

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
FACULTAD DE CIENCIAS



**ESTIMACIÓN DEL TIEMPO DE ORIGEN DEL CICLO ACTUAL DE
TRYPANOSOMA CRUZI: UNA APROXIMACIÓN A PARTIR DEL ANÁLISIS
EVOLUTIVO DE LA FAMILIA MASP**

T E S I S

**QUE PARA OBTENER EL GRADO DE
B I Ó L O G O**

PRESENTA

Metzli Cadena Soto

DIRECTOR DE TESIS

Dr. Carlos Alberto Flores López

Ensenada, B.C.

Noviembre de 2023



UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
FACULTAD DE CIENCIAS,
CAMPUS ENSENADA.



**“ESTIMACIÓN DEL TIEMPO DE ORIGEN DEL CICLO ACTUAL DE *TRYPANOSOMA CRUZI*:
UNA APROXIMACIÓN A PARTIR DEL ANÁLISIS EVOLUTIVO DE LA FAMILIA MASP”**

TESIS

PARA CUBRIR LOS REQUISITOS NECESARIOS PARA OBTENER EL TÍTULO DE

Biólogo

PRESENTA

**Metzli Cadena Soto
354247**

A quien el Comité de Tesis autoriza el trabajo terminal y de acuerdo con el
Art. 19 del R.G.E.P.E.P, emite los siguientes votos aprobatorios mediante rubrica:

Carlos Alberto Flores López
DIRECTOR

Manuel Alejandro Carballo Amador
SINODAL

Andrés Martínez Aquino
SINODAL

“Por la Realización Plena del Ser”

RESUMEN

Se propone una estimación del tiempo en el que evolucionó el ciclo de vida actual de *Trypanosoma cruzi*, parásito causante de la enfermedad de Chagas. Estimación basada en el análisis evolutivo de una familia de proteínas de superficie asociadas a mucina (MASPs, por sus siglas en inglés). Las MASPs solamente están presentes en *T. cruzi* y debido a la alta diversificación genética de las mismas y la gran cantidad de copias presentes en el genoma, se especula que juegan un papel importante en la defensa y escape del sistema inmune del hospedero y vector, así como una posible adaptación ligada al ciclo actual del parásito. El tiempo de origen del ciclo se puede estimar de forma indirecta a partir del estudio del tiempo en que inició la diversificación de las MASPs, basado en los fundamentos del reloj molecular. Utilizando la minería de genomas de diversos linajes de *T. cruzi* se alinearon las secuencias parálogas de nucleótidos de los genes de la familia de MASP. Se estimó la tasa de sustituciones sinónimas (Ks) entre secuencias génicas y se calculó el tiempo de divergencia de las mismas en los linajes genéticos analizados. Como resultado se encontró que la diversificación de las MASPs se llevó a cabo a partir de dos eventos evolutivos en *T. cruzi*, que ocurrieron hace aproximadamente 200,000 y 2,150,000 años. Estas estimaciones concuerdan con las fechas propuestas por otros autores sobre el tiempo de la ancestría de *T. cruzi* a partir de la estimación de la divergencia de genes ortólogos. En el presente estudio se aborda una aproximación al tiempo en el que *T. cruzi* comenzó a adaptarse a su ciclo actual. Hasta donde sabemos, este es el primer estudio en utilizar la evolución de una familia de genes como base para la estimación del tiempo de origen del parásito.

Palabras clave: *Trypanosoma cruzi*, Evolución, MASPs (proteínas de superficie asociadas a mucina), Divergencia.

ABSTRACT

We propose an estimation of the timeframe during which the current life cycle of *Trypanosoma cruzi*, the parasite responsible for Chagas disease, evolved. The estimation is based on the evolutionary analysis of a family of mucin-associated surface proteins (MASPs). MASPs are exclusively present in *T. cruzi*, and due to their high genetic diversification and the large number of copies within the genome, they are speculated to play a significant role in defense and evasion of the host and vector immune systems, as well as potential adaptation linked to the parasite's current cycle. The origin of the cycle can be indirectly estimated by studying the time of MASP diversification initiation, based on the principles of the molecular clock. By mining genomes from various *T. cruzi* lineages, the nucleotide paralog sequences of the MASP gene family were aligned. The synonymous substitution rate (Ks) between gene sequences was estimated, also the divergence time of these sequences in the genetic lineages analyzed was calculated. As a result, it was found that the diversification of MASPs occurred through two evolutionary events in *T. cruzi*, taking place approximately 200,000 and 2,150,000 years ago. These estimates correspond with dates proposed by other authors regarding the ancestry time of *T. cruzi* based on orthologous gene divergence. This study provides an approach to the time when *T. cruzi* began to adapt to its current cycle. To our knowledge, this is the first study to utilize the evolution of a gene family as a foundation for estimating the parasite's origin time.

Keywords: *Trypanosoma cruzi*, Evolution, MASPs (mucin-associated surface proteins), Divergence.

AGRADECIMIENTOS

A mi familia, agradezco a mis padres, que me enseñaron con su ejemplo el valor del esfuerzo y la perseverancia, gracias por apoyarme y darme todo su amor a lo largo de mi vida. También agradezco a mis hermanos que, aunque no siempre me entiendan, se toman el tiempo para escucharme. Gracias por los momentos que compartimos, los cuales me hacen olvidar mis preocupaciones.

Gracias a mis amigos que ahora también serán mis colegas biólogos, ha sido un gusto coincidir con ustedes. Hicieron que mis días siendo universitaria fueran mil veces mejor, agradezco que siempre me sentí comprendida y querida por ustedes. Espero que la amistad que formamos gracias a la Biología se prolongue para toda la vida.

A los profesores de la carrera que mostraron su pasión por la biología y daban las clases con mucho entusiasmo, hacían que las horas pasaran volando y me inspiraron a continuar aprendiendo de todas las áreas; en especial al Dr. Carlos Flores, que gracias a su paciencia y dedicación al enseñarnos me inspiró a querer conocer más de la genética y evolución. Agradezco la oportunidad de ser su tesista, pero sobre todo su gran compromiso y el apoyo que me dio a lo largo de este trabajo.

A la Universidad Autónoma de Baja California, que desde el momento en el que fui admitida me otorgó todas las herramientas para mi formación profesional, desde las instalaciones para hacer nuestras prácticas de laboratorio como el apoyo para nuestras salidas de campo; sin duda siempre estaré orgullosa de ser cimarrona egresada de la Licenciatura en Biología.

Por último, quisiera agradecer a Mario, a quien conocí al entrar a la universidad y se volvió un pilar en mi vida. Gracias por el equipo que hemos formado a lo largo de estos años; crecer contigo en todos los aspectos ha sido lo mejor de este viaje. Gracias por darme ánimos cuando más lo necesito, por enseñarme a creer en mí y, sobre todo, gracias por tu amor incondicional.

ÍNDICE DE CONTENIDO

RESUMEN	III
ABSTRACT	IV
AGRADECIMIENTOS	V
INTRODUCCIÓN	1
ANTECEDENTES	5
HIPÓTESIS	9
OBJETIVO GENERAL	9
OBJETIVOS ESPECÍFICOS	9
METODOLOGÍA	10
Búsqueda y descripción de genomas	10
Identificación de las proteínas MASP en <i>Trypanosoma cruzi</i>	12
Estimación de la tasa de sustitución sinónima	14
Búsqueda de tiempos de replicación de <i>Trypanosoma cruzi</i>	17
Estimación de tiempos de divergencia	18
RESULTADOS	20
Análisis de los alineamientos	20
Análisis de datos de sustituciones sinónimas (Ks)	22
Estimación de tiempos de divergencia	24
DISCUSIÓN	26
Estimación de tiempos de divergencia de las MASPs	26
Limitantes de las pruebas de análisis teórico-bioinformático	26
Calidad de ensamblaje de genomas de <i>Trypanosoma cruzi</i>	29
Evolución del ciclo actual de <i>Trypanosoma cruzi</i>	30
CONCLUSIÓN	33
BIBLIOGRAFÍA	35

ÍNDICE DE TABLAS

Tabla I. Número de copias parálogas de genes que codifican proteínas expresadas en la superficie de <i>Trypanosoma cruzi</i> reportadas en los genomas de diversas especies de kinetopláستidos, incluyendo diversos linajes de <i>T. cruzi</i>	4
Tabla II. Recopilación de características de linajes de <i>T. cruzi</i>	13
Tabla III. Datos de sustituciones sinónimas (Ks) obtenidos posterior al uso de CODEML.....	17
Tabla IV. Resumen de tiempos de replicación de <i>T. cruzi</i>	18
Tabla V. Resumen de las características de los alineamientos	20

ÍNDICE DE FIGURAS

Figura 1. Ciclo de vida <i>Trypanosoma cruzi</i>	3
Figura 2. Distribuciones teóricas de edades de pares de genes duplicados en un genoma	8
Figura 3. Fragmento que representa una región mayormente conservada en el linaje TCC vista como aminoácidos en Seaview.....	21
Figura 4. Fragmento que representa una región divergente en el linaje TCC vistas como aminoácidos en Seaview	21
Figura 5. Región conservada en linaje TCC obtenidas a partir del uso de Gblocks .	22
Figura 6. Distribución de la tasa de sustitución sinónima (Ks) estimada entre las copias parálogas de MASP de cada linaje	23
Figura 7. Rango de tiempos de divergencia estimados para linajes de <i>T. cruzi</i>	25

INTRODUCCIÓN

La enfermedad de Chagas causada por el parásito protozoario *Trypanosoma cruzi* en humanos fue descrita en Brasil en 1909 por el médico Carlos Chagas (Chagas, 1909). Esta enfermedad es endémica del continente americano, encontrándose en 21 países (OMS, 2021). De acuerdo con la Organización Panamericana de la Salud esta enfermedad causa en promedio 12,000 muertes al año en América (OPS, s/f). El parásito causante es parte del género *Trypanosoma* el cual pertenece a la clase Kinetoplastea (a los miembros de esta clase se les conoce como kinetoplastidos), los cuales son protozoarios flagelados que se caracterizan por la presencia de una gran mitocondria que se distribuye a lo largo del parásito. Dentro de esta particular mitocondria se encuentra una región llamada kinetoplasto; la cual contiene alrededor del 25% del total del ADN del parásito (Silber, 2000; Stuart et al., 2008). Actualmente no hay vacuna contra esta enfermedad, sin embargo, puede tratarse con benznidazol o nifurtimox, los cuales son medicamentos que pueden eliminar al parásito si son suministrados al inicio de la etapa aguda de la infección; no obstante, conforme transcurre la infección la eficacia del tratamiento disminuye (OMS, 2021). En México se estima que hay 4 millones de personas infectadas con el parásito *T. cruzi* (Arnal et al., 2019). Otros autores estiman que 88% de la población mexicana habita en áreas con potencial riesgo de contraer la infección (Ramsey et al., 2015).

La gran mayoría de las especies del género *Trypanosoma* (a excepción de *Trypanosoma equiperdum*) tienen un ciclo de vida heteroxeno, es decir, requieren de dos tipos de hospedero para completar su ciclo de vida; donde la mitad del ciclo se lleva a cabo en el hospedero vertebrado y la otra mitad en el vector invertebrado (principalmente insectos) (Claes et al., 2005; Hamilton & Stevens, 2017; Kaufer et al., 2017). En el caso de la enfermedad de Chagas, se requiere la presencia de un insecto vector de la subfamilia Triatominae. La infección en mamíferos se conoce como silvática y a la fecha se ha reportado en más de 150 especies, pertenecientes a los órdenes Chiroptera, Lagomorpha, Rodentia, Primates, Carnivora, Didelphimorphia y Cingulata (Guhl, 2017; Jansen et al., 2017). No obstante, también existen otras posibles vías de infección, como la transmisión oral a través de los alimentos, por transfusiones sanguíneas, de forma congénita, por trasplante de órganos y por accidentes de laboratorio (OPS, s/f).

El ciclo natural de *T. cruzi* posee cuatro fases de desarrollo, dos que se llevan a cabo en el vector y otras dos en el mamífero (ver Figura 1). En el mamífero, comienza cuando los parásitos entran al torrente sanguíneo a través de la herida provocada por la picadura, o comúnmente por el contacto con alguna mucosa (e.g., los ojos, la boca, etc.). Al provenir del insecto, los parásitos se encuentran en la fase llamada tripomastigotes metacíclicos, los cuales se introducen en las células para diferenciarse en amastigotes y continuar el ciclo. Una vez que esto sucede inician su proceso de división binaria intracelular. Cuando las células quedan llenas de parásitos, los amastigotes se transforman en tripomastigotes sanguíneos y posteriormente la membrana celular se rompe, liberándolos al torrente sanguíneo. Estos infectan a las células vecinas o ingresan al flujo circulatorio, esparciéndose a células de diferentes órganos y tejidos. El ciclo continúa cuando un vector se alimenta de un mamífero infectado e ingiere los tripomastigotes, los cuales se transforman en epimastigotes a lo largo del recorrido por el tracto digestivo del vector. Los epimastigotes pasan al intestino medio donde se multiplican, y finalmente llegan al recto, donde se diferencian en tripomastigotes metacíclicos, que se eliminan con los excrementos del insecto repitiendo en ellos el ciclo descrito anteriormente (Brener, 1997).

La evolución del origen de dicho ciclo ha sido difícil de estudiar, debido al escaso registro fósil de estos parásitos unicelulares. Sin embargo, la importancia del estudio evolutivo de las enfermedades infecciosas tiene un papel crucial para la comprensión de los factores genéticos y ecológicos que juegan un rol en la aparición de nuevos patógenos de importancia médica. Algunos patógenos pueden dar lugar a pandemias como la que se está viviendo actualmente debido a la evolución de un patógeno humano nuevo (i.e., SARS-CoV-2) (Sánchez et al., 2022). El análisis evolutivo de la enfermedad de Chagas ha sido enfoque de diversos estudios, aunque, no fue hasta finales de la década de 1980 y principios de la de 1990 cuando las técnicas de secuenciación de ácidos nucleicos comenzaron a producir datos moleculares que permitieron responder preguntas como la de su origen evolutivo. No obstante, las estimaciones del tiempo de aparición del parásito varían dependiendo de los datos utilizados, i.e., 37 - 88 millones de años (mda) (Briones et al., 1999), 3 - 16 mda (Machado & Ayala, 2001), 55 - 70 mda (Teeling et al., 2005), 1.36 - 2.18 mda (Flores-López & Machado, 2011)

y 3.35-4 mda (Lewis et al., 2011)). Por tanto, el tiempo de origen del ciclo actual de *T. cruzi*, aún permanece siendo cuestión de debate.

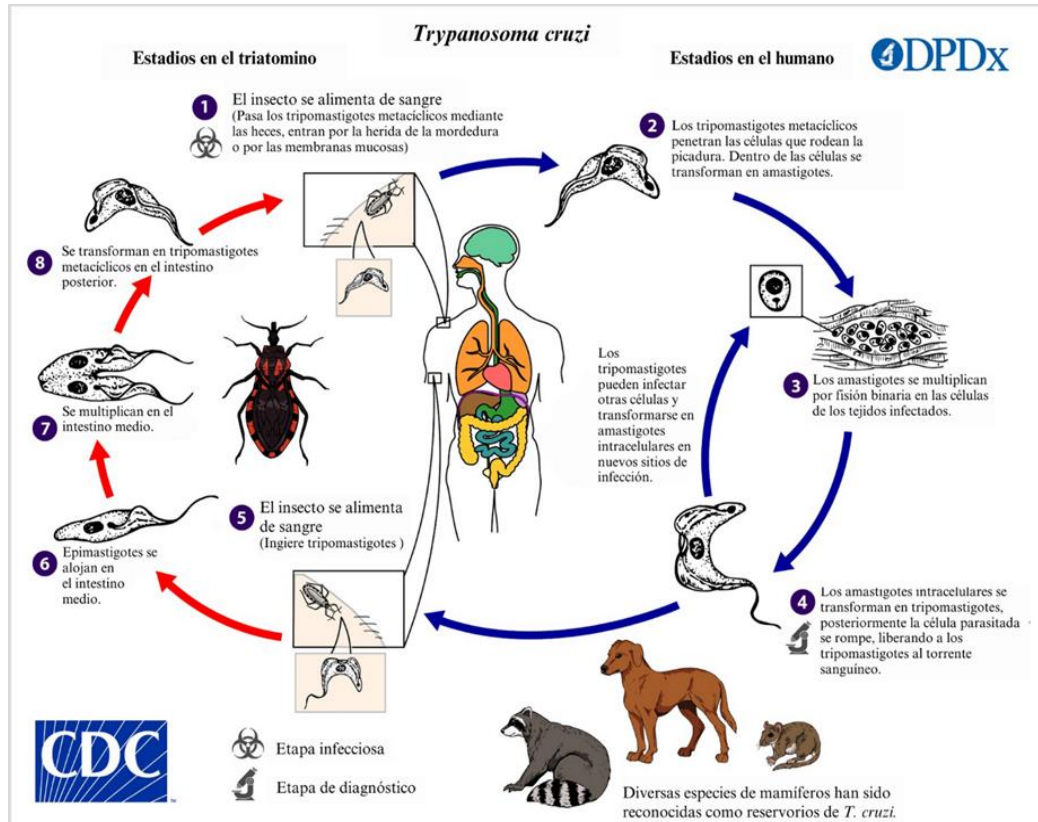


Figura 1. Ciclo de vida *Trypanosoma cruzi*. Tomado de “Centers for Disease Control and Prevention (2021)”.

