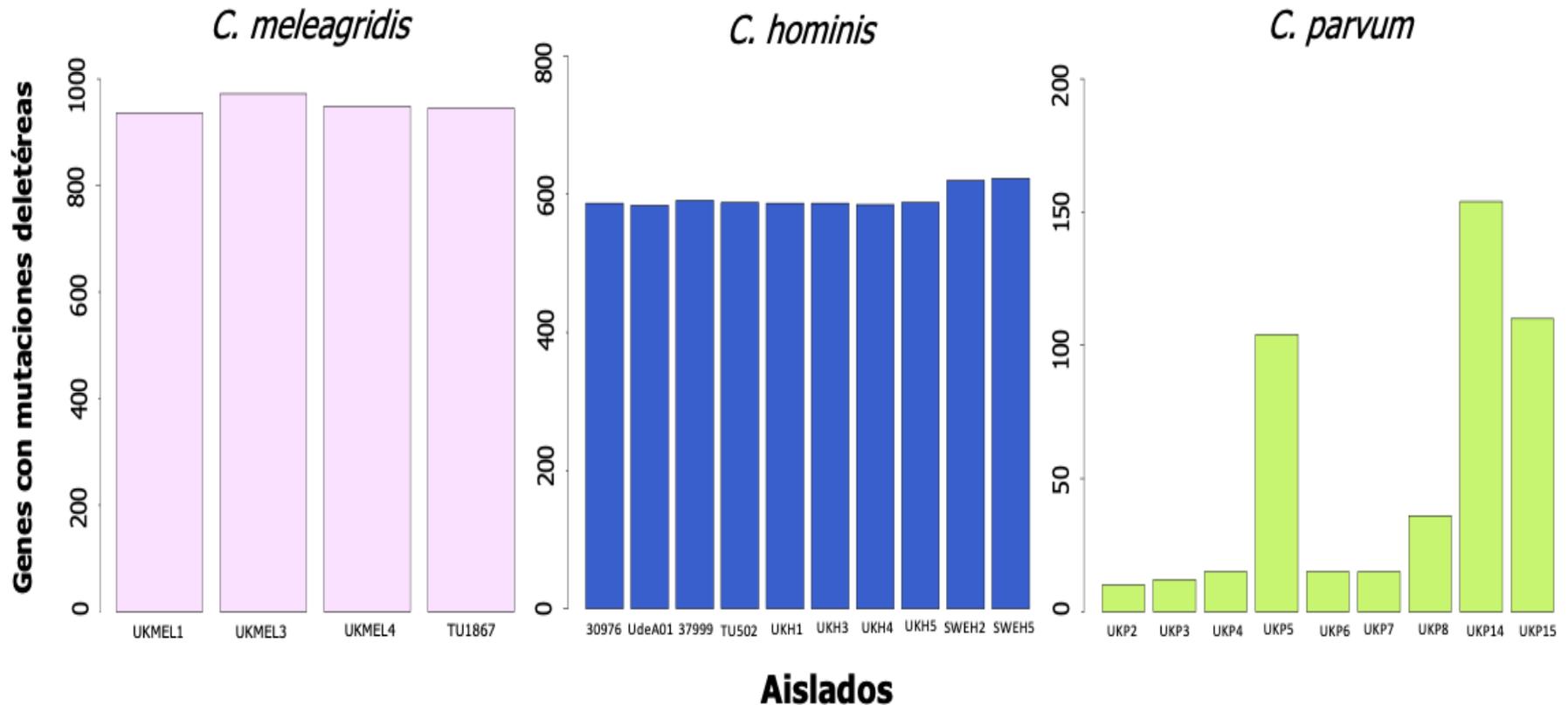
**Figura 8. Análisis de ortología KEGG en los genes con mutaciones deletéreas exclusivas de especie.** Asignación de términos KO a los genes con mutaciones deletéreas exclusivos de especie a través del servidor KAAS en la base de datos KEGG.

#### **6.4.1. Acumulación de genes con mutaciones deletéreas intraespecie y por cromosoma**

Con respecto a la acumulación de mutaciones deletéreas al interior de cada especie, *C. hominis* y *C. meleagris* tuvieron un comportamiento homogéneo con un aumento en los aislados SWEH2 y SWEH5. Mientras que *C. parvum* mostró una mayor acumulación de los genes mutados en los aislados antroponóticos y el aislado zoonótico UKP5 (**Figura 9**). Sin embargo, no se encontraron diferencias significativas en ninguna de las tres especies: *C. meleagris* (p-value 0,3916), *C. hominis* (p-value 0,4439) y *C. parvum* (p-value 0,2873).

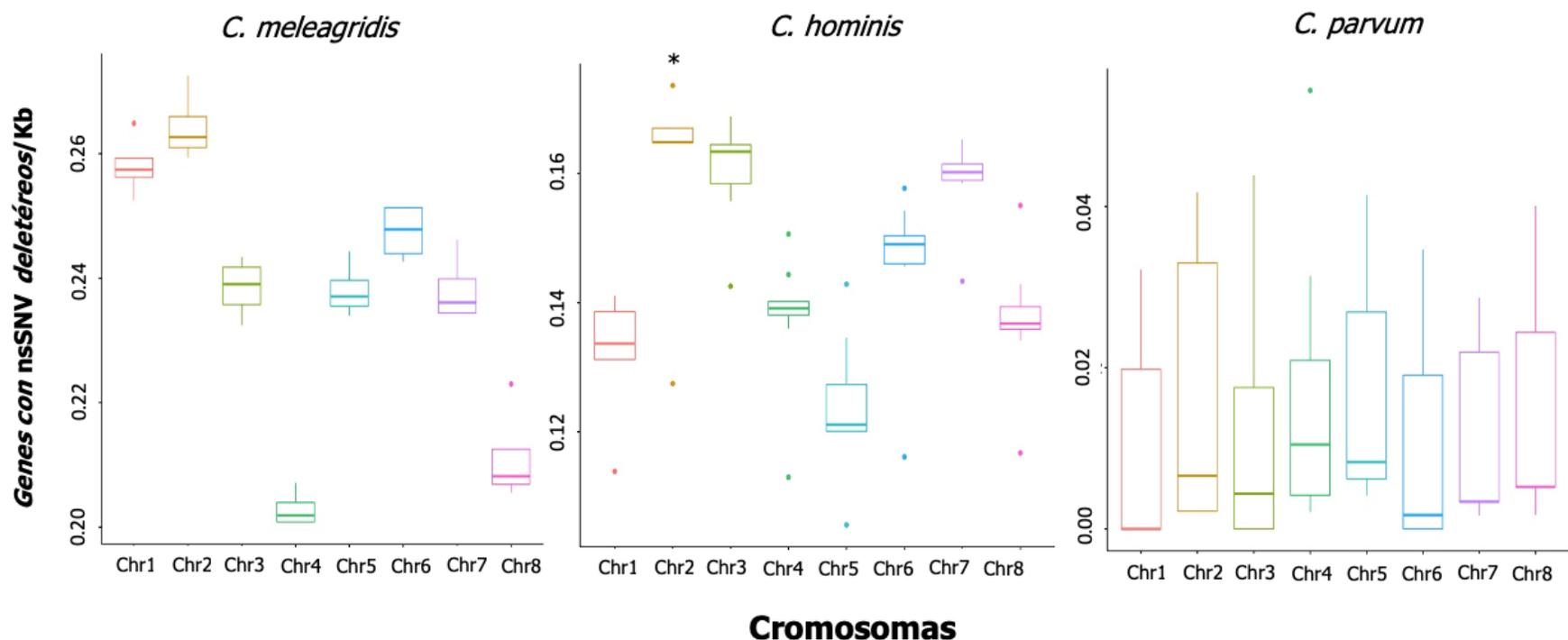
En cuanto a su distribución por cromosoma, en *C. hominis* se evidenció un aumento de mutaciones deletéreas por Kilobase (Kb) en el cromosoma 2 (p-value 0,002158\*). Las otras especies no presentaron diferencias estadísticamente significativas: *C. parvum* (p-value 0,05588) y *C. meleagris* (p-value 0,2711) (**Figura 10**).

## Acumulación de genes con mutaciones deletéreas intraespecie



**Figura 9. Acumulación de genes con mutaciones deletéreas intraespecie.** Gráficas de barras que muestran la acumulación de genes con mutaciones deletéreas en los aislados de cada una de las especies evaluadas. No se encontraron diferencias estadísticamente significativas. Figuras realizadas con R-studio versión 3.3.3.

## Acumulación de genes con *nsSNVs* deletéreos/Kb en cada cromosoma

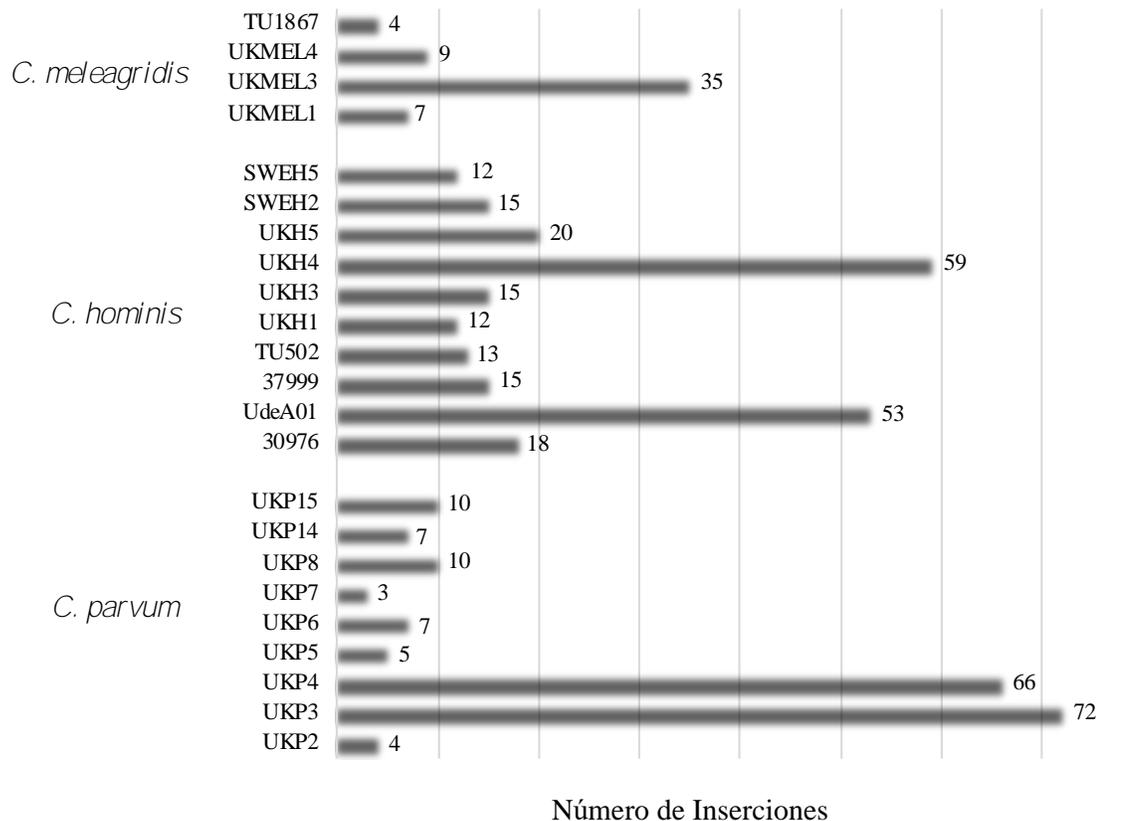


**Figura 10. Acumulación de genes con *nsSNVs* deletéreos/Kb en cada cromosoma.** Diagramas de cajas y bigotes que representan la acumulación de genes con cambios no sinónimos deletéreos por Kb en cada uno de los cromosomas de las tres especies evaluadas. Los conteos de los *SNVs* fueron normalizados con las longitudes de los cromosomas en miles de pares de bases. Los valores atípicos corresponden a los aislados que presentaron un mayor o menor número de genes con *nsSNVs* deletéreos en ese cromosoma. Se encontró una diferencia estadísticamente significativa en la especie *C. hominis* (p-value 0,002158\*). Figuras realizadas con R-studio versión 3.3.3.

## 6.5 Inserciones y deleciones – *Indels*

### 6.5.1 Acumulación de Inserciones inter e intraespecie

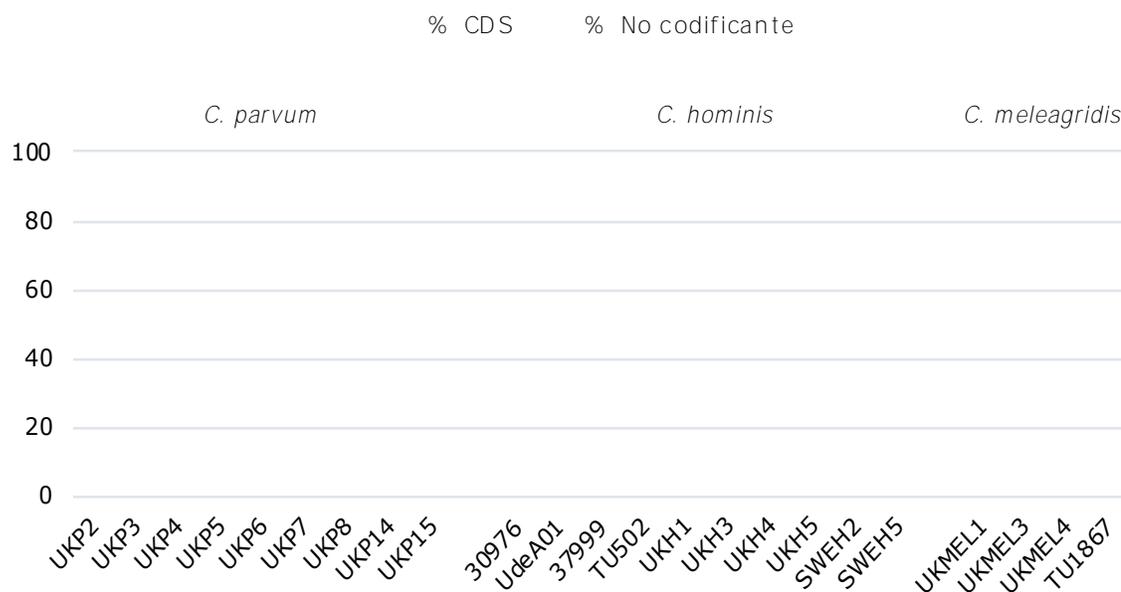
Se identificaron las regiones que presentaban inserciones respecto al genoma de *C. parvum* Iowa II, de las cuales la mayoría estuvieron en el rango de 50 a 500 nucleótidos. En el caso de *C. parvum*, el mayor número de inserciones se detectó en los aislados zoonóticos UKP3 y UKP4. En la Figura 11 se presenta la acumulación de inserciones en las tres especies, el eje "Y" corresponde a los aislados y el eje "X" al número de inserciones. No se encontraron diferencias significativas en la acumulación de inserciones por especie: *C. meleagris* (p-value 0,6519), *C. hominis* (p-value 0,3289) y *C. parvum* (p-value 0,9876).



**Figura 11. Acumulación de inserciones intraespecie.** Gráfica de barras horizontales que muestra el número de inserciones en cada uno de los aislados. No se encontraron diferencias estadísticamente significativas.

### 6.5.2. Identificación de inserciones en regiones codificantes y no codificantes

Para las tres especies, más del 60% de las inserciones se ubicaron en regiones codificantes, con excepción del aislado UKP5 de *C. parvum* en cuyo caso el 100% de las inserciones se ubicaron en regiones no codificantes (5 de 5). En contraste, los aislados UKMEL1 y TU1867 de *C. meleagridis* no presentaron inserciones en zonas no codificantes (**Figura 12**).



