



Uso de codones en la dinámica traduccional de *Trypanosoma cruzi*



Br. Lucas Inchausti

Tutor: Dr. Pablo Smircich

Co-tutora: Lic. Lucía Bilbao

Tesina de Grado

Licenciatura en Ciencias Biológicas

Departamento de Genómica – Laboratorio de Interacciones Moleculares

IIBCE – Facultad de Ciencias (UdelaR)

Montevideo, Uruguay

2020

“No estalla como las bombas ni suena como los tiros. Como el hambre, mata callando. Como el hambre, mata a los callados: a los que viven condenados al silencio y mueren condenados al olvido. Tragedia que no suena, enfermos que no pagan, enfermedad que no vende. [...] Elige a sus víctimas en el pobrerío. Las muerde y lentamente, poquito a poquito, va acabando con ellas. Sus víctimas no tienen derechos, ni dinero para comprar los derechos que no tienen. Ni siquiera tienen el derecho de saber de qué mueren.”

Eduardo Galeano

TABLA DE CONTENIDO

1	Introducción.....	1
1.1	Kinetoplástidos.....	1
1.2	<i>Trypanosoma cruzi</i>	2
1.2.1	Características clínicas y epidemiológicas.....	2
1.2.2	Características estructurales y ciclo de vida.....	4
1.2.3	Organización genómica.....	6
1.3	Regulación de la expresión génica.....	8
1.4	Uso diferencial de codones.....	11
2	Materiales y Métodos.....	13
2.1	Medios de cultivo.....	13
2.2	Cultivo <i>in-vitro</i> de epimastigotas.....	13
2.3	Diferenciación <i>in-vitro</i> de epimastigotas a tripomastigotas metacíclicos	13
2.4	Extracción y cuantificación de ARN total	13
2.5	Obtención de datos genómicos, transcriptómicos y traductómicos.....	14
2.6	Estudio de composición de codones.....	14
2.7	Construcción y análisis de perfiles de expresión de genes en transcriptoma y traductoma.....	15
2.8	Verificación de estimación de niveles de expresión mediante qPCR de genes seleccionados.....	16
2.9	Análisis de correspondencia del uso de codones y niveles de expresión.....	17
2.10	Identificación de ARNt en el genoma de <i>T. cruzi</i> y <i>T. brucei</i>	18
2.11	Alineamiento de secuencias de ARNt de <i>T. cruzi</i> y <i>T. brucei</i>	18
2.12	Determinación de sitios de enlentecimiento ribosomal en <i>T. cruzi</i>	18
3	Objetivos.....	20
	Objetivo general.....	20
	Objetivos específicos.....	20
4	Resultados y Discusión.....	21
4.1	Uso general de codones en <i>Trypanosoma cruzi</i>	21
4.1.1	Estudio de selección en el sesgo de uso de codones en transcriptoma y traductoma.....	21
4.1.2	Estimación de niveles de expresión génica mediante RNAseq y Ribosome profiling	23
4.1.3	Validación de niveles de expresión de genes seleccionados mediante qPCR.....	25

4.1.4	Asociación entre el uso diferencial de codones y niveles de expresión génica..	28
4.2	Determinación de sitios de enlentecimiento ribosomal a partir de datos de <i>Ribosome-profiling</i>	30
4.3	Análisis de ARNt.....	35
4.3.1	Análisis de secuencias de ARNt en el genoma de <i>T. cruzi</i>	35
4.3.2	Estrategias de cuantificación de ARNt.....	36
4.4	Correlación entre codones diferencialmente representados en sitios de enlentecimiento ribosomal y niveles esperados de ARNt.....	39
5	Conclusiones.....	42
6	Perspectivas.....	43
7	Bibliografía.....	44
8	Anexo.....	52

RESUMEN

Trypanosoma cruzi es un parásito protozoo agente causante de la enfermedad de Chagas. Esta patología afecta a millones de personas en el mundo y es considerada una “enfermedad tropical desatendida”, caracterizada por proliferar en entornos empobrecidos. *T. cruzi* presenta un ciclo de vida complejo, que involucra distintas etapas de diferenciación, cada una con características particulares, tanto en el vector triatmino, como en sus hospederos vertebrados.