A pesar de esto, la disponibilidad actual de diversos genomas de *T. cruzi* abre una gran oportunidad para explorar distintas preguntas, incluyendo la historia evolutiva del ciclo actual de *T. cruzi* (Callejas-Hernández et al., 2018a; Callejas-Hernández et al., 2018b; El-Sayed et al., 2005a; Grisard et al., 2014). Fue durante la secuenciación del genoma completo de *T. cruzi*, realizada por El-Sayed et al. (2005a) que se describió la gran diversificación genética que habían sufrido ciertas familias de proteínas que son expresadas en la membrana del parásito. Cabe señalar que estas familias ya habían sido ubicadas y descritas, aunque no se había profundizado en su análisis previamente (Buscaglia et al., 2004; El-Sayed et al., 2005a; Frasch, 2000). Estas proteínas juegan un papel importante en la defensa del parásito contra los constantes ataques del sistema inmune de las más de 150 especies de mamíferos y las aproximadamente 150 especies de vectores triatominos que actúan como hospederos

(Guhl, 2017; Jansen et al., 2017). Esto sugiere que tuvieron esta significativa diversificación durante la adaptación del ancestro de *T. cruzi* al ciclo actual del parásito (Flores-López, 2013). Dentro de estos ejemplos destaca el caso de la familia de proteínas llamada MASP (acrónimo de “Mucin-associated surface proteins”, por sus siglas en inglés); proteínas que solo han sido encontradas en *T. cruzi* (Tabla I). Esta familia de proteínas tiene su origen en un tipo de genes homólogos denominados parálogos, debido a la homología derivada a partir de un evento de duplicación génica. Tales genes pueden o no ubicarse dentro de un mismo genoma, así como tener distintas edades de duplicación (Koonin, 2005). Por lo que los datos genómicos actuales, correspondientes a las secuencias de las copias de genes MASP que se distribuyen en el genoma *T. cruzi* son una opción alternativa para estudiar la evolución del ciclo actual de *T. cruzi*. Específicamente la estimación del tiempo del comienzo de la diversificación genética de la familia de proteínas MASP.

Copias parálogas de algunas familias de proteínas de superficie

		Proteínas de superficie				
		Subfamilias de mucinas				
		TcSMUG	TcMUCI	TcMUCII	TS	MASP
<i>Trypanosoma cruzi</i>	<i>T. cruzi</i> Esmeraldo like	14	0	223	268	412
	<i>T. cruzi</i> Non-Esmeraldo like	12	0	418	272	321
	<i>T. cruzi</i> Dm28c2018	78	7	364	709	736
	<i>T. cruzi</i> Brazil A4	26	84	179	455	482
	<i>T. cruzi</i> TCC	81	41	625	688	941
	<i>T. cruzi</i> Y C6	14	0	418	616	703
	<i>T. cruzi</i> G	2	5	97	691	196
	<i>T. cruzi marinkellei</i> B7	0	0	62	841	249
	<i>T. evansi</i> STIB 805	0	0	0	6	0
<i>Trypanosoma spp.</i>	<i>T. grayi</i> ANR4	0	0	0	14	6
	<i>T. rangeli</i> SC58	0	0	1	17	0
	<i>T. theileri</i> isolate Edinburgh	0	0	2	32	5
	<i>T. vivax</i> Y486	0	0	0	6	0
	<i>T. brucei brucei</i> TREU927	0	0	0	8	0
	<i>T. congolense</i>	0	0	0	11	0

<i>Leishmania</i> spp.	<i>Leishmania major</i> strain Friedlin	0	0	0	0	0
	<i>L. braziliensis</i> MHOM/BR/75/M2904	0	0	0	0	0
	<i>L. donovani</i> BPK282A1	0	0	0	0	0
	<i>L. infantum</i> JPCM5	0	0	0	0	0

Tabla I. Número de copias parálogas de genes que codifican proteínas expresadas en la superficie de *Trypanosoma cruzi* reportadas en los genomas de diversas especies de kinetoplástidos, incluyendo diversos linajes de *T. cruzi*. Datos minados de TriTrypDB (Amos et al., 2022; Shanmugasundram et al., 2023) . MASP: Mucin-associated surface protein; TcSMUG: pequeña familia de genes tipo mucinas de *T. cruzi*; TcMUC, familia de genes de mucinas de *T. cruzi*, se clasifica en dos grupos (TcMUCI y TcMUCII) según la presencia o ausencia de repeticiones en tándem en la región central del gen; TS: Trans-Sialidase (Buscaglia et al., 2006).

ANTECEDENTES

En 2005 la revista Science dedicó un número especial a la exploración de los genomas de las tres especies de mayor importancia médica de kinetoplástidos: (a) *Leishmania major*, una de las principales especies responsables de la leishmaniasis, la cual es transmitida a humanos por moscas hembras de la subfamilia Phlebotominae y tiene una incidencia anual de 2 millones de casos, distribuidos en 88 países (Ivens et al., 2005; OMS, 2023); (b) *Trypanosoma brucei*, que provoca la enfermedad del sueño en África, la cual es transmitida al ser humano por la picadura de las moscas tsetsé (*Glossina spp.*), con una incidencia estimada que varía entre 300,000 y 500,000 casos por año (Berriman et al., 2005); (c) *Trypanosoma cruzi*, causante de la enfermedad de Chagas en el continente americano, con una incidencia anual de 30,000 casos (OPS, s/f). La exploración de genomas surgió del esfuerzo colectivo entre instituciones de diversos países: el Instituto J. Craig Venter, el Departamento de Microbiología y Medicina Tropical de la Universidad George Washington, el Instituto de Investigación Biomédica de Seattle, el Departamento de Patobiología de la Universidad de Washington, la División de Informática Biomédica y de Salud de la Universidad de Washington, el Instituto Sanger, el Centro de Genómica y Bioinformática del Instituto Karolinska, la Escuela de Biología de la Universidad de Newcastle, el Laboratorio de Genómica Funcional de Tripanosomátidos, el Departamento de Bioquímica e Inmunología de la Universidad Federal de Minas Gerais, el Laboratorio de Biología Molecular del Consejo de Investigación Médica, el Departamento de Patología de la

Universidad de Cambridge y el Departamento de Bioquímica de la Universidad de Iowa (El-Sayed et al., 2005b). Lo que abrió las puertas para mejorar nuestra comprensión de las bases genéticas y evolutivas de los ciclos de vida de estos parásitos. A partir de entonces ha ido aumentando paulatinamente el número de genomas descritos de diversos linajes de *T. cruzi* (Berná et al., 2018; Callejas-Hernández et al., 2018a; Franzén et al., 2011; Weatherly et al., 2009), las cuales se encuentran disponibles en las bases de datos del Centro Nacional de Información Biotecnológica (Sayers et al., 2022) y TriTrypDB (Amos et al., 2022; Shanmugasundram et al., 2023).

Actualmente la diversidad genética de *T. cruzi* se clasifica en un total de seis unidades discretas de tipificación (Discrete Typing Units, DTU); clasificación que está basada en la similitud genética que tienen entre si los linajes, las cuales pueden identificarse con ayuda de marcadores genéticos (Tibayrenc, 2003). De acuerdo con un consenso realizado en 2009, los DTU I y II corresponden a los dos linajes más ancestrales, mientras que los DTUs III y IV pasaron por al menos un evento de recombinación entre los primeros dos (híbridos homocigotos), por otro lado, los DTUs V y VI corresponden a híbridos heterocigotos derivados a partir de ancestros de los DTU II y III. Finalmente es importante mencionar el linaje TcBat, que en el mismo consenso de 2009 fue añadido a la clasificación como un séptimo DTU (Herrerros-Cabello et al., 2020). Este linaje fue descubierto en murciélagos de diversas regiones de América, incluyendo Brasil (*Myotis spp.* de la familia Vespertilionidae, *Noctilio albiventris* de la familia Noctylionidae, *Thyroptera tricolor* de la familia Tryropteridae, *Carollia perspicillata*, *Glossophaga soricina* y *Platyrrhinus lineatus* de la familia Phyllostomidae), Panamá (*Artibeus jamaicensis* de la familia Phyllostomidae) y Colombia (*M. oxyotus* de la familia Vespertilionidae, *Rhynchonycteris naso* de la familia Emballonuridae, *Carollia perspicillata*, *Artibeus lituratus*, *Artibeus planirostris* y *G. soricina* de la familia Phyllostomidae) (Herrerros-Cabello et al., 2020; Lima et al., 2015; Marcili et al., 2009).

Con el objetivo de coadyuvar en la comprensión del papel que juegan la gran cantidad de diferencias genéticas observadas entre DTUs, se han secuenciado una significativa cantidad de distintos linajes de *T. cruzi*. Como producto de estos proyectos de secuenciación genómica de los diversos linajes genéticos de *T. cruzi* se identificó en 2005 la presencia de una nueva

familia de genes llamada MASP (El-Sayed et al., 2005a), la cual es la segunda más grande (1377 genes y 433 pseudogenes), ocupando aproximadamente 6% del genoma (Serna et al., 2014), aunque, cabe señalar que el porcentaje genómico que ocupan las MASPs depende del linaje genético en cuestión. Las MASPs son glicoproteínas ancladas a la membrana celular, las cuales tienen extremos N- y C-terminal altamente conservados que codifican un péptido señal y un sitio de adición de anclaje GPI (glicosilfosfatidilinositol) respectivamente, además de una región central variable en cuanto a longitud y a la secuencia de aminoácidos que la componen, con una gran variedad de patrones repetitivos en su estructura (Bartholomeu et al., 2009; De Pablos & Osuna, 2012).

Diversos estudios han confirmado una gran diversidad en secuencias nucleotídicas entre las distintas copias génicas de la familia MASP. Esta diversidad sugiere que la capacidad de *T. cruzi* para invadir y multiplicarse en el citoplasma de diferentes tipos de células huésped está asociada al amplio repertorio de genes MASP (Bartholomeu et al., 2009; dos Santos et al., 2012). Posterior a la publicación del genoma completo de *T. cruzi*, investigadores han registrado que hay una mayor expresión de las MASPs durante los estadios infecciosos de *T. cruzi* (i.e. tripomastigotes que circulan por el torrente sanguíneo, así como los tripomastigotes metacíclicos) (Bartholomeu et al., 2009; De Pablos et al., 2011; dos Santos et al., 2012; Serna et al., 2014).

La significativa diversificación de MASP que está presente en los genomas de los distintos linajes genéticos de *T. cruzi* se llevó a cabo por duplicación de genes, mecanismo a nivel de evolución molecular que promueve la diversificación de cualquier familia de genes (Zhang, 2003). La duplicación de genes puede estudiarse con el objetivo de estimar el tiempo de origen de dicho fenómeno, basado en los principios del reloj molecular (Zuckerandl & Pauling, 1965). Motoo Kimura, propuso en 1968 la teoría neutral de evolución molecular, en donde explicaba que a nivel molecular (i.e., al comparar secuencias de ADN) la gran mayoría de cambios observados son neutrales con respecto a la aptitud del individuo (Kimura, 1968). Dentro de los múltiples principios de esta teoría, se encuentra el fundamento del reloj molecular (Zuckerandl & Pauling, 1965). Esto implica que al comparar la tasa de sustituciones neutrales entre un grupo de secuencias ortólogas (donde la homología es derivada de un evento de especiación) o parálogas (donde la homología es derivada de un

evento de duplicación) se pueden estimar tiempos de divergencia de una familia de genes o en su caso de un grupo taxonómico (ver Figura 2) (Faria et al., 2014; Flores-López & Machado, 2011; Puillandre et al., 2010; Viana et al., 2022; L. Zhang et al., 2020).

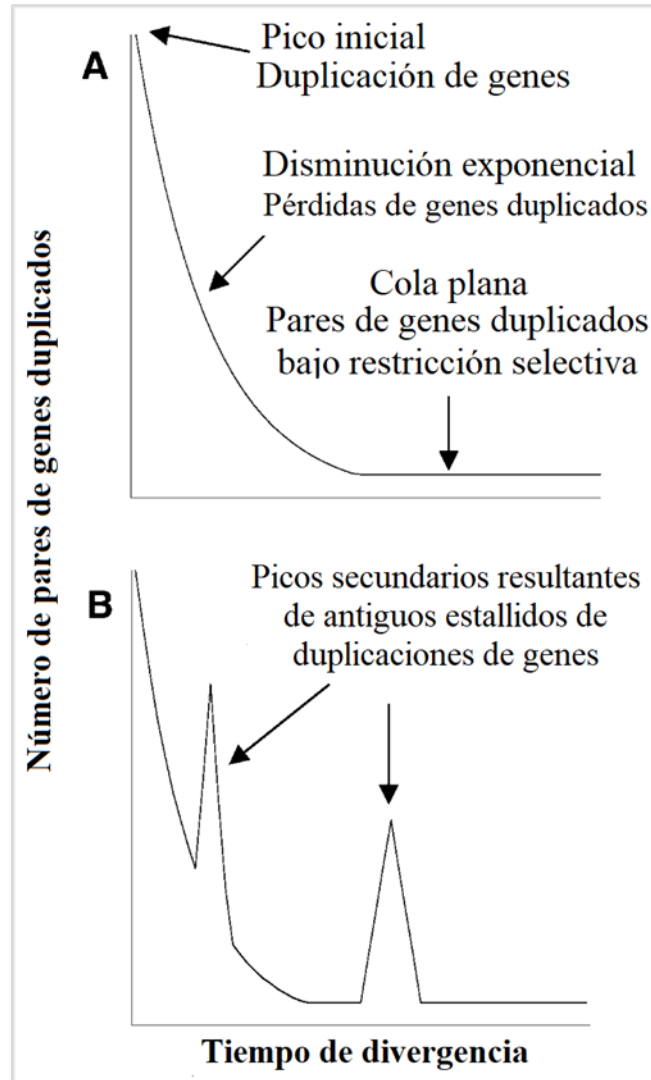


Figura 2. Distribuciones teóricas de edades de pares de genes duplicados en un genoma. Tomado de Blanc & Wolfe, 2004. (A) Distribución de edades de pares de genes parálogos esperada bajo tasas constantes de duplicación génica y eliminación de genes duplicados. Se señalan tres fases. Un pico inicial donde se ubican los genes duplicados recientemente. La disminución exponencial debido a la eliminación de genes duplicados que no están bajo restricciones selectivas. Una larga cola correspondiente a pares más antiguos de duplicados, donde ambos genes evolucionan bajo restricciones selectivas. (B) Distribución de edades de pares de genes parálogos esperada cuando ocurrieron dos eventos antiguos de duplicación a gran escala, tal suceso da lugar a dos picos secundarios.

Estos principios han sido aplicados en diversos grupos taxonómicos (Ho & Duchêne, 2014; Villarreal & Renner, 2014). No obstante, es necesario calibrar el reloj molecular para poder estimar los tiempos de divergencia de una serie de secuencias ortólogas o parálogas y para ello se cuenta con dos fuentes que brindan información sobre el pasado: la edad de fósiles y la tasa de sustitución de nucleótidos en los grupos taxonómicos de estudio (Kumar, 2005). Desafortunadamente, algunos grupos de organismos poseen un pobre registro fósil, principalmente debido a poseer tejidos con bajas probabilidades de fosilización, como es el caso de muchos microorganismos, incluyendo a los kinetoplastidos (Futuyma & Kirkpatrick, 2017). Sin embargo, se puede utilizar la tasa de mutación de nucleótidos que se ha calculado para *Trypanosoma brucei* como una aproximación a la tasa de mutación de *T. cruzi* (Lynch, 2010; Valdés et al., 1996).

HIPÓTESIS

El tiempo de origen del ciclo actual de *Trypanosoma cruzi* estimado con base en las secuencias de las proteínas MASP, arrojará tiempos de divergencia más ancestrales que los estimados previamente a partir de un enfoque filogenético, dado que estos últimos sufren del posible sesgo de no incluir linajes ancestrales de *T. cruzi* que se hayan extinguido. Sesgo que no existe al estimar tiempos de divergencia partiendo de una familia de proteínas.