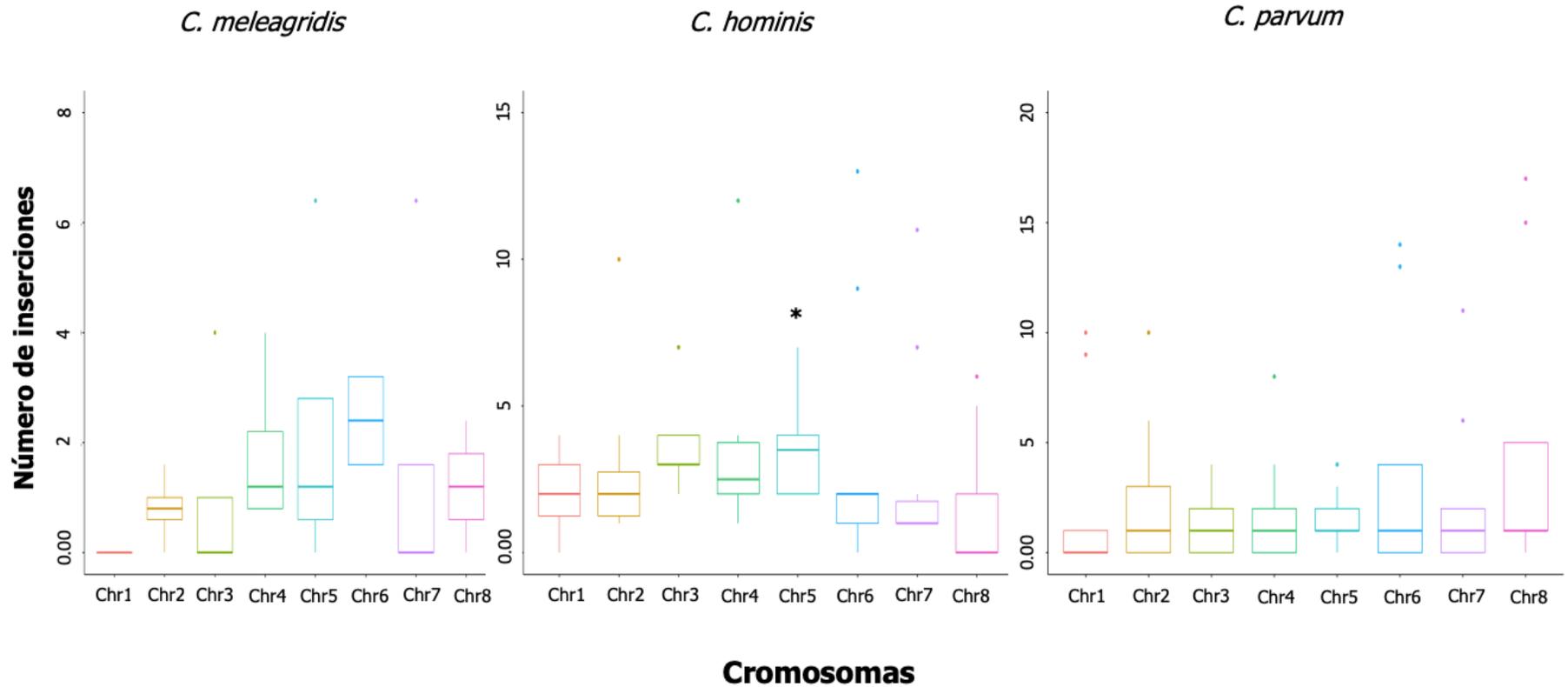
**Figura 12. Acumulación de inserciones intraespecie en regiones codificantes y no codificantes.** Gráficos de barras que muestran la acumulación porcentual de inserciones en CDSs y regiones no codificantes para cada uno de los aislados evaluados. No se encontraron diferencias estadísticamente significativas.

De las inserciones que presentaron una mayor longitud, dos se ubicaron en regiones codificantes en el aislado de *C. meleagridis* UKMEL1 con 1.035 pb en el cromosoma 5 y el aislado antroponótico de *C. parvum* UKP15 con 241 pb en el cromosoma 6. En cuanto a las regiones no codificantes, la inserción más grande se identificó en el aislado de *C. hominis* 30976 con 283 pb en el cromosoma 7.

### **6.5.3. Acumulación de inserciones por cromosoma**

Al evaluar la acumulación de inserciones por cromosoma, no se encontraron diferencias significativas en las especies *C. parvum* (p-value 0,25) y *C. meleagris* (p-value 0,4026), conservando una distribución homogénea. En contraste, en *C. hominis* se presentó una mayor acumulación de inserciones en el cromosoma 5, lo cual fue estadísticamente significativo (p-value 0,01575\*) (**Figura 13**).

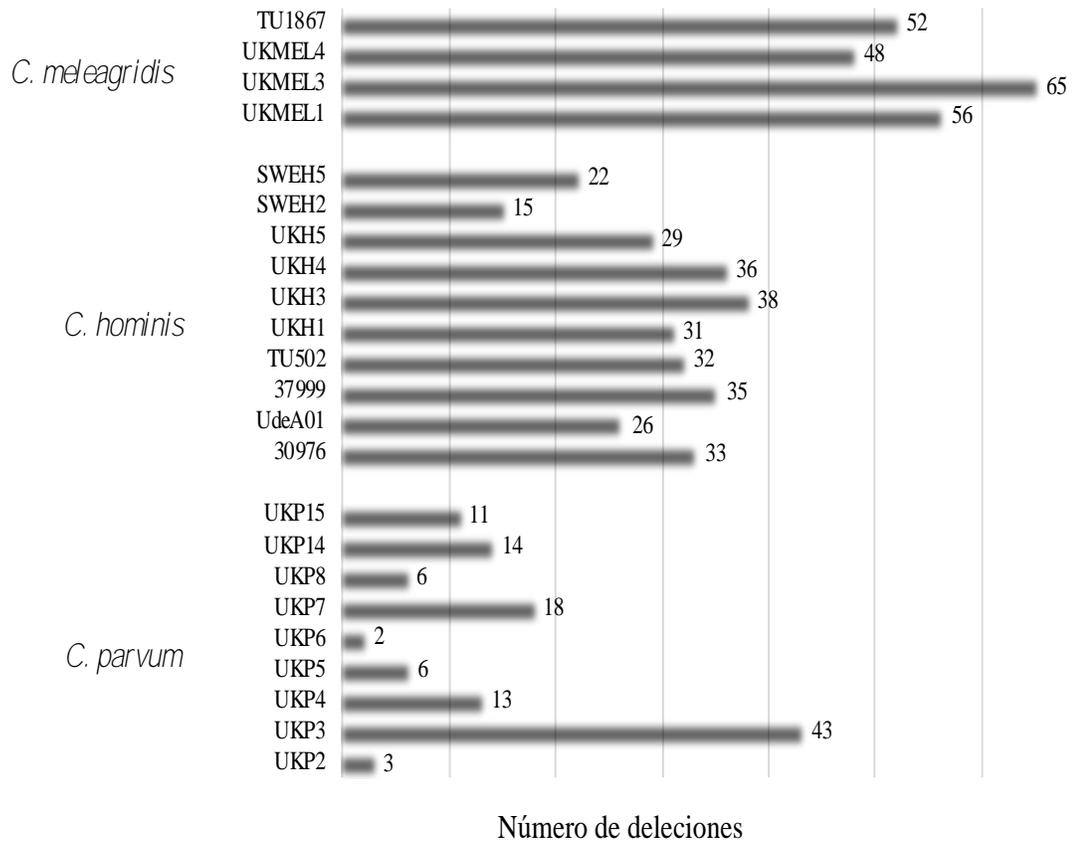
## Acumulación de inserciones por cromosoma



**Figura 13. Acumulación de inserciones por cromosoma.** Diagramas de cajas y bigotes que representan la acumulación de las inserciones en cada uno de los cromosomas de las tres especies evaluadas. No se encontraron inserciones en el cromosoma 1 de *C. meleagris*. Los valores atípicos corresponden a los aislados que presentaron un mayor número de inserciones en ese cromosoma. Se evidenció una diferencia estadísticamente significativa en la especie *C. hominis* (p-value 0,01575\*). Figuras realizadas con R-studio versión 3.3.3.

#### 6.5.4. Acumulación de deleciones inter e intraespecie

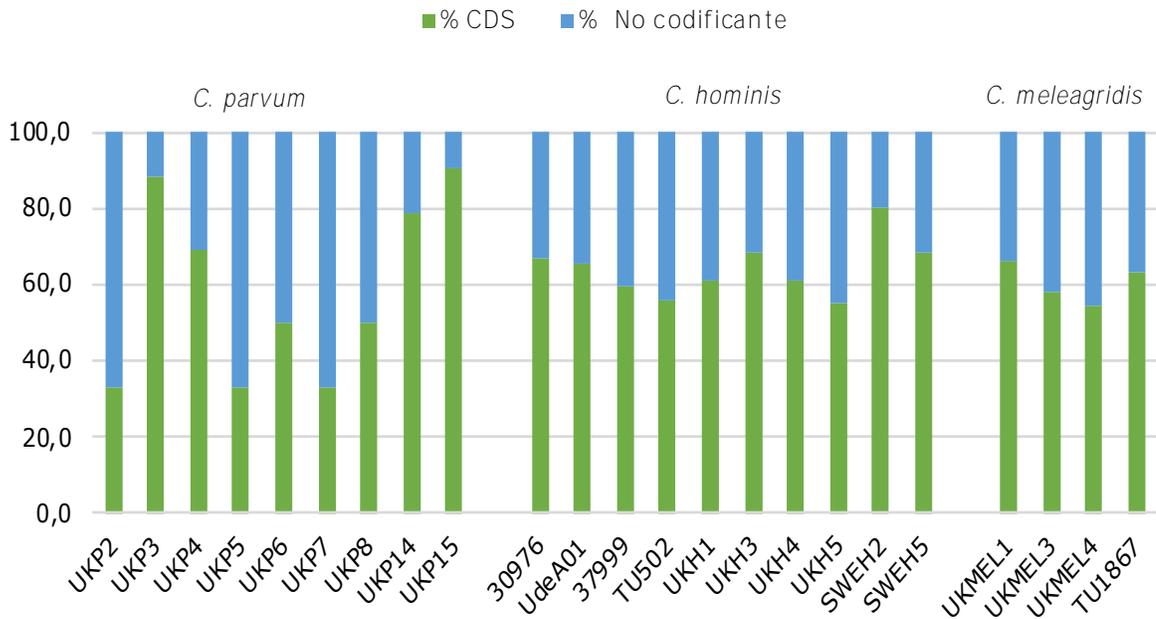
En cuanto a las deleciones, la mayoría estuvieron en el rango de 50 a 500 nucleótidos observándose el mayor número en los aislados de *C. meleagridis*, seguido por *C. hominis*. En la Figura 14 se presenta la acumulación de deleciones en las tres especies evaluadas, el eje "Y" corresponde a los aislados y el eje "X" al número de deleciones. No se encontraron diferencias en la acumulación de deleciones intraespecie: *C. meleagridis* (p-value 0,7926), *C. hominis* (p-value 0,605) y *C. parvum* (p-value 0,3183).



**Figura 14. Acumulación de deleciones intraespecie.** Gráfica de barras horizontales que muestra el número de deleciones en cada uno de los aislados. No se encontraron diferencias estadísticamente significativas en la acumulación de deleciones intraespecie.

### 6.5.5. Identificación de deleciones en regiones codificantes y no codificantes

Para la mayoría de los aislados de las especies *C. hominis* y *C. meleagridis*, la distribución de las deleciones en regiones codificantes y no codificantes fue similar, donde cerca del 60% se ubicaron en CDSs, con excepción del aislado SWEH2 de *C. hominis*, en cuyo caso el 80% de las deleciones se ubicaron en regiones codificantes. Por el contrario, *C. parvum* muestra un patrón intraespecie heterogéneo con un mayor porcentaje de deleciones en CDSs en el aislado zoonótico UKP3 y el antroponótico UKP15 (más del 80%), en comparación con los aislados zoonóticos UKP2, UKP5 y UKP7 en cuyo caso fueron menores al 40% (**Figura 15**).



**Figura 15. Acumulación de deleciones intraespecie en regiones codificantes y no codificantes.** Gráficos de barras que muestran la acumulación porcentual de deleciones en CDSs y regiones no codificantes para cada uno de los aislados evaluados. No se encontraron diferencias estadísticamente significativas.

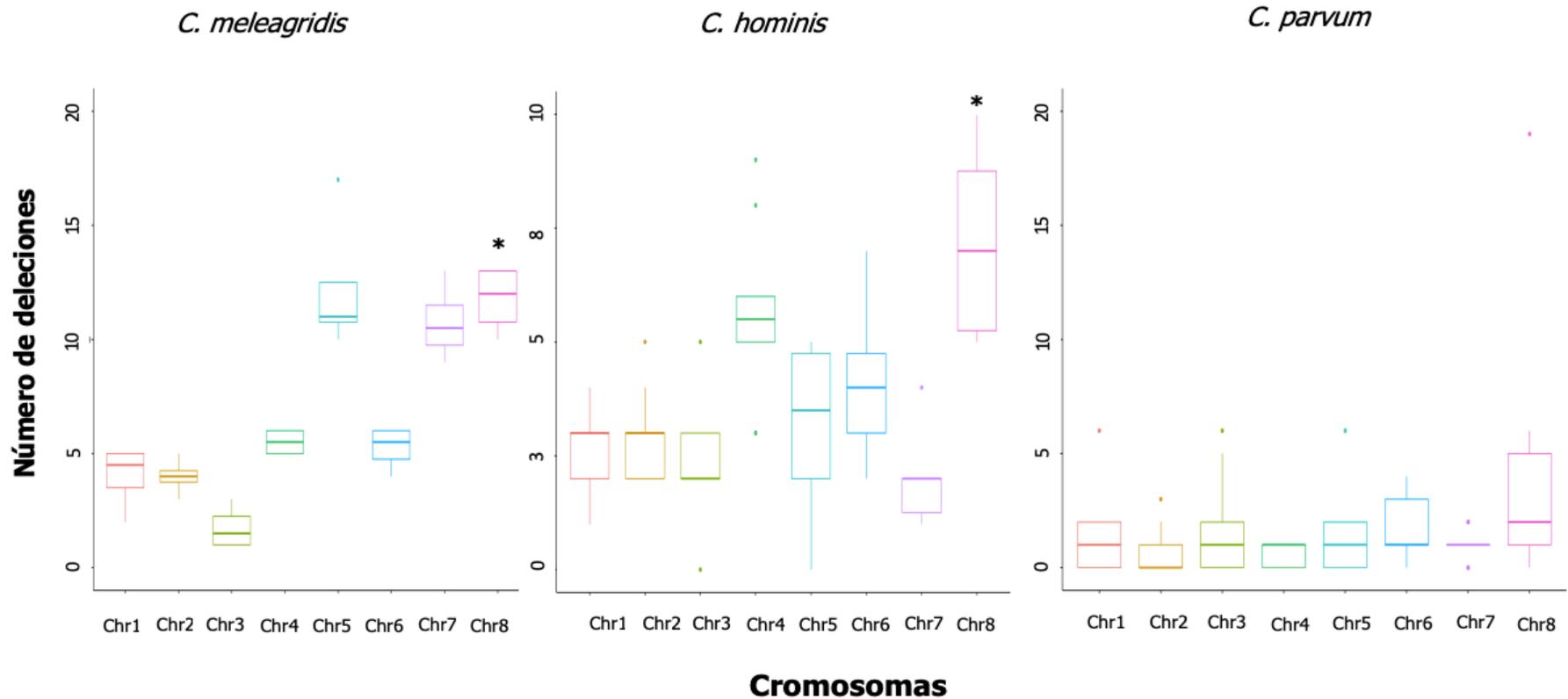
Las deleciones de mayor longitud se ubicaron en regiones codificantes, con 1.142 pb; en el cromosoma 2 del aislado antroponótico de *C. parvum* UKP14; y deleciones

con un tamaño de 1.011 pb y 644 pb en *C. hominis* UdeA01 y *C. meleagridis* UKMEL3 respectivamente, ambas ubicadas en el cromosoma 7 de estos aislados.

#### **6.5.6. Acumulación de deleciones por cromosoma**

La acumulación de deleciones por cromosoma fue similar en *C. parvum* (p-value 0,2753) mientras que *C. hominis* (p-value 0,008838\*) y *C. meleagridis* (p-value 0,02001\*) presentaron diferencias estadísticamente significativas con una mayor acumulación de estas variantes en el cromosoma 8 (**Figura 16**).

## Acumulación de deleciones por cromosoma



**Figura 16. Acumulación de deleciones por cromosoma.** Diagramas de cajas y bigotes que representan la acumulación de las deleciones en cada uno de los cromosomas de las tres especies evaluadas. Los valores atípicos corresponden a los aislados que presentaron un mayor número de deleciones en ese cromosoma. Se encontraron diferencias estadísticamente significativas en las especies *C. hominis* (p-value 0,008838\*) y *C. meleagris* (p-value 0,02001\*). Figuras realizadas con R-studio versión 3.3.3.

### **6.5.7. Identificación de genes con eventos de inserción y delección inter e intraespecie.**

Se identificaron y caracterizaron los genes con inserciones y deleciones en los aislados de las tres especies evaluadas respecto al genoma de *C. parvum* Iowa II, encontrando 322, 215 y 176 genes con inserciones y deleciones en *C. hominis*, *C. parvum* y *C. meleagridis* respectivamente. Tanto en *C. meleagridis* como *C. hominis* las deleciones fueron el *indel* más frecuente, correspondiendo al 76,1% y 59% de todas las variantes, con excepción del aislado UKH4 en el cual las inserciones predominaron. Por el contrario, en *C. parvum* las deleciones solo representaron el 38% de las variantes estructurales detectadas en los genes (**Tabla 12**).