Los genes de *T. cruzi* carecen de intrones, y su expresión se da de forma policistronica, con escasa participación de factores transcripcionales reguladores. Los principales mecanismos de regulación de la expresión de los genes codificantes se dan a nivel post-transcripcional.

Nuestro grupo de investigación ha realizado aportes en el estudio de la regulación post-transcripcional de *T. cruzi*. En particular, el trabajo de (Smircich et al., 2015) refuerza la idea de cómo la traducibilidad y la eficiencia traduccional de un ARNm son pasos regulatorios clave en este organismo. Continuando sobre esta línea, en este trabajo pretendemos ahondar en el estudio de la traducción como mecanismo de control de los niveles de la expresión génica.

Sesenta y un tripletes de bases alternativos (codones) en el ADN y el ARNm codifican para veinte aminoácidos diferentes, de manera que varios aminoácidos están codificados por dos o hasta seis codones distintos o “sinónimos”. El sesgo en el uso de codones (CUB) influye la tasa de traducción de ARNm, fenómeno que estaría ligado a los niveles de ARNt presentes en la célula; codones menos frecuentes estarían asociados a ARNt menos frecuentes. La hipótesis establece que en una región que presente codones “raros”, el ribosoma deberá esperar más tiempo a que llegue el ARNt cargado correcto, produciendo un enlentecimiento de su movimiento. Este fenómeno podría estar asociado a la optimización de los tiempos de síntesis proteica con el fin de permitir el correcto plegamiento de la proteína sintetizada, así como a la estabilización de los ARNm.

Por lo tanto, aquí caracterizamos el uso de codones sobre regiones donde la traducción presente un enlentecimiento, como un posible mecanismo de regulación post-

transcripcional de la expresión génica en *T. cruzi*, a través del estudio de perfiles de huellas ribosomales y su vinculación con los niveles esperados de ARNt en el organismo.

Mediante el cálculo de varias métricas diferentes, mostramos la existencia de presión selectiva para determinar el CUB en tripanosomátidos. Por otra parte, a partir de datos de *RNA-seq* obtenidos previamente, fuimos capaces de establecer el vínculo entre el uso de codones y las tasas de expresión génica a nivel global en *T. cruzi*. La disponibilidad de datos de *Ribo-seq* nos permitió estudiar la dinámica traduccional sobre los ARNm e inferir regiones en donde el movimiento del ribosoma se encuentra enlentecido. Pudimos establecer los codones que se encuentran diferencialmente representados en estas regiones y correlacionar este resultado con los niveles esperados de ARNt inferidos a partir del número de copias en el genoma.

1 INTRODUCCIÓN

1.1 KINETOPLÁSTIDOS

Los kinetoplástidos son un grupo de protistas unicelulares flagelados pertenecientes al filo Euglenozoa. Están caracterizados por la presencia de una única gran mitocondria conocida como “kinetoplasto”, que le da nombre a este grupo. Aunque los distintos organismos del grupo presentan muchas similitudes, como la presencia de un único flagelo que se origina cerca del kinetoplasto, una organización genómica y estructuras celulares similares, y cambios morfológicos drásticos durante su ciclo de vida, estos producen diversas enfermedades en humanos y son transmitidos por distintos vectores (Stuart et al., 2008).

Mediante estudios de ARNr 18S, la clase Kinetoplastea ha sido dividida en dos subclases: Prokinetoplastina y Metakinetoplastina, presentando este último cuatro órdenes: Eubodonidae, Parabodonidae, Neobodonidae y Trypanosomatidae, siendo este último el más estudiado (Moreira et al., 2004; d'Avila-Levy et al., 2015) (**Figura 1**).