OBJETIVO GENERAL

Estimar el tiempo de origen del ciclo actual de *Trypanosoma cruzi*, a partir del cálculo del tiempo de diversificación de la familia de proteínas MASP.

OBJETIVOS ESPECÍFICOS

- 1) Identificar las copias parálogas de secuencias génicas de MASP de *T. cruzi*.
- 2) Extraer y alinear las secuencias parálogas identificadas del genoma de los linajes TCC, Dm28c2018, BrazilA4, CL Brener Esmeraldo-like, CL Brener Non-Esmeraldo-Like, Sylvio-X10-2012 y G, utilizando los datos disponibles en TriTrypDB, una base de datos de libre acceso que se centra específicamente en los parásitos tripanosomátidos.
- 3) Estimar el tiempo de diversificación de las copias de MASP de forma independiente para cada linaje genético (i.e., cepas TCC, Dm28c2018, BrazilA4, CL Brener Esmeraldo-

like y CL Brener Non-Esmeraldo-Like) a partir del cálculo de la tasa de sustituciones silenciosas (Ks) y su correlación con la tasa de mutación de *T. brucei*.

4) Comparar los tiempos de divergencia estimados a partir del análisis de las proteínas MASP con las fechas propuestas previamente.

METODOLOGÍA

Búsqueda y descripción de genomas

Usando la base de datos de libre acceso TriTrypDB (<https://tritrypdb.org/tritrypdb/app>) (Amos et al., 2022; Shanmugasundram et al., 2023) se eligieron siete genomas, perteneciente a linajes de *T. cruzi*: TCC, Dm28c2018, BrazilA4, CL Brener Esmeraldo-like, CL Brener Non-Esmeraldo-like, SylvioX10-1-2012 y G. Tales genomas fueron seleccionados según su nivel de anotación, es decir, se priorizaron aquellos en los cuales se han identificado y caracterizado con precisión los genes y otras características funcionales. Esto basado en la disponibilidad de archivos fasta de los genomas anotados, tanto la versión en nucleótidos de las secuencias codificantes (CDCs, por sus siglas en inglés) como la versión traducida a aminoácidos (AA) de las mismas.

A la par de la selección de los genomas, se hizo una recopilación de las características de relevancia de cada genoma, con el propósito de conocer aspectos generales de los linajes de *T. cruzi*, objetos de este estudio. La información fue resumida en una tabla (ver Tabla II). Principalmente se centra en las características relacionadas con la secuenciación y ensamblaje de los genomas (Berná et al., 2018; Bradwell et al., 2018; El-Sayed, Myler, Bartholomeu, et al., 2005; Franzén et al., 2011, 2012; Wang et al., 2021; Weatherly et al., 2009).

De acuerdo con la información recabada, la obtención de los genomas fue realizada a partir de distintas estrategias de secuenciación: (a) SMRT, la secuenciación en tiempo real de molécula única, también conocida como SMRT por sus siglas en inglés, es una tecnología desarrollada por Pacific BioSciences. Implica la adhesión de una molécula de ADN o ARN a una polimerasa que incorpora nucleótidos marcados con fluoróforos, generando pulsos de luz en tiempo real que se traducen en una secuencia completa durante el avance de la enzima (Rhoads & Au, 2015); (b) Illumina, en este método de secuenciación, el ADN se fragmenta

en segmentos pequeños que se adhieren a una placa. A través de PCR en puente, estos fragmentos se amplifican, formando grupos de copias idénticas. Después, se introducen en la placa nucleótidos marcados con etiquetas fluorescentes distintas para cada tipo. Así la emisión de fluorescencia al unirse cada base posibilita su identificación, logrando una longitud de hasta 300 pares de bases por lectura (Rubio et al., 2020); (c) WGS, la estrategia de secuenciación completa del genoma por disparos o Whole-Genome Shotgun, implica la fragmentación aleatoria del genoma en pequeñas regiones, las cuales, posterior a la fragmentación, son clonadas en vectores para crear una biblioteca genómica. Los fragmentos de ADN clonados se secuencian individualmente utilizando el método Sanger (Sanger et al., 1977). La secuencia se ensambla mediante los solapamientos entre clones de distintas bibliotecas, esto mediante un programa de ensamblaje (Venter et al., 1998); (d) Roche 454, este método emplea secuenciación de ADN por síntesis, fragmentando el ADN en pequeños segmentos y registrando la emisión de luz al incorporar nucleótidos, permitiendo la lectura simultánea de múltiples secuencias cortas (EBI, s.f.). Es importante mencionar que la confiabilidad de los datos obtenidos puede ser evaluada con base en el nivel de cobertura de secuenciación o profundidad de secuenciación. Esto se refiere a la cantidad de lecturas de secuenciación únicas que se alinean con una región en un genoma de referencia o un ensamblaje *de novo* (Sims et al., 2014).

El nivel de ensamblaje se refiere al grado de complejidad con el cual se ha reconstruido la secuencia genómica a partir de los datos de secuenciación. En orden de jerarquía, los cóntigos (o *contigs* en inglés) son el primer nivel. Se trata de un conjunto de lecturas de ADN sin ordenar, que al superponerse parcialmente dan una representación continua de una región genómica (Alhakami et al., 2017; NIH, 2023). El siguiente nivel son los andamios, que son un conjunto de contigs unidos por espacios, de longitud conocida o estimada (McCartney et al., 2019). Finalmente, los andamios se ensamblan a nivel de cromosomas, de esta forma obtiene el tamaño de genoma, que es la longitud total del material genético contenido en el conjunto completo de cromosomas de un organismo, que se mide en unidades de pares de bases (nucleótidos) (Sessions, 2013). El ensamblaje de genomas puede evaluarse según varios parámetros. Una métrica ampliamente utilizada para evaluar la calidad del ensamblaje es el valor N50. Un indicador de la contigüidad de un ensamblaje, que puede aplicarse ya sea a contigs o andamios. Se obtiene al alinear los contigs en orden de sus longitudes (total de

bases), donde el tamaño del contig en el cual el 50% de la región de interés se encuentra en contigs será definido como el N50 (Jauhal & Newcomb, 2021). Posterior a al ensamblaje, la anotación de genes es lo siguiente, donde los genes son identificados caracterizados en términos de su función y estructura (Harbola et al., 2022). Es importante especificar que cuando se secuenció por primera vez el genoma del linaje CLBrener, se descubrió que este linaje era producto de un evento de hibridación ancestral entre ancestros de los DTUs TcII y TcIII, dando lugar a un linaje híbrido. Por lo que CLBrener Esmeraldo-like tiene su ascendencia del ancestro TcII, mientras que el linaje CLBrener Non-Esmeraldo-like tiene su ascendencia del ancestro TcIII (El-Sayed et al., 2005a; Machado & Ayala, 2001).

Identificación de las proteínas MASP en Trypanosoma cruzi

Dentro de los genomas anotados (en la modalidad nucleótidos) se procedió a localizar aquellos descritos como miembros de la familia MASP, utilizando como filtro la descripción génica “product=mucin-associated surface protein (MASP)”. Generando así un repositorio con los nombres de todos aquellos genes dentro de cada genoma que cumplan con el criterio anterior en cada uno de los linajes genéticos incluidos en este estudio.

Con el propósito de extraer las secuencias de nucleótidos de cada uno de los genes identificados previamente como pertenecientes a la familia de las MASP, se editaron los genomas de cada linaje genético usando el algoritmo BLAST+ (Altschul et al., 1990), haciendo uso de la función `./makeblastdb`, la cual crea las librerías apropiadas de cada genoma con el fin de que las diversas funciones del algoritmo BLAST puedan ser utilizadas desde la terminal de la computadora en uso. Realizado lo anterior se ejecutó el comando `./blastdbcmd`, el cual se utiliza para minar los genomas y extraer específicamente las secuencias deseadas a partir del listado de genes generado previamente. Este proceso se realizó para cada uno de los siete genomas y sus respectivas listas de genes MASP que habían sido previamente identificados.

Una vez obtenidas las secuencias de nucleótidos de los genes MASP identificados dentro de cada uno de los linajes genéticos utilizados, fueron alineadas dentro del editor de secuencias SEAVIEW (Gouy et al., 2021). Anterior a la ejecución del algoritmo para el alineamiento, las secuencias de nucleótidos fueron transformadas en aminoácidos, con el fin de disminuir la probabilidad de alinear regiones no-homólogas.

Características de linajes de *Trypanosoma cruzi*

Linaje de <i>T. cruzi</i>	Primer aislamiento	DTU	Plataforma de secuenciación	Nivel de cobertura	Nivel de ensamblaje			Tamaño (Mpb)	Genes anotados
					Cóntigos	Andamios	Cromosomas		
TCC	Humano (Argentina)	TcVI	PacBio	~60X	1,142 (N50: 264,196)	N.D.	N.D.	86.72	24,191
Dm28c2018	Humano (Argentina)	TcI	PacBio	~76X	636 (N50: 317,638)	N.D.	N.D.	53.16	15,319
BrazilA4	<i>Rattus sp.</i> (Brasil)	TcI	PacBio e Illumina	~200X	697 (N50: 191,353)	402 (N50: 914,771)	43	45.56	14,233
CLBrener Esmeraldo-like	<i>Triatoma infestans</i> (Brasil)	TcIV	Sanger	~8X	8,740 (N50: 29,950)	5,489 (N50: 151,000)	41	67.01	9,039
CLBrener Non-Esmeraldo-like	<i>Triatoma infestans</i> (Brasil)	TcIV	Sanger	~8X	8,740 (N50: 29,950)	5,489 (N50: 151,000)	41	67.06	9,385
SylvioX10-1-2012	Humano (Brasil)	TcI	Roche 454 e Illumina	Roche 454: ~9X Illumina: ~103X	7,092 (N50: 2,307)	N.D.	N.D.	38.58	10,847
G	<i>Didelphis marsupialis</i> (Brasil)	TcI	Roche 454	~37X	5,531(N50: 6,701)	1,450 (N50: 74,655)	N.D.	25.17	12,708

Tabla II. Recopilación de características de linajes de *T. cruzi*. Los datos fueron obtenidos a partir de una búsqueda bibliográfica (Berná et al., 2018; Bradwell et al., 2018; El-Sayed et al., 2005a; Franzén et al., 2011, 2012; Wang et al., 2021; Weatherly et al., 2009). El valor de N50 es equivalente a longitud (número de bases). Los datos faltantes se indican con N.D. (No Disponible). Mpb: millones de pares de bases.

Esto debido a que la probabilidad de alinear secuencias no-homólogas se disminuye al utilizar aminoácidos, por la sencilla razón de que al analizar secuencias proteínicas estas están conformadas por un total de 20 posibles variantes (i.e., 20 posibles aminoácidos), en comparación de solamente 4 posibles moléculas (i.e., 4 posibles nucleótidos) al analizar secuencias de ADN.

Utilizando los archivos de secuencias de MASP traducidos a aminoácidos, se procedió a alinear cada uno de los siete archivos. El alineamiento se realizó utilizando el algoritmo Multiple Sequence Comparison by Log-Expectation, más conocido por su acrónimo MUSCLE (Edgar, 2004). Finalmente se obtuvieron siete alineamientos con las secuencias de MASP para cada genoma de *T. cruzi* utilizado en este estudio.

Estimación de la tasa de sustitución sinónima

Para calcular la tasa de divergencia entre cada uno de los genes MASP en los genomas analizados, se realizó la estimación de la tasa de sustituciones sinónimas (Ks), las cuales corresponden a las sustituciones que no modifican el aminoácido codificado. Esto se debe a que estas sustituciones se consideran neutrales en términos de aptitud. Sin embargo, para llevar a cabo la estimación de Ks, fue necesario curar las bases de datos generadas. El aspecto principal que se requirió atención fue la veracidad de los alineamientos entre los sitios altamente divergentes de MASP.

Derivado del papel que se estima juegan las MASP en la biología de *T. cruzi* (posible rol en la evasión del sistema inmune dentro del vector y hospedero) (Bartholomeu et al., 2009), esta familia de genes ha experimentado una inmensa diversificación a nivel molecular. Misma que dificulta la veracidad en la identificación de los sitios homólogos al intentar alinear los genes. Afortunadamente las proteínas pertenecientes a la familia MASP tienen dos motivos proteínicos altamente conservados en los extremos de estas. Estos motivos proteínicos permiten identificar a las MASP como pertenecientes a esta familia. Sin embargo, muchas de las otras regiones al ser tan divergentes entre sí complican su veracidad,

Con el fin de evitar incluir sitios no-homólogos dentro de los análisis para estimar la tasa de sustitución sinónima, se debe eliminar de forma sistemática y conservadora aquellos sitios divergentes. Dicha eliminación de sitios divergentes se realizó a partir del uso del algoritmo

Gblocks (Castresana, 2000). Siguiendo el principio de reducir la probabilidad de inferir homología entre sitios a partir del alineamiento de nucleótidos versus aminoácidos, se transformaron las secuencias de las MASP a aminoácidos y se ejecutó el algoritmo Gblocks en el servidor NGPhylogeny.fr. (<https://ngphylogeny.fr/>) (Lemoine et al., 2019). Como resultado se obtuvo un alineamiento que contenía únicamente las regiones conservadas de las secuencias de MASP, las cuales variaron en longitud dependiendo del linaje de *T. cruzi* al que pertenecían.

La estimación de la tasa de divergencia neutral en las MASP, se realizó utilizando el programa CODEML dentro del software PAML (Phylogenetic Analysis by Maximum Likelihood, por sus siglas en inglés), el cual estima las tasas de sustitución sinónimas y no sinónimas entre secuencias homólogas (Yang, 2007). Específicamente para este análisis, se utilizó la tasa de sustituciones sinónimas, esto a partir de los alineamientos de las secuencias MASP que se obtuvieron a partir del uso de Gblocks. Para el linaje de *T. cruzi* TCC se obtuvieron un total de 47,586 comparaciones y por ende 47,586 estimaciones de Ks; para Dm28c2018 se obtuvieron un total de 264,628 estimaciones de Ks; para BrazilA4 88,831; para CLBrener Esmeraldo-like 64,261; para CLBrener Non-Esmeraldo-like 121,768; para el genoma del linaje SylvioX10-1-2012 2,278 y para el linaje G se obtuvieron 351 estimaciones de Ks. Los datos desglosados fueron clasificados en 40 categorías de Ks, las cuales se organizaron en un rango de 0 a >3.9, con incrementos de 0.1 entre cada categoría (Tabla III). La clasificación de los datos en las diferentes categorías se realizó mediante el cálculo de frecuencias, donde los valores más frecuentes oscilaban entre 0 – 0.4.

Frecuencia de comparaciones entre copias parálogas

Ks	Linajes de <i>T. cruzi</i>						
	TCC	Dm28c2018	BrazilA4	CLBrener Esmeraldo- like	CLBrener Non- Esmeraldo- like	SylvioX10- 1-2012	G
0-0.1	42,617	128,447	78,254	41,770	53,870	1,584	78
0.101-0.2	923	70,488	4,378	12,469	6,995	412	195
0.201-0.3	706	36,148	3,128	6,340	45,139	19	64
0.301-0.4	26	13,271	465	2,620	5,366	8	13
0.401-0.5	10	5,625	435	694	6,953	71	1
0.501-0.6	26	6,456	412	204	1,263	108	0
0.601-0.7	54	1,132	21	45	137	68	0
0.701-0.8	283	1,495	440	21	102	7	0
0.801-0.9	1,550	237	8	38	500	0	0
0.901-1	61	227	18	24	272	1	0
1.01-1.1	88	242	15	19	72	0	0
1.101-1.2	228	383	776	8	238	0	0
1.201-1.3	67	85	400	1	10	0	0
1.301-1.4	50	56	18	2	16	0	0
1.401-1.5	54	27	11	4	29	0	0
1.501-1.6	241	43	7	1	6	0	0
1.601-1.7	19	41	0	0	15	0	0
1.701-1.8	53	78	1	0	50	0	0
1.801-1.9	453	44	6	0	3	0	0
1.901-2	23	13	4	0	4	0	0
2.01-2.1	18	19	5	0	23	0	0
2.101-2.2	7	21	0	0	0	0	0
2.201-2.3	8	23	0	0	0	0	0
2.301-2.4	3	3	2	0	0	0	0
2.401-2.5	3	4	10	1	3	0	0
2.501-2.6	1	3	0	0	687	0	0
2.601-2.7	2	1	0	0	9	0	0
2.701-2.8	0	9	2	0	0	0	0
2.801-2.9	0	2	0	0	3	0	0
2.901-3	0	0	1	0	0	0	0
3.01-3.1	1	0	0	0	0	0	0
3.101-3.2	6	2	5	0	3	0	0
3.201-3.3	1	0	0	0	0	0	0
3.301-3.4	0	0	5	0	0	0	0
3.401-3.5	1	2	1	0	0	0	0
3.501-3.6	0	0	0	0	0	0	0
3.601-3.7	0	0	0	0	0	0	0
3.701-3.8	0	0	0	0	0	0	0

3.801-3.9	2	0	0	0	0	0	0
>3.9	1	1	3	0	0	0	0

Tabla III. Datos de sustituciones sinónimas (Ks) obtenidos posterior al uso de CODEML. Se observan los datos de la tasa de sustitución sinónima (Ks) calculada para cada comparación paróloga entre cada uno de los genes MASP dentro de cada linaje genético usando el programa CODEML, los números en cada columna indican la frecuencia con la que se presentaron los datos de cada categoría de Ks.