**Tabla 12. Número de genes con eventos de inserción y deleción respecto al genoma de *C. parvum* Iowa II**

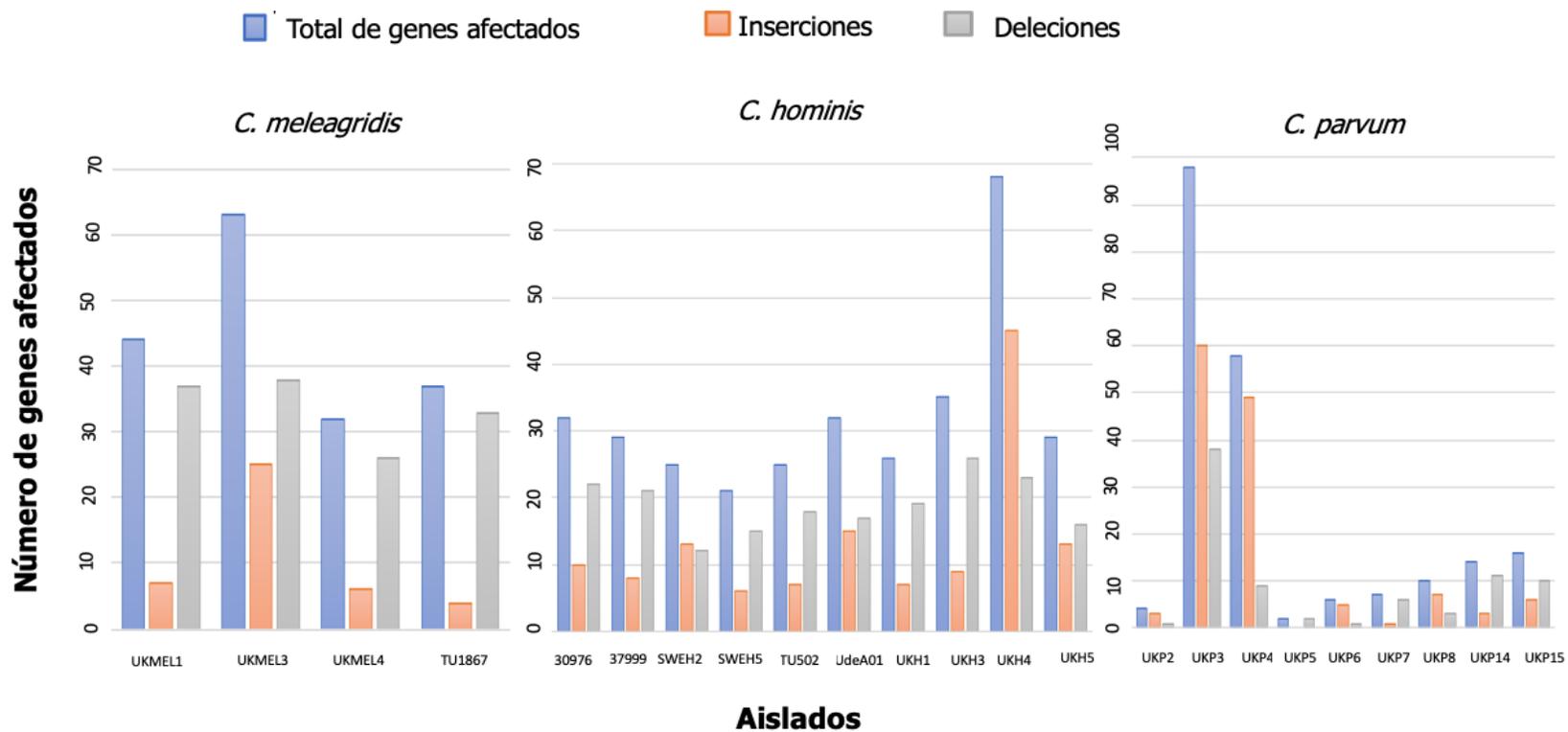
<b>Especie</b>	<b>Aislado</b>	<b>Número de genes</b>	<b>Inserciones</b>	<b>Deleciones</b>
<b><i>C. meleagridis</i></b>	UKMEL1	44	7	37
	UKMEL3	63	25	38
	UKMEL4	32	6	26
	TU1867	37	4	33
	<b>Total</b>	<b>176</b>	<b>42</b>	<b>134</b>
<b><i>C. hominis</i></b>	30976	32	10	22
	37999	29	8	21
	UdeA01	32	15	17
	TU502	25	7	18
	UKH1	26	7	19
	UKH3	35	9	26
	UKH4	68	45	23
	UKH5	29	13	16
	SWEH2	25	13	12
	SWEH5	21	6	15
	<b>Total</b>	<b>322</b>	<b>133</b>	<b>189</b>
<b><i>C. parvum</i></b>	UKP2	4	3	1
	UKP3	98	60	38
	UKP4	58	49	9
	UKP5	2	0	2
	UKP6	6	5	1
	UKP7	7	1	6
	UKP8	10	7	3
	UKP14	14	3	11
	UKP15	16	6	10
	<b>Total</b>	<b>215</b>	<b>134</b>	<b>81</b>

En *C. meleagridis* y *C. hominis* se observó un comportamiento homogéneo en la acumulación de *indels* con hasta 68 genes involucrados, siendo los aislados UKH4 y UKMEL3 los que presentaron más variantes estructurales respecto al genoma de referencia. En el caso de *C. parvum*, se evidencia heterogeneidad intragenotípica con rangos que oscilaron entre 2 y 98 genes, con la mayor acumulación en los aislados zoonóticos UKP3 y UKP4 (**Figura 17**).

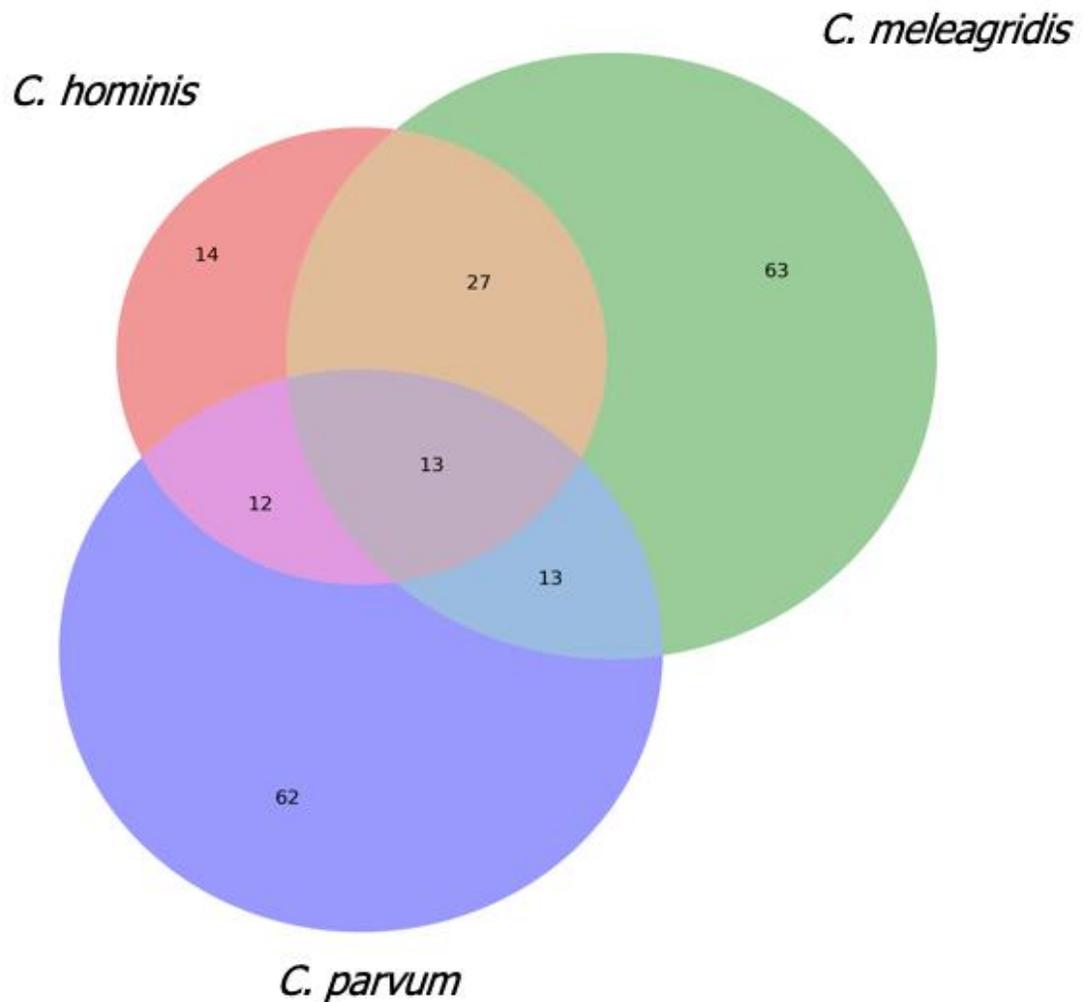
La deleción de mayor longitud se identificó en el aislado antroponótico UKP14 con 1.142 pb en un gen del cromosoma 2 que codifica una oxidoreductasa (cgd2\_2510), seguido de *C. hominis* UdeA01 con 1.011 pb en un gen codificante para una mucina criptosporidial (ortólogo de cgd7\_4020). Las inserciones de mayor longitud fueron identificadas en genes que codifican proteínas no caracterizadas de *C. meleagridis* UKMEL1 y UKMEL3 con 1.035 pb y 644 pb respectivamente.

A partir de la identificación de los genes que presentaron fenómenos de inserción y deleción, se caracterizaron aquellos que estuviesen presentes en por lo menos dos aislados de cada especie, con esta estrategia fueron predichos un total de 116 genes en *C. meleagridis*, 100 en *C. parvum* y 66 en *C. hominis*; de los cuales 13 son compartidos, 63 son exclusivos de *C. meleagridis*, 62 de *C. parvum* y 14 de *C. hominis* (**Figura 18**).

**Número de genes afectados por eventos *indels* en las tres especies de *Cryptosporidium***



**Figura 17. Número de genes afectados por eventos *indels* en las tres especies de *Cryptosporidium*.** Gráficos de barra para representar el total de genes afectados por inserciones y delecciones en los aislados de *Cryptosporidium* evaluados.



**Figura 18. Genes con eventos de inserción y deleción en las especies evaluadas.** Diagrama de Venn que esquematiza los genes con eventos *indels* que son compartidos y exclusivos en las tres especies de *Cryptosporidium* respecto al genoma de referencia Iowa II.

De acuerdo con la anotación depositada en CryptoDB Versión 46, de los 13 genes con eventos de inserción y deleción compartidos por las tres especies, el 40% codifica para proteínas no caracterizadas y los restantes para hidrolasas

(cgd8\_1220), proteínas con motivos de reconocimiento de ARN (cgd3\_4150), mucinas (cgd7\_4020), entre otros.

Se identificaron 27 genes con *indels* en todos los aislados de *C. meleagridis*, de los cuales el 56% equivalen a proteínas no caracterizadas. En cuanto a *C. parvum*, se presentaron inserciones en los genes cgd3\_190 (implicada en la formación de fibrilina) y cgd3\_3900 (proteína no caracterizada) en el 56% de los aislados evaluados. Un hallazgo interesante fue que todos los genomas de *C. hominis* presentaron *indels* en tres genes (ortólogos de cgd6\_4290, cgd7\_420 y cgd7\_500), de los cuales cgd6\_4290 y cgd7\_420 también se ven afectados en todos los aislados de *C. meleagridis* (**Tabla 13**).

**Tabla 13. Anotación de los genes afectados por eventos de *indels***

<i>C. meleagridis indels</i> en el 100% de los aislados (4 de 4)				
Gen ortólogo en <i>C. parvum</i>	Cromosoma	Anotación CryptoDB Versión 46	<i>Indel</i>	Longitud (pb)
<b>cgd2_540</b>	2	<i>Uncharacterized Protein</i>	Delección	132
<b>cgd2_690 *</b>	2	<i>Uncharacterized Protein</i>	2 Deleciones	70 y 93
<b>cgd3_1540</b>	3	<i>Signal peptide containing protein</i>	Delección	78
<b>cgd4_200</b>	4	<i>Uncharacterized Protein</i>	Delección	73
<b>cgd4_770</b>	4	<i>Uncharacterized Protein</i>	Delección	148
<b>cgd4_1010</b>	4	<i>Galactose-binding domain-containing protein</i>	Delección	108
<b>cgd5_220</b>	5	<i>Uncharacterized Protein</i>	Delección	96
<b>cgd5_550</b>	5	<i>Integrator complex subunit 9-like protein</i>	Delección	54
<b>cgd5_1940</b>	5	<i>Uncharacterized Protein</i>	Inserción	278
<b>cgd5_2180 †</b>	5	<i>Uncharacterized Protein</i>	Inserción	1035
<b>cgd5_2250</b>	5	<i>Uncharacterized Protein</i>	Delección	114
<b>cgd5_4490</b>	5	<i>Uncharacterized Protein</i>	Delección	54
<b>cgd6_830</b>	6	<i>Uncharacterized Protein</i>	Inserción	105
<b>cgd6_4290 ±</b>	6	<i>MATH/TRAF domain/WD40 repeat containing protein</i>	Delección	99
<b>cgd6_5400</b>	6	<i>Uncharacterized Protein</i>	Delección	59
<b>cgd7_420 ±</b>	7	<i>Protein with DEXDc ring HELICc possible SNF2 domain</i>	Delección	144
<b>cgd7_890</b>	7	<i>WD40/YVTN repeat-like-containing protein</i>	Delección	108
<b>cgd7_1300</b>	7	<i>Protein kinase-like domain containing protein</i>	Delección	123
<b>cgd7_1310</b>	7	<i>Glycosyltransferase 2-like protein</i>	Delección	259
<b>cgd7_1330</b>	7	<i>Serine/Threonine protein kinase</i>	Delección	96
<b>cgd7_4020</b>	7	<i>Cryptosporidial mucin</i>	Delección	79
<b>cgd8_410</b>	8	<i>Uncharacterized Protein</i>	Delección	72
<b>cgd8_680</b>	8	<i>Uncharacterized Protein</i>	Delección	120
<b>cgd8_1570</b>	8	<i>CCCH like finger domain nucleoporin</i>	Delección	171
<b>cgd8_2040</b>	8	<i>Potassium channel tetramerisation-type BTB domain</i>	Delección	120

<b>cgd8_2770</b>	8	<i>Uncharacterized Protein</i>	Delección	73
<b>cgd8_5190</b>	8	<i>Uncharacterized Protein</i>	Delección	225
<b><i>C. hominis</i> indels en el 100% de los aislados (10 de 10)</b>				
<b>cgd6_4290</b>	6	<i>Delección MATH/TRAF domain/WD40 repeat containing protein</i>	Delección	99
<b>cgd7_420</b>	7	<i>Protein with DEXDc ring HELICc possible SNF2 domain</i>	Delección	144
<b>cgd7_500</b>	7	<i>RTR1-type zinc finger</i>	Inserción	54
<b><i>C. parvum</i> indels en el 56% de los aislados (5 de 9)</b>				
<b>cgd3_190</b>	3	<i>Fibrillin/EGF-like calcium-binding conserved site-containing protein</i>	Inserción	100
<b>cgd3_3900</b>	3	<i>Uncharacterized Protein</i>	Inserción	100

\* Único gen en *C. meleagridis* que presentó dos delecciones en diferentes coordenadas. † Inserción más grande detectada en los genomas de esta especie. ‡ Delecciones compartidas por todos los aislados de *C. meleagridis* y *C. hominis*.

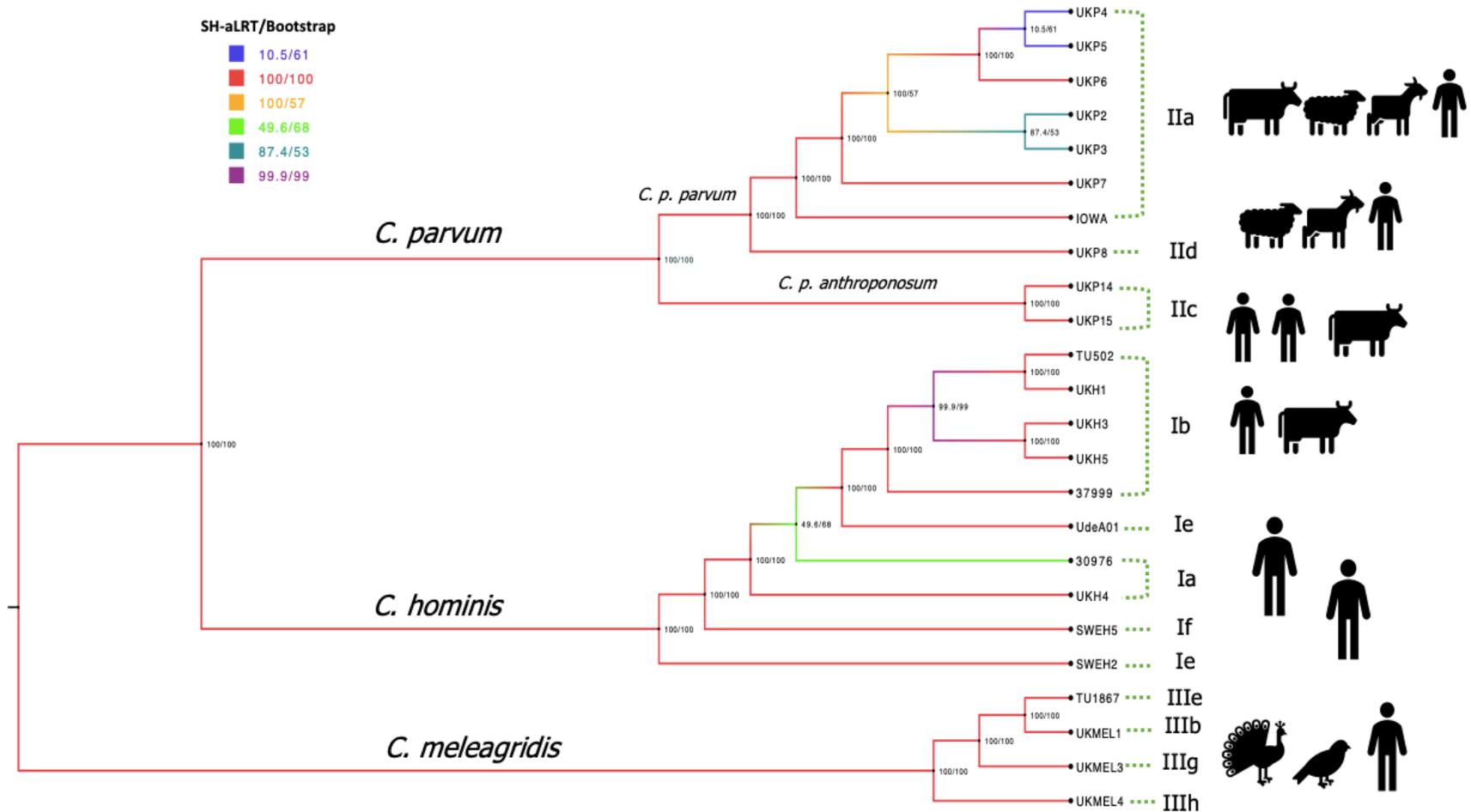
## **6.6. Análisis filogenómico con las tres principales especies de *Cryptosporidium* asociadas con enfermedad en el humano**

Con la finalidad de evaluar el comportamiento de aislados de diferentes subgenotipos y especies de *Cryptosporidium*, se realizó un análisis filogenómico de máxima verosimilitud (ML) a partir de los *SNVs* detectados entre los 24 genomas evaluados (23 genomas ensamblados *de novo* y el genoma de referencia *C. parvum* Iowa II). Para ello se construyó una matriz completa con un total de 800.861 *SNVs* y posteriormente se cargó en el programa IQ-TREE.

