

Análisis filogenético de las proteínas que integran los complejos replicativos del DNA pre-RC y pre-IC en las principales especies de *Cryptosporidium* asociadas con infección en humanos

Daniel A. Gómez-Chavarría¹ y Juan F. Alzate¹

¹ Centro Nacional de Secuenciación Genómica - CNSG, Sede de Investigación Universitaria - SIU, Departamento de Microbiología y Parasitología, Facultad de Medicina, Universidad de Antioquia, Medellín, Antioquia, Colombia

RESUMEN

Los parásitos del género *Cryptosporidium* pueden infectar un amplio rango de vertebrados y se han asociado con frecuencia a brotes de parásitos transmitidos por el agua, e infecciones en pacientes VIH y niños menores de 5 años causando cuadros diarreicos al infectar los enterocitos. Dentro de este género se sabe que 3 especies en particular se encuentran fuertemente asociadas con infecciones en humanos: *C. hominis*, *C. parvum* y *C. meleagridis*. Aun cuando se han presentado grandes avances en la secuenciación de genomas para este protozoo, sigue habiendo baches en la anotación de las proteínas de la maquinaria replicativa, en especial las que conforman los complejos de pre-RC y pre-IC. En este estudio, se descargaron 21 genomas de apicomplexas disponibles en bases de datos públicas, entre ellos 4 corresponden a especies de *Cryptosporidium* y usando proteínas de un genoma eucariota modelo curado, se buscaron las proteínas de los complejos pre-RC y pre-IC en *Cryptosporidium* mediante árboles filogenéticos de máxima verosimilitud que reprodujeran la historia evolutiva del filo apicomplexa. Se pudo evidenciar que los complejos de pre-RC y pre-IC en *Cryptosporidium* se encuentran constituidos por genes conservados de la maquinaria replicativa típica eucariota.

Palabras clave: *Cryptosporidium*, apicomplexa, replicación, maquinaria replicativa, filogenia, complejo de pre-RC y complejo de pre-IC.

INTRODUCCIÓN:

Cryptosporidium es un parásito intestinal intracelular obligado, perteneciente al filo de los Apicomplexa, tiene un ciclo de vida monoxeno con fases de reproducción sexual y asexual [1]. Se han descrito más de 30 especies de *Cryptosporidium*, pero sólo 3 especies: *C. parvum*, *C. hominis* y *C. meleagridis* son las más prevalentes causando enfermedad en humanos [2].

Este protozoo es el agente causal de la criptosporidiosis, una infección antropozoonótica caracterizada por un cuadro diarreico que puede afectar tanto a humanos como animales. Su principal vía de infección es la oro-fecal, al ingerir ooquistes esporulados de origen animal o humano expulsados en la materia fecal. Los ooquistes pueden contaminar fuentes hídricas, alimentos o aguas de uso recreacional [2]

En personas inmunocompetentes la enfermedad se manifiesta como un cuadro diarreico agudo autolimitado que se resuelve espontáneamente, caso contrario con las personas inmunocomprometidas, especialmente pacientes VIH/SIDA, donde se presenta como un síndrome diarreico grave con tendencia a la cronicidad lo cual puede comprometer la vida del paciente [3] [4]. Los niños menores de 5 años son otra población bastante afectada por esta enfermedad, en donde la criptosporidiosis representa la quinta causa de mortalidad por diarrea; además, se ha asociado con desnutrición y retraso en el crecimiento y desarrollo [4] [5]. En Colombia se estima una prevalencia de 19,7% en niños, 20,4% en animales y 38,9% en muestras de agua, lo cual aumenta el riesgo de infección por este protozoo en el país [6].

Con el fin de entender mejor la biología de este parásito, es fundamental continuar avanzando en el estudio de sus características genómicas y evolutivas. De particular interés es analizar la evolución de los genes que codifican proteínas que intervienen durante el proceso de replicación del genoma. La replicación del ADN es una de las tareas más importantes que llevan a cabo los organismos vivos, con esto se aseguran el paso correcto de la información de una generación a otra y la integridad funcional. [7]

Para que se dé adecuadamente este proceso se requiere de la intervención de múltiples proteínas y complejos proteicos, las cuales entran a cumplir su función en diferentes etapas. En eucariotas la replicación se da en dos pasos separados temporalmente, el primero inicia desde la fase M tardía hasta la G1 y comprende los procesos de ensamblaje de los complejos de pre-replicación (pre-RC), este se encuentra conformado por: el complejo de reconocimiento de origen (ORC), Cdt1 y Cdc6 los cuales tienen como función cargar la helicasa replicativa (MCM2-7) al ADN en su forma inactiva. Durante la fase S se da el segundo momento, llevándose a cabo la activación de las helicasas por parte de Cdc45 y un complejo tetramérico llamado GINS. El producto de la unión Cdc45 – Mcm 2–7 – GINS (CMG), constituye el complejo de pre-iniciación (pre-IC), el cual posee actividad helicasa y su formación se asocia con la fosforilación de MCM por el complejo quinasa Cdc7-Dbf4. Una vez formado, el complejo CMG es capaz de reclutar los otros componentes del replisoma al ADN fundido en el origen. Este proceso es bien conocido en eucariotas modelo, sin embargo, en modelos parasitarios aún no se conoce a detalle cómo se dan estos procesos

replicativos y si existen ortólogos de todas las proteínas descritas en este proceso en los eucariotas modelo. [8]

Gracias a la disponibilidad pública de decenas de genomas de *Cryptosporidium*, se abre la oportunidad de desarrollar análisis bioinformáticos y filogenéticos más detallados que permitan encontrar los ortólogos de la maquinaria replicativa, determinar inicialmente presencia y/o ausencia de estas proteínas durante el proceso evolutivo del parásito y así alcanzar una mejor comprensión de los procesos moleculares y biológicos involucrados en el ciclo de vida de este parásito, al tiempo que se abren nuevas oportunidades para blancos terapéuticos los cuales escasean actualmente en las parasitosis. [9]

MÉTODOS:

Búsqueda y muestreo de genomas

Se recurrió a VEuPathDB, el cual es un centro de recursos bioinformáticos [10]. Se usaron en total 20 genomas de apicomplexas, cubriendo así algunos géneros y especies de los 5 linajes mayores dentro de este filo. Haemospororida (*Plasmodium falciparum* 3D7, *Plasmodium vivax* Sal-1, *Plasmodium knowlesi* strain H, *Plasmodium chabaudi* chabaudi, *Plasmodium yoelii* yoelii 17XNL, *Plasmodium berghei* ANKA); Piroplasmida (*Babesia microti* strain RI, *Babesia bovis* T2Bo, *Theileria equi* strain WA, *Theileria orientalis* strain Shintoku, *Theileria annulata* strain Ankara), Coccidia (*Toxoplasma gondii* ME49, *Eimeria falciformis* Bayer Haberkorn, *Eimeria brunetti* Houghton y *Neospora caninum* Liverpool), Gregarinasina (*Gregarina niphandrodes* Unknown strain) y Cryptosporidium (*Cryptosporidium parvum* Iowa II, *Cryptosporidium hominis* UdeA01, *Cryptosporidium muris* RN66, *Cryptosporidium meleagridis* strain UKMEL1). Estos organismos tienen en común que son eucariotas, unicelulares, parásitos [11]. También se tomó también el genoma de *Chromera velia* (*Chromera velia* CCMP2878), el cual es un alga unicelular fotosintética perteneciente al subfilum alveolata y el pariente más cercano conocido a los parásitos apicomplexanos[12]. Este último junto con *Gregarina niphandrodes* servirán como grupos externos en las filogenias.

Secuencias de las proteínas que conforman los pre-RC y pre-IC

La base de datos UniProt es un sitio web que contiene secuencias de proteínas y datos de anotación [13]. De esta se obtuvieron 19 proteínas en total, las cuales conforman los complejos de pre-replicación y pre-iniciación en eucariotas; se escogió como organismo modelo a *Saccharomyces cerevisiae*, código de acceso S288C, al ser un eucariota, unicelular cuyo genoma se encuentra completamente secuenciado, anotado y curado[14].

Búsqueda en genomas de apicomplexas

Mediante la estrategia BLASTP se buscó cada una de las secuencias de las proteínas referentes en todos los genomas apicomplexanos y se extrajeron aquellas cuyos valores de e-value fueran $\leq 1e-6$ y bit score ≥ 300 , estos valores darán indicios de relaciones biológicamente significativas, que luego pasarían a verificarse en los árboles filogenéticos.

Reconstrucción de las filogenias

Una vez obtenidas las secuencias de cada proteína mediante el BLASTp, se procedió a alinearlas con ayuda de la herramienta bioinformática MAFFT [15]. Cada alineamiento se usó para reconstruir árboles filogenéticos usando la estrategia de máxima verosimilitud (maximum likelihood) con 1000 repeticiones por árbol en el programa IQ-TREE2 *Efficient software for phylogenomic inference* [16]. De los resultados obtenidos, se tomaron los árboles consenso (archivos .contree) que fueron visualizados y editados con el programa FigTree[17].

Análisis de señales de localización nuclear (NLS) de las proteínas encontradas en *Cryptosporidium*

Para dicho análisis se usó el servidor web NLStradamus, el cual permite predecir en que posición de las secuencias, se encuentra la señalización que ayuda a identificar y/o confirmar que proteínas van al núcleo [18].

RESULTADOS

Características genómicas generales de *Cryptosporidium* respecto a otros apicomplexas

Durante el proceso de recolección de los genomas se pudo apreciar de forma general las características de estos. Si bien todos los genomas pertenecen a parásitos apicomplexas se encuentran diferencias entre ellos, la más notorias y evidentes fueron el tamaño del genoma y la cantidad genes: Por ejemplo, para el caso de *Cryptosporidium hominis* UdeA01 es de 9.04 Mb con 3869 genes, tamaño doblado por el genoma de *Plasmodium berghei* ANKA de 18,78 Mb con 5254 genes y 7 veces más pequeño que el genoma de *Toxoplasma gondii* ME49 de 65.67 Mb con 8920 genes (Ver Tabla 1). Esto ya nos da indicios de divergencias genéticas muy marcadas en la línea evolutiva de estos protozoos y puede deberse a la pérdida y ganancia de genes y/o la reducción genómica dependiendo de donde se observe.