Particularmente, todos los miembros del orden Trypanosomatidia pertenecen a una única familia Trypanosomatidae. Dentro de esta familia se encuentran los denominados “TriTryps”, agentes causantes de la enfermedad del sueño africana (*Trypanosoma brucei*), la enfermedad de Chagas (*Trypanosoma cruzi*) y diversas formas de leishmaniasis (*Leishmania spp.*). Los TriTryps se caracterizan por ser organismos monoflagelados y diexénicos, con ciclos de vida y estrategias de supervivencia variados, y que sufren diversos cambios morfológicos al ser transmitidos a hospederos vertebrados mediante un vector invertebrado, principalmente insectos (Lukes et al., 2018).

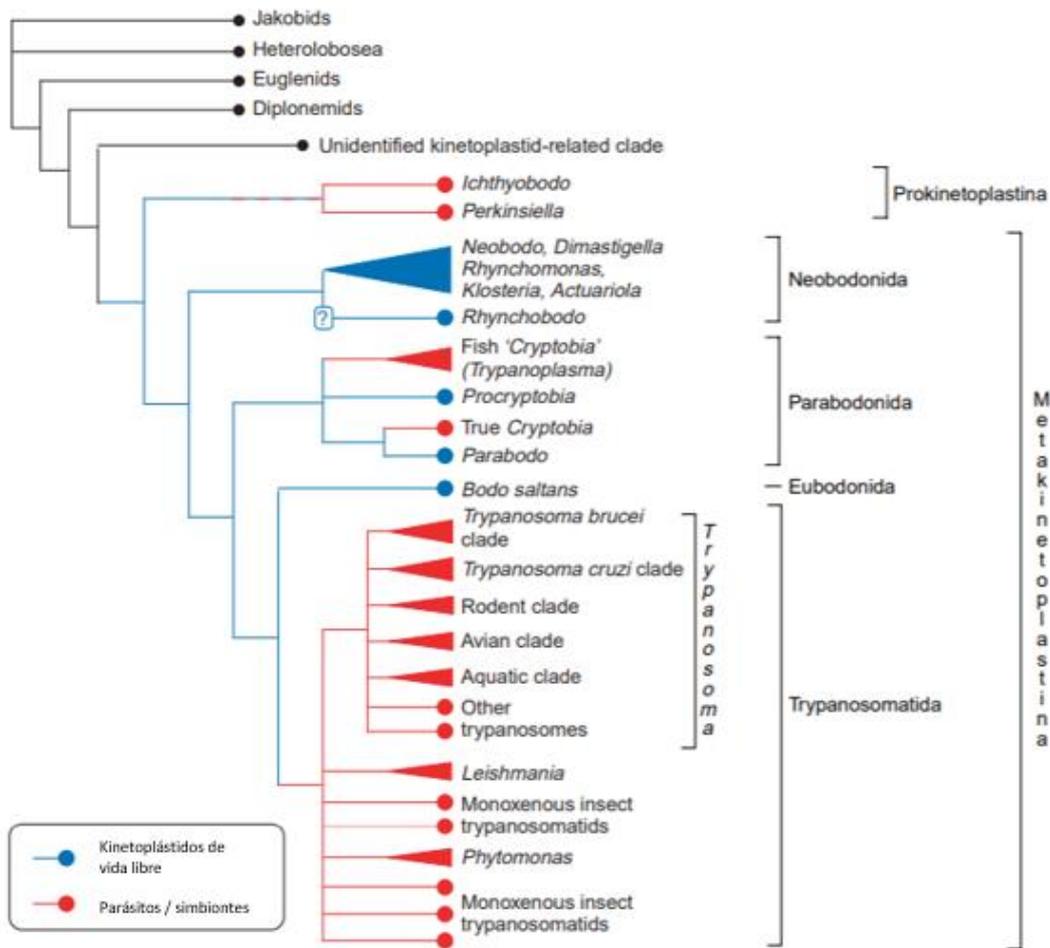


Figura 1. Árbol filogenético de Kinetoplástidos. Extraído y modificado de (Simpson et al., 2006)

1.2 *TRYPANOSOMA CRUZI*

1.2.1 Características clínicas y epidemiológicas

1.2.1.1 *Enfermedad de Chagas*

La tripanosomiasis americana, o enfermedad de Chagas, es una enfermedad causada por el parásito *Trypanosoma cruzi* (*T. cruzi*), transmitida principalmente por contacto con las heces o la orina infectadas de triatominos que se alimentan de sangre. En general, pican y defecan/orinan cerca de la picadura, y los parásitos ingresan al organismo cuando la persona picada se frota instintivamente y empuja las heces o la orina hacia la picadura, los ojos, la boca o alguna lesión cutánea abierta.