El árbol obtenido presenta tres cladas monofiléticas que agrupa los 24 aislados genomas para cada una de las tres principales especies de *Cryptosporidium* que infectan al hombre. Los soportes de rama son óptimos y presentan una concordancia del 100% entre SH-aLRT y *bootstrap* ultra rápido para la mayoría de las ramas, con discrepancias en tres nodos internos de *C. parvum* y un nodo interno de *C. hominis*.

En la especie *C. parvum*, se observó la presencia de dos ramas separadas y con buenos soportes estadísticos las cuales permitieron ubicar los aislados antroponóticos UKP14 y UKP15 en una rama diferente a los aislados zoonóticos.

Todos los genomas de las especies *C. hominis* y *C. parvum* se agruparon de acuerdo con la familia de subgenotipos dada por el marcador *gp60*. En el caso de *C. meleagridis* no fue posible evaluar dicha agrupación debido a la ausencia de varios aislados pertenecientes a la misma familia alélica. No se evidenció una relación entre la topología y la región geográfica de donde fueron obtenidos (**Figura 19**).



**Figura 19. Árbol filogenómico de las tres principales especies de *Cryptosporidium*.** El árbol de máxima verosimilitud (ML) se basó en los SNVs entre los 24 genomas de *Cryptosporidium* analizados. Se utilizó el TVM como modelo evolutivo. Los soportes de las ramas se basan en un análisis de 1000 réplicas **SH-aLRT** (%) /réplicas de **Bootstrap** ultra-rápido (%). Las ramas rojas tienen una concordancia entre SH-aLRT y bootstrap de 100%. Las ramas se identifican con el nombre del aislado, la familia alélica *gp60* y los principales hospederos reportados.

## **7. DISCUSIÓN**

En este estudio se realizó un análisis comparativo de la secuencia completa de 24 genomas de diferentes aislados de las principales especies del género *Cryptosporidium* asociadas con infección en humanos (*C. hominis*, *C. parvum* y *C. meleagridis*) con el objetivo de identificar diferencias genéticas estructurales y funcionales entre dichas especies y explorar su relación con características fenotípicas particulares tales como especificidad de hospedero, distribución geográfica y virulencia, entre otros.

A partir de los datos NGS, todos los proyectos genómicos fueron ensamblados *de novo*, reconstruyendo los genomas a partir de los *reads* superpuestos y siguiendo una estrategia metagenómica con el ensamblador *SPAdes*. Los resultados obtenidos después del ensamblaje revelaron un comportamiento homogéneo, con genomas cercanos a las 9.0 Mb en la mayoría de los aislados, lo que concuerda con estudios genómicos previos realizados con especies de este género (94,101,106,119). Seis aislados presentaron genomas con tamaños cercanos a 8,2 Mb, dos de estos genomas correspondían a los aislados SWEH2 y SWEH5 de *C. hominis*, los cuales fueron secuenciados con Ion Torrent. El menor tamaño obtenido en este caso podría deberse a las longitudes variables de los *reads* y al menor rendimiento reportado para esta plataforma en comparación con Illumina (145). Además de los principios de secuenciación, también existen diferencias en los datos que generan las dos plataformas. En el caso de Illumina, todos los *reads* obtenidos durante un experimento poseen las mismas longitudes y adicionalmente tiene la facilidad de generar lecturas ya sea de secuencia pareada (en inglés paired - end) o de extremo único (en inglés single - end). Mientras que en Ion Torrent las longitudes de los *reads* son variables y solo genera lecturas de extremo único (146). Con respecto a los cuatro genomas restantes, el menor tamaño obtenido podría ser el resultado de las diferentes estrategias metodológicas de preparación y secuenciación de las librerías.

En cuanto a la fragmentación de los ensamblajes, los resultados mostraron la presencia de siete aislados que presentaban más de 1.300 contigs. Estudios previos han descrito que un ensamblaje fragmentado puede estar relacionado con genomas altamente heterocigotos y con un alto contenido de repeticiones (147,148). No obstante, los análisis realizados en este apicomplexa se han enfocado principalmente en los estadios parasitarios haploides y se ha estimado una baja frecuencia de secuencias repetitivas, las cuales representan <1% del genoma (101). Por consiguiente, es posible que la fragmentación evidenciada en siete de los 23 genomas se relacione con la calidad de los *reads* y/o una cobertura superficial de los aislados. A pesar de que el panorama ideal al realizar un ensamblaje *de novo* es obtener el genoma con el menor número posible de secuencias, es usual que algunas regiones estén alineadas con más confianza que otras (149). Sin embargo, la obtención de diferentes niveles de fragmentación en los genomas analizados no generó problemas metodológicos ni sesgos en los alineamientos realizados. Los bloques con una alta similitud fueron alineados a lo largo de toda su longitud, y esto fue independiente del número de secuencias en consulta, alineándose más del 96% de las bases totales en todos los genomas.

A partir de los alineamientos globales realizados se encontraron porcentajes de identidad de secuencia de 96.85% en los aislados de *C. hominis* contra el genoma de referencia *C. parvum* Iowa II, lo que confirma la alta similitud entre ambas especies, y concuerda con estudios previos que reportan diferencias en máximo un 3% (92,104,150). Estos estudios describen que ambos genomas exhiben una notable conservación estructural, y algunos autores sugieren que las diferencias fenotípicas se deben a variaciones sutiles en las secuencias de los genes que codifican para proteínas del parásito que participan en la interacción con la célula hospedadora (92,104,150). Se han reportado diferencias en las copias de genes

subteloméricos que codifican para familias de proteínas tipo mucinas y proteasas, lo que indicaría que están bajo una fuerte selección (103).

En el caso de *C. parvum* se encontraron diferencias entre los aislados evaluados, con un menor porcentaje de bloques alineados en los aislados antroponóticos en comparación con los zoonóticos. Este hallazgo confirma lo descrito por diferentes autores (70,106,109), quienes han encontrado diferencias en cuanto a la estructura genómica y contenido de nucleótidos entre los aislados zoonóticos y antroponóticos de *C. parvum*. Los genomas de *C. parvum* pertenecientes a la familia alélica IIc son más similares a *C. hominis* que a otros aislados zoonóticos de *C. parvum*. Nader y col. (70) encontraron un mayor número de eventos de recombinación entre aislados antroponóticos de *C. parvum* y *C. hominis*, en comparación con los aislados zoonóticos. Adicionalmente, demostraron que los genomas de *C. parvum* que pertenecen a la familia alélica IIc comparten un subconjunto de locus con *C. hominis*, y que la diversidad de nucleótidos entre los aislados zoonóticos y antroponóticos de *C. parvum* es potenciada por el intercambio de información genética mediante la recombinación de genes bajo selección positiva. Estos autores concluyeron que la recombinación juega un papel importante en la biología de *Cryptosporidium*, y que las especies están coevolucionando en presencia de un intercambio genético continuo.

Al realizar el análisis de identidad entre los genomas de *C. meleagridis* versus *C. parvum*, se encontró un porcentaje menor al observado con *C. hominis*, lo que corrobora análisis filogenéticos en los que se muestra una mayor distancia de esta especie con *C. parvum* y *C. hominis* (15,151). Este hallazgo puede relacionarse con las diferencias en la especificidad de hospedador reportadas en las tres especies, donde *C. hominis* y *C. parvum* tienen rangos de hospedador delimitados a mamíferos, mientras que *C. meleagridis* es descrita tanto en mamíferos como en

aves (15,151). Se ha sugerido que posiblemente los mamíferos fueron los hospedadores originales para *C. meleagridis* y que luego la especie se adaptó a las aves (17).

En todos los aislados evaluados los cambios más frecuentes fueron las transiciones de C >T y G >A; este hallazgo concuerda con lo reportado por varios autores (118,152) que encontraron el mismo patrón, donde estas transiciones representaban el 93% de todas las sustituciones; no obstante, no están claras las implicaciones biológicas de dichos cambios. El predominio del cambio C>T se ha reportado en otros apicomplexa, incluyendo *Neospora caninum*, modelo en el que este tipo de transiciones constituyeron uno de los principales cambios de un solo nucleótido (153).

En cuanto a los resultados obtenidos en la identificación de las variantes de nucleótido único, se observó un comportamiento similar en el número y frecuencia en los aislados de *C. hominis* con alrededor de 220.000 variantes, resultados que difieren de lo reportado en estudios comparativos previos (93,113). Isaza y col. (93) a partir de comparaciones interespecie, identificaron en promedio 43.258 *SNVs* entre el genoma de referencia *C. parvum* Iowa II y los aislados de *C. hominis* UdeA01, UKH1 y TU502\_2012, depositados en CryptoDB versión 8. Es posible que el menor número de *SNVs* detectados en el trabajo de Isaza y col. se relacione directamente con el uso de aproximaciones metodológicas diferentes a las que se usaron en el presente trabajo, ya que en el trabajo de Isaza solo tuvieron en cuenta aquellos *SNVs* localizados a una distancia de 30 nucleótidos o más de regiones donde el alineamiento fallaba, pudiendo subestimar el total de las variantes encontradas. Adicionalmente, los genomas utilizados en el estudio de Isaza fueron obtenidos de la base de datos CryptoDB versión 8, cuyos genomas tenían un menor grado de depuración en comparación con los utilizados en este estudio (versión 43). Otra

comparación interespecie en *C. hominis* fue realizada por Gilchrist y col. (113), quienes al mapear los *reads* cortos de 32 genomas de *C. hominis* contra un genoma de *C. parvum* Iowa secuenciado con Pacific Biosciences (PacBio), encontraron un total de 36.780 *SNVs*. El ensamblaje de un genoma a partir de *reads* largos puede influir en la organización estructural y por ende afectar el número de variantes detectadas. Aunque no está claro cómo los métodos de secuenciación de *reads* largos afectan la precisión del ensamblaje (154), se ha evidenciado que las tasas de error en esta plataforma son mucho mayores que las registradas para Illumina (15% versus 0,1%)(155).

Con respecto a las variantes de un solo nucleótido intraespecie en *C. parvum*, se identificó un número mayor de *SNVs* en los aislados antroponóticos con más de 18.000 en comparación con los aislados zoonóticos, que oscilaron entre 1.595 y 5.752 *SNVs*. Este resultado se relaciona con lo reportado por Widmer y col. (106), que identificaron alrededor de 16.606 *SNVs* al comparar los genomas de dos aislados de *C. parvum* con diferentes rangos de hospedadores, el antroponótico *C. parvum* TU114 y el genoma del aislado zoonótico de referencia Iowa II. La acumulación diferencial de *SNVs* intragenotípicas en *C. parvum* refleja una mayor diversidad genética, la cual posiblemente se encuentre relacionada con la habilidad del parásito para adaptarse a la presión de selección que encuentra en los diferentes hospedadores ya sea humanos y/o animales, por esta razón se ha postulado que la acumulación de variantes podría influir en el rango de hospedador (156–158).

En cuanto a *C. meleagridis* se identificaron más de 600.000 *SNVs* en los aislados comparados contra el genoma de referencia; es importante resaltar que actualmente no hay publicaciones que den cuenta de análisis genómicos en esta especie. Esto refleja el carácter innovador de esta investigación, siendo el primer estudio que

realiza una aproximación a la caracterización genómica estructural y/o funcional en esta especie.

En el presente análisis se encontró una acumulación de variantes homogénea en los cromosomas de *C. parvum*, mientras que en *C. hominis* y *C. meleagridis* se observó un mayor número en los cromosomas 1 y 3. Estos hallazgos difieren de reportes previos que sugieren una distribución uniforme de los *SNVs* en *C. hominis* (113) y la presencia de locus altamente divergentes en los cromosomas 1, 4 y 6 de *C. parvum* (106,118). Estas discrepancias pueden explicarse por el mayor número de genomas analizados en este estudio en comparación con las investigaciones anteriores.

Análisis de genómica comparativa realizados en otros apicomplexas han identificado más de 112.000 *SNVs* en genomas de *Plasmodium falciparum*, 70.000 entre aislados de *Toxoplasma gondii*, cerca de 3.130 *SNVs* en genomas de *Neospora caninum* y valores que oscilaban entre 46.000 y 120.000 en *Theileria parva*. Estos análisis han contribuido sustancialmente en la comprensión de la diversidad genética al interior de estos parásitos, así como en la determinación de nuevos marcadores de virulencia. Adicionalmente, la identificación de *SNVs* tiene el potencial de complementar los análisis realizados con marcadores micro y mini satélites para explorar la diversidad intraespecie (153,159–161).

El análisis de *SNVs* ha sido útil en la evaluación de las relaciones filogenéticas entre diferentes organismos, proporcionando información sobre la evolución (162–164). No obstante, la relación filogenética dada por un número limitado de locus debe analizarse con cautela en el caso de patógenos que adquieren diversidad genética por recombinación, en lugar de la acumulación de mutaciones (161). Teniendo en cuenta lo anterior, los enfoques filogenómicos han ganado protagonismo al estudiar los genomas completos desde una perspectiva evolutiva con el objetivo de entender

los mecanismos de los cambios genéticos, así como su interrelación con la evolución de los organismos y sus características fenotípicas (165). Esta aproximación constituye una herramienta potente para la interpretación de los sistemas biológicos, llevando los análisis filogenéticos a escalas que cubren todo el genoma, en lugar de solo unas pocas familias de genes (165). Estos análisis se han utilizado en el estudio del metabolismo, la patogenicidad, la fisiología y el comportamiento de organismos en general (162). Adicionalmente, pueden ser una herramienta importante para la solución de problemas taxonómicos entre organismos estrechamente relacionados, el desarrollo de nuevos marcadores genéticos y la toma de decisiones de tratamiento en enfermedades infecciosas (166–168).

Los estudios filogenómicos en parásitos apicomplexa son escasos. En el modelo de *Theileria* spp. (161,169), el análisis de *SNVs* distribuidos en todo el genoma, demostró que las cepas analizadas presentan divergencia genética y que la secuencia del genoma completo podría usarse para determinar con mayor claridad las relaciones filogenéticas entre especies de este género. El análisis filogenómico basado en un alto número de CDSs en aislados de *Babesia microti* (170), propuso la reubicación taxonómica de esta especie en un nuevo género dentro de los apicomplexa. Con respecto a *Plasmodium* (171), un estudio realizado a partir del análisis de secuencia de múltiples locus en todo el genoma, fue de utilidad para dilucidar las relaciones filogenéticas en el orden Haemosporida.

En cuanto al género *Cryptosporidium* los análisis con enfoque filogenómico se han centrado en *C. hominis* y *C. parvum*; principalmente mediante el estudio de múltiples locus concatenados (70,118), o de diferentes posiciones variables que no alcanzan a representar el 1% del genoma (113). Teniendo en cuenta lo anterior, y con la finalidad de conocer las relaciones evolutivas y confirmar la ubicación taxonómica de las especies evaluadas, se realizó un análisis filogenómico a partir de una matriz

generada con las coordenadas de todos los *SNVs* identificados en los 24 genomas seleccionados (23 genomas analizados y el genoma de referencia *C. parvum* Iowa II). Para la realización del árbol se implementó el método de máxima verosimilitud (ML) que si bien es uno de los análisis más costosos computacionalmente (requiere mayor disponibilidad de RAM), es considerado uno de los más fiables, y es frecuentemente utilizado en el campo de la filogenómica (149). El árbol generado conserva la topología previamente reportada para las tres especies intestinales de *Cryptosporidium*, a partir de estudios filogenéticos unilocus y de múltiples locus con marcadores tales como 18S rRNA, el gen *gp60* y genes altamente polimórficos; donde se evidencia la presencia de tres cladas diferentes y se confirma la cercanía entre *C. hominis* y *C. parvum*, siendo *C. meleagridis* una especie más divergente (70,151,172).

Las dos pruebas estadísticas utilizadas para dar soporte a las ramas concuerdan en la mayoría de los casos, soportadas con una señal filogenética alta; sin embargo, se presentaron discrepancias en tres nodos internos de *C. parvum* y un nodo interno de *C. hominis*. Es importante indicar que las diferencias mencionadas no alteran la topología global del árbol filogenómico obtenido. En literatura relacionada, se ha descrito que las discordancias en los soportes de las ramas obtenidos mediante bootstrap y pruebas de razón de probabilidad pueden surgir como consecuencia de realizar el análisis con muestras pequeñas y con modelos de sustitución de nucleótidos altamente heterogéneos (173).