Especies	Código de acceso	Tamaño (Mb)
<i>Cryptosporidium hominis</i> UDEA01	LN877954	9.04
<i>Cryptosporidium parvum</i> Iowa II	CM000436	9.10
<i>Cryptosporidium muris</i> RN66	DS989726	9.25
<i>Cryptosporidium meleagridis</i> strain UKMEL1	JIBK01000002	8.97
<i>Babesia bovis</i> T2Bo	AAXT01000001	8.18
<i>Babesia microti</i> strain RI	LN871599	6.43
<i>Theileria annulata</i> strain Ankara	tannA_chr01	8.36
<i>Theileria equi</i> strain WA	CP001669	11.67
<i>Theileria orientalis</i> strain Shintoku	AP011946	9.01
<i>Plasmodium falciparum</i> 3D7	Pf3D7_14_v3	23.33
<i>Plasmodium berghei</i> ANKA	PbANKA_14_v3	18.78
<i>Plasmodium chabaudi</i> chabaudi	PCHAS_13_v3	18.97
<i>Plasmodium knowlesi</i> strain H	PKNH_14_v2	24.40
<i>Plasmodium vivax</i> Sal-1	Pv_Sal1_chr14	27.01
<i>Plasmodium yoelii</i> yoelii 17XNL	AABL01005066	22.94
<i>Neospora caninum</i> Liverpool	FR823391	59.10
<i>Toxoplasma gondii</i> ME49	TGME49_chrX	65.67
<i>Eimeria falciformis</i> Bayer Haberkorn 1970	EfaB_Contig_56725	43.67
<i>Eimeria brunetti</i> Houghton	HG710173	66.89
<i>Chromera velia</i> CCMP2878	CCMP2878	193.89
<i>Gregarina niphadores</i> Unknown strain	KK211454	14.01

Tabla 1. Resumen de los genomas incluidos en el análisis

Árboles filogenéticos de máxima verosimilitud

Se pudieron reconstruir un total de 10 árboles filogenéticos distribuidos de la siguiente manera: 1 de Cdc6, 1 de Cdc45, 1 de Cdt1, 1 árbol consenso con las 6 subunidades MCM (MCM2-7), 6 árboles de ORC; uno por cada subunidad (ORC1-6). De las proteínas que conforman el complejo GINS (Sld5, Psf1, Psf2 y Psf3) no se pudo obtener los árboles filogenéticos con la metodología usada. En las demás filogenias se logró reconstruir árboles con todas las especies del linaje *Cryptosporidium* estudiadas y representantes de los demás linajes Piroplasmida, Haemosporida, Coccidia y Gregarinasina. Todos los árboles se encuentran enraizados con *S. cerevisiae*, además cuentan con *Chromera velia*, pariente más

cercano conocido a los parásitos apicomplexanos [12]. Este último junto con *Gregarina niphandrodes* servirán como grupos externos en las filogenias.

Las filogenias esperadas debían conservar el patrón topológico que muestra las relaciones evolutivas dentro del nicho de los apicomplexa (ver Figura complementaria 1). En donde se aprecia la distribución en taxones formando cladas bien definidas con sus respectivos miembros. A su vez en el linaje del taxón de *Cryptosporidium* se espera encontrar a *C. muris* como linaje ancestral, seguido de *C. meleagridis* y finalmente un nodo que se bifurca en las dos especies mayor emparentadas *C. hominis* y *C. parvum*, con soportes muy altos. Un dato interesante entre *C. hominis* y *C. parvum* es que ambos conservan el mismo patrón de nomenclatura numérica en los genes, lo que ayuda y da fuerza a la relación de los ortólogos, esto puede relacionarse con el alto nivel de conservación entre los genomas ambas especies.

La reconstrucción filogenética correspondiente al complejo hexamérico MCM2-7 (Ver figura1) muestra que esta proteína se encuentra conservada en todo el Filo y se logra reproducir la topología en todas las subunidades y fue la única proteína multimérica que permitió la reconstrucción de un árbol que incluyera todas las subunidades sin generar interferencias entre las secuencias que pudiera alterar los soportes o la topología. En general se pudo apreciar que los nodos internos o ancestrales tienen soportes de Bootstrap menores que los nodos terminales, en estos últimos se da un predominio de valores entre 90 y 100. Este patrón de soportes se conservó para todos los demás obtenidos.

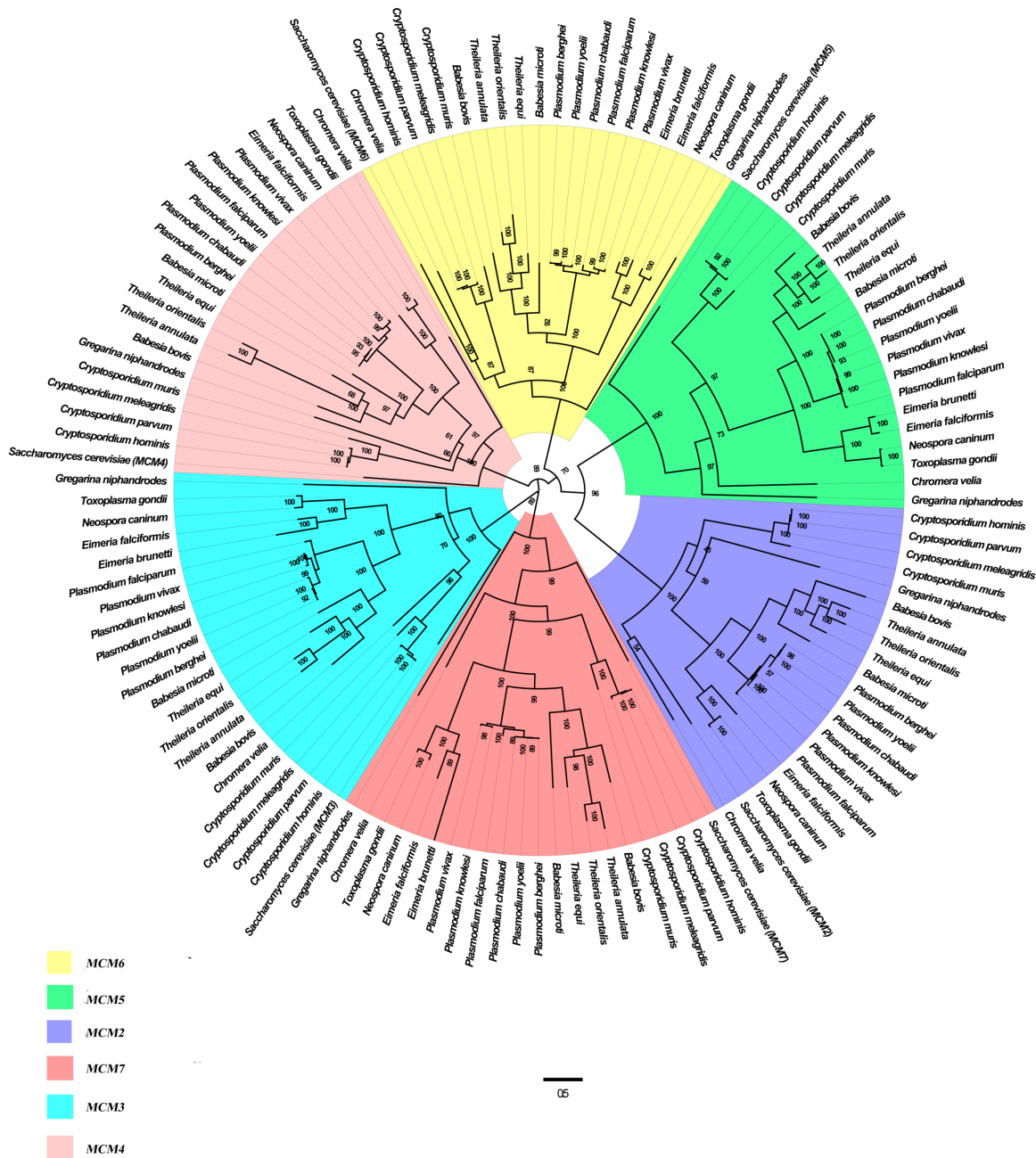


Figura 1. Árbol consenso de máxima verosimilitud del complejo Mcm2-7. Cada sub-árbol reproduce la topología del filo apicomplexa (Ver figuras complementarias 2, 3, 4, 5, 6 y 7).

Las proteínas del reconocimiento de origen (ORC) también son un complejo hexamérico, sin embargo, estas no permitieron hacer una reconstrucción en forma de general de todas las subunidades ya que al unir los árboles individuales estos perdían la topología esperada y se agrupaban de forma inespecífica con bajos soportes. Se presentan árboles individuales de cada una de las secuencias ORC1-6. (Ver figuras 2 a 7). La filogenia correspondiente a la subunidad ORC1 (Ver figura 2) es la que reproduce por completo una topología ideal, sin embargo, en las demás subunidades algunos de los taxones pertenecientes a los Piroplasmida

comienzan a saltar entre los diferentes linajes y los grupos externos *C. velia* y *G. niphandrodes*, en algunas filogenias pierden su posición ancestral como por ejemplo en la subunidad ORC3 (ver figura 4). Se sigue conservando el patrón de los soportes apicales altos y basales bajos.

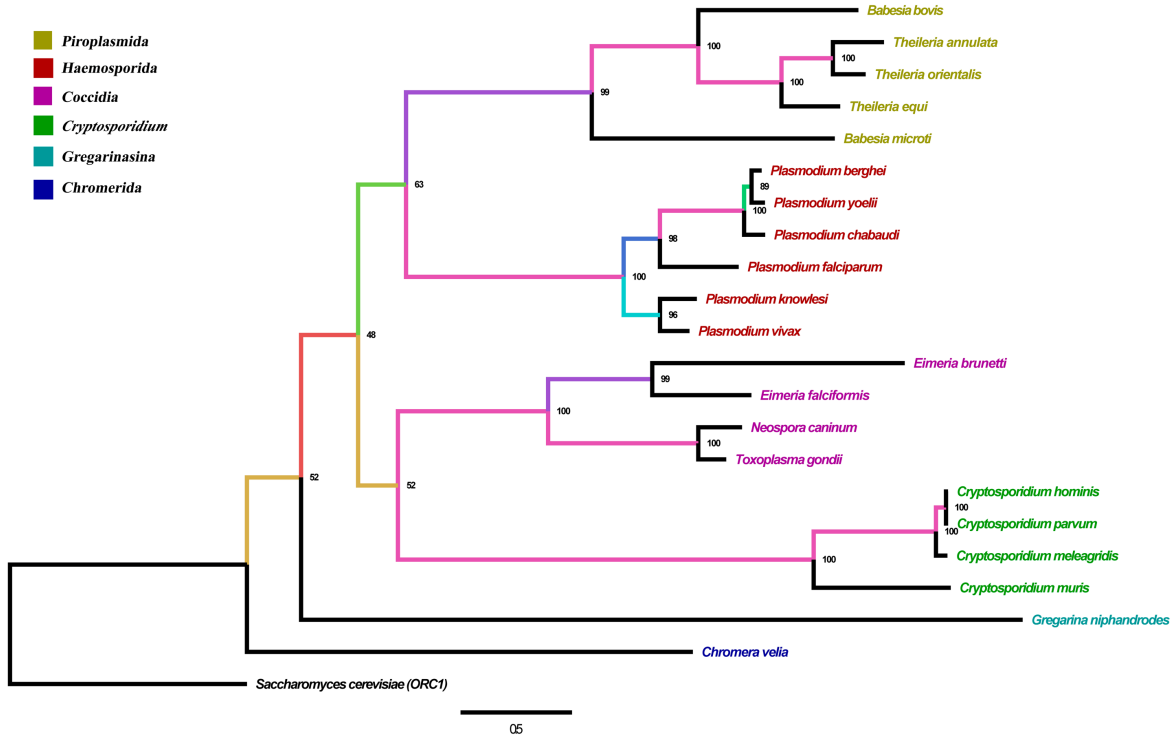


Figura 2. Árbol consenso de máxima verosimilitud de la subunidad ORC1.