Búsqueda de tiempos de replicación de Trypanosoma cruzi

Se procedió a realizar la búsqueda bibliográfica sobre el ciclo de vida de *T. cruzi* con especial énfasis en los tiempos de replicación del parásito, los cuales permitieron estimar un promedio, y así contar con una tasa de mutación confiable para ser utilizada en la fórmula. Se eligieron seis artículos, donde los autores trabajaron con distintos linajes del parásito. Esto a partir de muestras aisladas de triatomíneos (*Rhodnius prolixus*, *Panstrongylus geniculatus* y *Triatoma brasiliensis*), marsupiales (*Didelphis albiventris*) y cultivos celulares que contenían a *T. cruzi*. Las muestras aisladas fueron cultivadas en medio LIT (Liver Infusion Tryptose o infusión de hígado y triptona), células del músculo esquelético de embrión bovino (BESM, por sus siglas en inglés), células HeLa y medio lacto (mezcla de hidrolizado de lactoalbúmina, suero y hemoglobina). El tiempo de replicación reportado en los distintos estudios fue variable entre linajes. El menor tiempo reportado fue de 7.5 horas para el linaje Limbo Tree Platform, mientras que los linajes que tuvieron un mayor tiempo de replicación fueron SN-5 y AF-1 con una duración que rondó entre 25 y 30 horas (Tabla IV). Los estadios morfológicos en los cuales se basaron las estimaciones del tiempo de replicación estudios fueron dos: epimastigotes y amastigotes. Ambos estadios morfológicos son formas replicativas del parásito, no obstante, tienen características que los diferencian. Los epimastigotes se alojan en el intestino medio del triatomino infectado, tienen forma alargada, un flagelo libre y mide entre 20 y 40 μm , mientras que los amastigotes se alojan al interior de las células del vertebrado, tienen forma redonda, están desprovistos de flagelo, y miden entre 2 y 6.5 μm de diámetro; en ambos estadios el cinetoplasto se ubica en posición anterior al núcleo (Peña-Callejas et al., 2022).

Tiempos de replicación de T. cruzi

Linaje	Tiempo (h)	Estadio morfológico	Medio de cultivo	Referencia
SN-5 ¹	25-30	Epimastigote	Medio LIT	Mejía-Jaramillo et al., 2008
AF-1 ²	25-30		Medio LIT	
M1/TcI ³	19,5	Epimastigote	Medio LIT	Araújo et al., 2007
JCPD4/TcI ⁴	9,6		Medio LIT	
Y	24	Epimastigote	Medio LIT	Elias et al., 2007
House 510	11,4	Amastigote	BESM	Luban & Dvorak, 1974
	8,6		BESM*	
Limbo Tree Platform	11,5		BESM	
	7,5		BESM*	
Ernestina	13,8	Amastigote	BESM	Dvorak & Hyde, 1973
	13,5		HeLa (Ohio)	
Y	14	N.D.	Medio LIT	Fernandes & Castellani, 1966
	24		Medio Lacto	

Tabla IV. Resumen de tiempos de replicación de *T. cruzi*. Esta tabla recopila los datos de seis artículos de investigación donde se reportan los tiempos promedios de la replicación de *T. cruzi*, así como los estadios en los que se basan las mismas. Los datos faltantes se indican con N.D. (No Disponible). ¹Linaje aislado de *Rhodnius prolixus*; ²Linaje aislado de *Panstrongylus geniculatus*; ³Linaje aislado de *Didelphis albiventris*; ⁴Linaje aislado de *Triatoma brasiliensis*.

Estimación de tiempos de divergencia

El tiempo de divergencia de las MASP se estimó dentro de cada linaje genético de *T. cruzi* usando la fórmula $t = K/(2r)$ (Goldman & Yangf, 1994; Hayashida & Miyata, 1983; Yang & Nielsen, 2000; Zuckerkandl & Pauling, 1965). En esta fórmula K es la divergencia entre las secuencias siendo comparadas, lo cual se traduce al número de sustituciones de nucleótidos entre las secuencias, que en este caso $K = Ks$ (dado a que se asume las sustituciones sinónimas son aquellas que tienden a actuar de forma neutral con respecto a la aptitud del individuo, y

por lo tanto se acumularán de forma constante, de acuerdo a los principios del reloj molecular) y r es equivalente a la tasa de mutación estimada para *T. cruzi*. Sin embargo, se carecen de estudios que hayan intentado estimar la tasa de mutación de *T. cruzi*, no obstante, se cuentan con estudios que estimaron la tasa de mutación de *T. brucei*, el cuál en términos filogenéticos nos brinda la estimación más cercana a *T. cruzi* (Stevens et al., 1999). En este caso la tasa de mutación estimada para *T. brucei* es 1.65×10^{-9} sustituciones por sitio por generación (Flores-López & Machado, 2011). A partir de este dato y posterior a una búsqueda bibliográfica del tiempo de replicación celular de *T. cruzi*, se generaron dos tasas de mutación promedio para *T. cruzi*, las cuales fueron 1.2045×10^{-6} para un ciclo de 12 horas y 6.022×10^{-7} para un ciclo de 24 horas.

Se seleccionaron de cada linaje los valores de K_s , en los que la mayoría de las comparaciones parecían mostrar el tiempo de ancestría común, esto con el fin de estimar los tiempos específicos en los que comenzó la diversificación de las MASP en *T. cruzi*. Para el linaje TCC se usaron los valores 0.9 y 1.9; para el linaje Dm28c2018 se usaron los valores 0.6 y 2.3; para el linaje BrazilA4 se usaron los valores 0.6 y 1.3; para el linaje CLBrenner Esmeraldo-like se usaron los valores 0.5 y 1.1 y para el linaje CLBrenner Non-Esmeraldo-like se usaron los valores 0.6 y 2.6. Dado que se asumió que $K = K_s$, y se estimó una tasa de mutación para *T. cruzi*, los datos obtenidos pudieron ser sustituidos por los elementos de la fórmula $t = K/(2r)$. Es así como se obtuvieron una serie de rangos de tiempos de divergencia, los cuales fueron representados de manera gráfica. Para la estimación de tiempos de divergencia se excluyeron dos de los siete linajes (SylvioX10-1-2012 y G) debido a la falta de datos. Finalmente, los tiempos de divergencia de las proteínas MASP fueron comparados con los tiempos de estimación del origen de *T. cruzi* propuestos por otros autores (Briones et al., 1999; Flores-López & Machado, 2011; Lewis et al., 2011; Machado & Ayala, 2001). Con el fin de poner en contexto las estimaciones de este estudio.

RESULTADOS

Análisis de los alineamientos

Se analizaron siete genomas pertenecientes a linajes de *T. cruzi*: TCC, Dm28c2018, BrazilA4, CL Brener Esmeraldo-like, SylvioX10-1-2012, CL Brener Non-Esmeraldo-like y G; el total de copias de genes MASP encontrados se muestran en la Tabla V. Al finalizar, los alineamientos de las MASP para cada genoma fueron reducidos drásticamente (eliminando entre 95-98% del alineamiento original).

Resumen de alineamientos de T. cruzi

Linajes de <i>T. cruzi</i>	Alineamiento inicial		Alineamiento posterior a Gblocks	
	Secuencias MASP	Longitud del alineamiento (aminoácidos)	Secuencias MASP	Longitud del alineamiento (aminoácidos)
TCC	913	2206	309	18
Dm28c2018	736	1979	728	34
BrazilA4	480	830	422	37
CLBrener Esmeraldo-like	412	1057	359	28
CLBrener Non-Esmeraldo-like	560	993	494	16
SylvioX10-1-2012	337	848	68	50
G	196	990	27	50

Tabla V. Resumen de las características de los alineamientos. Se presentan la diferencia entre el alineamiento inicial y el alineamiento posterior al uso de Gblocks, el cual elimina los sitios altamente divergentes del alineamiento, se observa una clara reducción tanto en el número de secuencias como en la longitud de los alineamientos.

Al hacer un análisis en Seaview de las secuencias MASP pertenecientes a los distintos linajes de *T. cruzi*, se observó que todas tenían una estructura similar, donde cerca de los extremos se ubicaban los sitios conservados, los cuales representan las regiones N- y C-terminal (ver Figura 3), y en la parte central las regiones divergentes, las cuales ya han sido reportadas en las MASPs (ver Figura 4). Por último, se observó el alineamiento generado a partir de Gblocks, donde en todos los casos se redujo la longitud de las secuencias (ver Figura 5).

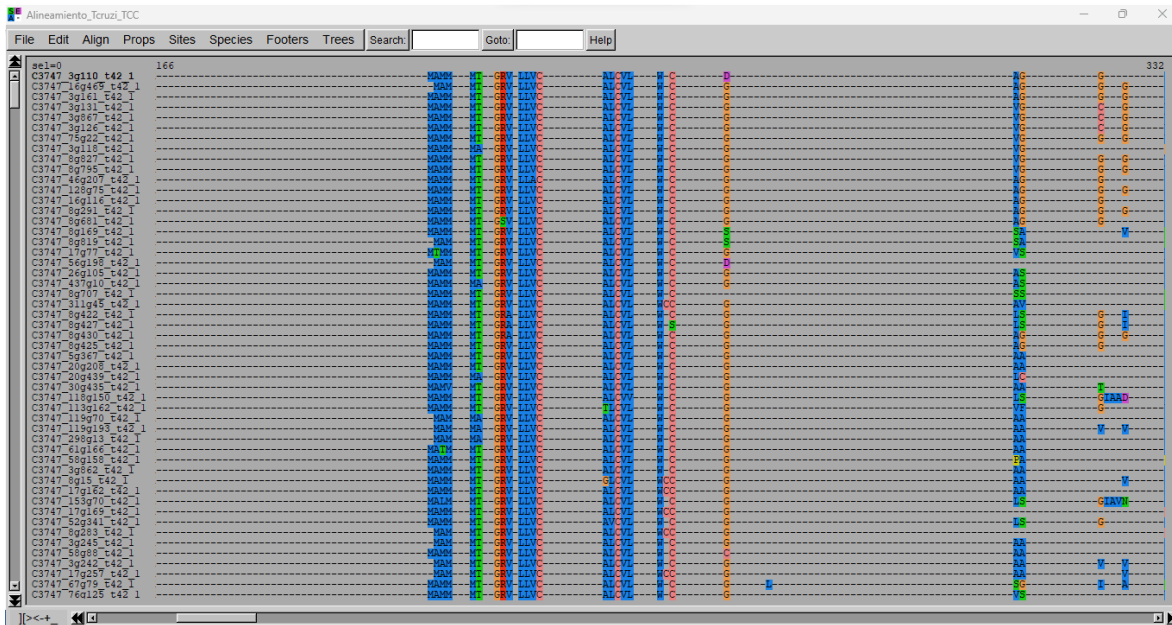


Figura 3. Fragmento que representa una región mayormente conservada en el linaje TCC vista como aminoácidos en Seaview. En la posición 211 se observa un sitio donde las secuencias tuvieron un buen alineamiento, indicando que estas regiones están altamente conservadas.

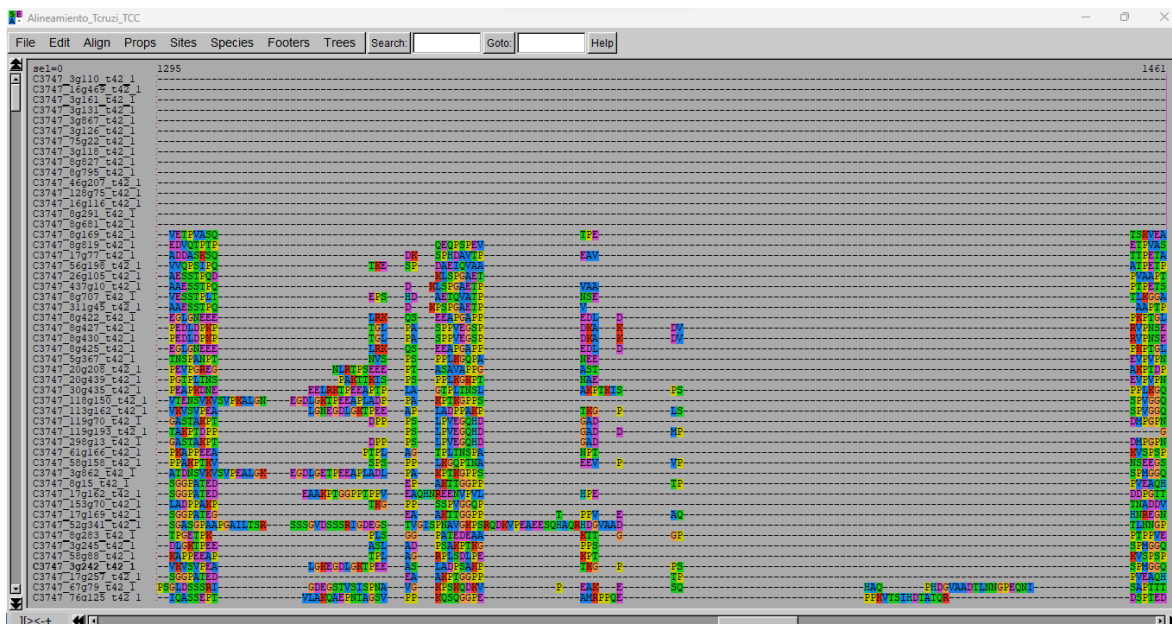


Figura 4. Fragmento que representa una región divergente en el linaje TCC vistas como aminoácidos en Seaview. En la posición 1297, cerca de la región central no se obtuvo un buen del alineamiento las secuencias, por lo que se asume que esta región es variable en las MASPs.

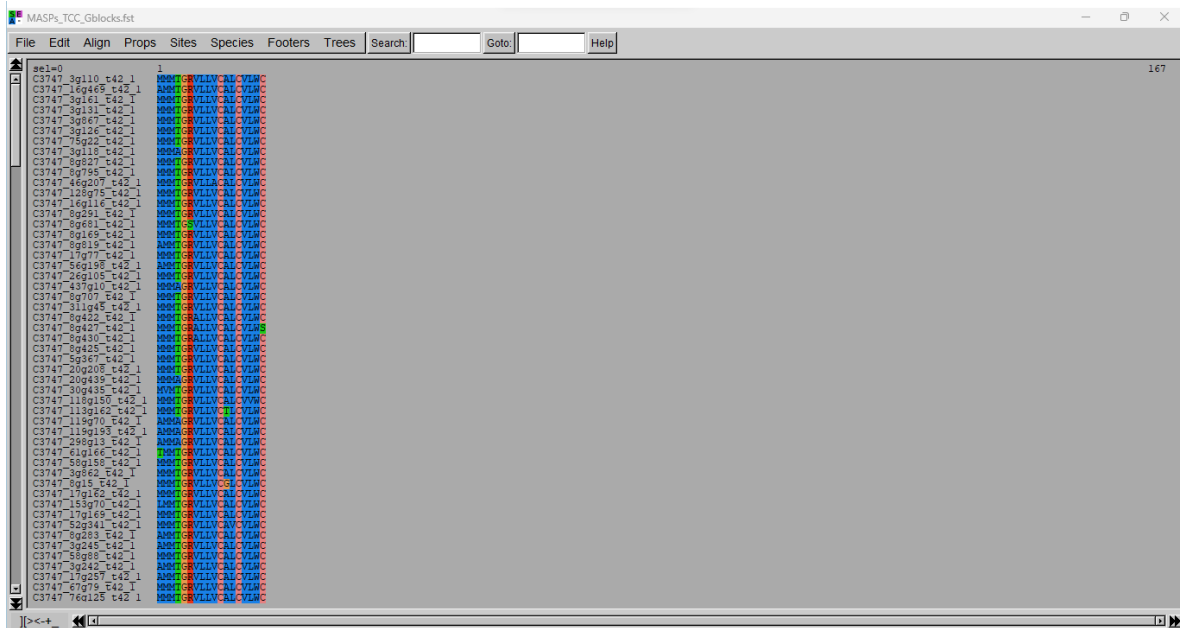


Figura 5. Región conservada en linaje TCC obtenidas a partir del uso de Gblocks. En el caso del linaje TCC la longitud de la secuencia fue de 18 aminoácidos.