Inicialmente, la enfermedad de Chagas estaba confinada a las zonas rurales de América Latina. Debido a la mayor movilidad de la población en los últimos decenios, la mayoría de las personas infectadas ha pasado a vivir en entornos urbanos. La enfermedad se ha ido detectando cada vez más en Estados Unidos y Canadá, en muchos países europeos y en algunos países africanos, principalmente del Mediterráneo Oriental y del Pacífico Occidental (Rassi et al., 2010).

La enfermedad de Chagas tiene dos fases: la fase inicial o aguda, que dura entre 4 y 8 semanas después de la infección, y la fase crónica. Durante la fase aguda circula una gran cantidad de parásitos en el torrente sanguíneo, pero no suelen haber síntomas claros ni específicos de la enfermedad. En menos del 50% de las personas infectadas por un triatomino, un signo inicial característico puede ser una lesión cutánea o la hinchazón amoratada de un párpado. Además, esas personas pueden presentar fiebre, dolor de cabeza, agrandamiento de ganglios linfáticos, palidez, dolores musculares, dificultad para respirar, hinchazón y dolor abdominal o torácico (Rassi et al., 2010).

Por otro lado, la fase crónica de la enfermedad dura toda la vida y puede causar la muerte del individuo infectado. Los parásitos permanecen ocultos principalmente en células musculares cardíacas y del aparato digestivo. Con el paso de los años, la infección puede causar arritmias o insuficiencia cardíaca progresiva como consecuencia de la destrucción del músculo cardíaco (*OMS*).

La enfermedad de Chagas puede tratarse con benznidazol, y con nifurtimox, que matan al parásito. Ambos medicamentos son sumamente eficaces si son administrados durante la fase aguda de la infección, incluso en los casos de transmisión congénita (Rassi et al., 2010). Sin embargo, su eficacia disminuye a medida que transcurre el tiempo de la infección, y las reacciones adversas son más frecuentes en edades avanzadas. Estos medicamentos son sumamente tóxicos, con severos efectos secundarios; en el caso de nifurtimox, la pérdida de peso, alteraciones psicológicas, excitación, somnolencia, vómitos, diarrea, mientras que en el caso del benznidazol, manifestaciones a nivel cutáneo tales como hipersensibilidad y dermatitis con erupciones y edemas, fiebre y dolores musculares (Castro et al., 2006). Ambos medicamentos mostraron también tener efectos mutagénicos y tumorigénicos (Castro et al., 2006; Teixeira et al., 1994).

1.2.2 Características estructurales y ciclo de vida

1.2.2.1 Características estructurales

Los tripanosomátidos presentan una considerable distancia filogenética respecto al resto de eucariotas, siendo uno de los géneros más ancestrales que se han estudiado. Debido a esto y a su riguroso estilo de vida, estos parásitos han desarrollado a lo largo de su evolución características adaptativas excepcionales (Smith & Parsons, 1996).

Una de las características más destacables de estos organismos es la presencia de una única y muy desarrollada mitocondria, que abarca gran parte del volumen celular. El ADN mitocondrial representa hasta un 30% del ADN celular total, y conforma una estructura particular denominada kinetoplasto, cuya localización varía dependiendo de la etapa del ciclo de vida del parásito y está físicamente ligada al cuerpo basal, en la base del flagelo (De Souza, 1984).