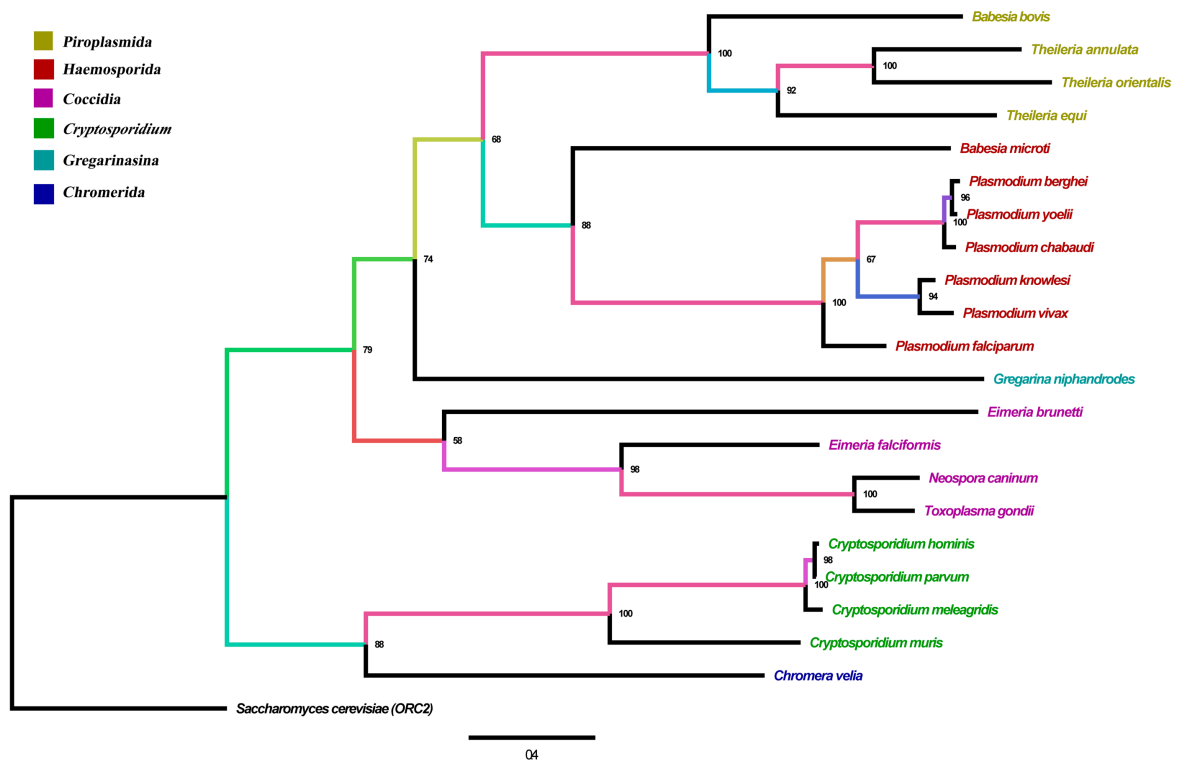


Figura 3. Árbol consenso de máxima verosimilitud de la subunidad ORC2.

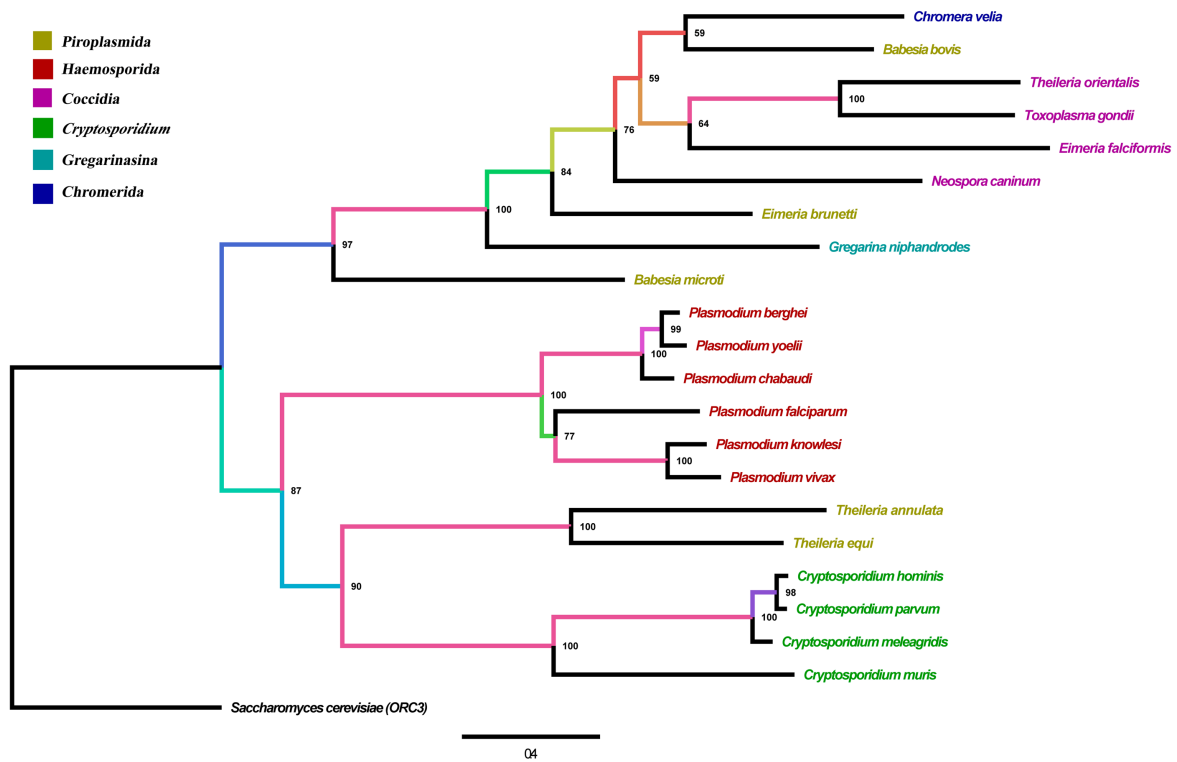


Figura 4. Árbol consenso de máxima verosimilitud de la subunidad ORC3.



Figura 5. Árbol consenso de máxima verosimilitud de la subunidad ORC4.

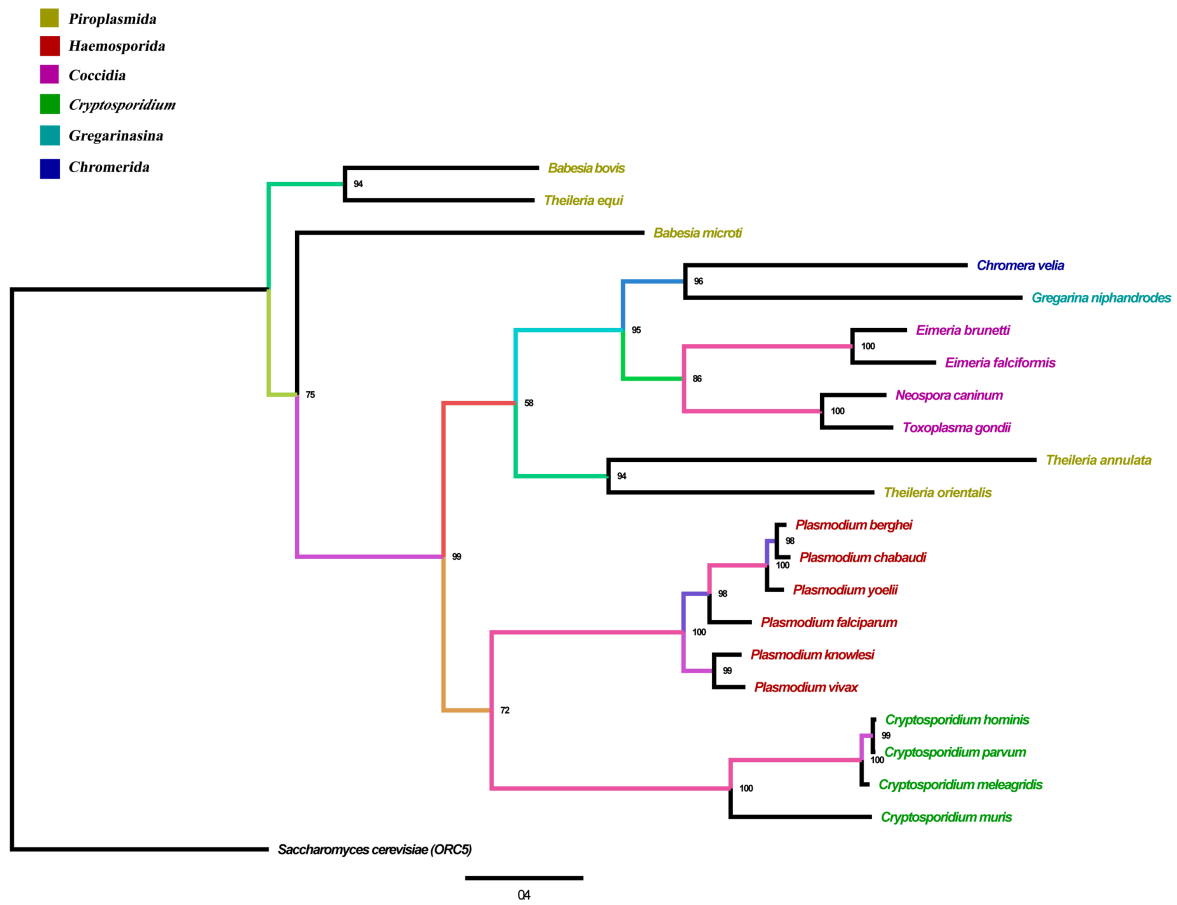


Figura 6. Árbol consenso de máxima verosimilitud de la subunidad ORC5.



Figura 7. Árbol consenso de máxima verosimilitud de la subunidad ORC6.

En resumen, de las proteínas que conforman el complejo de reconocimiento de origen (ORC) se hallaron todas las secuencias en las 4 especies de *Cryptosporidium* estudiadas. Respecto a los demás linajes, los haemosporida y las coccidias mantienen sus miembros agrupados en una misma clada, caso contrario de los piroplasmida, donde sus miembros se distribuyen entre los Haemosporida y las coccidias.

Por otra parte, de la proteína Cdc45 se encontró una filogenia con la topología esperada para *Cryptosporidium*. Respecto a los demás linajes se encuentran cambios de posiciones entre algunos miembros de los haemosporida y los piroplasmida, además, se pierde la señal de dos coccidias. Se sigue conservando el mismo patrón de los soportes apicales altos y basales bajos (Ver figura 8).