Análisis de datos de sustituciones sinónimas (Ks)

Los datos de Ks obtenidos a partir de las comparaciones entre cada copia paróloga de genes MASP para cada linaje fueron graficados con el fin de poder visualizar los tiempos en los que hubo una mayor diversificación de dichos genes en el pasado (ver Figura 5). El patrón obtenido permitió observar la acumulación de una gran cantidad de genes duplicados en la categoría “Ks = 0”, indicándonos una reciente duplicación entre la gran mayoría de genes MASP, lo cual es el patrón esperado de cualquier genoma (Lynch & Conery, 2000). Esto debido a la naturaleza de las duplicaciones, en donde eventualmente, la mayoría son removidas por selección natural, por lo que como resultado de tal remoción las comparaciones entre genes parálogos suelen tener pocas sustituciones silenciosas (i.e., Ks) indicándonos que son duplicaciones de origen reciente, y que todavía no ha pasado suficiente tiempo para que selección negativa las remueva del genoma en cuestión.

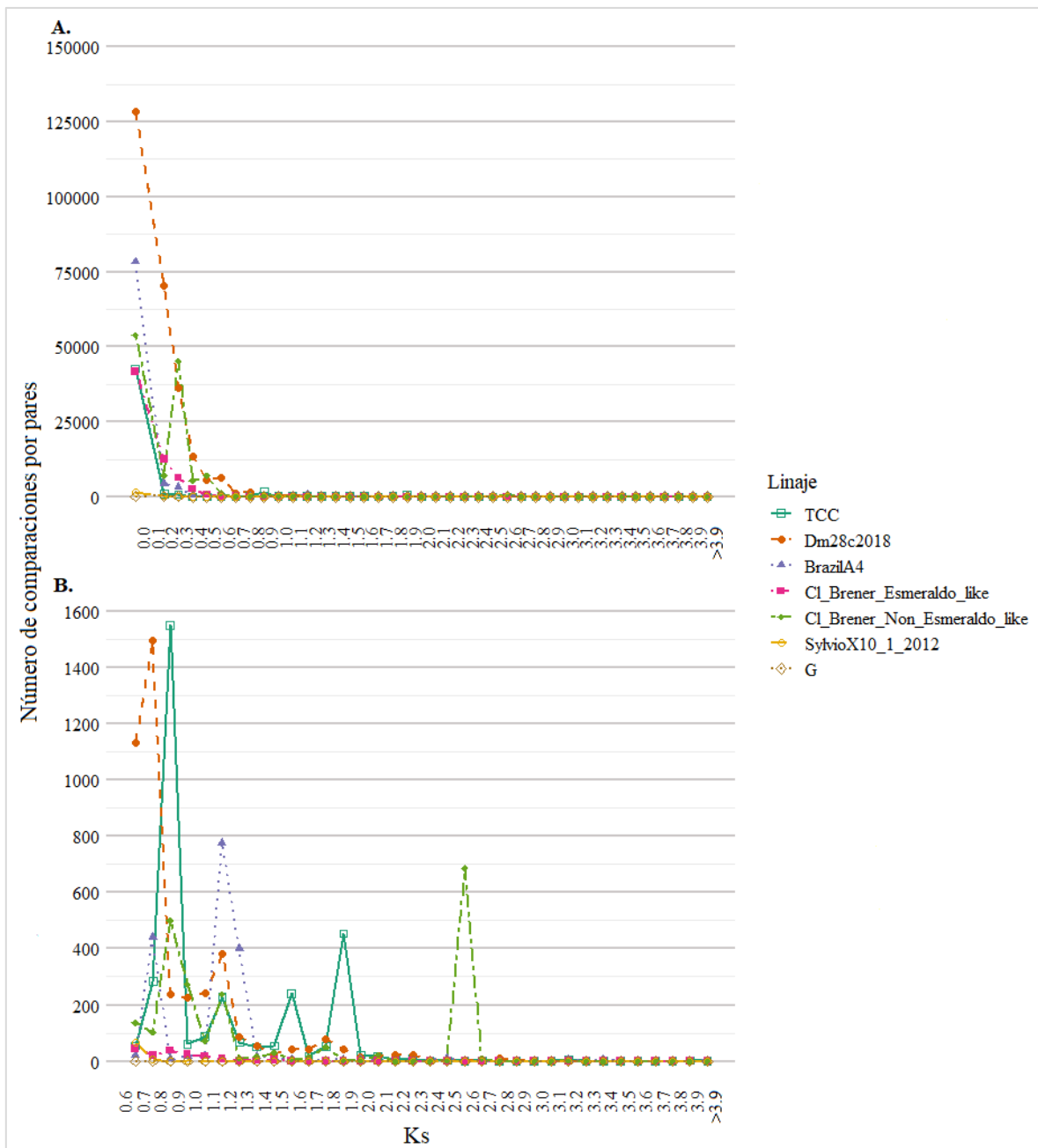


Figura 6. Distribución de la tasa de sustitución sinónima (K_s) estimada entre las copias parálogas de MASP de cada linaje. **A.** Se observa el panorama general de las comparaciones entre cada copia paróloga de genes MASP donde se observa una acumulación de genes duplicados en las primeras categorías. **B.** Acercamiento a las categorías de K_s 0.6 - >3.9, se aprecia de manera más clara los picos que representan eventos de duplicación.

Estimación de tiempos de divergencia

La estimación de los tiempos de divergencia se realizó sustituyendo v en la fórmula $t = K/(2v)$, por dos valores, correspondientes a las tasas de sustitución estimadas para *T. cruzi*, los cuales fueron 1.2045×10^{-6} para un ciclo de 12 horas y 6.022×10^{-7} para un ciclo de 24 horas. Es así como se obtuvo un rango de tiempos de divergencia (ver Figura 7).

Con respecto a la divergencia de las MASP más ancestrales, en todos los genomas analizados aparecían distintas cúspides ancestrales (ver Figura 5). Para el linaje TCC los valores de mayor divergencia fueron dos, los cuales fueron estimados con $K_s=0.9$ (donde aproximadamente 1550 comparaciones tenían una divergencia en ese rango) y $K_s=1.9$ (donde 453 comparaciones tenían una divergencia en ese rango), para el linaje Dm28c2018 los valores de mayor divergencia fueron estimados con $K_s=1.2$ (donde aproximadamente 383 comparaciones tenían una divergencia en ese rango) y $K_s=0.6$ (donde 3456 comparaciones tenían una divergencia en ese rango), para el linaje BrazilA4 los valores de mayor divergencia fueron estimados con $K_s=0.6$ (donde aproximadamente 412 comparaciones tenían una divergencia en ese rango) y $K_s=1.3$ (donde 400 comparaciones tenían una divergencia en ese rango), para el linaje CLBrenner Esmeraldo-like los valores de mayor divergencia fueron estimados con $K_s=0.5$ (donde aproximadamente 694 comparaciones tenían una divergencia en ese rango) y $K_s=1.1$ (donde 19 comparaciones tenían una divergencia en ese rango), por último, para el linaje CLBrenner Non-Esmeraldo-like los valores de mayor divergencia fueron estimados con $K_s=0.6$ (donde aproximadamente 1263 comparaciones tenían una divergencia en ese rango) y $K_s=2.6$ (donde 687 comparaciones tenían una divergencia en ese rango).

En resumen, se estimó el tiempo de divergencia considerando que hubo dos periodos de diversificación uno reciente y uno más ancestral. Los rangos de tiempos de divergencia estimados considerando una diversificación reciente fueron los siguientes para cada linaje de *T. cruzi*: 1) TCC de 373,599 – 788,709 años, 2) Dm28c2018 de 249,066 – 498,132 años, 3) BrazilA4 de 249,066 – 539,643 años, 4) CL Brener Esmeraldo-like de 207,555 – 456,621 años y CL Brener Non-Esmeraldo-like de 249,066 – 1,079,286 años. Por otro lado, los rangos de tiempo de divergencia estimados considerando una diversificación más ancestral fueron los siguientes para cada linaje: 1) TCC de 747,260 – 1,577,549 años, para Dm28c2018 fue de 498,173 – 996,347 años, para BrazilA4 fue de 498,173 – 1,079,376 años, para CL Brener

Esmeraldo-like fue de 415,144 – 913, 318 años y para CL Brener Non-Esmeraldo-like fue de 498,173 – 2,158,751 años.

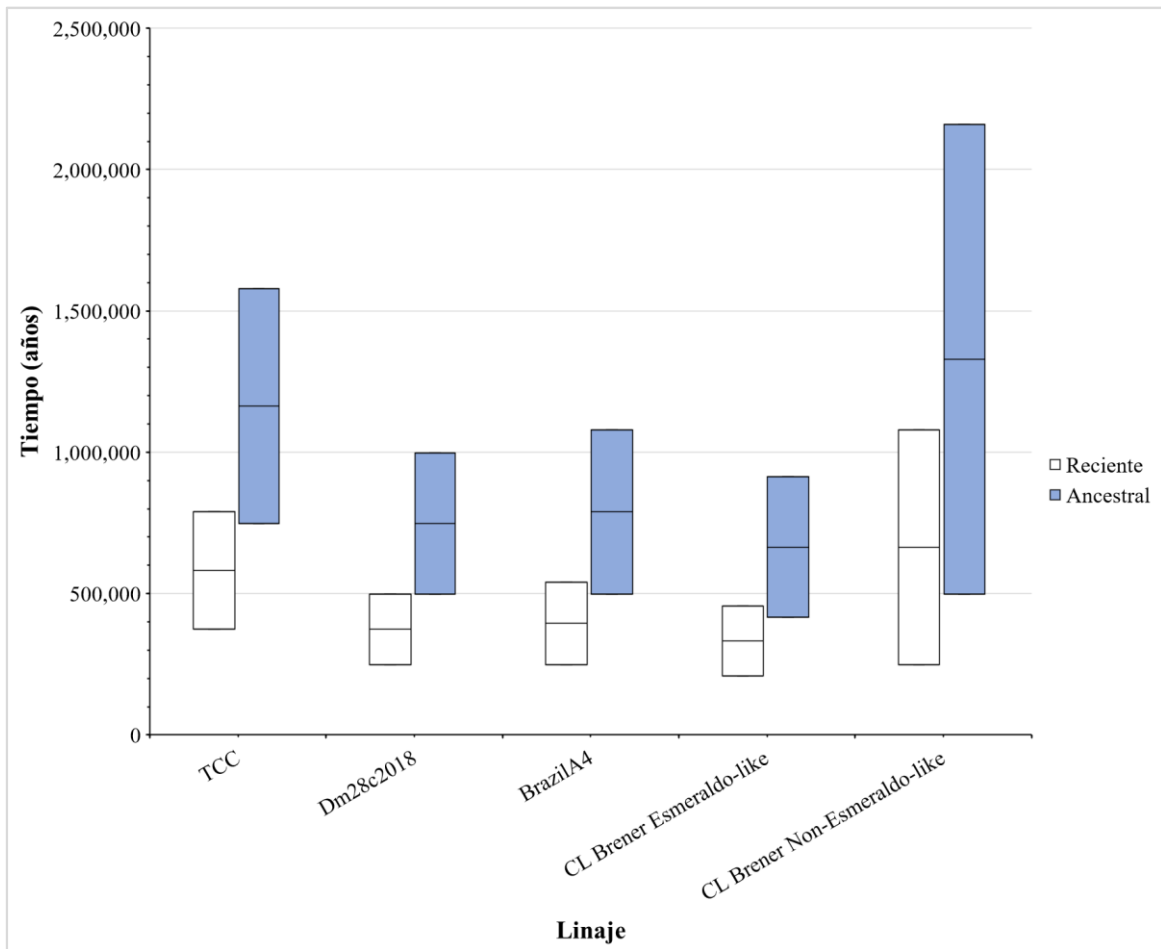


Figura 7. Rango de tiempos de divergencia estimados para linajes de *T. cruzi*. Se observa para cada linaje dos columnas correspondientes a los tiempos de divergencia calculados considerando una diversificación ancestral y reciente. Las columnas se extienden desde el tiempo mínimo hasta el tiempo máximo de divergencia, donde CLBrener Non-Esmeraldo-like y CLBrener Esmeraldo-like corresponden a los linajes con los rangos de tiempo máximos y mínimos más amplios respectivamente. Se identifican cuáles linajes tuvieron divergencias tempranas o tardías, así como las diferencias en la variabilidad temporal entre los linajes.

DISCUSIÓN

Estimación de tiempos de divergencia de las MASPs

Los tiempos de divergencia estimados fueron variables entre los linajes de *T. cruzi*. En general, el tiempo de divergencia estimado basado en el análisis bioinformático de las secuencias de las proteínas MASP a través de la fórmula $t = K/(2v)$, sugiere que ocurrieron dos grandes eventos de diversificación de las MASPs. El primero, ocurrió entre hace 415,000 a 2,150,000 años y el segundo entre 207,000 a 1,079,000. Tales eventos de diversificación deberían de ser estudiados más a fondo, para esclarecer las posibles razones que podrían estar asociadas con estos eventos (e.g., adaptación a nuevos hospederos y/o vectores).

Estos tiempos son más recientes en comparación con los tiempos de divergencia presentados por Machado & Ayala (2001), estimados entre 3 y 16 millones de años (mda) a partir de secuencias ortólogas de genes nucleares y secuencias mitocondriales (región COII-ND1, citocromo oxidasa subunidad II – NADH deshidrogenasa subunidad 1). Este estudio fue uno de los primeros en informar una edad considerablemente reciente para la evolución de *T. cruzi* en comparación con estudios anteriores (Briones et al., 1999). Briones y colaboradores estimaron tiempos de divergencia entre 37 y 88 mda a partir de secuencias de ARN ribosómico de subunidades pequeñas (SSU rRNA). Nuestros resultados concuerdan con los obtenidos por Flores-López & Machado (2011), los cuales se estimaron entre 1.36 y 2.18 mda mediante la comparación de 31 secuencias ortólogas de loci nucleares y mitocondriales. A su vez, los resultados son similares a los presentados por Lewis et al. (2011), donde el tiempo de divergencia de *T. cruzi* se estimó entre 3.35 y 4.0 mda a partir de secuencias ortólogas nucleares (glucosa-6-fosfato isomerasa, GPI) y secuencias mitocondriales (región COII-ND1, citocromo oxidasa subunidad II – NADH deshidrogenasa subunidad 1).

Limitantes de las pruebas de análisis teórico-bioinformático

Con respecto al diseño experimental utilizado en este estudio, resulta fundamental especificar los posibles sesgos del mismo. Entre ellos se encuentra la reducción de la cantidad de nucleótidos después de obtener las regiones conservadas mediante el uso del algoritmo Gblocks. Dada la amplia diversificación adaptativa que ha ocurrido en la familia de las MASP, el alineamiento correcto de regiones homólogas entre la gran cantidad de secuencias parálogas resulta difícil de obtener. Esto se evidencia cuando se utilizan dos algoritmos

distintos para el alineamiento de las secuencias, los cuales generan alineamientos distintos, a causa de la gran cantidad de regiones divergentes entre las copias génicas. En teoría, cuando las secuencias son lo suficientemente similares entre sí (es decir, con poca divergencia entre ellas), no deberían generarse alineamientos distintos entre ambos algoritmos. Sin embargo, en el caso de las MASP al comparar los alineamientos generados a partir de la utilización de los algoritmos Clustal y MUSCLE, generaron alineamientos distintos.

Debido a las discrepancias en los alineamientos generados por los algoritmos Clustal y MUSCLE, se tomó la decisión de hacer uso del algoritmo Gblocks para eliminar los sitios altamente divergentes del alineamiento. Como resultado, la cantidad de datos en el alineamiento se redujo significativamente. No hacerlo implicaría la inclusión de datos no homólogos en el alineamiento, lo cual distorsionaría los tiempos de divergencia estimados, ya que los sitios no homólogos se considerarían como sitios divergentes. Es importante mencionar que también existe la posibilidad de que, al eliminar numerosos sitios divergentes mediante Gblocks, se hayan eliminado algunos sitios que originalmente eran homólogos. Esto reduciría la cantidad de sitios homólogos a compararse, lo que podría sesgar la estimación de los tiempos evolutivos de las MASP.

Otra limitante (que no es única de nuestro diseño experimental), es el desconocimiento de la tasa de mutación de *T. cruzi*, ya que se ha visto que también afecta a otros estudios que han intentado estimar el tiempo de evolución de *T. cruzi* (Flores-López & Machado, 2011; Lewis et al., 2011; Machado & Ayala, 2001). En este estudio, se utilizó la tasa de mutación de *T. brucei* como una estimación cercana en términos filogenéticos a *T. cruzi* (Stevens et al., 1999). De cualquier modo, es importante tener en cuenta que, aunque esta estimación se asume que es muy similar a la de *T. cruzi*, ningún estudio previo se ha planteado estimar dicha tasa de mutación a partir de un diseño experimental generado para esta particular cuestión. Además, es necesario destacar que se consideraron iguales la tasa de mutación total y la tasa de mutación neutral, no obstante, esta última solo representa una fracción de todas las mutaciones que tienen lugar en el genoma.