El análisis filogenómico de los aislados de *C. parvum* incluidos en el estudio evidenció la separación del clado principal en dos ramas con un soporte estadístico de 100%, ubicándose en una rama los aislados zoonóticos y en otra los antroponóticos. Este hallazgo concuerda con lo reportado por varios autores (70,106,118,174), que a través de análisis filogenéticos unilocus y de múltiples locus entre aislados

antroponóticos y zoonóticos, determinaron que los aislados antroponóticos de *C. parvum* formaban un grupo separado de los aislados zoonóticos de esta especie. Nader y col. (70) a través de un análisis filogenético de 61 locus con evolución neutral en 21 genomas de *Cryptosporidium*, identificaron dos linajes de *C. parvum* con especificidad de hospedador diferente, basados en lo cual propusieron la designación de *Cryptosporidium parvum parvum* para aislados zoonóticos, y *Cryptosporidium parvum anthroponosum* para antroponóticos.

Al realizar un árbol filogenómico incluyendo 800.861 posiciones variables entre 24 genomas de diferentes especies, se observó una agrupación que coincidía con la clasificación de las familias alélicas de *gp60* de los aislados de *C. hominis* y *C. parvum* evaluados en este trabajo; sin embargo, este comportamiento no se observa en otros estudios. Gilchrist y col. (113) analizaron 1.582 *SNVs* en 32 genomas de *C. hominis* circulantes en Bangladesh. En el árbol obtenido los investigadores no encontraron agrupaciones en cuanto a la familia de subgenotipos, concluyendo que el uso de un solo marcador (*gp60*) no refleja los cambios evolutivos de todo el genoma; y confirmando a su vez la debilidad de la tipificación con marcadores únicos en las asignaciones taxonómicas dentro de este género. Nuestros hallazgos permiten confirmar que al realizar un análisis con un mayor número de posiciones en diferentes aislados se mejora considerablemente el poder resolutivo en comparación con el obtenido a partir de análisis unilocus, de múltiples locus o con fragmentos parciales del genoma.

Adicional a la fuerte selección positiva que se ha descrito para el locus *gp60*, otro aspecto que pudo influir en la topología del árbol filogenómico obtenido por Gilchrist y col. (113), es la alta tasa de recombinación en el cromosoma 6 reportada entre aislados circulantes de países endémicos para *C. hominis* (150,175). Lo anterior se asocia con una mayor variabilidad genética, generando subgenotipos

hipertransmisibles y favoreciendo que las familias alélicas basadas en gp60 se distribuyan ampliamente en el árbol filogenético sin mostrar agrupación.

Teniendo en cuenta que los únicos cuatro genomas en la especie *C. meleagridis* pertenecen a subgenotipos diferentes no fue posible evaluar el agrupamiento de los aislados en cuanto a las familias alélicas dadas por el marcador *gp60*. Es indudable la necesidad de aumentar el número de proyectos genómicos de diferentes familias de subgenotipos no solo en *C. meleagridis* sino también en *C. hominis* y *C. parvum*, para facilitar la realización de análisis adicionales con miras a reflejar la variabilidad genética ya descrita al interior de las especies de este género y enriquecer las conclusiones taxonómicas y evolutivas derivadas de este tipo de estudios.

Diversos autores han propuesto que las diferencias fenotípicas entre las especies de *Cryptosporidium* están relacionadas con polimorfismos sobre regiones codificantes para proteínas (37,70,92,176). En relación con esta aproximación, en nuestro estudio se analizaron las variantes de un solo nucleótido ubicadas en CDSs. Adicionalmente, y teniendo en cuenta que una sustitución no sinónima altera la secuencia de la proteína, y por ende puede llegar a causar cambios en la actividad bioquímica, las interacciones proteína-proteína y la función molecular, entre otros (153), también se identificaron los cambios sinónimos y no sinónimos en todas las variantes detectadas.

Las especies que presentaron el mayor número de *SNVs* en regiones codificantes para proteínas fueron *C. meleagridis* y *C. hominis* con valores superiores a 400.000 y 150.000 variantes respectivamente, de las cuales el 42% correspondieron a cambios no sinónimos. En cuanto al número de *SNVs* en regiones codificantes para la especie *C. hominis*, nuestros resultados difieren de lo descrito por Isaza y col (93), quienes mediante un análisis comparativo de cuatro aislados de *C. hominis* (versión

8 de CryptoDB) frente a *C. parvum* Iowa II, identificaron 36.753 *SNVs* localizados sobre regiones codificantes. Es posible que estas discordancias en cuanto al número total de *SNVs* identificados estén relacionadas con diferencias en las estrategias metodológicas utilizadas. Aunque en nuestro estudio no se identificaron los genes que presentaron una mayor cantidad de *SNVs*, análisis comparativos realizados con anterioridad han encontrado que los genes con mayor número de *SNVs* en *C. hominis* y *C. parvum*, fueron aquellos relacionados con el ensamblaje de ribosomas, los procesos de traducción y los genes codificantes para proteínas con dominios transmembrana (93,106).

En el caso de *C. parvum* el número de variantes en regiones codificantes fue menor, manteniéndose por debajo de los 20.000 *SNVs* y mostrando diferencias intraespecie relacionadas con un comportamiento heterogéneo entre los aislados zoonóticos y los antroponóticos. Adicionalmente, se identificó que más del 50% de los *SNVs* localizados sobre regiones codificantes corresponden a mutaciones no sinónimas. Esto podría indicar una mayor divergencia genética y la presencia de una fuerza selectiva que favoreciese el cambio de la secuencia ocasionando el truncamiento de un mayor número de genes. En un estudio previo realizado por Widmer y Col. (106), a partir de análisis comparativos entre *C. parvum* Iowa y el aislado antroponótico *C. parvum* TU114, se observó que los *SNVs* no sinónimos se presentaban en un rango de 28% al 32% de todos los *SNVs* y el 60% de todas las posiciones de nucleótidos en los dos genomas no son sinónimos.

En otros apicomplexa también se han estudiado las variantes en CDSs, incluyendo especies de los géneros *Plasmodium*, *Neospora* y *Theileria* (153,161,169,177). De Oliveira y col. (177) describieron que el 36.2% de los *SNVs* ubicados en regiones codificantes en *Plasmodium vivax* correspondían a mutaciones no sinónimas, y que la presencia de estos cambios podían estar implicados en la resistencia a los

medicamentos antimaláricos. En el caso de *Neospora caninum*, un estudio previo determinó que el 6.7% de los genes tenían por lo menos una variante no sinónima (153).

Con respecto a los genes con mutaciones deletéreas, Sikora y col. (51), realizaron un análisis intraespecie comparando 14 genomas de *C. hominis*, en el cual encontraron 18 genes con mutaciones no sinónimas, de los cuales solo el gen que codifica para una proteína de la pared del ooquiste COWP9 (ortólogo de *cgd6\_210*) tenía mutaciones perjudiciales en ocho de los catorce genomas, sugiriendo una modificación que la convertía en una proteína no funcional. En el presente estudio, fueron predichos 183 genes con mutaciones que están presentes en por lo menos un aislado de cada especie, 377 exclusivos de *C. meleagridis*, 103 de *C. hominis* y 33 de *C. parvum*. Al evaluar la anotación de los genes compartidos y exclusivos de especie, se encontró que la mayoría codifican para enzimas y proteínas implicadas procesos de reparación y recombinación del ADN, así como la biogénesis de ARN de transferencia y de ribosomas. No obstante, se requieren estudios complementarios de expresión génica desde la activación del gen hasta que la proteína realiza su función, utilizando organismos silvestres y *knockout*. Asimismo, identificar si estos genes presentan transcripción activa mediante análisis de transcriptómica. Estos enfoques adicionales serían de utilidad para determinar si tales sustituciones de aminoácidos realmente tienen un efecto en la biología del parásito.

Los genes con mutaciones deletéreas compartidos y exclusivos de especie identificados en este estudio, codifican en su mayoría para proteínas de superficie con dominios transmembrana y proteínas secretadas por vías no clásicas; sugiriendo que se pueden afectar los procesos de interacción entre el parásito y las células del hospedero. Este hallazgo concuerda con lo reportado por otros autores que han descrito que los procesos que se afectan principalmente son los de adhesión e

invasión (93,113,119,175,178). Se ha determinado que las proteínas secretadas por vías no clásicas, suelen ser factores de crecimiento, citocinas inflamatorias, componentes de la matriz extracelular que regulan la diferenciación celular, la proliferación y apoptosis, así como proteínas de la superficie en los parásitos involucradas en las fases iniciales de la interacción con el hospedador (179). Calarco y col. (153) enfatizaron la importancia de no solo identificar y comparar las variantes de secuencia presentes en los genes, sino también comprender si tales mutaciones tienen el potencial de interrumpir la función de la proteína resultante. No obstante, en el presente análisis comparativo, el 53,5% de todos los genes con cambios deletéreos se ubicaron en regiones codificantes no caracterizadas (proteínas hipotéticas), dificultando en gran medida la asignación de un contexto biológico a estas mutaciones; es por ello que debido a la ausencia de anotación en los diferentes genomas publicados de *Cryptosporidium*, varios autores han sugerido la necesidad de implementar estrategias adicionales para mejorar la anotación existente e identificar las funciones de las proteínas hipotéticas predichas con el fin de mejorar la comprensión de la virulencia y la patogenicidad en este género (93,94).

En respuesta a lo anterior, algunos esfuerzos se han realizado en este campo. Isaza y col. (93) realizaron la re-anotación del genoma de referencia *C. parvum* Iowa II, utilizando transcriptómica y la anotación automática con posterior curación manual. Se evidenció un aumento en el recuento de genes codificantes, pasando de 3.805 a 3.865 genes, y también cambió el porcentaje de genes con intrones, pasando de 4,3% a 10,84%. Sin embargo, en la actualidad no se cuenta con publicaciones cuyo principal objetivo sea mejorar la anotación estructural y funcional en este género, por lo tanto, se requieren más investigaciones con miras a realizar la curación manual de las regiones codificantes, puesto que cuanto mejor sea la anotación de los genomas, más información proporcionará a la comunidad científica (180).

Posterior al análisis de las variantes de un solo nucleótido y su aplicación en el campo de la filogenómica, se evaluaron las inserciones y deleciones en los 23 aislados estudiados respecto al genoma de referencia *C. parvum* Iowa II. Se observó que más del 60% de todos los *indels* detectados se ubicaron en CDSs, lo cual se esperaba, teniendo en cuenta que varios autores han descrito que el genoma de *Cryptosporidium* posee un porcentaje de regiones codificantes mayor al 70% (70,119). El mayor número de deleciones se presentó en los aislados de *C. meleagridis*, seguido de *C. hominis*, lo que podría sugerir la pérdida de fragmentos parciales del genoma en dichas especies respecto a *C. parvum*. Los resultados obtenidos en este trabajo demuestran que en los 23 genomas analizados los *indels* son menos abundantes que los *SNVs*, contrario a lo reportado en otros apicomplexas como *Plasmodium falciparum*, en el cual son abundantes y por lo tanto, el mecanismo dominante de polimorfismo dentro del genoma (181).

Feng y col. (118), al comparar cuatro genomas de *C. parvum*, identificaron alrededor de 1.200 eventos de inserción y 1.500 deleciones, los cuales tenían una distribución homogénea en todos los cromosomas, con un predominio de los *indels* de mayor longitud en el cromosoma 6. En contraste, en nuestro estudio al incluir un mayor número de genomas pertenecientes a esta especie se pudo determinar que se presentaban menos de 100 variaciones estructurales en cada uno de los aislados tanto zoonóticos como antroponóticos, siendo el cromosoma 2 el que contenía las variantes de mayor longitud.

A pesar de que en el presente análisis comparativo no se identificaron las regiones de los cromosomas en las que se presentan con mayor frecuencia las inserciones y deleciones, estudios previos han determinado que los *indels* son significativamente más frecuentes en las regiones peri teloméricas y subteloméricas (70). Guo y col. (103) al comparar cinco genomas de *C. hominis* con la referencia *C. parvum* Iowa,

identificaron varias inserciones y deleciones cerca de los telómeros en el cromosoma 6, lo cual podría indicar que la duplicación o delección de genes subteloméricos está involucrada en las diferencias en cuanto a la especificidad de hospedador entre especies de *Cryptosporidium*. Posteriormente, la caracterización molecular de dichas regiones permitió determinar la participación de miembros de familias de genes multicopia y bajo una fuerte selección positiva, como las proteínas MEDLE, proteasas similares a la insulina y glucoproteínas de tipo mucina, relacionadas con la interacción parásito - hospedador (109,119,182).

Aunque se ha descrito la presencia de deleciones en los genes que codifican las proteínas MEDLE en *C. parvum* y *C. hominis* (106,117) en el presente estudio no se evidenciaron variantes estructurales en estos genes o los que codifican para proteasas similares a insulina. Sin embargo, se encontraron deleciones en todos los genomas de *C. meleagridis* y en por lo menos en dos aislados de *C. hominis* y *C. parvum* en un gen codificante para una mucina criptosporidial (ortólogo de cgd7\_4020), también conocida como gp900. Se ha detallado que la glicoproteína gp900 almacenada en los micronemas y presente en la superficie del parásito está implicada en la adhesión de la porción apical de esporozoítos y merozoítos a la superficie de los enterocitos para iniciar el proceso de invasión (183,184). Estudios *in vitro* (185,186) con dos aislados de *C. parvum* encontraron que gp900 estaba codificada en un solo locus, no presentaba reordenamientos genómicos significativos y que, adicionalmente, actuaba como blanco para anticuerpos que inhiben la invasión, sugiriendo que esta glicoproteína es altamente inmunogénica en etapas invasivas del parásito. Aunque se conoce la importancia de gp900 como mediador de la adhesión en el ciclo de vida de *Cryptosporidium*, no es la única molécula que participa en este proceso, ya que las proteínas relacionadas con la trombospondina, p23, el complejo proteico gp40 / p30 y la proteína de tipo circunsporozoito - CSL, también tienen un rol importante en este proceso (27,37). Esta redundancia en el

repertorio de proteínas implicadas en los procesos de adhesión sugiere que las alteraciones identificadas en el gen de la gp900 no afectarían los procesos de interacción con la célula hospedera. Sin embargo, se requieren análisis adicionales para determinar las implicaciones biológicas de estas deleciones en el proceso de unión e invasión principalmente en aislados de *C. meleagridis*.

Otro hallazgo interesante en este estudio fue la identificación de deleciones en todos los genomas de *C. meleagridis* y *C. hominis* para un gen con dominio WD-40 (ortólogo de cgd6\_4290) y otro con dominios SNF2/DEXDc/HELICc (ortólogo de cdg7\_420). El dominio WD-40 tiene repeticiones de triptófano-Ácido aspártico (WD) de aproximadamente 40 aminoácidos y es considerado uno de los diez dominios proteicos más abundantes en los eucariotas (187,188). Las proteínas que contienen estas repeticiones están involucradas en diversos procesos celulares, actuando como un adaptador para las interacciones proteína-proteína y proteína-ADN, en la transducción de señales, transcripción, regulación del ciclo celular y apoptosis; sin embargo, no se le ha atribuido ninguna actividad enzimática intrínseca (188,189). Estos dominios se han reportado como altamente polimórficos en otros protozoos apicomplexas como *Plasmodium falciparum*, sugiriendo la participación de los WD-40 en procesos celulares y metabólicos básicos (190).

Por su parte, los miembros del complejo SWI2/SNF2 utilizan el ATP para remodelar la cromatina y reposicionar los nucleosomas; tienen un dominio N-terminal DEXDc y C-terminal HELICc (191). Reportes previos indican la presencia de homólogos de estos dominios en diferentes apicomplexas, incluidos *Plasmodium falciparum* (192) y *Toxoplasma gondii* (193). Se ha descrito que *Cryptosporidium* posee una versión de esta ATPasa con una arquitectura única que contiene dos dominios como N-terminales y un dominio bromo. Por tal motivo se ha sugerido que esta enzima podría

modificar los grupos amino de las histonas, pero su función específica no ha sido determinada (194).

Se identificaron inserciones exclusivas en el 50% de los aislados de *C. parvum* en el gen *cgd3\_190*. Este gen codifica una proteína almacenada en los micronemas con dominios similares al factor de crecimiento epidérmico - EGF involucrados en la señalización celular (183). En otros apicomplexa, como *Toxoplasma gondii*, estos dominios se han asociado con los procesos de adhesión a la célula hospedadora (195). También se ha demostrado, que en presencia de calcio los dominios EGF adoptan una estructura extendida y resistente a proteasas favoreciendo la interacción de la porción N-terminal de la molécula con los ligandos de la célula hospedadora favoreciendo la invasión (183).