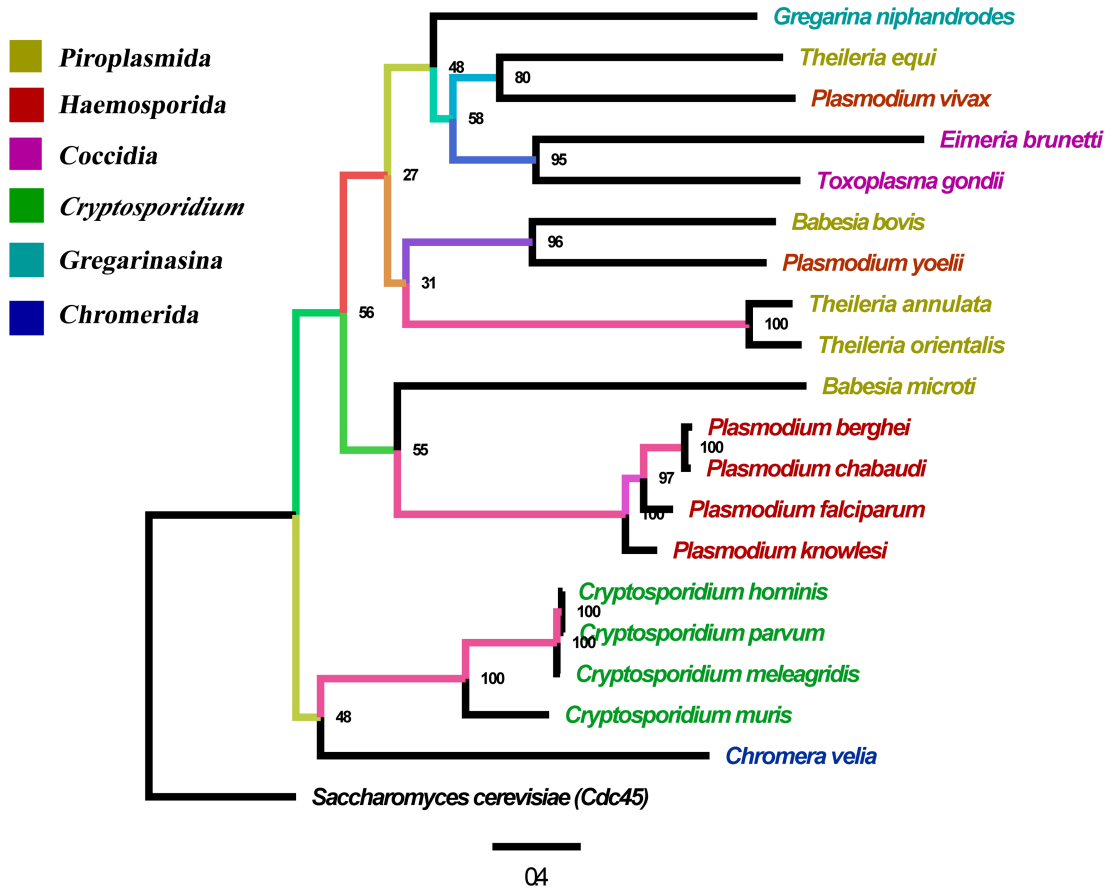


Figura 8. Árbol consenso de máxima verosimilitud de la subunidad Cdc45.

Para el caso de la proteína Cdc6, se encuentra un movimiento en la clada de *Cryptosporidium*, donde *C. meleagridis* que debería estar basal respecto a *C. hominis* y *C. parvum*, cambia posición con *C. parvum* (Ver figura 9). Esto hace que el soporte de este nodo decaiga ligeramente. Las demás ramas se mantienen bien respecto a la topología esperada con soportes apicales altos y basales bajos. Es importante resaltar que la relación entre la nomenclatura de los genes es igual para *C. hominis* y *C. parvum*. (Ver tabla 3)

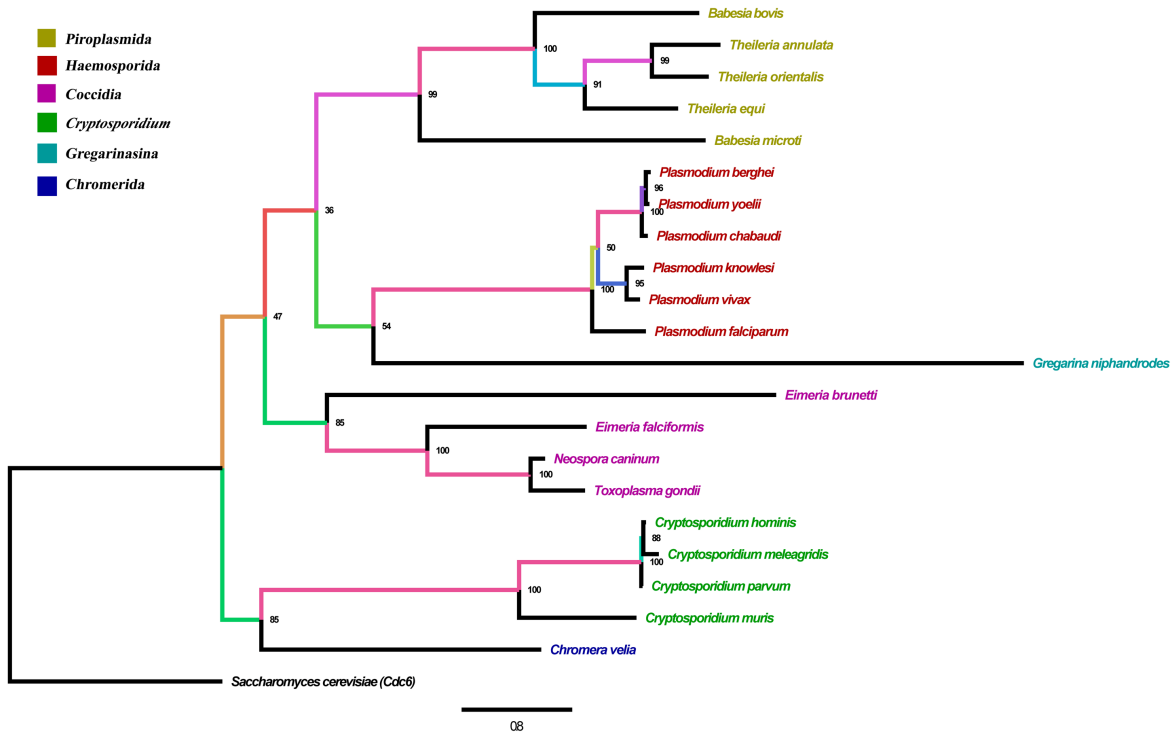


Figura 9. Árbol consenso de máxima verosimilitud de la proteína Cdc6.

Siguiendo con la proteína Cdt1 se da un hallazgo bastante interesante, pues es la única proteína que mostró la existencia de parálogos por eventos de duplicación del gen en algún momento de su evolución, este fenómeno se observó en especial es taxones de los Piroplasmida: *B. bovis*, *T. annulata*, *T. orientalis* y *T. equi*, y en las coccidias: *T. gondii*. Se conserva la topología del árbol, manteniéndose el linaje que contiene las especies de *Cryptosporidium*. Se pierde la señal de *C. velia* y las dos especies de *Eimeria*. Se conserva el patrón de los soportes apicales altos y basales bajos. (Ver figura 10)

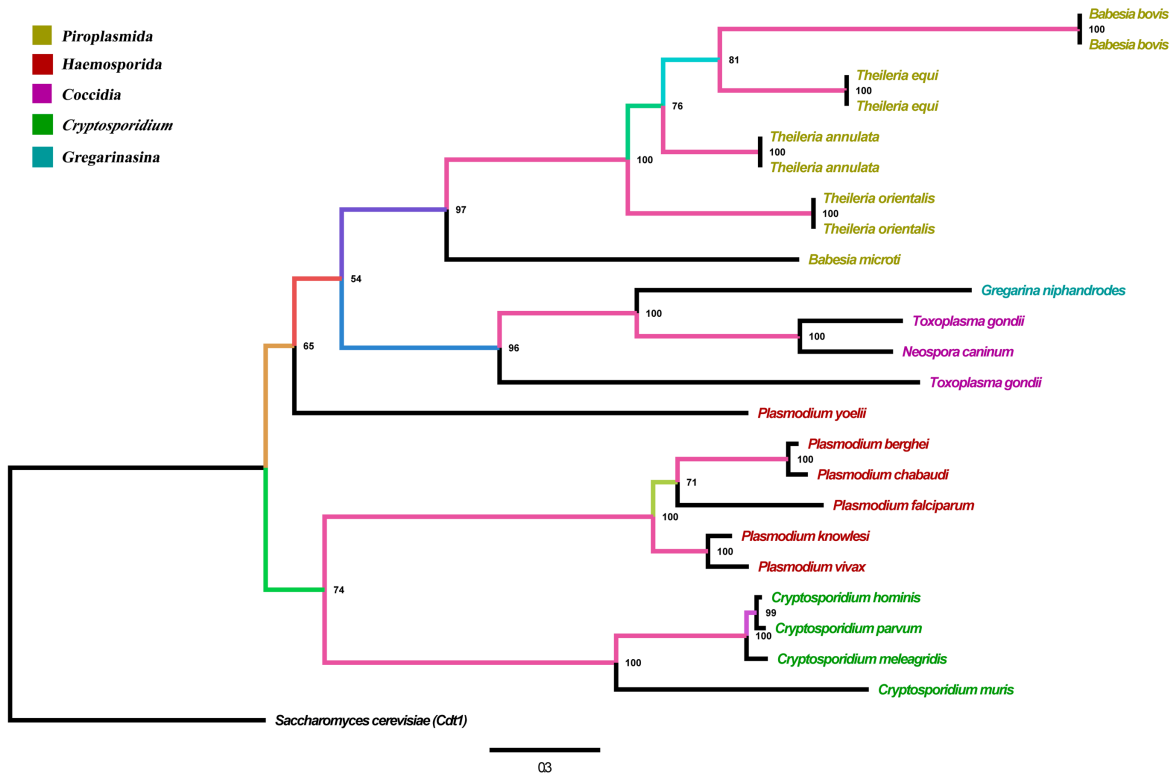


Figura 10. Árbol consenso de máxima verosimilitud de la proteína Cdt1.

Las proteínas del complejo GINS fueron las más difíciles de rastrear, y no se pudo obtener filogenias de estas, si bien se tuvo resultados de las búsquedas mediante BLASTP, al momento de reconstruir las filogenias no hubo ninguna relación biológicamente significativa y no fue posible discriminar entre las 4 posibles subunidades.

De forma general, se observó un relación significativa entre las proteínas encontradas en *C. parvum* y en *C. hominis*, donde los códigos numéricos de las anotaciones coinciden; en estos casos los nodos de estos linajes arrojaron valores entre 98 y 100 demostrando una fuerte filiación entre las proteínas. Se encontraron y se obtuvieron filogenias de todas las proteínas buscadas con excepción a las 4 que conforman el complejo GINS.

Predicción de señales de localización nuclear (NLS)

Posterior a los análisis filogenéticos se procedió buscar en las secuencias de las proteínas encontradas las señales de localización nuclear en las 4 especies de *Cryptosporidium* analizadas. No todas las búsquedas fueron exitosas. De un total de 60 secuencias proteicas solo en 21 se pudo predecir las señales de importación a núcleo. (Ver tabla 2). La proteína más consistente en sus predicciones fue la subunidad McM2, esta se encontró en las 4 especies, por otro lado, y al ser las más abundantes, se encontraron varias de las subunidades de McM2-7 y ORC1-6. A su vez, las secuencias señales se encuentran constituidas en su

mayoría por aminoácidos básicos principalmente Lisina (K) y Arginina (R); en algunos casos hay un predominio total de uno de estos dos aminoácidos.