Un factor importante a considerar es que la tasa de mutación puede variar según el tamaño efectivo poblacional. Especies con tamaños efectivos poblacionales más grandes tienden a tener tasas de mutación más altas, mientras que las especies con tamaños efectivos

poblacionales más pequeños tienden a tener tasas de mutación más bajas (Lynch & Conery, 2003). En el caso de *Trypanosoma cruzi*, aunque resulta difícil determinar el tamaño efectivo poblacional en un hospedero, este puede variar dependiendo de diversos factores, como la respuesta inmune del hospedero, la cepa involucrada, los tejidos afectados y la etapa de la infección (Messenger et al., 2015).

Otro factor importante que debe considerarse al intentar estimar el tiempo de divergencia entre linajes al utilizar secuencias de nucleótidos es la saturación de sustituciones, ya que pueden causar una subestimación en el tiempo de las divergencias estimadas. Esta saturación ocurre cuando se han producido múltiples sustituciones de nucleótidos en una región o secuencia específica, o que ocurre principalmente cuando se comparan secuencias que divergieron hace mucho tiempo (Li, 1997).

Dado que la selección natural actúa sobre el fenotipo, a nivel molecular esta selección actuaría principalmente sobre aquellos cambios que alteren la codificación de una proteína (e.g., las sustituciones no-sinónimas). Mientras que aquellas sustituciones que no alteran la codificación de un aminoácido estarían evolucionando de manera neutral, a una tasa similar a la de mutación, ya que se considera que están libres de selección. Asimismo, se asume que el número de sustituciones sinónimas (K_s) entre dos secuencias homólogas aumenta de manera lineal con el tiempo, al menos cuando la divergencia de las secuencias es relativamente baja, y no se haya llegado a un nivel de saturación de sustituciones (Li, 1997). Saturación que ocasionaría una subestimación en el tiempo de divergencia que realmente ha ocurrido entre dos secuencias homólogas. Por tanto, la selección natural a nivel molecular puede ser definida como: el proceso en el que ciertas variantes de genes o moléculas biológicas se propagan en una población debido a su ventaja funcional o adaptativa. Tales variantes tienen una mayor probabilidad de ser transmitidos a las generaciones futuras (Harris & Meyer, 2006). En lo que respecta a la selección negativa, se refiere al proceso mediante el cual se eliminan las variantes genéticas o moleculares que tienen efectos perjudiciales o desventajosos en un organismo (Loewe, 2008).

Para este estudio, se utilizaron los valores de diversificación de MASP más ancestrales, dado que estos representan el tiempo de origen de la diversificación de las mismas. Tiempo que asumimos está relacionado con el tiempo de adaptación del ciclo actual de *T. cruzi*. Los

cuales rondaron los valores de $K_s > 2.0$ (ver Figura 5). Por lo que se asume tienen cierto nivel de saturación, causando un sesgo en nuestras estimaciones hacia tiempos más recientes de los que realmente pudieran ser. Este tipo de saturación ha sido previamente reportado por algunos estudios (Blanc & Wolfe, 2004; Gojobori, 1983; Li, 1997).

Calidad de ensamblaje de genomas de Trypanosoma cruzi

Es necesario considerar la variación que pudiera existir entre los diversos genomas con respecto al número real de copias de los genes MASP presentes en cada genoma. Dado a que las secuencias fueron extraídas de una base de datos (TriTrypDB), donde se recopilaban las secuencias genómicas del grupo de los tripanosomas. Estos datos fueron generados por diferentes proyectos de investigación, lo que implica que los ensamblajes y anotaciones de genomas varían entre sí, ya que se han derivado a partir de distintas plataformas de secuenciación.

Cabe destacar que el ensamblaje de genomas es un proceso complejo que puede verse afectado por diversos errores. Por ejemplo, los genomas con una baja calidad de ensamblaje pueden tener un número menor de MASP's predichos en comparación con los que realmente existen en el genoma en cuestión. A continuación, se describen algunos de los errores comunes que pueden ocurrir durante el ensamblaje de genomas (Salzberg et al., 2012; Schirmer et al., 2016; Treangen & Salzberg, 2012).

1. Errores de secuenciación: Las tecnologías de secuenciación pueden introducir errores de lectura, especialmente en regiones repetitivas o difíciles de secuenciar; afectando la calidad de los datos y, a su vez, el ensamblaje del genoma.
2. Errores de inserción y eliminación (indels): Durante el proceso de ensamblaje, es posible que se introduzcan errores en la secuencia del genoma, ya sea la inserción o eliminación de fragmentos de ADN. Estos errores pueden ser originados debido a la fragmentación o secuenciación incorrecta de los fragmentos de ADN durante la extracción.
3. Conexiones erróneas (quimeras): Es posible que se produzcan conexiones incorrectas entre fragmentos de ADN durante el ensamblaje, dando como resultado la generación de secuencias híbridas llamadas quimeras, las cuales pueden generar información errónea sobre la estructura del genoma.

4. Problemas con repeticiones y regiones genómicas complejas: La complejidad de estas regiones dificultan el proceso de ensamblaje. Dando lugar a ensamblajes incorrectos, duplicaciones o colapsos de estas regiones repetitivas, debido a que no se pueden distinguir correctamente entre repeticiones similares o idénticas en el genoma.

Es importante considerar que las tecnologías de secuenciación y los métodos utilizados en el ensamblaje de genomas están en constante mejora, por consiguiente, se espera que con el paso del tiempo la calidad y precisión de los ensamblajes mejoren de manera significativa, lo que a su vez aumentará la confianza en los datos y estudios derivados de estos.

Evolución del ciclo actual de *Trypanosoma cruzi*

Con respecto al ciclo actual de *T. cruzi*, se han propuesto diversas hipótesis. Una de ellas es la llamada hipótesis del "supercontinente" del sur. La cual sugiere que la evolución temprana del clado *T. cruzi* estuvo posiblemente asociada a la fauna presente en el antiguo supercontinente Gondwana, específicamente en los marsupiales. Tras la separación de los continentes que conformaban Gondwana (Sudamérica, África, Antártica, Australia, India, entre otros), el clado *T. cruzi* se diversificó en los Xenarthra (cingulados y vermilinguos) y, al igual que los marsupiales, se convirtieron en los primeros hospederos de *T. cruzi*. Un ejemplo que ilustra la relación entre *T. cruzi* y los marsupiales es la interacción del parásito con las especies del género *Didelphis*. En donde las especies de este género desempeñan un doble papel, actuando simultáneamente como vectores y reservorios de *T. cruzi*, lo que sugiere un mecanismo de transmisión ancestral previo al papel de vector que posteriormente tomaron los triatomíneos dentro del ciclo de *T. cruzi* (Noireau et al., 2009; Stevens et al., 1999).

Otra hipótesis respaldada con datos moleculares en las últimas décadas es la denominada "Bat-seeding", la cual sugiere que el clado *T. cruzi* evolucionó a partir de un clado de tripanosomas que se diversificaron exclusivamente en murciélagos. Dada la capacidad de vuelo de estos mamíferos se logró una distribución transcontinental, donde se produjeron varios saltos hacia mamíferos terrestres tanto en el Nuevo Mundo como en el Viejo Mundo, dando lugar a la diversificación de este clado (Hamilton et al., 2009, 2012; Hamilton & Stevens, 2017; Lima et al., 2013, 2015; Steverding, 2014). En relación a los eventos evolutivos en el Nuevo Mundo, se estima que la expansión de murciélagos infectados con

tripanosomas hacia Sudamérica ocurrió hace aproximadamente 7-10 millones de años a través de América del Norte. Esta estimación, realizada por Stadelmann et al., 2007, se basó en el género de murciélagos *Myotis*, puesto que tuvo una de las radicaciones más diversas entre los mamíferos y representa un modelo singular para investigar la especiación y la diversificación. Las especies de *Myotis* analizadas pertenecían a los dos reinos biogeográficos del continente americano (Neártico y Neotropical); de acuerdo con datos moleculares, el tiempo de divergencia del clado *Myotis spp.* en el Nuevo Mundo fue hace aproximadamente 10-14 mda, mientras que el tiempo en el que ocurrió la diversificación del subclado compuesto por especies de la región Neotropical (Sudamérica) fue hace aproximadamente 7-10 mda (Stadelmann et al., 2007). Tal evento pudo propiciar el traslape de nichos entre murciélagos, mamíferos terrestres y vectores invertebrados, dando eventualmente origen a *T. cruzi* (Hamilton et al., 2012; Steverding, 2014). Es importante señalar que la literatura disponible acerca de la dinámica poblacional tanto de vectores, como hospederos de *T. cruzi* está orientada al ciclo de transmisión de la enfermedad de Chagas, en otras palabras, se centra en la infección en humanos, con énfasis en el ciclo doméstico (Kribs & Mitchell, 2015), por lo que es un área que aún necesita ser explorada con el fin de comprender otros factores ecológicos y biológicos relacionados con los vectores y reservorios de *T. cruzi*.

Para comprender el ciclo actual de *T. cruzi*, es indispensable hablar sobre los vectores triatominos, que desempeñan un papel crucial como vectores del parásito y son fundamentales en el ciclo de desarrollo del mismo. A diferencia de los marsupiales y murciélagos, los triatominos tienen un origen relativamente reciente, pues se cree que surgieron a partir de ancestros depredadores de origen silvático (chinchas reduvidas) que adoptaron un estilo de vida hematófago hace aproximadamente 1-5 millones de años (Gorla et al., 1997; Noireau et al., 2009; Stevens et al., 1999). Lo que descartaría la suposición hecha por algunos autores sobre la coevolución del parásito y los triatominos. Aun así, es necesario una mayor exploración acerca de cuáles fueron los factores que pudieron influir para que se diera la estrecha relación entre triatominos y *T. cruzi* (e.g. presión selectiva, radiación de triatominos, coexistencia en el mismo hábitat, entre otros).

Por último, es importante mencionar el momento en que el ser humano se incorporó a este ciclo. Se estima que los humanos estuvieron expuestos por primera vez a *T. cruzi* cuando migraron al Nuevo Mundo hace no más de 30,000 años y comenzaron a tener contacto con los vectores artrópodos del parásito. Se considera que para ese entonces ya existía un ciclo silvestre de *T. cruzi* bien establecido en América. Pero, no fue hasta el establecimiento de los primeros asentamientos permanentes y el inicio de la domesticación de plantas y animales (hace 8,000-6,000 años) que los humanos fueron introducidos en el ciclo de transmisión de *T. cruzi* (Araújo et al., 2009; Ferreira et al., 2011; Hamilton et al., 2012; Steverding, 2014).

Por otro lado, una hipótesis plantea que la adaptación de *T. cruzi* a diferentes hospedadores y vectores pudo haber sido impulsada por la diversificación de las proteínas MASP, específicas de este parásito. Estas proteínas desempeñan un papel crucial en la interacción del parásito con las células del hospedero y en su evasión del sistema inmunológico (Buscaglia et al., 2006; Fonseca et al., 2019; Solís-Oviedo et al., 2018).

La diversificación de las MASP probablemente ha permitido a *T. cruzi* interactuar de manera específica con diferentes tejidos y células de los hospederos, facilitando su colonización y supervivencia en diversos ambientes. Además, estas proteínas podrían desempeñar un papel crucial en la modulación de la respuesta inmunológica del hospedador y en la evasión de la vigilancia del sistema inmunológico. Aunque se requiere más investigación para comprender completamente el papel de las MASP en la adaptación de *T. cruzi*, su estudio representa una contribución importante para entender la evolución de este parásito y su capacidad para infectar y persistir en diferentes hospederos. Cabe destacar que cada linaje de tripanosomátidos tiene una superficie celular única, en la cual se ubican otras familias de proteínas asociadas a la interacción con el hospedero y el vector. En *Trypanosoma cruzi*, además de las MASPs se encuentra otra gran familia de proteínas, las trans-sialidasas (TS) que están involucradas en la modulación de la respuesta inmunológica, en *Trypanosoma brucei* se encuentran las glicoproteínas de superficie variante (VSGs, por sus siglas en inglés), las cuales producen un "cambio de antígeno" evitando así el reconocimiento por parte del sistema inmune del hospedero, mientras que en especies del género *Leishmania* la molécula más abundante es el lipofosfoglicano (LPG), el LPG está relacionado con el desarrollo y supervivencia en el intestino de los vectores y la modulación de la respuesta inmunológica.

Dado lo anterior, se especula que esta diversificación en las familias de genes fue clave para la adaptación a un estilo de vida parasitario por parte de los kinetoplástidos, por lo que es necesario continuar ampliando el conocimiento acerca de las características específicas que se relacionan con el parasitismo en kinetoplástidos (Finamore, 2010; Jackson et al., 2016; Manna et al., 2013; Venkatesh et al., 2018).

CONCLUSIÓN

Como resultado de esta investigación, y mediante el análisis de secuencias de proteínas MASP y el uso de los principios del reloj molecular, se reportan dos eventos de diversificación de MASP en los principales linajes de *T. cruzi*, los cuales se sitúan en un rango aproximado de 200,000 años a 2,150,000 millones de años atrás. Aunque aún no se disponen de datos concretos sobre los eventos asociados a la diversificación de las MASP, se sugiere que esta diversificación puede ser resultado tanto de eventos genéticos (presión selectiva, plasticidad genómica y recombinación), como de eventos ecológicos (adaptación a nuevos nichos ecológicos, interacciones con múltiples hospederos y cambios en la disponibilidad de recursos).

Con respecto a la hipótesis planteada, los resultados no apoyan la misma. Los tiempos de estimación a partir de la diversificación de las MASP respaldan una reciente diversificación del clado *T. cruzi*. Tiempos que son consistentes con investigaciones anteriores realizadas por Flores-López y Machado (2011), Lewis et al. (2011) y Machado y Ayala (2001) a partir del análisis de secuencias ortólogas. Lo que sugiere que la evolución del ciclo actual del parásito *T. cruzi*, no es tan ancestral como otros autores llegaron a sugerir (Clément et al., 2020; Hamilton et al., 2012; Stevens et al., 2001), cabe señalar que dichas diferencias pueden deberse a la distinta fuente de origen de datos analizadas.

Por otro lado, es oportuno mencionar que el diseño experimental de este estudio no está ausente de limitantes. Entre estas se encuentra la dificultad en la generación de alineamientos con una alta certeza de ser correcto, debido a la amplia diversificación que ha sufrido la familia de las MASP. Lo que obligó a utilizar solamente las regiones seleccionadas por un algoritmo diseñado para descartar las regiones altamente divergentes de alineamientos de nucleótidos (Gblocks), que en este caso limitó significativamente la cantidad de sitios en los alineamientos a utilizarse para las estimaciones del reloj molecular (ver Figura 5).

Adicionalmente, el desconocimiento de aspectos generales del tiempo de replicación ancestral de *T. cruzi*, y si la tasa de mutación ha sufrido cambios a través del tiempo, son limitaciones que pudieran entrar bajo cuestionamiento. Estas limitaciones representan desafíos que deben ser abordados para explorar aspectos no considerados en esta investigación.

En resumen, este estudio contribuye al conocimiento sobre la diversificación de los principales linajes de *T. cruzi*, aportando información valiosa sobre el ciclo de transmisión de este parásito. Aunque se han identificado limitaciones, los resultados obtenidos abren nuevas perspectivas para el estudio de las enfermedades zoonóticas y proporcionan una base sólida para futuras investigaciones en este campo.