## **8. CONCLUSIONES**

1. Los genomas estudiados presentan fenómenos de variación genómica, que van desde el cambio de un solo nucleótido hasta inserciones y deleciones mayores o iguales a 50 nucleótidos. Además presentaron eventos de variabilidad genética interespecie e intragenotípicas, siendo los *SNVs* el principal polimorfismo detectado.
2. Las especies *Cryptosporidium hominis* y *Cryptosporidium meleagridis* presentaron el mayor número tanto de variaciones de un solo nucleótido como de inserciones y deleciones con diferencias en la acumulación de estas variantes en los cromosomas.
3. El análisis de los *SNVs* detectados contribuye sustancialmente en la comprensión de la diversidad genética dentro de las especies de *Cryptosporidium* lo que será de utilidad en el desarrollo de pruebas de identificación molecular y de genotipado.
4. La ausencia de anotación de la mayoría de los genomas de *Cryptosporidium* disponibles en la actualidad dificulta la exploración de las implicaciones biológicas de los cambios observados en los genomas.
5. Se presenta el análisis filogenómico más completo a partir de 800.861 posiciones variables, con el que se generó un árbol más robusto con diferentes especies familias alélicas y subgenotipos del género *Cryptosporidium*.
6. El análisis genómico comparativo realizado en este estudio estableció diferencias estructurales y funcionales entre las principales especies de *Cryptosporidium* de importancia en la salud pública humana. Asimismo, se identificaron los genes con mutaciones deletéreas y afectados por fenómenos *indels*, demostrando la alteración de genes relacionados con funciones enzimáticas, adhesión, invasión y posterior establecimiento de la infección. Estudios adicionales son requeridos para relacionar las variantes detectadas con las diferencias fenotípicas al interior del género.

7. Este es el estudio de genómica comparativo más robusto realizado en el género *Cryptosporidium*, en el que se incluyeron genomas de diferentes aislados, familias alélicas y subgenotipos de las tres principales especies intestinales de *Cryptosporidium*.

## **9. MATERIAL SUPLEMENTARIO Y ANEXOS**

**Tabla suplementaria 1. Transiciones en los aislados**

Especie	Aislado	Transiciones							
		AG	% AG	GA	% GA	TC	% TC	CT	% CT
<i>C. meleagridis</i>	UKMEL1	94.251	15,09	126.448	20,25	93.965	15,05	126.189	20,20
	UKMEL3	93.525	15,11	125.507	20,27	93.233	15,06	125.098	20,21
	UKMEL4	94.510	15,13	126.311	20,22	94.062	15,06	126.161	20,20
	TU1867	94.408	15,10	126.607	20,25	93.978	15,03	126.187	20,18
		<b>AG</b>	<b>% AG</b>	<b>GA</b>	<b>% GA</b>	<b>TC</b>	<b>% TC</b>	<b>CT</b>	<b>% CT</b>
<i>C. hominis</i>	30976	43.416	19,41	40.396	18,06	43.486	19,44	40.506	18,11
	UdeA01	43.265	19,40	40.370	18,10	43.345	19,43	40.421	18,12
	37999	43.351	19,39	40.432	18,08	43.407	19,41	40.469	18,10
	TU502	43.438	19,35	40.541	18,06	43.516	19,39	40.561	18,07
	UKH1	43.341	19,36	40.445	18,06	43.456	19,41	40.503	18,09
	UKH3	43.419	19,34	40.532	18,06	43.505	19,38	40.550	18,06
	UKH4	42.960	19,38	40.097	18,08	43.169	19,47	40.150	18,11
	UKH5	43.281	19,34	40.441	18,07	43.420	19,40	40.495	18,09
	SWEH2	41.430	19,43	38.714	18,16	41.419	19,43	38.855	18,22
	SWEH5	41.739	19,40	39.171	18,21	41.746	19,40	39.223	18,23
		<b>AG</b>	<b>% AG</b>	<b>GA</b>	<b>% GA</b>	<b>TC</b>	<b>% TC</b>	<b>CT</b>	<b>% CT</b>
<i>C. parvum</i>	UKP2	219	13,70	169	10,57	220	13,76	162	10,13
	UKP3	280	14,13	201	10,15	275	13,88	197	9,94
	UKP4	307	15,87	226	11,69	276	14,27	188	9,72
	UKP5	351	13,11	274	10,24	317	11,84	230	8,59
	UKP6	284	17,59	164	10,15	224	13,87	153	9,47
	UKP7	284	13,40	236	11,14	276	13,03	196	9,25
	UKP8	953	16,56	876	15,22	885	15,38	817	14,19
	UKP14	3.246	17,57	3436	18,60	3.195	17,29	3.404	18,42
	UKP15	3.615	18,14	3520	17,66	3.573	17,93	3.529	17,71

**Tabla suplementaria 2. Transversiones en los aislados**

Especie	Aislado	Transversiones							
		AT	% AT	AC	% AC	TA	%TA	TG	% TG
<i>C. meleagridis</i>	UKMEL1	40.928	6,55	22.654	3,63	41.156	6,59	22.636	3,62
	UKMEL3	40.371	6,52	22.418	3,62	40.727	6,58	22.431	3,62
	UKMEL4	40.830	6,54	22.584	3,62	41.204	6,60	22.554	3,61
	TU1867	40.972	6,55	22.708	3,63	41.338	6,61	22.676	3,63
		<b>AT</b>	<b>% AT</b>	<b>AC</b>	<b>% AC</b>	41 338	<b>%TA</b>	<b>TG</b>	<b>% TG</b>
<i>C. hominis</i>	30976	10.182	4,55	6.895	3,08	10.440	4,67	6.918	3,09
	UdeA01	10.152	4,55	6.869	3,08	10.343	4,64	6.852	3,07
	37999	10.210	4,57	6.898	3,09	10.419	4,66	6.890	3,08
	TU502	10.338	4,61	6.935	3,09	10.518	4,69	6.931	3,09
	UKH1	10.297	4,60	6.918	3,09	10.483	4,68	6.895	3,08
	UKH3	10.373	4,62	6.956	3,10	10.554	4,70	6.944	3,09
	UKH4	10.095	4,55	6.827	3,08	10.297	4,64	6.858	3,09
	UKH5	10.313	4,61	6.941	3,10	10.492	4,69	6.915	3,09
	SWEH2	9.509	4,46	6.495	3,05	9.757	4,58	6.481	3,04
	SWEH5	9.576	4,45	6.529	3,03	9.792	4,55	6.567	3,05
		<b>AT</b>	<b>% AT</b>	<b>AC</b>	<b>% AC</b>	<b>TA</b>	<b>%TA</b>	<b>TG</b>	<b>% TG</b>
<i>C. parvum</i>	UKP2	138	8,63	129	8,07	153	9,57	185	11,57
	UKP3	205	10,35	159	8,03	211	10,65	201	10,15
	UKP4	158	8,17	136	7,03	192	9,93	186	9,62
	UKP5	271	10,12	238	8,89	289	10,80	344	12,85
	UKP6	124	7,68	139	8,61	134	8,30	192	11,89
	UKP7	185	8,73	153	7,22	259	12,22	200	9,44
	UKP8	415	7,21	320	5,56	404	7,02	379	6,58
	UKP14	958	5,19	649	3,51	870	4,71	681	3,69
	UKP15	974	4,89	737	3,70	1.044	5,24	731	3,67

**Tabla suplementaria 3. Continuación transversiones en los aislados**

Especie	Aislado	Transversiones continuación							
		CA	% CA	CG	% CG	GT	% GT	GC	% GC
<i>C. meleagridis</i>	UKMEL1	18.339	2,94	9.762	1,60	18.205	2,91	9.762	1,56
	UKMEL3	18.150	2,93	9.660	1,61	17.970	2,90	9.660	1,56
	UKMEL4	18.338	2,94	9.784	1,61	18.264	2,92	9.784	1,57
	TU1867	18.389	2,94	9.763	1,60	18.206	2,91	9.763	1,56
		<b>CA</b>	<b>% CA</b>	<b>CG</b>	<b>% CG</b>	<b>GT</b>	<b>% GT</b>	<b>GC</b>	<b>% GC</b>
<i>C. hominis</i>	30976	7.169	3,20	3.535	1,58	7.203	3,22	3.566	1,59
	UdeA01	7.154	3,21	3.541	1,59	7.171	3,22	3.548	1,59
	37999	7.173	3,21	3.529	1,58	7.227	3,23	3.563	1,59
	TU502	7.233	3,22	3.538	1,58	7.311	3,26	3.570	1,59
	UKH1	7.201	3,22	3.542	1,58	7.257	3,24	3.563	1,59
	UKH3	7.234	3,22	3.529	1,57	7.295	3,25	3.569	1,59
	UKH4	7.126	3,21	3.477	1,57	7.133	3,22	3.525	1,59
	UKH5	7.198	3,22	3.529	1,58	7.228	3,23	3.566	1,59
	SWEH2	6.838	3,21	3.388	1,59	6.884	3,23	3.436	1,61
	SWEH5	6.962	3,24	3.381	1,57	7.008	3,26	3.442	1,60
		<b>CA</b>	<b>% CA</b>	<b>CG</b>	<b>% CG</b>	<b>GT</b>	<b>% GT</b>	<b>GC</b>	<b>% GC</b>
<i>C. parvum</i>	UKP2	84	5,25	25	1,56	82	5,13	29	1,81
	UKP3	100	5,05	27	1,36	87	4,39	35	1,77
	UKP4	95	4,91	32	1,65	85	4,40	49	2,53
	UKP5	149	5,57	40	1,49	130	4,86	41	1,53
	UKP6	81	5,02	28	1,73	59	3,65	30	1,86
	UKP7	122	5,76	36	1,70	126	5,95	43	2,03
	UKP8	259	4,50	90	1,56	255	4,43	99	1,72
	UKP14	717	3,88	333	1,80	668	3,62	314	1,70
	UKP15	737	3,70	335	1,68	772	3,87	356	1,79

## **10. BIBLIOGRAFÍA**

1. Fayer R, Xiao L. *Cryptosporidium* and *Cryptosporidiosis*. Second edi. 2008. 527 p.
2. Tzipori S, Griffiths JK. Natural history and biology of *Cryptosporidium parvum*. *Adv Parasitol*. 1998;40:5–36.
3. Tzipori S, Widmer G. A hundred-year retrospective on cryptosporidiosis. *Trends Parasitol*. 2009;24(4):184–9.
4. Nime, F.A. Burek, et al. Acute enterocolitis in a human being infected with the protozoan *Cryptosporidium*. *Gastroenterology*. 1976;70:592–598.
5. Meisel JL et al. Overwhelming watery diarrhea associated with a *Cryptosporidium* in an immunosuppressed patient. *Gastroenterology*. 1976;70:1156–1160.
6. Mackenzie WR et al. A massive outbreak in Milwaukee of *Cryptosporidium* infection transmitted through the public water supply. *N Eng J Med*. 1994;331:161–167.
7. Barta JR. Phylogenetic Analysis of the Class Sporozoa (Phylum Apicomplexa Levine, 1970): Evidence for the Independent Evolution of Heteroxenous Life Cycles. *J Parasitol* [Internet]. 1989 Feb 3;75(2):195–206. Available from: <http://www.jstor.org/stable/3282766>
8. Bull S et al. Cross-reaction of an anti-*Cryptosporidium* monoclonal antibody with sporocysts of *Monocystis* species. *Vet Parasitol*. 1998;77:195–197.
9. Aldeyarbi HM, Karanis P. The Ultra-Structural Similarities between *Cryptosporidium parvum* and the Gregarines. *J Eukaryot Microbiol*. 2016;63(1):79–85.
10. Carreno RA, Martin DS, Barta JR. *Cryptosporidium* is more closely related to the gregarines than to coccidia as shown by phylogenetic analysis of apicomplexan parasites inferred using small-subunit ribosomal RNA gene sequences. *Parasitol Res*. 1999;85:899–904.
11. Clode PL, Koh WH, Thompson RCA. Life without a Host Cell: What is

- Cryptosporidium?. Trends Parasitol [Internet]. 2015;31(12):614–24. Available from: <http://dx.doi.org/10.1016/j.pt.2015.08.005>
12. Ryan U, Paparini A, Monis P, Hijjawi N. It's official – Cryptosporidium is a gregarine: What are the implications for the water industry? Water Res [Internet]. 2016;105:305–13. Available from: <http://dx.doi.org/10.1016/j.watres.2016.09.013>
  13. Cavalier-smith T. Gregarine site-heterogeneous 18S rDNA trees , revision of gregarine higher classification , and the evolutionary diversification of Sporozoa. Eur J Protistol [Internet]. 2014;50(5):472–95. Available from: <http://dx.doi.org/10.1016/j.ejop.2014.07.002>
  14. Thompson RA, Koh WH, Clode PL. Cryptosporidium — What is it? Food Waterborne Parasitol [Internet]. 2016;4:54–61. Available from: <http://dx.doi.org/10.1016/j.fawpar.2016.08.004>
  15. Khan A, Shaik JS, Grigg ME. Genomics and molecular epidemiology of Cryptosporidium species. Acta Trop. 2017;1–0.
  16. Xiao L. Genetic Diversity and Population Structure of Cryptosporidium. Trends Parasitol [Internet]. 2018;31–68. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0065308X16300094>
  17. Caccio SM, Widmer G. Cryptosporidium: parasite and disease [Internet]. 2014. 563 p. Available from: 10.1007/978-3-7091-1562-6
  18. Mohammed A, Degefu H, Jilo K. Cryptosporidium and Its Public Health Importance: Review. Int J Res Stud Microbiol Biotechnol. 2017;3(4).
  19. Ryan U, Fayer R, Xiao L. Cryptosporidium species in humans and animals: current understanding and research needs. Parasitology [Internet]. 2014;141(13):1667–85. Available from: [http://www.journals.cambridge.org/abstract\\_S0031182014001085](http://www.journals.cambridge.org/abstract_S0031182014001085)
  20. Nichols RA, Connelly L, Sullivan C, Smith H. Identification of Cryptosporidium species and genotypes in Scottish raw and drinking waters during a one-year

- monitoring period. *Appl Environ Microbiol*. 2010;76(17):5977–86.
21. Plutzer J, Karanis P. Genetic polymorphism in *Cryptosporidium* species: An update. *Vet Parasitol*. 2009;165(3–4):187–99.
  22. Pagoso EJ, Rivera WL. *Cryptosporidium* species from common edible bivalves in Manila Bay, Philippines. *Mar Pollut Bull* [Internet]. 2017;119(1):31–9. Available from: <http://dx.doi.org/10.1016/j.marpolbul.2017.03.005>
  23. Gómez Couso H. *Cryptosporidium* en moluscos bivalvos. Universidad de Santiago de Compostela; 2005.
  24. Hijjawi N, Estcourt A, Yang R, Monis P, Ryan U. Complete development and multiplication of *Cryptosporidium hominis* in cell-free culture. *Vet Parasitol* [Internet]. 2010;169(1–2):29–36. Available from: <http://dx.doi.org/10.1016/j.vetpar.2009.12.021>
  25. O'Hara SP, Chen XM. The Cell Biology of *Cryptosporidium* infection. *Microbes Infect*. 2011;13(9):721–30.
  26. Wetzel DM, Schmidt J, Kuhlenschmidt MS, Dubey JP, Sibley LD. Gliding Motility Leads to Active Cellular Invasion by *Cryptosporidium parvum* Sporozoites. *Infect Immun*. 2005;73(9):5379–5387.
  27. Langer-Curry R, Riggs MW. *Cryptosporidium parvum* apical complex glycoprotein CSL contains a sporozoite ligand for intestinal epithelial cells. *Infect Immun*. 1999;67:5282–5291.
  28. Chalmers RM, Davies AP, Tyler K. *Cryptosporidium*. *Microbiol (United Kingdom)*. 2019;165(5):500–2.
  29. Current WL. Cryptosporidiosis, a protozoologist's view of an emerging zoonosis. *Microecol Ther*. 1985;15:65–200.
  30. Gerace E, Lo Presti VDM, Biondo C. *Cryptosporidium* Infection: Epidemiology, Pathogenesis, and Differential Diagnosis. *Eur J of Microbiology Immunol*. 2019;9(October):119–23.
  31. Koh W, Clode PL, Monis P, Thompson RCA. Multiplication of the waterborne

- pathogen *Cryptosporidium parvum* in an aquatic biofilm system. *Parasit Vectors* [Internet]. 2013;6(1):270. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3848567&tool=pmcentrez&rendertype=abstract>
32. Koh W, Thompson A, Edwards H, Monis P, Clode PL. Extracellular excystation and development of *Cryptosporidium*: tracing the fate of oocysts within *Pseudomonas* aquatic biofilm systems. *BMC Microbiol* [Internet]. 2014;14:281. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4236811&tool=pmcentrez&rendertype=abstract>
  33. DiCesare EW, Hargreaves BR, Jellison KL. Biofilm roughness determines *Cryptosporidium parvum* retention in environmental biofilms. *Appl Environ Microbiol*. 2012;78(12):4187–93.
  34. Hijjawi N. *Cryptosporidium*: New developments in cell culture. *Exp Parasitol* [Internet]. 2010;124(1):54–60. Available from: <http://dx.doi.org/10.1016/j.exppara.2009.05.015>
  35. Rosales MJ, Pérez Cordon G, Sánchez Moreno M, Marín Sánchez C, Mascaró C. Extracellular like-gregarine stages of *Cryptosporidium parvum*. *Acta Trop*. 2005;95(1):74–8.
  36. Cama VA, Ross JM, Crawford S, Kawai V, Chavez-Valdez R, Vargas D, et al. Differences in Clinical Manifestations among *Cryptosporidium* Species and Subtypes in HIV-Infected Persons. *J Infect Dis* [Internet]. 2007;196(5):684–91. Available from: <https://academic.oup.com/jid/article-lookup/doi/10.1086/519842>
  37. Bouzid M, Hunter PR, Chalmers RM, Tyler KM. *Cryptosporidium* pathogenicity and virulence. *Clin Microbiol Rev*. 2013;26(1):115–34.
  38. Toro-Londono MA, Bedoya-Urrego K, Garcia-Montoya GM, Galvan-Diaz AL, Alzate JF. Intestinal parasitic infection alters bacterial gut microbiota in