Organismo	Gen ID	Proteína	Predicción de NLS
C. hominis UDEA01	CHUDEA2_1100-t30_1	MCM2	148 - RRKRR - 152
	CHUDEA4_4320-t30_1	ORC1	34 - KRGRRK - 40
			70 - KKRKV - 74
			83 - IKNKRGRKGKGVK - 95
	CHUDEA4_1930-t30_1	ORC2	33 - RNAAIEKLDKIRREK - 48
	CHUDEA4_430-t30_1	ORC5	560 - KLKSKRGRKKINKQSSEKVKSKNK - 583
			171 - KKRR - 174
CHUDEA4_2150-t30_1	CDT1	89 - KKKSKTRLKN - 98	
C. parvum Iowa II	cgd2_1100-RA	MCM2	148 - RRKRR - 152
	cgd4_970-RA	MCM7	628 - R - 628
	cgd4_4320-RA	ORC1	34 - KRGRRK - 40
			83 - KNKRGRKGK - 91
	cgd4_430-RA	ORC5	559 - KLKSKRGRRKINKQSSEKVKSRNK - 582
			158 - RKKLIKKGIFNSKKRRSR - 175
			559 - KLKSKRGRRKINKQSSEKVKSR - 580
cgd8_2380-RA	CDC45	166 - KK - 167	
cgd4_2150-RA	CDT1	89 - KKKSKTRLKN - 98	
C. meleagridis UKMEL I	CmeUKMEL1_18425-t37_1	MCM2	147 - KRRKR - 152
	CmeUKMEL1_01710-t37_1	ORC1	34 - KRGRKNE - 41
			84 - KNKRGRKGK - 93
	CmeUKMEL1_02085-t37_1	ORC2	33 - RNAAIERKLDKIRKEKR - 49
	CmeUKMEL1_08310-t37_1	ORC5	159 - RKKLIKKGIFNSKKRRSR - 176
			560 - KLKSKRGRRKINKQSS - 575
CmeUKMEL1_02205-t37_1	CDT1	89 - KKKSKTRLK - 97	
C. muris RN66	CMU_009270-t26_1	MCM2	150 - KRRKR - 154
	CMU_013220-t26_1	MCM3	724 - KKS - 727
	CMU_019090-t26_1	MCM7	616 - RKGRAKVR - 624
	CMU_007950-t26_1	ORC3	942 - RRAAELNQGKIKKLELKRERHKLAINS - 969

Tabla 2. Predicción de NLS en *Cryptosporidium*.

Secuencias proteicas encontradas

En total se encontraron 60 proteínas correspondientes a las subunidades MCM2-7, ORC1-6, Cdc45, Cdc6 y Cdt1 de los 4 especies de *Cryptosporidium* analizados. En la Tabla 3 se muestra el código de la proteína, el nombre con el cual se encuentran anotadas en la base de datos, el número de aminoácidos y la asociación filogenética correspondiente.

Gen ID	Longitud	Nombre de la proteína	Asociación filogenética
CHUDEA2_1100-t30_1	970	unspecified product	MCM2
CHUDEA2_1600-t30_1	862	unspecified product	MCM3
CHUDEA2_1250-t30_1	894	unspecified product	MCM4
CHUDEA7_2920-t30_1	791	unspecified product	MCM5
CHUDEA6_240-t30_1	1082	unspecified product	MCM6
CHUDEA4_970-t30_1	857	unspecified product	MCM7
cgd2_1100-RA	970	MCM domain containing protein	MCM2
cgd2_1600-RA	862	MCM domain containing protein	MCM3
cgd2_1250-RA	894	MCM domain containing protein	MCM4
cgd7_2920-RA	791	DNA replication licensing factor MCM5 like AAA+ ATPase	MCM5
cgd6_240-RA	1053	DNA replication licensing factor MCM6-like AAA Atpase	MCM6
cgd4_970-RA	857	DNA replication licensing factor MCM7	MCM7
CmeUKMEL1_18425-t37_1	970	MCM2/3/5 family protein	MCM2
CmeUKMEL1_16415-t37_1	860	MCM2/3/5 family protein	MCM3
CmeUKMEL1_18495-t37_1	894	MCM2/3/5 family protein	MCM4
CmeUKMEL1_03920-t37_1	791	MCM2/3/5 family protein	MCM5
CmeUKMEL1_17720-t37_1	1040	MCM2/3/5 family protein	MCM6
CmeUKMEL1_06380-t37_1	858	MCM2/3/5 family protein	MCM7
CMU_009270-t26_1	971	DNA replication licencing factor MCM2, putative	MCM2
CMU_013220-t26_2	872	DNA replication licencing factor MCM3, putative	MCM3
CMU_009090-t26_1	929	cell division control protein 54, putative	MCM4
CMU_022890-t26_1	807	DNA replication licensing factor MCM5, putative	MCM5
CMU_027970-t26_1	955	DNA replication licensing factor MCM6, putative	MCM6
CMU_019090-t26_1	833	DNA replication licencing factor MCM7, putative	MCM7
CHUDEA4_4320-t30_1	869	unspecified product protein	ORC1
CHUDEA4_1930-t30_1	355	unspecified product protein	ORC2
CHUDEA6_1940-t30_1	945	unspecified product protein	ORC3
CHUDEA2_1550-t3	495	unspecified product protein	ORC4
CHUDEA4_430-t30_1	740	unspecified product protein	ORC5
CHUDEA4_2420-t30_1	599	unspecified product protein	ORC6
cgd4_4320-RA	868	ORC/CDC6 like AAA+ Atpase	ORC1
cgd4_1940-RA	150	Nucleoside diphosphate kinase	ORC2
cgd6_1940-RA	645	Origin recognition complex subunit 3	ORC3
cgd2_1550-RA	432	P-loop containing nucleoside triphosphate hydrolase	ORC4
cgd4_430-RA	739	Origin recognition complex, subunit 5	ORC5
cgd4_2420-RA	599	Sushi/SCR/CCP domain containing protein	ORC6
CmeUKMEL1_01710-t37_1	869	AAA domain protein	ORC1
CmeUKMEL1_02085-t37_1	355	Origin recognition complex subunit 2 family protein	ORC2
CmeUKMEL1_12770-t37_1	932	Origin recognition complex (ORC) subunit 3 N-terminus family p	ORC3
CmeUKMEL1_16390-t37_1	455	hypothetical protein	ORC4
CmeUKMEL1_08310-t37_1	740	hypothetical protein	ORC5
CmeUKMEL1_02335-t37_1	599	Ankyrin repeats (3 copies) family protein	ORC6
CMU_002090-t26_1	779	hypothetical protein, conserved	ORC1
CMU_016600-t26_1	336	origin recognition complex subunit 2 family protein	ORC2
CMU_007950-t26_1	979	hypothetical protein, conserved	ORC3
CMU_013280-t26_1	527	hypothetical protein, conserved	ORC4
CMU_012540-t26_1	692	hypothetical protein, conserved	ORC5
CMU_017430-t26_1	585	SUSHI domain-containing protein	ORC6
CHUDEA7_2310-t30_1	550	unspecified product	CDC6
cgd7_2310-RA	551	ORC/CDC6 like AAA ATPase	CDC6
CmeUKMEL1_04210-t37_1	552	ATPase family associated with various cellular activities (AAA) fa	CDC6
CMU_034360-t26_1	631	hypothetical protein, conserved	CDC6
CHUDEA8_2380-t30_1	706	unspecified product	CDC45
cgd8_2380-RA	706	CDC45-like protein	CDC45
CmeUKMEL1_14865-t37_1	706	CDC45-like protein family protein	CDC45
CMU_026380-t26_1	730	hypothetical protein, conserved	CDC45
CHUDEA4_2150-t30_1	727	unspecified product	CDT1
cgd4_2150-RA	727	DNA replication factor CDT1 like protein	CDT1
CmeUKMEL1_02205-t37_1	726	DNA replication factor CDT1 like family protein	CDT1
CMU_017130-t26_1	737	hypothetical protein, conserved	CDT1

Tabla 3. Resumen de todas las proteínas encontradas en *Cryptosporidium* con su respectiva asociación filogenética: (*C. hominis* UDEA01), cgd (*C. parvum* Iowa II), CmeUKMEL strain UKMEL1 (*C. meleagridis*) y CMU (*C. muris* RN66)

De forma general se puede apreciar que algunas de las proteínas encontradas cuentan con anotaciones acordes a las proteínas referentes con las cuales se buscaron y reconstruyeron las filogenias, algunas muy específicas como el nombre de las subunidad a la que pertenece cuando se trata de complejos proteicos y otras muy generales cuando lo asocian a un complejo, sin mencionar la subunidad específica. Para el aislado *C. hominis* UDEA01 predomina el nombre “unspecified product” para las proteínas anotadas, es importante señalar la correspondencia total entre la identificación del gen de *C. hominis* UDEA01 y *C. parvum* Iowa II. Este último, entre las 4 especies, es quien cuenta con la nomenclatura, según los resultados encontrados, más acertada con muy pocas ambigüedades o diferencias. También es importante recalcar la similitud entre las longitudes entre los ortólogos, resultado reflejado en la tabla 3 si se comparan los tamaños de las secuencias.

DISCUSIÓN

Los avances a nivel bioinformático y la secuenciación de genomas han permitido la realización de estudios sobre microorganismos no modelo, esto a su vez abre puertas a otras disciplinas como la filogenética, la cual usando múltiples datos permite el análisis de las relaciones filogenéticas y los procesos evolutivos a nivel molecular de genes o taxones mejorando la comprensión no solo molecular sino sobre la biología de estos [19].

En este caso se recurrió a los análisis filogenéticos para contrastar la maquinaria de replicación, específicamente los complejos pre-RC y pre-IC de un eucariota modelo sobre modelos apicomplexanos; enfocado a 3 cepas de las principales especies de *Cryptosporidium* que afecta a humanos [2].

De las filogenias obtenidas se puede inferir que los apicomplexa analizados poseen complejos de pre-RC y pre-IC constituidos por ortólogos de las proteínas conocidas en eucariotas modelo. Se sabe que el complejo de pre-RC se encuentra conformado por la interacción estable y productiva del complejo proteico ORC que cuenta con 6 subunidades, la proteína Cdc6 la cual modula la actividad de ORC, la helicasa replicativa conformada también por 6 subunidades MCM y el mediador Cdt1 (20). Por otra parte, el complejo de pre-IC está conformado por la proteína Cdc45, la cual activa las subunidades de la helicasa y el complejo GINS formado por 4 subunidades [8]. Todas las proteínas mencionadas pudieron ser rastreadas, excepto las proteínas del complejo GINS. Si bien se obtuvieron algunas secuencias mediante BLASTP las filogenias no mostraron capacidad de discernir entre las diferentes secuencias y, por ende, estas no fueron completamente concluyentes.