BIBLIOGRAFÍA

- Alhakami, H., Mirebrahim, H., & Lonardi, S. (2017). A comparative evaluation of genome assembly reconciliation tools. *Genome Biology*, *18*(1), 1–14.
<https://doi.org/10.1186/s13059-017-1213-3>
- Altschul, S. F., Warren, G., Webb, M., Myers, E. W., & Lipman, D. J. (1990). Altschul 1990.pdf. En *Journal of molecular biology* (Vol. 215, pp. 403–410).
- Amos, B., Aurrecochea, C., Barba, M., Barreto, A., Basenko, E. Y., Bazant, W., Belnap, R., Blevins, A. S., Böhme, U., Brestelli, J., Brunk, B. P., Caddick, M., Callan, D., Campbell, L., Christensen, M. B., Christophides, G. K., Crouch, K., Davis, K., Debarry, J., ... Zheng, J. (2022). VEuPathDB: The eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Research*, *50*(D1), D898–D911.
<https://doi.org/10.1093/nar/gkab929>
- Araújo, A., Jansen, A. M., Reinhard, K., & Ferreira, L. F. (2009). Paleoparasitology of Chagas disease: a review. *Memórias do Instituto Oswaldo Cruz*, *104*(suppl 1), 9–16.
<https://doi.org/10.1590/S0074-02762009000900004>
- Arnal, A., Waleckx, E., Rico-Chávez, O., Herrera, C., & Dumonteil, E. (2019). Estimating the current burden of Chagas disease in Mexico: A systematic review and meta-analysis of epidemiological surveys from 2006 to 2017. *PLoS Neglected Tropical Diseases*, *13*(4), 1–19. <https://doi.org/10.1371/journal.pntd.0006859>
- Bartholomeu, D., Cerqueira, G., Leão, A. C., DaRocha, W., Pais, F., Macedo, C., Djikeng, A., Teixeira, S., & El-Sayed, N. (2009). Genomic organization and expression profile of the mucin-associated surface protein (masp) family of the human pathogen *Trypanosoma cruzi*. *Nucleic Acids Research*, *37*(10), 3407–3417.
<https://doi.org/10.1093/nar/gkp172>
- Berná, L., Rodríguez, M., Chiribao, M. L., Parodi-Talice, A., Pita, S., Rijo, G., Alvarez-Valin, F., & Robello, C. (2018). Expanding an expanded genome: long-read sequencing of *Trypanosoma cruzi*. *Microbial genomics*, *4*(5).
<https://doi.org/10.1099/mgen.0.000177>
- Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C.,

- Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Böhme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C. M., Arrowsmith, C., ... El-Sayed, N. M. (2005). The genome of the African trypanosome *Trypanosoma brucei*. *Science*, *309*(5733), 416–422. <https://doi.org/10.1126/science.1112642>
- Blanc, G., & Wolfe, K. H. (2004). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell*, *16*(7), 1667–1678. <https://doi.org/10.1105/tpc.021345>
- Bradwell, K. R., Koparde, V. N., Matveyev, A. V., Serrano, M. G., Alves, J. M. P., Parikh, H., Huang, B., Lee, V., Espinosa-Alvarez, O., Ortiz, P. A., Costa-Martins, A. G., Teixeira, M. M. G., & Buck, G. A. (2018). Genomic comparison of *Trypanosoma conorhini* and *Trypanosoma rangeli* to *Trypanosoma cruzi* strains of high and low virulence. *BMC Genomics*, *19*(1), 1–20. <https://doi.org/10.1186/s12864-018-5112-0>
- Brener, Z. (1997). *Trypanosoma cruzi*: Morgologia e ciclo evolutivo. En J. C. P. Dias & J. R. Coura (Eds.), *Clínica e terapêutica da doença de Chagas: uma abordagem prática para o clínico geral* (1a ed., pp. 25–31). Editora FIOCRUZ. <https://doi.org/10.7476/9788575412435>
- Briones, M. R. S., Souto, R. P., Stolf, B. S., & Zingales, B. (1999). The evolution of two *Trypanosoma cruzi* subgroups inferred from rRNA genes can be correlated with the interchange of American mammalian faunas in the Cenozoic and has implications to pathogenicity and host specificity. *Molecular and Biochemical Parasitology*, *104*(2), 219–232. [https://doi.org/10.1016/S0166-6851\(99\)00155-3](https://doi.org/10.1016/S0166-6851(99)00155-3)
- Buscaglia, C. A., Campo, V. A., Di Noia, J. M., Torrecilhas, A. C. T., De Marchi, C. R., Ferguson, M. A. J., Frasch, A. C. C., & Almeida, I. C. (2004). The Surface Coat of the Mammal-dwelling Infective Trypomastigote Stage of *Trypanosoma cruzi* Is Formed by Highly Diverse Immunogenic Mucins. *Journal of Biological Chemistry*, *279*(16), 15860–15869. <https://doi.org/10.1074/jbc.M314051200>
- Buscaglia, C. A., Campo, V. A., Frasch, A. C. C., & Di Noia, J. M. (2006). *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. *Nature reviews. Microbiology*, *4*(3), 229–236. <https://doi.org/10.1038/nrmicro1351>

- Callejas-Hernández, F., Gironès, N., & Fresno, M. (2018). Genome sequence of *Trypanosoma cruzi* strain Bug2148. *Genome Announcements*, 6(3), 1–2.
<https://doi.org/10.1128/genomeA.01497-17>
- Callejas-Hernández, F., Rastrojo, A., Poveda, C., Gironès, N., & Fresno, M. (2018). Genomic assemblies of newly sequenced *Trypanosoma cruzi* strains reveal new genomic expansion and greater complexity. *Scientific Reports*, 8(1), 1–13.
<https://doi.org/10.1038/s41598-018-32877-2>
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, 17(4), 540–552.
<https://doi.org/10.1093/oxfordjournals.molbev.a026334>
- Chagas, C. (1909). Nova tripanozomíaze humana: estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi* n. gen., n. sp., agente etiológico de nova entidade morbida do homem. *Memórias do Instituto Oswaldo Cruz*, 1(2), 159–218.
<https://doi.org/10.1590/S0074-02761909000200008>
- Claes, F., Büscher, P., Touratier, L., & Goddeeris, B. M. (2005). *Trypanosoma equiperdum*: Master of disguise or historical mistake? *Trends in Parasitology*, 21(7), 316–321. <https://doi.org/10.1016/j.pt.2005.05.010>
- Clément, L., Dietrich, M., Markotter, W., Fasel, N. J., Monadjem, A., López-Baucells, A., Scaravelli, D., Théou, P., Pigeault, R., Ruedi, M., & Christe, P. (2020). Out of Africa: The origins of the protozoan blood parasites of the *Trypanosoma cruzi* clade found in bats from Africa. *Molecular Phylogenetics and Evolution*, 145, 106705.
<https://doi.org/10.1016/J.YMPEV.2019.106705>
- De Pablos, L. M., González, G. G., Parada, J. S., Hidalgo, V. S., Lozano, I. M. D., Samblás, M. M. G., Bustos, T. C., & Osuna, A. (2011). Differential expression and characterization of a member of the mucin-associated surface protein family secreted by *Trypanosoma cruzi*. *Infection and Immunity*, 79(10), 3993–4001.
<https://doi.org/10.1128/IAI.05329-11>
- De Pablos, L. M., & Osuna, A. (2012). Multigene families in *Trypanosoma cruzi* and their role in infectivity. *Infection and Immunity*, 80(7), 2258–2264.

<https://doi.org/10.1128/IAI.06225-11>

dos Santos, S. L., Freitas, L. M., Lobo, F. P., Rodrigues-Luiz, G. F., Mendes, T. A. de O., Oliveira, A. C. S., Andrade, L. O., Chiari, É., Gazzinelli, R. T., Teixeira, S. M. R., Fujiwara, R. T., & Bartholomeu, D. C. (2012). The MASP Family of Trypanosoma cruzi: Changes in Gene Expression and Antigenic Profile during the Acute Phase of Experimental Infection. *PLoS Neglected Tropical Diseases*, 6(8), e1779.

<https://doi.org/10.1371/journal.pntd.0001779>

EBI. (s/f). *454 sequencing / Functional genomics II*. EMBL's European Bioinformatics Institute. Recuperado el 24 de agosto de 2023, de

<https://www.ebi.ac.uk/training/online/courses/functional-genomics-ii-common-technologies-and-data-analysis-methods/next-generation-sequencing/454-sequencing/>

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797.

<https://doi.org/10.1093/nar/gkh340>

El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A.-N., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler, E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Åslund, L., Attipoe, P., ... Andersson, B. (2005). The Genome Sequence of Trypanosoma cruzi, Etiologic Agent of Chagas Disease. *Science*, 309(5733), 409–415.

<https://doi.org/10.1126/science.1112631>

El-Sayed, N. M., Myler, P. J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renauld, H., Worthey, E. A., Hertz-Fowler, C., Ghedin, E., Peacock, C., Bartholomeu, D. C., Haas, B. J., Tran, A., Wortman, J. R., Alsmark, U. C. M., Angiuoli, S., Anupama, A., ... Hall, N. (2005). Comparative Genomics of Trypanosomatid Parasitic Protozoa. *Science*, 309(5733), 404–409.

<https://doi.org/10.1126/science.1112181>

Faria, N. R., Rambaut, A., Suchard, M. A., Baele, G., Bedford, T., Ward, M. J., Tatem, A. J., Sousa, J. D., Arinaminpathy, N., Pèpin, J., Posada, D., Peeters, M., Pybus, O. G., & Lemey, P. (2014). The early spread and epidemic ignition of HIV-1 in human

- populations. *Science (New York, N.Y.)*, 346(6205), 56.
<https://doi.org/10.1126/SCIENCE.1256739>
- Ferreira, L. F., Jansen, A. M., & Araújo, A. (2011). Chagas disease in prehistory. *Anais da Academia Brasileira de Ciências*, 83(3), 1041–1044. <https://doi.org/10.1590/S0001-37652011005000013>
- Finamore, J. M. C. (2010). *Caracterização do polimorfismo intraespecífico do lipofosfoglicano (LPG) de Leishmania infantum Nicolle, 1908 e seu papel na interação com Lutzomyia longipalpis e macrófagos murinos*. Universidade Federal do Maranhão.
- Flores-López, C. A. (2013). *Molecular evolutionary studies on Trypanosoma cruzi, the agent of Chagas disease*. [University of Maryland].
<https://drum.lib.umd.edu/handle/1903/14530>
- Flores-López, C. A., & Machado, C. A. (2011). Analyses of 32 loci clarify phylogenetic relationships among Trypanosoma cruzi lineages and support a single hybridization prior to human contact. *PLoS Neglected Tropical Diseases*, 5(8).
<https://doi.org/10.1371/journal.pntd.0001272>
- Fonseca, L. M., da Costa, K. M., Chaves, V. de S., Freire-De-Lima, C. G., Morrot, A., Mendonça-Previato, L., Previato, J. O., & Freire-De-Lima, L. (2019). Theft and reception of host cell's sialic acid: Dynamics of trypanosoma cruzi trans-sialidases and mucin-like molecules on chagas' disease immunomodulation. *Frontiers in Immunology*, 10. <https://doi.org/10.3389/fimmu.2019.00164>
- Franzén, O., Ochaya, S., Sherwood, E., Lewis, M. D., Llewellyn, M. S., Miles, M. A., & Andersson, B. (2011). Shotgun sequencing analysis of Trypanosoma cruzi i Sylvio X10/1 and comparison with T. cruzi VI CL Brener. *PLoS Neglected Tropical Diseases*, 5(3), 1–9. <https://doi.org/10.1371/journal.pntd.0000984>
- Franzén, O., Talavera-López, C., Ochaya, S., Butler, C. E., Messenger, L. A., Lewis, M. D., Llewellyn, M. S., Marinkelle, C. J., Tyler, K. M., Miles, M. A., & Andersson, B. (2012). Comparative genomic analysis of human infective Trypanosoma cruzi lineages with the bat-restricted subspecies T. cruzi marinkellei. *BMC Genomics*, 13(1), 7–9.

<https://doi.org/10.1186/1471-2164-13-531>

- Frasch, A. C. C. (2000). Functional diversity in the trans-sialidase and mucin families in *Trypanosoma cruzi*. *Parasitology Today*, *16*(7), 282–286.
[https://doi.org/10.1016/S0169-4758\(00\)01698-7](https://doi.org/10.1016/S0169-4758(00)01698-7)
- Futuyma, D., & Kirkpatrick, M. (2017). The Tree of Life. En D. Futuyma & M. Kirkpatrick (Eds.), *Evolution* (4a ed., p. 725). Sinauer Associates, Inc.
- Gojobori, T. (1983). CODON SUBSTITUTION IN EVOLUTION AND THE “SATURATION” OF SYNONYMOUS CHANGES. *Genetics*, *105*(4), 1011–1027.
<https://doi.org/10.1093/genetics/105.4.1011>
- Goldman, N., & Yangf, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*.
<https://doi.org/10.1093/oxfordjournals.molbev.a040153>
- Gorla, D. E., Dujardin, J. P., & Schofield, C. J. (1997). Biosystematics of Old World triatominae. *Acta Tropica*, *63*(2–3), 127–140. [https://doi.org/10.1016/S0001-706X\(97\)87188-4](https://doi.org/10.1016/S0001-706X(97)87188-4)
- Gouy, M., Tannier, E., Comte, N., & Parsons, D. P. (2021). Seaview Version 5: A Multiplatform Software for Multiple Sequence Alignment, Molecular Phylogenetic Analyses, and Tree Reconciliation. En K. Katoh (Ed.), *Multiple Sequence Alignment* (1a ed., Vol. 2231, pp. 241–260). Humana New York, NY.
- Grisard, E. C., Teixeira, S. M. R., de Almeida, L. G. P., Stoco, P. H., Gerber, A. L., Talavera-López, C., Lima, O. C., Andersson, B., & de Vasconcelos, A. T. R. (2014). *Trypanosoma cruzi* clone Dm28c draft genome sequence. *Genome Announcements*, *2*(1), 0–1. <https://doi.org/10.1128/genomeA.e011114-13>
- Guhl, F. (2017). Geographical distribution of Chagas disease. En J. Telleria & M. Tibayrenc (Eds.), *American Trypanosomiasis Chagas Disease* (2a ed., pp. 89–112). Elsevier Inc.
- Hamilton, P. B., Adams, E. R., Njiokou, F., Gibson, W. C., Cuny, G., & Herder, S. (2009). Phylogenetic analysis reveals the presence of the *Trypanosoma cruzi* clade in African

- terrestrial mammals. *Infection, Genetics and Evolution*, 9(1), 81–86.
<https://doi.org/10.1016/j.meegid.2008.10.011>
- Hamilton, P. B., & Stevens, J. R. (2017). Classification and phylogeny of *Trypanosoma cruzi*. En *American Trypanosomiasis Chagas Disease: One Hundred Years of Research: Second Edition* (Second Edi). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-801029-7.00015-0>
- Hamilton, P. B., Teixeira, M. M. G., & Stevens, J. R. (2012). The evolution of *Trypanosoma cruzi*: The “bat seeding” hypothesis. *Trends in Parasitology*, 28(4), 136–141. <https://doi.org/10.1016/j.pt.2012.01.006>
- Harbola, A., Negi, D., Manchanda, M., & Kesharwani, R. K. (2022). Bioinformatics and biological data mining. En D. B. Singh & R. K. Pathak (Eds.), *Bioinformatics* (pp. 457–471). Elsevier. <https://doi.org/10.1016/B978-0-323-89775-4.00019-5>
- Harris, E. E., & Meyer, D. (2006). The molecular signature of selection underlying human adaptations. *American Journal of Physical Anthropology*, 131(S43), 89–130.
<https://doi.org/10.1002/ajpa.20518>
- Hayashida, H., & Miyata, T. (1983). Unusual evolutionary conservation and frequent DNA segment exchange in class I genes of the major histocompatibility complex. *Proceedings of the National Academy of Sciences of the United States of America*, 80(9 I), 2671–2675. <https://doi.org/10.1073/pnas.80.9.2671>
- Herreros-Cabello, A., Callejas-Hernández, F., Gironès, N., & Fresno, M. (2020). *Trypanosoma cruzi* genome: Organization, multi-gene families, transcription, and biological implications. *Genes*, 11(10), 1–26. <https://doi.org/10.3390/genes11101196>
- Ho, S. Y. W., & Duchêne, S. (2014). Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular Ecology*, 23(24), 5947–5965.
<https://doi.org/10.1111/mec.12953>
- Ivens, A. C., Peacock, C. S., Worthey, E. A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S. M., Bianchetti, G., Borzým, K., ...