- children [Internet]. Vol. 7, PeerJ. 2019. p. e6200. Available from: <https://peerj.com/articles/6200>
39. Xiao L. Molecular epidemiology of cryptosporidiosis: An update. *Exp Parasitol* [Internet]. 2010;124(1):80–9. Available from: <http://dx.doi.org/10.1016/j.exppara.2009.03.018>
  40. Palmieri F, Cicalini S, Froio N, Rizzi E, Goletti D, Festa A et al. Pulmonary cryptosporidiosis in an AIDS patient: successful treatment with paromomycin plus azithromycin. *Int J STD AIDS* [Internet]. 2005 Jul 1;16(7):515–7. Available from: <https://doi.org/10.1258/0956462054308332>
  41. Sponseller JK, Griffiths JK, Tzipori S. The evolution of respiratory cryptosporidiosis: Evidence for transmission by inhalation. *Clin Microbiol Rev.* 2014;27(3):575–86.
  42. Bushen OY, Kohli A, Pinkerton RC, Dupnik K, Newman RD, Sears CL, et al. Heavy cryptosporidial infections in children in northeast Brazil : comparison of *Cryptosporidium hominis* and *Cryptosporidium parvum*. 2007;378–84.
  43. Houpt ER, Bushen OY, Sam NE, Kohli A, Asgharpour A, Ng CT, et al. Short report: Asymptomatic cryptosporidium hominis infection among human immunodeficiency virus-infected patients in Tanzania. 2005;73(3):520–2.
  44. Cama VA, Bern C, Roberts J, Cabrera L, Sterling CR, Ortega Y, et al. *Cryptosporidium* species and subtypes and clinical manifestations in children, Peru. *Emerg Infect Dis.* 2008;14(10):1567–74.
  45. Jex AR, Gasser RB. Genetic richness and diversity in *Cryptosporidium hominis* and *C. parvum* reveals major knowledge gaps and a need for the application of “next generation” technologies - Research review. *Biotechnol Adv* [Internet]. 2010;28(1):17–26. Available from: <http://dx.doi.org/10.1016/j.biotechadv.2009.08.003>
  46. Zhang W, Wang R, Yang F, Zhang L, Cao J, Zhang X, et al. Distribution and Genetic Characterizations of *Cryptosporidium* spp. in Pre-Weaned Dairy Calves

- in Northeastern China's Heilongjiang Province. *PLoS One*. 2013;8(1):1–6.
47. Strong WB, Gut J, Nelson RG. Cloning and sequence analysis of a highly polymorphic *Cryptosporidium parvum* gene encoding a 60-kilodalton glycoprotein and characterization of its 15- and 45-kilodalton zoite surface antigen products. *Infect Immun*. 2000;68(7):4117–34.
  48. Karim MR, Zhang S, Jian F, Li J, Zhou C, Zhang L, et al. Multilocus typing of *Cryptosporidium* spp. and *Giardia duodenalis* from non-human primates in China. *Int J Parasitol* [Internet]. 2014;44(13):1039–47. Available from: <http://dx.doi.org/10.1016/j.ijpara.2014.07.006>
  49. Xiao L, Fayer R, Ryan U, Upton SJ. *Cryptosporidium* taxonomy: recent advances and implications for public health. *Clin Microbiol Rev* [Internet]. 2004;17(1):72–97. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14726456>  
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC321466>
  50. Avendaño Valenzuela C, Amaya Martínez A. Caracterización molecular de los subtipos de la GP60 de *Cryptosporidium parvum* y *Cryptosporidium hominis* alrededor del mundo. *Rev MVZ Córdoba* [Internet]. 2017;22(3):6339. Available from: <http://revistas.unicordoba.edu.co/index.php/revistamvz/article/view/1138>
  51. Sikora P, Arrighi RBG, Beser J, Andersson S. Genomic variation in IbA10G2 and other patient-derived *cryptosporidium hominis* subtypes. *J Clin Microbiol*. 2017;55(3):844–58.
  52. Alves M, Xiao L, Sulaiman I, Lal AA, Matos O, Antunes F. Subgenotype analysis of *Cryptosporidium* isolates from humans, cattle, and zoo ruminants in Portugal. *J Clin Microbiol*. 2003;41(6):2744–7.
  53. Abe N, Makino I. Multilocus genotypic analysis of *Cryptosporidium* isolates from cockatiels, Japan. *Parasitol Res*. 2010;106(6):1491–7.
  54. Abal-Fabeiro JL, Maside X, Bello X, Llovo J, Bartolomé C. Multilocus patterns

- of genetic variation across *Cryptosporidium* species suggest balancing selection at the gp60 locus. *Mol Ecol*. 2013;22(18):4723–32.
55. Glaberman S, Sulaiman IM, Bern C, Limor J, Peng MM, Morgan U, et al. A multilocus genotypic analysis of *Cryptosporidium meleagridis*. *J Eukaryot Microbiol*. 2001;Suppl:19S-22S.
  56. Robinson G, Chalmers RM. Assessment of polymorphic genetic markers for multi-locus typing of *Cryptosporidium parvum* and *Cryptosporidium hominis*. *Exp Parasitol* [Internet]. 2012;132(2):200–15. Available from: <http://dx.doi.org/10.1016/j.exppara.2012.06.016>
  57. Hernández Gallo N, Hernández Flórez LJ, Cortés Vecino JA. Criptosporidiosis y «Una Salud». *Rev Salud Pública*. 2018;20(1):138–43.
  58. Guadalupe OP, Caccio SM, Fayer R, Mank TG, Smith HV, Thompson RC. *Giardia* and *Cryptosporidium*: From Molecules to Disease. 2009. 502 p.
  59. de Arango M, Rodríguez DA, Prada NE. Frecuencia de *Cryptosporidium* spp en materia fecal de niños entre un mes y trece años en un hospital local colombiano. *Colomb Med*. 2006;37(2):121–5.
  60. Pijnacker R, Mughini-gras L, Vennema H, Enserink R, Van Den Wijngaard CC, Kortbeek T, et al. Characteristics of child day care centres associated with clustering of major enteropathogens. *Epidemiol Infect* [Internet]. 2016;144(12):2527–39. Available from: [http://www.journals.cambridge.org/abstract\\_S0950268816001011](http://www.journals.cambridge.org/abstract_S0950268816001011)
  61. Kotloff KL, Blackwelder WC, Nasrin D, Nataro JP, Farag TH, Van Eijk A, et al. The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: Epidemiologic and clinical methods of the case/control study. *Clin Infect Dis*. 2012;55(SUPPL. 4).
  62. Platts-mills JA, Babji S, Bodhidatta L, Gratz J, Haque R, Havt A, et al. Pathogen-specific burdens of community diarrhoea in developing countries: a multisite birth cohort study ( MAL-ED ). *Lancet Glob Heal*. 2015;3:564–75.

63. Putignani L, Menichella D. Global Distribution, Public Health and Clinical Impact of the Protozoan Pathogen *Cryptosporidium*. *Interdiscip Perspect Infect Dis*. 2010;2010:1–39.
64. Chappell CL, Okhuysen PC, Langer-Curry RC, Akiyoshi DE, Widmer G, Tzipori S. *Cryptosporidium meleagridis*: Infectivity in healthy adult volunteers. *Am J Trop Med Hyg*. 2011;85(2):238–42.
65. Feltus DC, Giddings CW, Schneck BL, Monson T, Warshauer D, McEvoy JM. Evidence supporting zoonotic transmission of *Cryptosporidium* spp. in Wisconsin. *J Clin Microbiol*. 2006;44(12):4303–8.
66. Chalmers RM, Elwin K, Hadfield SJ, Robinson G. Sporadic human cryptosporidiosis caused by *Cryptosporidium cuniculus*, United Kingdom, 2007-2008. *Emerg Infect Dis*. 2011;17(3):536–8.
67. Xiao L, Feng Y. Molecular epidemiologic tools for waterborne pathogens *Cryptosporidium* spp. and *Giardia duodenalis*. *Food Waterborne Parasitol* [Internet]. 2017;8–9(September):14–32. Available from: <https://doi.org/10.1016/j.fawpar.2017.09.002>
68. Sulaiman IM, Hira PR, Zhou L, Al-Ali FM, Al-Shelahi FA, Shweiki HM, et al. Unique endemicity of cryptosporidiosis in children in Kuwait. *J Clin Microbiol*. 2005;43(6):2805–9.
69. Caccio SM, Thompson RA, Mclauchlin J, Smith H. Unravelling *Cryptosporidium* and *Giardia* epidemiology. 2005;21(9).
70. Nader JL, Mathers TC, Ward BJ, Pachebat JA, Martin T, Robinson G, et al. Evolutionary genomics of anthroponosis in *Cryptosporidium*. *Nat Microbiol* [Internet]. 2019;4(5):826–36. Available from: <https://doi.org/10.1038/s41564-019-0377-x>
71. Chalmers RM, Smith R, Elwin K, Clifton-Hadley FA, Giles M. Epidemiology of anthroponotic and zoonotic human cryptosporidiosis in England and Wales, 2004–2006. *Epidemiol Infect*. 2011;139(5):700–12.

72. Tanriverdi S, Grinberg A, Chalmers RM, Hunter PR, Petrovic Z, Akiyoshi DE, et al. Inferences about the global population structures of *Cryptosporidium parvum* and *Cryptosporidium hominis*. *Appl Environ Microbiol*. 2008;74(23):7227–34.
73. Medema G, Teunis P, Blokker M, Deere D, Davison A, Charles P, et al. WHO Guidelines for Drinking Water Quality. 2006.
74. Efstratiou A, Ongerth JE, Karanis P. Waterborne transmission of protozoan parasites: Review of worldwide outbreaks - An update 2011–2016. *Water Res* [Internet]. 2017;114:14–22. Available from: <http://dx.doi.org/10.1016/j.watres.2017.01.036>
75. Omarova A, Tussupova K, Berndtsson R, Kalishev M, Sharapatova K. Protozoan parasites in drinking water: A system approach for improved water, sanitation and hygiene in developing countries. *Int J Environ Res Public Health*. 2018;15(3):1–18.
76. Medina LA, Angel V, Franco L, Jaramillo JC, Ochoa FL, Vasquez AM. Prevalencia de *Cryptosporidium* en muestras fecales en Medellin. Estudio de 10 casos. *Salud Uninorte*. 1988. p. 45–52.
77. Arroyo B, Arzuza O, Villegas S, Rocha A, Díaz H. Infecciones parasitarias intestinales en pacientes positivos para el Virus de la Inmunodeficiencia Humana (VIH) en la ciudad de Cartagena de Indias, Colombia. *Asoc Colomb Infectología*. 2003;7–2:58–63.
78. Botero JH, Castaño A, Montoya MN, Ocampo NE, Hurtado MI, Lopera MM. A preliminary study of the prevalence of intestinal parasites in immunocompromised patients with and without gastrointestinal manifestations. *Rev Inst Med Trop Sao Paulo*. 2003;45(4):197–200.
79. Siuffi M, Angulo M, Velasco CA, López P, Dueñas VH, Rojas C. Relación entre los niveles de carga viral y los niveles de linfocitos CD4 en el diagnóstico de *Cryptosporidium* spp. *Corporación Ed Médica del Val*. 2006;37(1):15–20.

80. Carreño M, Velasco CA, Rueda E. Prevalencia de *Cryptosporidium* spp en niños menores de 13 años con afecciones oncológicas. 2005;36(Supl 1):6–9.
81. Bayona M, Avendaño C, Amaya Á. Epidemiologic Characterization of Cryptosporidiosis in Infant Population At the Sabana Centro (Cundinamarca). *Rev UDCA Actual Divulg Científica*. 2011;14(1):7–13.
82. Galván-Díaz AL. Cryptosporidiosis in Colombia: a Systematic Review. *Curr Trop Med Reports*. 2018;5(3):144–53.
83. Mead J, Arrowood M, Current W, Sterling C. Field inversion gel electrophoretic separation of *Cryptosporidium* spp. chromosome-size DNA. *J Parasitol*. 1988;74:366-369.
84. Petersen C. Cellular biology of *Cryptosporidium parvum*. *Parasitol Today*. 1993;9(3):87–91.
85. Hays MP, Mosier DA, Oberst RD. Enhanced karyotype resolution of *Cryptosporidium parvum* by contour-clamped homogeneous electric fields. *Vet Parasitol*. 1995;58(4):273–80.
86. Blunt D, Khramtsov N, Upton S, Montelone B. Molecular karyotype analysis of *Cryptosporidium parvum*: evidence for eight chromosomes and a low-molecular-size molecule. *Clin Diagn Lab Immunol*. 1997;4(1):11–3.
87. Cacciò S, Camilli R, La Rosa G, Pozio E. Establishing the *Cryptosporidium parvum* karyotype by NotI and SfiI restriction analysis and Southern hybridization. *Gene*. 1998;219(1–2):73–9.
88. Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, et al. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science*. 2004;304(5669):441–5.
89. Spano F, Crisanti A. *Cryptosporidium parvum*: the many secrets of a small genome. *International Journal for Parasitology*. 2000.
90. Heiges M, Wang H, Robinson E, Aurrecochea C, Gao X, Kaluskar N, et al. CryptoDB: a *Cryptosporidium* bioinformatics resource update. *Nucleic Acids*

- Res. 2006;34(90001):D419–22.
91. Liu C, Vigdorovich V, Kapur V, Abrahamsen MS. A random survey of the *Cryptosporidium parvum* genome. *Infect Immun*. 1999;67(8):3960–9.
  92. Xu P, Widmer G, Wang Y, Ozaki LS, Alves JM, Serrano MG, et al. The Genome of *Cryptosporidium hominis*. 2004;431.
  93. Isaza JP, Galvan AL, Polanco V, Huang B, Matveyev A, Serrano MG, et al. Revisiting the reference genomes of human pathogenic *Cryptosporidium* species: Reannotation of *C. parvum* Iowa and a new *C. hominis* reference. *Sci Rep*. 2015;5(October):1–14.
  94. Ifeonu OO, Chibucos MC, Orvis J, Su Q, Elwin K, Guo F, et al. Annotated draft genome sequences of three species of *Cryptosporidium*: *Cryptosporidium meleagridis* isolate UKMEL1, *C. baileyi* isolate TAMU-09Q1 and *C. hominis* isolates TU502 2012 and UKH1. *Pathog Dis*. 2016;74(7):1–5.
  95. Morris A, Robinson G, Swain MT, Chalmers RM. Direct sequencing of *Cryptosporidium* in stool samples for public health. *Front Pathog Genomics*. 2019;In Press.
  96. Widmer G, Sullivan S. Genomics and population biology of *Cryptosporidium* species. *Parasite Immunol*. 2012;34(2–3):61–71.
  97. Huang J, Mullapudi N, Lancto CA, Scott M, Abrahamsen MS, Kissinger JC. Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum*. 2004;5(11).
  98. Stanley D, Zhu G. *Cryptosporidium*: Genomic and Biochemical Features. 2011;124(1):1–12.
  99. Swapna LS, Parkinson J. Genomics of apicomplexan parasites. *Crit Rev Biochem Mol Biol*. 2017;52(3):254–73.
  100. Tanriverdi S, Blain JC, Deng B, Ferdig MT, Widmer G. Genetic crosses in the apicomplexan parasite *Cryptosporidium parvum* define recombination parameters. *Mol Microbiol*. 2007;63(5):1432–9.

101. Abrahamsen MS. *Cryptosporidium parvum* genome project. *Comp Funct Genomics*. 2001;2(1):19–21.
102. Striepen B, Pruijssers A, Huang J, Li C, Gubbels M, Umejiego N, et al. Gene transfer in the evolution of parasite nucleotide biosynthesis. *Proc Natl Acad Sci*. 2004;101:3154–3159.
103. Guo Y, Tang K, Rowe LA, Li N, Roellig DM, Knipe K, et al. Comparative genomic analysis reveals occurrence of genetic recombination in virulent *Cryptosporidium hominis* subtypes and telomeric gene duplications in *Cryptosporidium parvum*. *BMC Genomics*. 2015;16(1).
104. Mazurie AJ, Alves JM, Ozaki LS, Zhou S, Schwartz DC, Buck GA. Comparative Genomics of *Cryptosporidium*. Hindawi Publ Corp. 2013;2013:1–9.
105. Anantharaman V, Abhiman S, de Souza RF, Aravind L. Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system. *Gene* [Internet]. 2011;475(2):63–78. Available from: <http://dx.doi.org/10.1016/j.gene.2010.12.001>
106. Widmer G, Lee Y, Hunt P, Martinelli A, Tolhoff M, Bodi K. Comparative genome analysis of two *Cryptosporidium parvum* isolates with different host range. *Infect Genet Evol*. 2012;12(6):1213–21.
107. Haubold B, Wiehe T. Comparative genomics: Methods and applications. *Naturwissenschaften*. 2004;91(9):405–21.
108. Mazurie AJ, Alves JM, Ozaki LS, Zhou S, Schwartz DC, Buck GA. Comparative genomics of *cryptosporidium*. *Int J Genomics*. 2013;2013.
109. Fan Y, Feng Y, Xiao L. Comparative genomics: how has it advanced our knowledge of cryptosporidiosis epidemiology?. *Genet Evol Phylogeny*. 2019 Dec;118(12):3195-3204.
110. Langsley G, van Noort V, Carret C, Meissner M, de Villiers EP, Bishop R, et al. Comparative genomics of the Rab protein family in Apicomplexan parasites. *Microbes Infect*. 2008;10(5):462–70.