En contraste con lo anterior, se ha mencionado en la literatura que la maquinaria de replicación y reparación del ADN en *C. parvum* se encuentra conformada por las proteínas típicas de los modelos eucariotas, sin embargo, estos hacen algunas salvedades, por ejemplo: pueden contener un número reducido de los componentes y en especial la diferencia más marcada se encuentra en las proteínas RPA, estas se encuentran en dos complejos heterotriméricos de proteína de replicación A (RPA) únicos que se expresan de manera diferencial durante el ciclo de vida del parásito y en respuesta al daño del ADN [20]. Esta

última característica se sale del alcance del análisis realizado en este estudio y hasta el momento no se ha hecho una descripción completa y detallada de todos los componentes de la maquinaria replicativa y mucho menos a nivel de los complejos de pre-RC y pre-IC en este parásito.

Al momento de analizar las filogenias, se puede apreciar en todos los árboles la formación de un linaje correspondiente al género *Cryptosporidium* que contiene 3 nodos internos con soportes entre 92 y 100% con un predominio de 100%. Comparado con los demás linajes, dicha clada se mantiene inalterada en cuanto a su topología en los diferentes árboles construidos. Esto se relaciona con lo reportado en la literatura, donde se encuentra a *C. muris* (especie principalmente zoonótica) como especie más ancestral, seguido de *C. meleagridis* y finalmente *C. hominis* y *C. parvum* [21] para nuestro caso con las especies muestreadas. La cercanía filogenética entre *C. hominis* UDEA 01 y *C. parvum* Iowa II se debe a que poseen una identidad genómica del 96,85% [19].

Para el caso particular de las proteínas MCM2-7 (Complejo de mantenimiento de minicromosomas), es importante recalcar que fue el único complejo proteico que permitió reconstruir una filogenia donde se recopilaran las 6 subunidades sin que se presentaran conflictos entre los diferentes sub-árboles que lo conforman; esto no se pudo realizar las subunidades ORC y mucho menos con el complejo GINS. Se sabe que este hexámero se encuentra ampliamente conservado dentro de los eucariotas [22] y adicionalmente se han descrito 3 parálogos más: MCM1 el cual es un factor de transcripción que no tiene relación con los eventos replicativos del DNA, mientras que MCM8 y MCM9 se han descrito y asociado con la replicación del genoma en vertebrados, aunque estas dos subunidades se puede encontrar de forma extendida dentro los eucariotas, no se encuentran presentes en levaduras y para nuestro caso en *Cryptosporidium* [22]. Esto corrobora los hallazgos encontrados, en donde se muestran las 6 subunidades más conservadas entre los eucariotas, todos estos como genes de copia única ya que no se encontraron parálogos dentro de ninguno de los apicomplexas analizadas; y aunque es poco común en algunos supergrupos eucariotas como Opisthokonta y Excavata se han reportado parálogos de las subunidades MCM2 y MCM3 en algunos de sus miembros [22].

En lo que respecta a las proteínas que conforman el complejo de reconocimiento de origen (ORC) se conoce por estudios en eucariotas modelo que consta de 6 subunidades (ORC1-6) entre las cuales a nivel individual se han encontrado diferencias en cuanto a la composición de sus dominios, por ejemplo: ORC1-5 posee un dominio de hélice alada (WHD) y un dominio AAA+ [23][24]. Esto por mencionar solo uno de ellos. Durante la reconstrucción de las filogenias, la subunidad ORC6 fue la que causó mayor dificultad y en la topología de dicha filogenia hubo cambios de posición entre algunas especies del linaje Haemosporida y Piroplasmida, estos movimientos también se pueden observar en menor medida en las demás subunidades, exceptuando ORC1, cuya topología era la esperada para todos los grupos apicomplexas. En general, las 6 filogenias cuentan con soportes muy altos en los nodos más apicales y en los basales se ve un predominio de soportes mayores a 50. Estos cambios posicionales pudieron deberse a las diferencias intrínsecas entre las subunidades, ya que no todas cuentan con los mismos dominios y por sus funciones presentan variaciones [23] [24]

[25]. Considerando lo anterior, *Cryptosporidium* y los demás Apicomplexas conservan el complejo de reconocimiento de origen típico descrito en eucariotas modelo.

En el caso de la proteína Cdc6, se sabe que es un importante regulador de la replicación en los eucariotas, esta se une a los complejos ORC y posteriormente ayudará a establecer el complejo de pre-RC [26] [27]. En la filogenia de esta proteína se dio un hallazgo particular y fue que en la clada de *Cryptosporidium* hubo mayor cercanía filogenética entre *C. hominis* y *C. meleagridis*, desplazando a *C. parvum*. Esto puede implicar, en este caso, una mayor semejanza a nivel de secuencia proteica entre *C. hominis* y *C. meleagridis* respecto a *C. hominis* y *C. parvum*. A pesar de esto, se sigue conservando el patrón de coincidencia entre los códigos de las proteínas de *C. hominis* CHUDEA7_2310 y para *C. parvum* cgd7_2310. El resto del árbol conserva la topología esperada.

En cuanto a la proteína Cdc45, se ha descrito en todos los modelos eucariotas, y su función se basa en la activación de las helicasas (MCM2-7) en un trabajo conjunto con las proteínas del complejo GINS [28] [29]. Se conoce que esta proteína es un ortólogo de *recJ* presente en Archaea; este último junto con Eukarya son linajes antiguos relacionados evolutivamente que alguna vez tuvieron un ancestro común antes de que se diera la divergencia., que para el caso de los eucariotas solo poseen una copia de este homólogo ancestral en la forma Cdc45 [29]. En la filogenia de esta proteína se reprodujo la topología esperada y se encontró en todos los organismos analizados. Esto da indicios de lo conservada que se encuentra esta proteína en los eucariotas.

En cuanto a la proteína Cdt1, su principal función es reclutar la helicasa replicativa en los orígenes de replicación en todos los eucariotas [30]. Al observar el árbol filogenético de Cdt1 nos llamó la atención que fue la única proteína en donde se encontraron parálogos, cabe resaltar que solo se encontró en miembros del linaje Piroplasmida (*B. bovis*, *T. annulata*, *T. orientalis* y *T. equi*) y una coccidia (*T. gondii*) con soportes de 100 en cada clada donde se observó este fenómeno. Los genes parálogos suelen provenir de procesos de duplicación génica, las nuevas copias ser perdidas o mantenerse en el genoma con una dirección evolutiva diferente al gen original; a su vez se pueden generar por cruce desigual, la retroposición o duplicación cromosómica (o genómica) [31]. Para el caso de Cdt1 se sugieren dos dominios Cdt1M y Cdt1C, el primero se une a la geminina y el segundo se une a MCM2-7; esto basado en que se usó el dominio de hélice alada de las arqueas como prototipo para un sistema de replicación temprano [32]. Sin embargo, faltan estudios posteriores que analicen a profundidad este suceso. El resto de la topología se mantiene.

Para el caso de las proteínas del complejo GINS no se pudo reconstruir filogenias, si bien se encontraron algunas secuencias mediante el BLASTP y se construyeron árboles filogenéticos con estas, ninguno de los obtenidos fue concluyente, ya que, en primer lugar, no se encontró en todos los apicomplexas analizados y en segundo lugar porque las filogenias no pudieron discernir entre las 4 proteínas que forman este complejo, dando resultados no concordantes con la topología esperada. Este complejo proteico se encuentra conformado por 4 subunidades Sld5, Psf1, Psf2 y Psf3. Este tetrámero tiene un papel crucial, junto con Cdc45, en la activación de la helicasa y el posterior desenrollamiento de la doble hélice de DNA en

eucariotas [33]. Además, estudios previos han mostrado gran similitud estructural entre Sld5 con Psf1 y Psf2 con Psf3, sugiriendo así un origen evolutivo común para las cuatro [34]. Adicional a esto, se sabe que las 4 subunidades se encuentran constituidas por 2 dominios los cuales se encuentran permutados entre ellas y que confirma su semejanza estructural [33] [34]. Con base en lo anterior, una posible explicación es que, si bien se encontraron algunas secuencias compatibles con estas proteínas, los árboles filogenéticos no pudieron separarlas de forma adecuada en los diferentes linajes, al contrario, se unían formando una clada entre ellas impidiendo identificar los posibles existentes en los organismos estudiados. Sin embargo, es importante resaltar que sí se encontraron secuencias compatibles y más aún, en la base de datos CryptoDB se pueden encontrar 3 proteínas anotadas dentro del genoma de referencia *C. parvum* Iowa II: “cgd7_1920Complex protein GINS subunit domain A”, “cgd5_4173GINS complex subunit Psf1” y “cgd8_1630GINS complex subunit Sld5” [10]. Todo lo anterior confirma la existencia de las proteínas de este complejo en *Cryptosporidium*, con la salvedad de que queda por demostrar o no la existencia de la cuarta restante con nuevos estudios.

En relación con lo anterior y en un intento por tratar de identificar las secuencias encontradas correspondientes al complejo GINS, se usaron predictores de dominio como Pfam [35] y EGNORG-MAPPER [36]. Esta estrategia no fue exitosa, pues en ambos predictores el reconocimiento se encontró limitado a la asociación con proteínas del complejo GINS sin decir a que subunidad específica pertenecían las secuencias o, en su defecto, no se encontraban resultados. Incluso tomando las referentes de *S. cerevisiae*, el predictor no pudo identificar específicamente cada una de las 4 subunidades del complejo.

Finalmente, con la ayuda de NLStradamus [18] se intentó reafirmar los hallazgos encontrados en *Cryptosporidium*, partiendo del hecho de que la replicación del DNA se lleva a cabo en el núcleo, y por ende todas las proteínas analizadas en este estudio deberían ir allí para realizar su función. Los resultados mostraron señales en menos de la mitad de las secuencias, sin embargo, esto tiene sentido en la medida que los organismos estudiados no son modelos y este tipo de programas, al igual que los anteriores mencionados, se nutren de los hallazgos encontrados en estudios que no incluyen estas especies. Si bien actualmente se cuenta con muchos genomas disponibles, aún quedan investigaciones pendientes por realizar para refinar las anotaciones al caracterizar adecuadamente las proteínas de los Apicomplexa.

Conclusiones

Filogenéticamente hablando se confirmó que los complejos de pre-RC y pre-IC en *Cryptosporidium* se encuentran conformados por las mismas proteínas que constituyen estos mismos complejos en eucariotas modelo, a pesar de esto, aun se requieren nuevos estudios para caracterizar, detallar y anotar adecuadamente estas proteínas para tener un mayor entendimiento de los procesos moleculares y sus implicaciones en la biología de estos parásitos.

Conflicto de intereses

Los autores declaran que no existen intereses en competencia.

Contribuciones de autor

- Daniel A. Gómez-Chavarría realizó los experimentos, analizó los datos, elaboró figuras y/o tablas, redactó o revisó borradores del documento y aprobó el borrador final.
- Juan F. Alzate concibió y diseñó los experimentos, realizó los experimentos, analizó los datos, y aprobó el borrador final.