El núcleo, a diferencia de la mitocondria, no presenta características distintivas con respecto al resto de las células eucariotas típicas. Sin embargo, *T. cruzi* sí presenta características particulares en sus procesos nucleares, tales como en la replicación, transcripción y reparación del ADN, considerados ancestrales con respecto al resto de eucariotas. Durante los estadios replicativos (epimastigota y amastigota), *T. cruzi* se reproduce mediante fisión binaria, y presenta un núcleo esférico y un evidente nucleolo central. Epimastigotas y amastigotas presentan gran diferencia a nivel de tamaño del núcleo, mientras que su forma es similar. Se desconoce aún con exactitud a qué se debe esta diferencia. Durante el estadio no replicativo tripomastigota existe una considerable disminución de la actividad transcripcional y una dramática reducción del tamaño celular. En este estadio el núcleo presenta una forma alargada, carente de nucleolo y con alto contenido de heterocromatina (Schenkman & Pascoalino, 2011).

En cuanto a la superficie celular, está conformada por una bicapa lipídica y otros componentes del lado extracelular que conforman la glicocálix. Las moléculas que se encuentran en la superficie celular incluyen las familias de glicoproteínas de mucinas, transialidasas y Tc85, entre otras. Estas características de su superficie celular son las que permiten que el parásito interactúe con las células de sus hospederos, en conjunto con la membrana de la vacuola parasitófora al momento de la infección (de Souza, 2009).

Un organelo característico de los tripanosomátidos es el glicosoma, donde tienen lugar varias etapas de la vía de la glucólisis. La compartimentalización de la glucólisis en estos organelos parece ser esencial para la regulación de dicho proceso, que le permite al parásito sobrellevar períodos de anaerobiosis que tienen lugar en determinados estadios del ciclo de vida, a través de la obtención de ATP mediante otras vías metabólicas (Michels et al., 2006). A su vez, el contenido enzimático de los glicosomas puede variar rápidamente durante los fenómenos de diferenciación celular.

Por último, otro organelo particular presente en todos los tripanosomátidos es el acidocalcisoma. Este organelo está involucrado en funciones tales como el almacenamiento de calcio, potasio, hierro, zinc, magnesio, la homeostasis del pH celular y la osmorregulación junto a la vacuola contráctil (Docampo & Moreno, 2011; Miranda et al., 2000; Rohloff et al., 2004).

1.2.2.2 Ciclo de vida de *Trypanosoma cruzi*

T. cruzi se caracteriza por tener un ciclo de vida complejo, que involucra distintas etapas de diferenciación celular, cada una con características particulares de capacidad infectiva y replicativa, tanto en su hospedero triatomino que funciona como vector, como en sus hospederos vertebrados. Los estadios amastigota y tripomastigota se desarrollan en el hospedero mamífero, mientras que los estadios epimastigota y tripomastigota metacíclico lo hacen en el vector triatomino.

El ciclo de vida inicia con la ingesta de tripomastigotas sanguíneos presentes en la sangre del hospedero mamífero infectado, por parte del vector triatomino. Generalmente, la mayor parte de los tripomastigotas mueren en el estómago, mientras que la fracción minoritaria que logra sobrevivir se diferencia a la forma epimastigota, proliferativa y no infectiva, en el tracto digestivo medio. Estos parásitos diferenciados migran a través del intestino, donde se replican intensamente y continúan migrando hasta las regiones más posteriores del intestino y la ampolla rectal del insecto, donde se da una nueva diferenciación a tripomastigotas metacíclicos, no proliferativos e infectivos. Cuando el insecto vuelve a alimentarse de la sangre de un nuevo mamífero, defeca cerca de la herida y libera junto a las heces tripomastigotas metacíclicos, que penetran al torrente sanguíneo a través de la herida. Una vez allí, los tripomastigotas infectan las células hospederas tales

como macrófagos, fibras cardíacas y músculo liso, y son internalizados por las vacuolas endocíticas de las mismas. A partir de la acidificación del medio y ruptura de las vacuolas, los tripomastigotas son liberados al medio y se diferencian en amastigotas, proliferativos y no infectivos. Dentro de la célula, los amastigotas sufren varias replications, produciendo la lisis celular y la liberación de los parásitos, que se diferencian una vez más en tripomastigotas, con la capacidad de infectar nuevas células o alcanzar el torrente sanguíneo del hospedero, donde podrá ser ingerido nuevamente por el vector triatomino, completando su ciclo de vida (**Figura 2**).