111. Hayano K, Tian F, Kambadakone AR, Yoon SS, Duda DG, Ganeshan B, et al. Comparative genomic analysis of the IID subtype family of *Cryptosporidium parvum*. 2016;39(4):607–12.
112. Guo Y, Tang K, Rowe LA, Li N, Roellig DM, Knipe K, et al. Comparative genomic analysis reveals occurrence of genetic recombination in virulent *Cryptosporidium hominis* subtypes and telomeric gene duplications in *Cryptosporidium parvum*. BMC Genomics. 2015;16(1).
113. Gilchrist CA, Cotton JA, Burkey C, Arju T, Gilmartin A, Lin Y, et al. Genetic diversity of *cryptosporidium hominis* in a bangladeshi community as revealed by whole-genome sequencing. J Infect Dis. 2018;218(2):259–64.
114. Sikora P, Andersson S, Winiiecka-krusnell J, Alsmark C, Troell K, Beser J, et al. Genomic Variation in IbA10G2 and Other Patient- Derived *Cryptosporidium hominis* Subtypes. 2017;55(3):844–58.
115. Ryan U, Hijjawi N. New developments in *Cryptosporidium* research. Int J Parasitol [Internet]. 2015;45(6):367–73. Available from: <http://dx.doi.org/10.1016/j.ijpara.2015.01.009>
116. Carlton JM. Genome sequencing and comparative genomics of tropical disease pathogens. Cell Microbiol. 2003;5(12):861–73.
117. Liu S, Roellig DM, Guo Y, Li N, Frace MA, Tang K, et al. Evolution of mitosome metabolism and invasion-related proteins in *Cryptosporidium*. BMC Genomics. 2016;17(1):1–16.
118. Feng Y, Li N, Roellig DM, Kelley A, Liu G, Amer S, et al. Comparative genomic analysis of the IID subtype family of *Cryptosporidium parvum*. Int J Parasitol. 2017;47(5):281–90.
119. Xu Z, Guo Y, Roellig DM, Feng Y, Xiao L. Comparative analysis reveals conservation in genome organization among intestinal *Cryptosporidium* species and sequence divergence in potential secreted pathogenesis determinants among major human-infecting species. BMC Genomics.

- 2019;20(1):1–15.
120. Bouzid M, Tyler KM, Christen R, Chalmers RM, Elwin K, Hunter PR. Multi-locus analysis of human infective *Cryptosporidium* species and subtypes using ten novel genetic loci. *BMC Microbiol.* 2010;10.
  121. Bouzid M, Hunter PR, McDonald V, Elwin K, Chalmers RM, Tyler KM. A new heterogeneous family of telomerically encoded *Cryptosporidium* proteins. *Evol Appl.* 2013;6(2):207–17.
  122. Leinonen R, Sugawara H, Shumway M. The sequence read archive. *Nucleic Acids Res.* 2011;39(SUPPL. 1):2010–2.
  123. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. 2011;27(21):2957–63.
  124. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. 2012;19(5):455–77.
  125. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10.
  126. Puiu D. CryptoDB: the *Cryptosporidium* genome resource. *Nucleic Acids Res* [Internet]. 2004;32(90001):329D – 331. Available from: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkh050>
  127. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2010;26(5):589–95.
  128. Delcher AL, Phillippy A, Carlton J, Salzberg SL. Fast algorithms for large-scale genome alignment and comparison. 2002;30(11):2478–83.
  129. Nattestad M, Schatz MC. Assemblytics: A web analytics tool for the detection of variants from an assembly. *Bioinformatics.* 2016;32(19):3021–3.
  130. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9.

131. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8.
132. Holder A, Simon J, Strauser J, Taylor J, Shibberu Y. Dynamic Programming Used to Align Protein Structures with a Spectrum Is Robust. *Biology (Basel)*. 2013;2(4):1296–310.
133. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng P. SIFT missense predictions for genomes. *Nat Protoc [Internet]*. 2016;11(1):1073–1081. Available from: <http://dx.doi.org/10.1038/nprot.2015-123>
134. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol*. 2017;34(8):2115–22.
135. Ye J, Zhang Y, Cui H, Liu J, Wu Y, Cheng Y, et al. WEGO 2.0: A web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Res*. 2018;46(W1):W71–5.
136. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 1999;27(1):29–34.
137. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol*. 2001;305(3):567–80.
138. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol [Internet]*. 2019;37(4):420–3. Available from: <http://dx.doi.org/10.1038/s41587-019-0036-z>
139. Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng Des Sel*. 2004;17(4):349–56.
140. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res*.

- 2016;44(W1):W232–5.
141. Andrew Rambaut. FigTree, version 1.4.4. Comput Progr Distrib by author, website <http://tree.bio.ed.ac.uk/software/figtree/> [accessed enero, 2020]. 2018;
  142. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013.
  143. Hadfield SJ, Pachebat JA, Swain MT, Robinson G, Cameron SJS, Alexander J, et al. Generation of whole genome sequences of new *Cryptosporidium hominis* and *Cryptosporidium parvum* isolates directly from stool samples. BMC Genomics [Internet]. 2015;16(1). Available from: <http://dx.doi.org/10.1186/s12864-015-1805-9>
  144. Akiyoshi DE, Dilo J, Pearson C, Tumwine J, Tzipori S, Chapman S, et al. Characterization of *Cryptosporidium meleagridis* of Human Origin Passaged through Different Host Species. Infect Immun. 2003;71(4):1828–32.
  145. Salipante SJ, Kawashima T, Rosenthal C, Hoogestraat DR, Cummings LA, Sengupta DJ, et al. Performance comparison of Illumina and Ion Torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. Appl Environ Microbiol. 2014;80(24):7583–91.
  146. Lahens NF, Ricciotti E, Smirnova O, Toorens E, Kim EJ, Baruzzo G, et al. A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression. BMC Genomics. 2017;18(1):602.
  147. Lantz H, Dominguez Del Angel V, Hjerde E, Sterck L, Capella-Gutierrez S, Notredame C, et al. Ten steps to get started in Genome Assembly and Annotation. F1000Research. 2018;7.
  148. Baptista RP, Kissinger JC. Is reliance on an inaccurate genome sequence sabotaging your experiments?. PLoS Pathog. 2019;15(9):1–11.
  149. Bleidorn C. Phylogenomics An Introduction. Madrid: Springer Nature; 2017. 1–219 p.

150. Zahedi A, Gofton AW, Jian F, Papparini A, Oskam C, Ball A, et al. Comparative genomic analysis reveals occurrence of genetic recombination in virulent *Cryptosporidium hominis* subtypes and telomeric gene duplications in *Cryptosporidium parvum*. *BMC Genomics*. 2013;5(1):1–14.
151. Šlapeta J. Cryptosporidiosis and *Cryptosporidium* species in animals and humans: A thirty colour rainbow?. *Int J Parasitol* [Internet]. 2013;43(12–13):957–70. Available from: <http://dx.doi.org/10.1016/j.ijpara.2013.07.005>
152. Troell K, Hallström B, Divne AM, Alsmark C, Arrighi R, Huss M, et al. *Cryptosporidium* as a testbed for single cell genome characterization of unicellular eukaryotes. *BMC Genomics* [Internet]. 2016;17(1):1–12. Available from: <http://dx.doi.org/10.1186/s12864-016-2815-y>
153. Calarco L, Barratt J, Ellis J. Genome wide identification of mutational hotspots in the apicomplexan parasite *neospora caninum* and the implications for virulence. *Genome Biol Evol*. 2018;10(9):2417–31.
154. De Maio N, Shaw LP, Hubbard A, George S, Sanderson ND, Swann J, et al. Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microb Genomics*. 2019;5(9):1–21.
155. Watson M. A simple test for uncorrected insertions and deletions (indels) in bacterial genomes [Internet]. 2018 [cited 2020 Mar 25]. Available from: <http://www.opiniomics.org/a-simple-test-for-uncorrected-insertions-and-deletions-indels-in-bacterial-genomes/>
156. Blake DP, Clark EL, Macdonald SE, Thenmozhi V, Kundu K, Garg R, et al. Population, genetic, and antigenic diversity of the apicomplexan *Eimeria tenella* and their relevance to vaccine development. *Proc Natl Acad Sci U S A*. 2015;112(38):E5343–50.
157. Grinberg A, Widmer G. *Cryptosporidium* within-host genetic diversity: Systematic bibliographical search and narrative overview. *Int J Parasitol* [Internet]. 2016;(March). Available from:

<http://dx.doi.org/10.1016/j.ijpara.2016.03.002>

158. Weir W, Karagenc T, Gharbi M, Simuunza M, Aypak S, Aysul N, et al. Population diversity and multiplicity of infection in *Theileria annulata*. *Int J Parasitol* [Internet]. 2011;41(2):193–203. Available from: <http://dx.doi.org/10.1016/j.ijpara.2010.08.004>
159. Carlton JM, Escalante AA, Neafsey D, Volkman SK. Comparative evolutionary genomics of human malaria parasites. *Trends Parasitol*. 2008;24(12):545–50.
160. Bontell IL, Hall N, Ashelford KE, Dubey JP, Boyle JP, Lindh J, et al. Whole genome sequencing of a natural recombinant *Toxoplasma gondii* strain reveals chromosome sorting and local allelic variants. *Genome Biol*. 2009;10(5).
161. Hayashida K, Abe T, Weir W, Nakao R, Ito K, Kajino K, et al. Whole-genome sequencing of *theileria parva* strains provides insight into parasite migration and diversification in the african continent. *DNA Res*. 2013;20(3):209–20.
162. Eisen JA, Fraser CM. Phylogenomics: Intersection of evolution and genomics. *Science* (80- ). 2003;300(5626):1706–7.
163. Lack JB, Reichard M V., Van Den Bussche RA. Phylogeny and evolution of the Piroplasmida as inferred from 18S rRNA sequences. *Int J Parasitol*. 2012;42(4):353–63.
164. Roy Amit RS. Molecular Markers in Phylogenetic Studies-A Review. *J Phylogenetics Evol Biol* [Internet]. 2014;02(02). Available from: <http://esciencecentral.org/journals/molecular-markers-in-phylogenetic-studiesa-review-2329-9002-2-131.php?aid=30965>
165. Gabaldón T. Filogenómica y evolución del genoma [Internet]. 2011 [cited 2020 Mar 23]. Available from: [https://www.institutoroche.es/biotecnologia/62/filogenomica\\_y\\_evolucion\\_de\\_l\\_genoma](https://www.institutoroche.es/biotecnologia/62/filogenomica_y_evolucion_de_l_genoma)
166. Read TD, Salzberg SL, Pop M, Shumway M, Umayam L, Jiang L, et al. Comparative genome sequencing for discovery of novel polymorphisms in

- Bacillus anthracis. *Science* (80- ). 2002;296(5575):2028–33.
167. Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, et al. Whole-Genome Comparison of Mycobacterium tuberculosis Clinical and Laboratory Strains. *Society*. 2002;184(19):5479–90.
168. Theys K, Lemey P, Vandamme AM, Baele G. Advances in Visualization Tools for Phylogenomic and Phylodynamic Studies of Viral Diseases. Vol. 7, *Frontiers in Public Health*. 2019.
169. Bogema DR, Micallef ML, Liu M, Padula MP, Djordjevic SP, Darling AE, et al. Analysis of Theileria orientalis draft genome sequences reveals potential species-level divergence of the Ikeda, Chitose and Buffeli genotypes. *BMC Genomics*. 2018;19(1):1–15.
170. Cornillot E, Hadj-Kaddour K, Dassouli A, Noel B, Ranwez V, Vacherie B, et al. Sequencing of the smallest Apicomplexan genome from the human pathogen Babesia microti. *Nucleic Acids Res*. 2012;40(18):9102–14.
171. Galen SC, Borner J, Martinsen ES, Schaer J, Austin CC, West CJ, et al. The polyphyly of Plasmodium: Comprehensive phylogenetic analyses of the malaria parasites (Order Haemosporida) reveal widespread taxonomic conflict. *R Soc Open Sci*. 2018;5(5).
172. Ren X, Zhao J, Zhang L, Ning C, Jian F, Wang R, et al. Cryptosporidium tyzzeri n. sp. (Apicomplexa: Cryptosporidiidae) in domestic mice (Mus musculus). *Exp Parasitol*. 2012;130(3):274–81.
173. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59(3):307–21.
174. Danišová O, Valenčáková A, Stanko M, Luptáková L, Hatalová E, Čanády A. Rodents as a reservoir of infection caused by multiple zoonotic species/genotypes of C. parvum, C. hominis, C. suis, C. scrofarum, and the first evidence of C. muskrat genotypes I and II of rodents in Europe. *Acta*

- Trop. 2017;172(October):29–35.
175. Li N, Xiao L, Cama VA, Ortega Y, Gilman RH, Guo M, et al. Genetic recombination and *Cryptosporidium hominis* virulent subtype IbA10G2. *Emerg Infect Dis.* 2013;19(10):1573–82.
  176. Pain A, Crossman L, Parkhill J. Comparative Apicomplexan genomics. *Nature.* 2005;3(June):454–5.
  177. C. de Oliveira T, Rodrigues PT, Menezes MJ, Gonçalves-Lopes RM, Bastos MS, Lima NF, et al. Genome-wide diversity and differentiation in New World populations of the human malaria parasite *Plasmodium vivax*. *PLoS Negl Trop Dis.* 2017;11(7):1–24.
  178. Widmer G. Diverse single-amino-acid repeat profiles in the genus *Cryptosporidium*. *Mineral Mag.* 2018;145(9):1151–60.
  179. Nickel W. The mystery of nonclassical protein secretion: A current view on cargo proteins and potential export routes. *Eur J Biochem.* 2003;270(10):2109–19.
  180. Omasits U, Varadarajan AR, Schmid M, Goetze S, Melidis D, Bourqui M, et al. An integrative strategy to identify the entire protein coding potential of prokaryotic genomes by proteogenomics. 2017;2083–95.
  181. Miles A, Iqbal Z, Vauterin P, Pearson R, Campino S, Theron M, et al. Indels, structural variation, and recombination drive genomic diversity in *Plasmodium falciparum*. *Genome Res.* 2016;26(9):1288–99.
  182. Feng Y, Xiao L. Differential Expression of Three *Cryptosporidium* Species-Specific MEDLE Proteins. 2019;10(May):1–11.
  183. Carruthers VB, Tomley FM. Microneme proteins in apicomplexans. *Subcell Biochem.* 2008;47:33–45.
  184. Okhuysen PC, Chappell CL. *Cryptosporidium* virulence determinants - Are we there yet?. *Int J Parasitol.* 2002;32(5):517–25.
  185. Bonnin A, Ojcius DM, Souque P, Barnes DA, Doyle PS, Gut J, et al.

- Characterization of a monoclonal antibody reacting with antigen-4 domain of gp900 in *Cryptosporidium parvum* invasive stages. *Parasitol Res.* 2001;87(8):589–92.
186. Barnes DA, Bonnin A, Huang JX, Gousset L, Wu J, Gut J, et al. A novel multi-domain mucin-like glycoprotein of *Cryptosporidium parvum* mediates invasion. *Mol Biochem Parasitol.* 1998;96(1–2):93–110.
  187. Jain BP, Pandey S. WD40 Repeat Proteins: Signalling Scaffold with Diverse Functions. *Protein J* [Internet]. 2018;37(5):391–406. Available from: <http://dx.doi.org/10.1007/s10930-018-9785-7>
  188. Xu C, Min J. Structure and function of WD40 domain proteins. *Protein Cell.* 2011;2(3):202–14.
  189. Stirnimann CU, Petsalaki E, Russell RB, Christoph WM. WD40 proteins propel cellular networks. *Rev Trends Biochem. Sci* Vol35. 2010;35(10):531–8.
  190. Chahar P, Kaushik M, Gill SS, Gakhar SK, Gopalan N, Datt M, et al. Genome-wide collation of the *Plasmodium falciparum* WDR protein superfamily reveals malarial parasite-specific features. *PLoS One.* 2015;10(6):1–24.
  191. Sullivan WJ, Radke JB, Kim K, White MW. Epigenetic and Genetic Factors that Regulate Gene Expression in *Toxoplasma gondii* [Internet]. Second Edi. Elsevier; 2013. 613–645 p. Available from: <http://dx.doi.org/10.1016/B978-0-12-396481-6.00018-0>
  192. Ji D-D, Arnot DE. A *Plasmodium falciparum* homologue of the ATPase subunit of a multi- protein complex involved in chromatin remodelling for transcription. *Mol Biochem Parasitol.* 1997;88(1–2):151–62.
  193. Sullivan WJ, Monroy MA, Bohne W, Nallani KC, Chrivia J, Yaciuk P, et al. Molecular cloning and characterization of an SRCAP chromatin remodeling homologue in *Toxoplasma gondii*. *Parasitol Res.* 2003;90(1):1–8.
  194. Templeton TJ, Iyer LM, Anantharaman V, Enomoto S, Abrahante JE, Subramanian GM, et al. Comparative analysis of apicomplexa and genomic

- diversity in eukaryotes. *Genome Res.* 2004;14(9):1686–95.
195. Huynh MH, Boulanger MJ, Carruthers VB. A conserved apicomplexan microneme protein contributes to *Toxoplasma gondii* invasion and virulence. *Infect Immun.* 2014;82(10):4358–68.