- Myler, P. J. (2005). The Genome of the Kinetoplastid Parasite, *Leishmania major*. *Science*, *309*(5733), 436–442. <https://doi.org/10.1126/science.1112680>
- Jackson, A. P., Otto, T. D., Aslett, M., Armstrong, S. D., Bringaud, F., Schlacht, A., Hartley, C., Sanders, M., Wastling, J. M., Dacks, J. B., Acosta-Serrano, A., Field, M. C., Ginger, M. L., & Berriman, M. (2016). Kinetoplastid Phylogenomics Reveals the Evolutionary Innovations Associated with the Origins of Parasitism. *Current Biology*, *26*(2), 161–172. <https://doi.org/10.1016/j.cub.2015.11.055>
- Jansen, A. M., Xavier, S. C. C., & Roque, A. L. R. (2017). Ecological aspects of *Trypanosoma cruzi* : wild hosts and reservoirs. En J. Telleria & M. Tibayrenc (Eds.), *American Trypanosomiasis Chagas Disease* (2a ed., pp. 243–264). Elsevier Inc.
- Jauhal, A. A., & Newcomb, R. D. (2021). Assessing genome assembly quality prior to downstream analysis: N50 versus BUSCO. *Molecular Ecology Resources*, *21*(5), 1416–1421. <https://doi.org/10.1111/1755-0998.13364>
- Kaufer, A., Ellis, J., Stark, D., & Barratt, J. (2017). The evolution of trypanosomatid taxonomy. *Parasites and Vectors*, *10*(1), 1–17. <https://doi.org/10.1186/s13071-017-2204-7>
- Kimura, M. (1968). Evolutionary Rate at the Molecular Level. *Nature*, *217*, 624–626. <https://doi.org/10.1006/rwgn.2001.0432>
- Koonin, E. V. (2005). Orthologs, paralogs, and evolutionary genomics. *Annual Review of Genetics*, *39*, 309–338. <https://doi.org/10.1146/annurev.genet.39.073003.114725>
- Kribs, C. M., & Mitchell, C. (2015). Host switching vs. host sharing in overlapping sylvatic *Trypanosoma cruzi* transmission cycles *Trypanosoma cruzi* transmission cycles. *Journal of Biological Dynamics*, 247–227. <https://doi.org/10.1080/17513758.2015.1075611>
- Kumar, S. (2005). Molecular clocks : four decades of evolution. *Nature Reviews Genetics*, *6*(August), 654–662.
- Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., & Gascuel, O. (2019). NGPhylogeny.fr : new generation phylogenetic services for

non-specialists Fred. *Nucleic Acids Research*, 47(April), 260–265.
<https://doi.org/10.1093/nar/gkz303>

Lewis, M. D., Llewellyn, M. S., Yeo, M., Acosta, N., Gaunt, M. W., & Miles, M. A. (2011). Recent, independent and anthropogenic origins of *Trypanosoma cruzi* hybrids. *PLoS Neglected Tropical Diseases*, 5(10).
<https://doi.org/10.1371/journal.pntd.0001363>

Li, W. H. (1997). *Molecular evolution* (1st ed.). Sinauer Associates, Inc., Publishers.

Lima, L., Espinosa-Álvarez, O., Hamilton, P. B., Neves, L., Takata, C. S., Campaner, M., Attias, M., de Souza, W., Camargo, E. P., & Teixeira, M. M. (2013). *Trypanosoma livingstonei*: a new species from African bats supports the bat seeding hypothesis for the *Trypanosoma cruzi* clade. *Parasites & Vectors*, 6(1), 221.
<https://doi.org/10.1186/1756-3305-6-221>

Lima, L., Espinosa-Álvarez, O., Ortiz, P. A., Trejo-Varón, J. A., Carranza, J. C., Pinto, C. M., Serrano, M. G., Buck, G. A., Camargo, E. P., & Teixeira, M. M. G. (2015). Genetic diversity of *Trypanosoma cruzi* in bats, and multilocus phylogenetic and phylogeographical analyses supporting Tcbat as an independent DTU (discrete typing unit). *Acta Tropica*, 151, 166–177. <https://doi.org/10.1016/j.actatropica.2015.07.015>

Loewe, L. (2008). *Negative Selection*. Nature Education.
<https://www.nature.com/scitable/topicpage/negative-selection-1136/>

Lynch, M. (2010). Evolution of the mutation rate Michael Lynch. *Trends Genet*, 26(8), 345–352. <https://doi.org/10.1016/j.tig.2010.05.003>.Evolution

Lynch, M., & Conery, J. S. (2000). The Evolutionary Fate and Consequences of Duplicate Genes. *Science*, 290(5494), 1151–1155.
<https://doi.org/10.1126/science.290.5494.1151>

Lynch, M., & Conery, J. S. (2003). The Origins of Genome Complexity. *Science*, 302(5649), 1401–1404. <https://doi.org/10.1126/science.1089370>

Machado, C. A., & Ayala, F. J. (2001). Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. *Proceedings of the*

National Academy of Sciences, 98(13), 7396–7401.

<https://doi.org/10.1073/pnas.121187198>

Manna, P. T., Kelly, S., & Field, M. C. (2013). Adaptin evolution in kinetoplastids and emergence of the variant surface glycoprotein coat in African trypanosomatids.

Molecular Phylogenetics and Evolution, 67(1), 123–128.

<https://doi.org/10.1016/j.ympev.2013.01.002>

Marcili, A., Lima, L., Cavazzana, M., Junqueira, A. C. V., Veludo, H. H., Maia Da Silva, F., Campaner, M., Paiva, F., Nunes, V. L. B., & Teixeira, M. M. G. (2009). A new genotype of *Trypanosoma cruzi* associated with bats evidenced by phylogenetic analyses using SSU rDNA, cytochrome b and Histone H2B genes and genotyping based on ITS1 rDNA. *Parasitology*, 136(6), 641–655.

<https://doi.org/10.1017/S0031182009005861>

McCartney, M. A., Mallez, S., & Gohl, D. M. (2019). Genome projects in invasion biology.

Conservation Genetics, 20(6), 1201–1222. [https://doi.org/10.1007/s10592-019-01224-](https://doi.org/10.1007/s10592-019-01224-x)

x

Messenger, L. A., Miles, M. A., & Bern, C. (2015). Between a bug and a hard place:

Trypanosoma cruzi genetic diversity and the clinical outcomes of Chagas disease.

Expert Review of Anti-Infective Therapy, 13(8), 995–1029.

<https://doi.org/10.1586/14787210.2015.1056158>

NIH. (2023). *Cóntigo*. National Human Genome Research Institute | Glosario parlante de

términos genómicos y genéticos. [https://www.genome.gov/es/genetics-](https://www.genome.gov/es/genetics-glossary/Cóntigo)

[glossary/Cóntigo](https://www.genome.gov/es/genetics-glossary/Cóntigo)

Noireau, F., Diosque, P., & Jansen, A. M. (2009). *Trypanosoma cruzi* : adaptation to its vectors and its hosts. *Veterinary Research*, 40(2), 26.

<https://doi.org/10.1051/vetres/2009009>

OMS. (2021, abril 1). *La enfermedad de Chagas (trypanosomiasis americana)*. Centro de

Prensa/OMS. [https://www.who.int/es/news-room/fact-sheets/detail/chagas-disease-](https://www.who.int/es/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis))

[\(american-trypanosomiasis\)](https://www.who.int/es/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis))

- OMS. (2023, enero 12). *Leishmaniasis*. Centro de Prensa/OMS.
<https://www.who.int/es/news-room/fact-sheets/detail/leishmaniasis>
- OPS. (s/f). *Enfermedad de Chagas*. Recuperado el 20 de noviembre de 2021, de
<https://www.paho.org/es/temas/enfermedad-chagas>
- Peña-Callejas, G., González, J., Jiménez-Cortés, J. G., De Fuentes-Vicente, J. A., Salazar-Schettino, P. M., Bucio-Torres, M. I., Cabrera-Bravo, M., & Flores-Villegas, A. L. (2022). Chagas disease: biology and transmission of *Trypanosoma cruzi*. *TIP Revista Especializada en Ciencias Químico-Biológicas*, 25(June).
<https://doi.org/10.22201/fesz.23958723e.2022>
- Puillandre, N., Watkins, M., & Olivera, B. M. (2010). Evolution of Conus peptide genes: Duplication and positive selection in the A-superfamily. *Journal of Molecular Evolution*, 70(2), 190–202. <https://doi.org/10.1007/s00239-010-9321-7>
- Ramsey, J. M., Townsend Peterson, A., Carmona-Castro, O., Moo-Llanes, D. A., Nakazawa, Y., Butrick, M., Tun-Ku, E., de la Cruz-Félix, K., & Ibarra-Cerdeña, C. N. (2015). Atlas of Mexican Triatominae (Reduviidae: Hemiptera) and vector transmission of Chagas disease. *Memorias do Instituto Oswaldo Cruz*, 110(3), 339–352. <https://doi.org/10.1590/0074-02760140404>
- Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and Its Applications. *Genomics, Proteomics and Bioinformatics*, 13(5), 278–289.
<https://doi.org/10.1016/j.gpb.2015.08.002>
- Rubio, S., Pacheco-Orozco, R. A., Gómez, A. M., Perdomo, S., & García-Robles, R. (2020). Secuenciación de nueva generación (NGS) de ADN: presente y futuro en la práctica clínica. *Universitas Médica*, 61(2).
<https://doi.org/10.11144/javeriana.umed61-2.sngs>
- Salzberg, S. L., Phillippy, A. M., Zimin, A., Puiu, D., Magoc, T., Koren, S., Treangen, T. J., Schatz, M. C., Delcher, A. L., Roberts, M., Marcxais, G., Pop, M., & Yorke, J. A. (2012). GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Research*, 22(3), 557–567. <https://doi.org/10.1101/gr.131383.111>

- Sánchez, A., Contreras, A., Corrales, J. C., & de la Fe, C. (2022). In the beginning it was zoonosis: One Health to combat this and future pandemics. *SESPAS Report 2022. Gaceta Sanitaria*, 36(tabla 1), S61–S67. <https://doi.org/10.1016/j.gaceta.2022.01.012>
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>
- Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., ... Sherry, S. T. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, 50. <https://doi.org/10.1093/nar/gkab1112>
- Schirmer, M., D'Amore, R., Ijaz, U. Z., Hall, N., & Quince, C. (2016). Illumina error profiles: Resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics*, 17(1), 1–15. <https://doi.org/10.1186/s12859-016-0976-y>
- Serna, C., Lara, J. A., Rodrigues, S. P., Marques, A. F., Almeida, I. C., & Maldonado, R. A. (2014). A synthetic peptide from *Trypanosoma cruzi* mucin-like associated surface protein as candidate for a vaccine against Chagas disease. *Vaccine*, 32(28), 3525–3532. <https://doi.org/10.1016/j.vaccine.2014.04.026>
- Sessions, S. K. (2013). Genome Size. En *Brenner's Encyclopedia of Genetics* (2a ed., pp. 301–305). Elsevier. <https://doi.org/10.1016/B978-0-12-374984-0.00639-2>
- Shanmugasundram, A., Starns, D., Böhme, U., Amos, B., Wilkinson, P. A., Harb, O. S., Warrenfeltz, S., Kissinger, J. C., McDowell, M. A., Roos, D. S., Crouch, K., & Jones, A. R. (2023). TriTrypDB: An integrated functional genomics resource for kinetoplastida. *PLOS Neglected Tropical Diseases*, 17(1), e0011058. <https://doi.org/10.1371/journal.pntd.0011058>
- Silber, A. (2000). *Identificación de dos moléculas de Trypanosoma cruzi involucradas en la invasión de las células del huésped*. Universidad de Buenos Aires.
- Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth

- and coverage: Key considerations in genomic analyses. *Nature Reviews Genetics*, 15(2), 121–132. <https://doi.org/10.1038/nrg3642>
- Solís-Oviedo, R. L., Monteon, V., López, R., & Pech-Canul, Á. de la C. (2018). Biochemical, Cellular, and Immunologic Aspects during Early Interaction between *Trypanosoma cruzi* and Host Cell. *Chagas Disease - Basic Investigations and Challenges*. <https://doi.org/10.5772/intechopen.77236>
- Stadelmann, B., Lin, L. K., Kunz, T. H., & Ruedi, M. (2007). Molecular phylogeny of New World Myotis (Chiroptera, Vespertilionidae) inferred from mitochondrial and nuclear DNA genes. *Molecular Phylogenetics and Evolution*, 43(1), 32–48. <https://doi.org/10.1016/j.ympev.2006.06.019>
- Stevens, J., Noyes, H. A., Dover, G. A., & Gibson, W. C. (1999). The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi*. *Parasitology*, 118(1), 107–116. <https://doi.org/10.1017/S0031182098003473>
- Stevens, J., Noyes, H., Schofield, C., & Gibson, W. (2001). The molecular evolution of trypanosomatidae. En *Advances in Parasitology Volume 48* (Vol. 48, pp. 1–53). Elsevier. [https://doi.org/10.1016/S0065-308X\(01\)48003-1](https://doi.org/10.1016/S0065-308X(01)48003-1)
- Steverding, D. (2014). The history of Chagas disease. *Parasites and Vectors*, 7(1), 1–8. <https://doi.org/10.1186/1756-3305-7-317>
- Stuart, K., Brun, R., Croft, S., Fairlamb, A., Gürtler, R. E., McKerrow, J., Reed, S., & Tarleton, R. (2008). Kinetoplastids: related protozoan pathogens, different diseases. *The Journal of Clinical Investigation*, 118(4), 1301. <https://doi.org/10.1172/JCI33945>
- Teeling, E. C., Springer, M. S., Madsen, O., Bates, P., O'Brien, S. J., & Murphy, W. J. (2005). A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science*, 307(5709), 580–584. <https://doi.org/10.1126/science.1105113>
- Tibayrenc, M. (2003). Genetic subdivisions within *Trypanosoma cruzi* (Discrete Typing Units) and their relevance for molecular epidemiology and experimental evolution. *Kinetoplastid Biology and Disease*, 2, 1–6. <https://doi.org/10.1186/1475-9292-2-12>
- Treangen, T. J., & Salzberg, S. L. (2012). Repetitive DNA and next-generation sequencing:

- computational challenges and solutions. *Nature Reviews Genetics*, *13*(1), 36–46.
<https://doi.org/10.1038/nrg3117>
- Valdés, J., Taylor, M. C., Cross, M. A., Ligtenberg, M. J. L., Rudenko, G., & Borst, P. (1996). The viral thymidine kinase gene as a tool for the study of mutagenesis in *Trypanosoma brucei*. *Nucleic Acids Research*, *24*(10), 1809–1815.
<https://doi.org/10.1093/nar/24.10.1809>
- Venkatesh, D., Zhang, N., Zoltner, M., del Pino, R. C., & Field, M. C. (2018). Evolution of protein trafficking in kinetoplastid parasites: Complexity and pathogenesis. *Traffic*, *19*(11), 803–812. <https://doi.org/10.1111/tra.12601>
- Venter, J. C., Adams, M. D., Sutton, G. G., Kerlavage, A. R., Smith, H. O., & Hunkapiller, M. (1998). Shotgun Sequencing of the Human Genome. *Science*, *280*(5369), 1540–1542. <http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/referer?http://www.sciencemag.org/cgi/content/full/280/5369/1540>
- Viana, R., Moyo, S., Amoako, D. G., Tegally, H., Scheepers, C., Althaus, C. L., Anyaneji, U. J., Bester, P. A., Boni, M. F., Chand, M., Choga, W. T., Colquhoun, R., Davids, M., Deforche, K., Doolabh, D., du Plessis, L., Engelbrecht, S., Everatt, J., Giandhari, J., ... de Oliveira, T. (2022). Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature*, *603*(7902), 679–686.
<https://doi.org/10.1038/s41586-022-04411-y>
- Villarreal, J. C., & Renner, S. S. (2014). A review of molecular-clock calibrations and substitution rates in liverworts, mosses, and hornworts, and a timeframe for a taxonomically cleaned-up genus *Nothoceros*. *Molecular Phylogenetics and Evolution*, *78*(1), 25–35. <https://doi.org/10.1016/j.ympev.2014.04.014>
- Wang, W., Peng, D., Baptista, R. P., Li, Y., Kissinger, J. C., & Tarleton, R. L. (2021). Strain-specific genome evolution in *Trypanosoma cruzi*, the agent of Chagas disease. *PLoS Pathogens*, *17*(1), 1–30. <https://doi.org/10.1371/JOURNAL.PPAT.1009254>
- Weatherly, D. B., Boehlke, C., & Tarleton, R. L. (2009). Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. *BMC Genomics*, *10*, 1–13.
<https://doi.org/10.1186/1471-2164-10-255>

- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8), 1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yang, Z., & Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution*, 17(1), 32–43. <https://doi.org/10.1093/oxfordjournals.molbev.a026236>
- Zhang, J. (2003). Evolution by gene duplication: an update. *Trends in Ecology & Evolution*, 18(6), 292–298. [https://doi.org/10.1016/S0169-5347\(03\)00033-8](https://doi.org/10.1016/S0169-5347(03)00033-8)
- Zhang, L., Chen, F., Zhang, X., Li, Z., Zhao, Y., Lohaus, R., Chang, X., Dong, W., Ho, S. Y. W., Liu, X., Song, A., Chen, J., Guo, W., Wang, Z., Zhuang, Y., Wang, H., Chen, X., Hu, J., Liu, Y., ... Tang, H. (2020). The water lily genome and the early evolution of flowering plants. *Nature*, 577(7788), 79–84. <https://doi.org/10.1038/s41586-019-1852-5>
- Zuckerandl, E., & Pauling, L. (1965). Evolutionary Divergence and Convergence in Proteins. En A. Press (Ed.), *Evolving Genes and Proteins* (1a ed., Vol. 42, Número 2, pp. 97–166). Elsevier. <https://doi.org/10.1016/B978-1-4832-2734-4.50017-6>