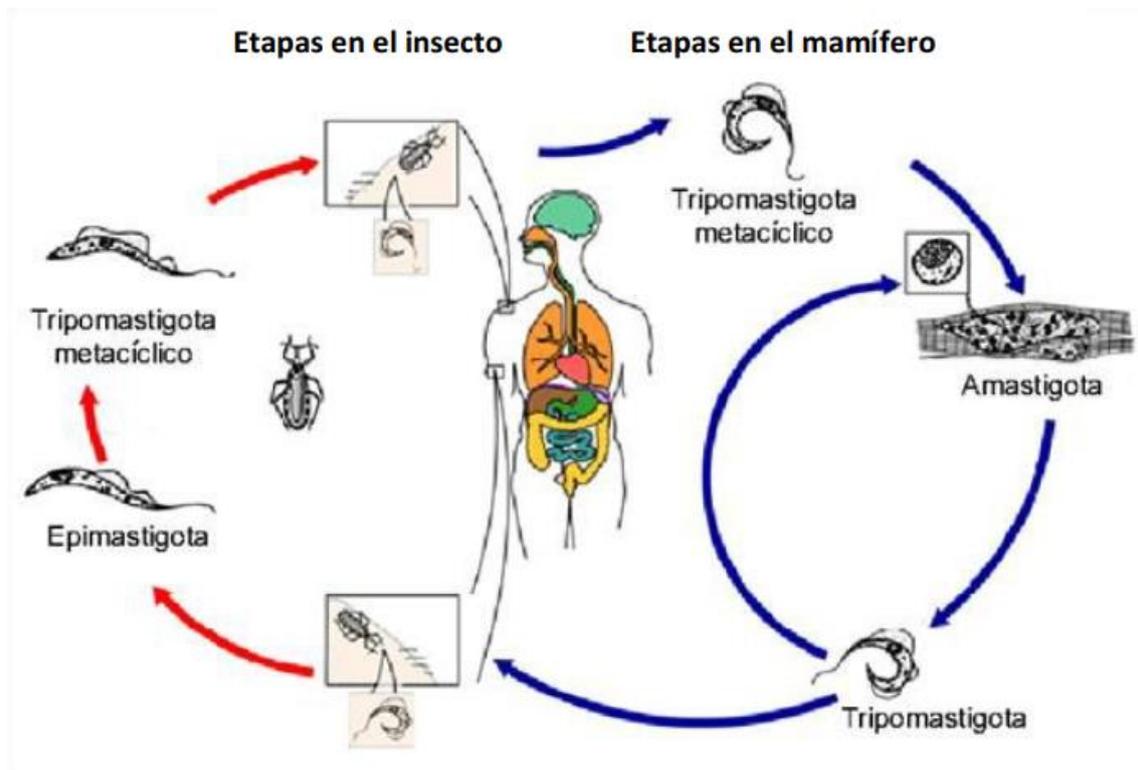


Figura 2. Ciclo de vida de *Trypanosoma cruzi*. Extraído y modificado del sitio web del *Centers for Disease Control and Prevention* (USA, www.cdc.gov).

1.2.3 Organización genómica

1.2.3.1 Genoma de *Trypanosoma cruzi*

En el año 2005, un consorcio internacional publica los genomas de los TriTryps, e inmediatamente es publicado un *special issue* en la revista *Science* (Ash & Jasny, 2005) donde se exploran las secuencias de estos organismos, estudios que han permitido desde entonces caracterizar más en profundidad todas las particularidades biológicas que

presentan. La cepa elegida para el ensamblado del genoma de *T. cruzi* (El-Sayed et al., 2005) fue la CL Brener (TcVI), híbrida entre TcII y TcIII. Aparte de las dificultades provenientes de la cepa elegida para el secuenciado, el ensamblado del genoma (de aproximadamente 55 Mb para el genoma haploide) presentó otras dificultades, principalmente asociadas a la enorme cantidad de secuencias repetidas.