Disponibilidad de los datos

Los genomas apicomplexas se pueden encontrar en VEuPathDB (PlasmoDB, ToxoDB, CryptoDB, PiroplasmaDB) y las proteínas de referencia de *S. cerevisiae* están disponibles en Uniprot

Bibliografía

1. Cryptosporidium spp. y Criptosporidiosis. http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1315-25562005000100003.
2. Gerace, E., Presti, V. D. M. lo & Biondo, C. Cryptosporidium Infection: Epidemiology, Pathogenesis, and Differential Diagnosis. *European Journal of Microbiology & Immunology* **9**, 119 (2019).
3. Vergara Castiblanco, C., Santos Núñez, S., Santos, F. F. & Ares Mazás. La criptosporidiosis en la región andina de Colombia: seroprevalencia y reconocimiento de antígenos. *Revista Panamericana de Salud Pública* **8**, 373–379 (2000).
4. Coutinho, B. P. *et al.* CRYPTOSPORIDIUM INFECTION CAUSES UNDERNUTRITION AND, CONVERSELY, WEANLING UNDERNUTRITION INTENSIFIES INFECTION. *The Journal of parasitology* **94**, 1225 (2008).
5. IA, K. *et al.* Morbidity, mortality, and long-term consequences associated with diarrhoea from Cryptosporidium infection in children younger than 5 years: a meta-analyses study. *The Lancet. Global health* **6**, e758–e768 (2018).
6. AL, G.-D. *et al.* Common occurrence of Cryptosporidium hominis in children attending day-care centers in Medellín, Colombia. *Parasitology research* **119**, 2935–2942 (2020).
7. Raia, P., Delarue, M. & Sauguet, L. An updated structural classification of replicative DNA polymerases. *Biochemical Society Transactions* **47**, 239–249 (2019).
8. da Silva, M. S. *et al.* Nuclear DNA Replication in Trypanosomatids: There Are No Easy Methods for Solving Difficult Problems. *Trends in Parasitology* **33**, 858–874 (2017).
9. Shrivastava, A. K., Kumar, S., Smith, W. A. & Sahu, P. S. Revisiting the global problem of cryptosporidiosis and recommendations. *Tropical Parasitology* **7**, 8 (2017).

10. Aurrecochea, C. *et al.* EuPathDB: the eukaryotic pathogen genomics database resource. *Nucleic Acids Research* **45**, D581 (2017).
11. JP, I. & JF, A. Genome microsatellite diversity within the Apicomplexa phylum. *Molecular genetics and genomics : MGG* **291**, 2117–2129 (2016).
12. J, V., Z, F., I, H., S, K. & M, O. Ploidy and Number of Chromosomes in the Alveolate Alga *Chromera velia*. *Protist* **169**, 53–63 (2018).
13. UniProt: the universal protein knowledgebase in 2021. *Nucleic acids research* **49**, D480–D489 (2021).
14. SR, E. *et al.* The reference genome sequence of *Saccharomyces cerevisiae*: then and now. *G3 (Bethesda, Md.)* **4**, 389–398 (2014).
15. J, R., S, L., KM, A., DM, S. & K, K. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic acids research* **47**, W5–W10 (2019).
16. BQ, M. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular biology and evolution* **37**, 1530–1534 (2020).
17. FigTree. <http://tree.bio.ed.ac.uk/software/figtree/>.
18. AN, N. B., A, P., N, P. & AM, M. NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC bioinformatics* **10**, (2009).
19. Arias-Agudelo, L. M., Garcia-Montoya, G., Cabarcas, F., Galvan-Diaz, A. L. & Alzate, J. F. Comparative genomic analysis of the principal *Cryptosporidium* species that infect humans. *PeerJ* **8**, (2020).
20. Rider, S. D., Jr. & Zhu, G. *Cryptosporidium*: Genomic and Biochemical Features. *Experimental parasitology* **124**, 2 (2010).
21. Xu, Z., Guo, Y., Roellig, D. M., 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** **20**, 1–15 (2019).
22. Y, L., TA, R. & SJ, A. Ancient diversification of eukaryotic MCM DNA replication proteins. *BMC evolutionary biology* **9**, (2009).
23. Jaremko, M. J., On, K. F., Thomas, D. R., Stillman, B. & Joshua-Tor, L. The dynamic nature of the human origin recognition complex revealed through five cryoem structures. *eLife* **9**, 1–19 (2020).
24. E, O.-P. *et al.* Origin Recognition Complex (ORC) Evolution Is Influenced by Global Gene Duplication/Loss Patterns in Eukaryotic Genomes. *Genome biology and evolution* **12**, 3878–3889 (2020).
25. BP, D., IN, C. & BJ, M. The origin recognition complex protein family. *Genome biology* **10**, 214 (2009).

26. Feng, X. *et al.* The structure of ORC–Cdc6 on an origin DNA reveals the mechanism of ORC activation by the replication initiator Cdc6. *Nature Communications* 2021 12:1 **12**, 1–12 (2021).
27. Borlado, L. R. & Méndez, J. CDC6: from DNA replication to cell cycle checkpoints and oncogenesis. *Carcinogenesis* **29**, 237–243 (2008).
28. T, P. *et al.* Cdc45 (cell division cycle protein 45) guards the gate of the Eukaryote Replisome helicase stabilizing leading strand engagement. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E249–E258 (2015).
29. L, S.-P. & CP, P. Cdc45: the missing RecJ ortholog in eukaryotes? *Bioinformatics (Oxford, England)* **27**, 1885–1888 (2011).
30. Pozo, P. N. & Cook, J. G. Regulation and Function of Cdt1; A Key Factor in Cell Proliferation and Genome Stability. *Genes* **8**, 2 (2017).
31. Zhang, J. Evolution by gene duplication: an update. *Trends in Ecology & Evolution* **18**, 292–298 (2003).
32. Khayrutdinov, B. I. *et al.* Structure of the Cdt1 C-terminal domain: Conservation of the winged helix fold in replication licensing factors. *Protein Science : A Publication of the Protein Society* **18**, 2252 (2009).
33. Carroni, M. *et al.* New insights into the GINS complex explain the controversy between existing structural models. *Scientific Reports* 2017 7:1 **7**, 1–7 (2017).
34. Choi, J. M., Lim, H. S., Kim, J. J., Song, O.-K. & Cho, Y. Crystal structure of the human GINS complex. *Genes & Development* **21**, 1316 (2007).
35. J, M. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Research* **49**, D412–D419 (2021).
36. Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* **47**, D309 (2019).

Anexos

Figuras complementarias

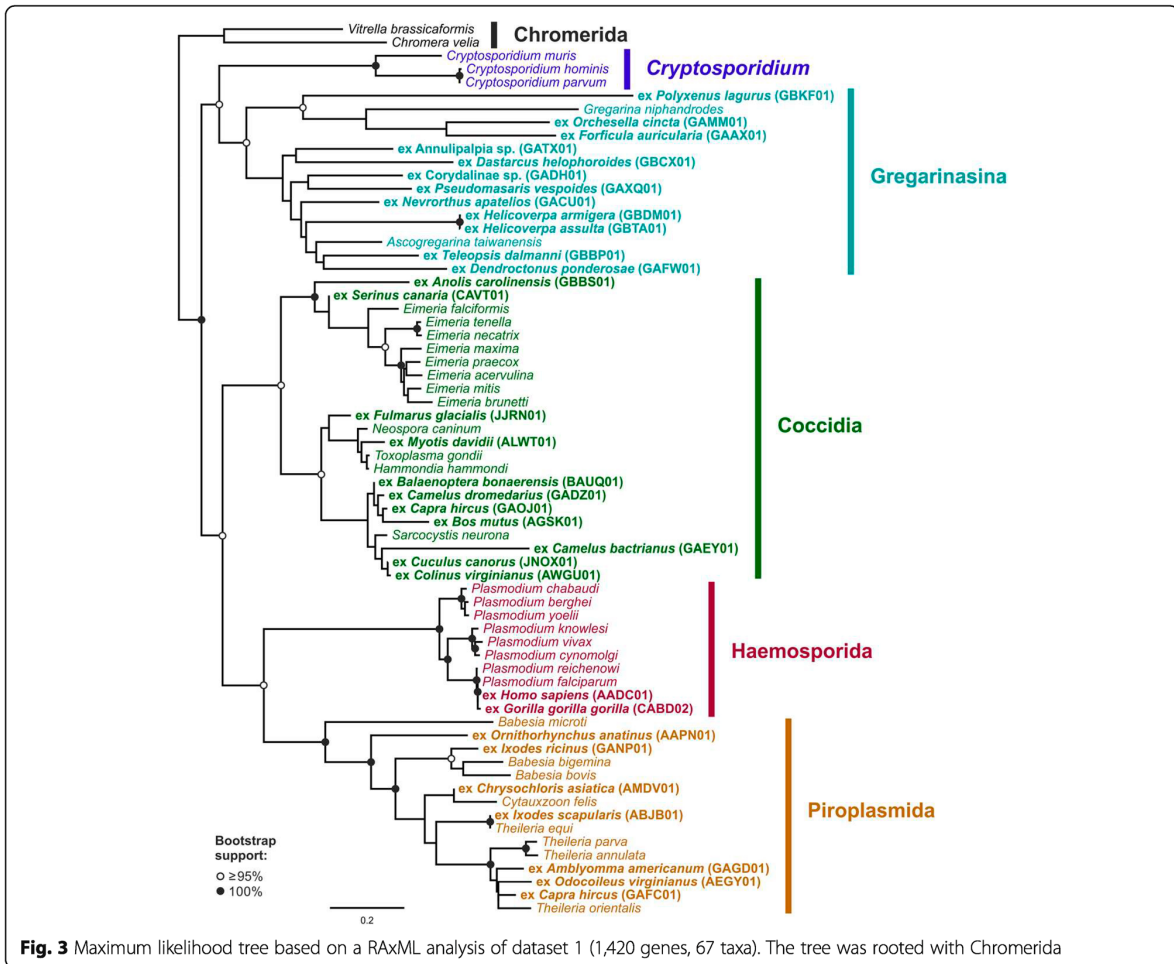


Fig. 3 Maximum likelihood tree based on a RAxML analysis of dataset 1 (1,420 genes, 67 taxa). The tree was rooted with Chromerida

Figura complementaria 1. Árbol filogenético de máxima verosimilitud que reconstruye el filo apicomplexa.

Tomado de: Borner, J. & Burmester, T. Parasite infection of public databases: a data mining approach to identify apicomplexan contaminations in animal genome and transcriptome assemblies. BMC Genomics 2017 18:1 18, 1–12 (2017).

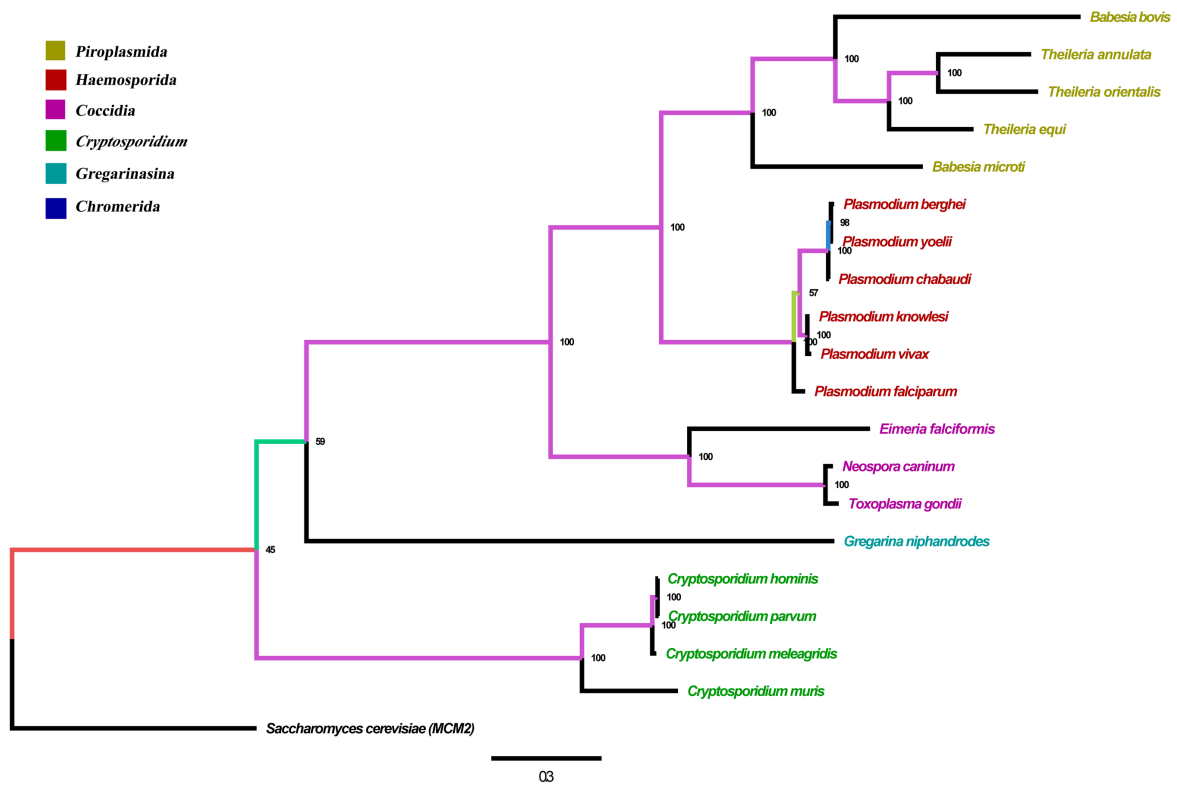


Figura complementaria 2. Árbol filogenético de máxima verosimilitud de la subunidad Mcm2

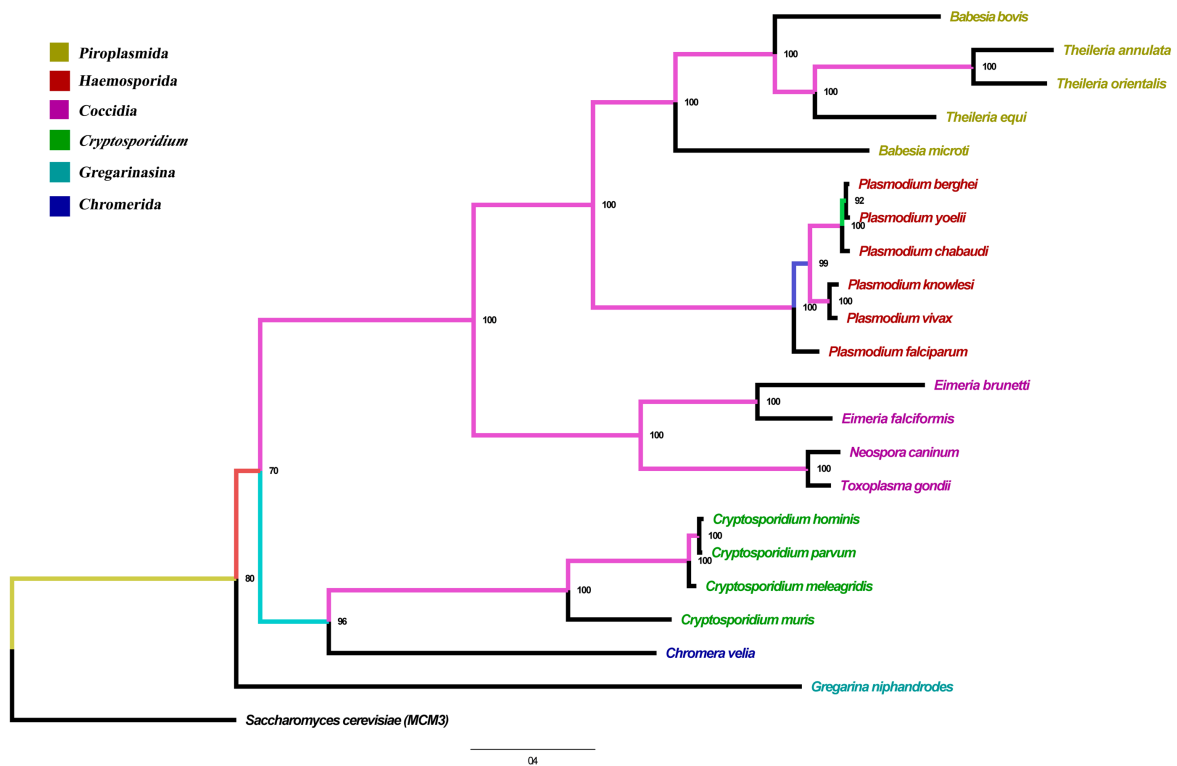


Figura complementaria 3. Árbol filogenético de máxima verosimilitud de la subunidad Mcm3

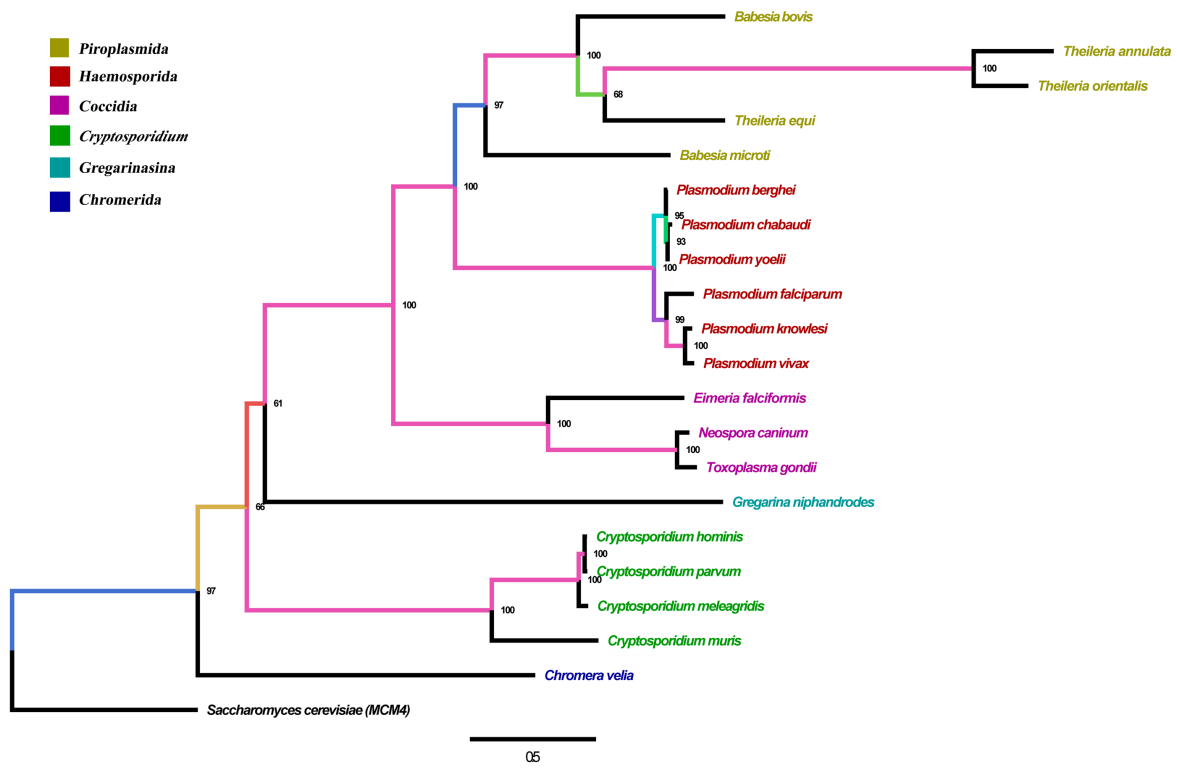


Figura complementaria 4. Árbol filogenético de máxima verosimilitud de la subunidad Mcm4

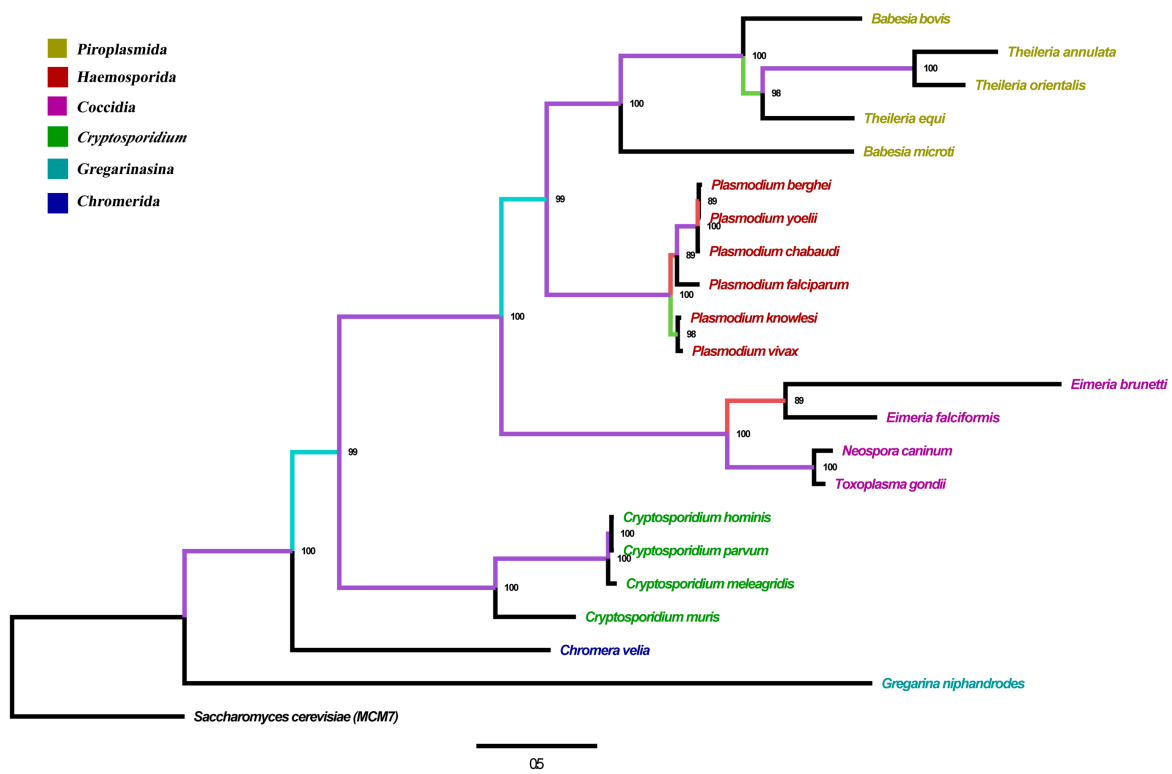


Figura complementaria 7. Árbol filogenético de máxima verosimilitud de la subunidad Mcm7