El genoma haploide contiene unos 12.000 genes codificantes de proteínas, de las cuales actualmente un 40% no tiene función conocida. El genoma contiene unos 1994 genes de ARN no codificante y 3590 pseudogenes. Más de la mitad del genoma está formado por secuencias repetidas, tales como retrotransposones, repetidos en tándem y subteloméricos y genes de familias multigénicas. Las principales familias multigénicas de este organismo corresponden a proteínas de tipo transialidasas, mucinas, metaloproteasas, DGF-1, proteínas RHS y las proteínas de superficie asociadas a mucinas (MASP). Algunas de las familias multigénicas codificantes para antígenos de superficie son compartidas en tripanosomátidos mientras que otras son exclusivas. *T. cruzi* no posee mecanismos de variación antigénica, sino que expresa varias proteínas de varias familias, que funciona como una estrategia fundamental para la evasión del sistema inmune del hospedero y demás procesos asociados con la infección.

La secuenciación de los genomas de TriTryps demostró la existencia de una organización particular de los genes, que se encuentran agrupados en grandes regiones con igual orientación a las que se denominó *directional gene clusters* (DGC), y que son característicos de estos organismos. Estos agrupamientos de genes recuerdan a los clásicos operones de los organismos procariotas, sin embargo, los genes incluidos en DGCs no presentan asociación funcional en tripanosomátidos (Palenchar & Bellofatto, 2006). A su vez, este tipo de organización de los genes determina la existencia de sitios denominados *strand switch regions* donde se invierte el sentido de la transcripción, y que juegan roles clave en el inicio de esta. La ausencia de promotores canónicos para la ARN polimerasa II y la escasa presencia de genes codificantes para factores de transcripción en el genoma (Palenchar & Bellofatto, 2006), sumado a este tipo de organización genómica, sugieren que la regulación de la expresión génica se da principalmente a nivel post-transcripcional (Kramer, 2012).

1.3 REGULACIÓN DE LA EXPRESIÓN GÉNICA

En la mayoría de los eucariotas, la primera etapa de regulación de la expresión génica está dada a nivel de la transcripción, mediante mecanismos tales como la alteración del estado de compactación de la cromatina, metilación del ADN, expresión de factores de transcripción, presencia de elementos potenciadores o represores, regulación de la formación del complejo de iniciación, etc.

En el caso de *T. cruzi*, así como de los tripanosomátidos en general, la regulación de la expresión génica a nivel transcripcional es prácticamente inexistente, por lo que la expresión de sus genes está regulada principalmente a nivel post-transcripcional. Esta afirmación está apoyada por el hecho de que genes que son transcritos en la misma unidad policistrónica pueden presentar diferentes niveles de ARNm en estado estacionario, evidenciando la existencia de mecanismos de regulación que operan luego de la transcripción. Estudios transcriptómicos mediante microarreglos en las diferentes etapas del ciclo de vida de *T. cruzi* encontraron diferencias en los niveles de estado estacionario de los transcritos (Minning et al., 2009).

La expresión de los genes codificantes de *T. cruzi* ocurre de forma bidireccional entre DGCs y de forma policistrónica, generando transcritos primarios que contienen secuencias codificantes para varias proteínas, sin intrones, todas en una misma molécula de ARN. Estos ARNm primarios son procesados co-transcripcionalmente mediante dos mecanismos moleculares para generar transcritos monocistrónicos maduros: *trans-splicing* y poliadenilación (Araújo & Teixeira, 2011). Por un lado, el *trans-splicing* consiste en la adición de un miniexón de 39 pares de bases, que incorpora una estructura de caperuza, a las regiones 5' de los diferentes genes incluidos en un mismo ARNm primario. Este miniexón proviene de un ARN SL de aproximadamente 120 pares de bases que se encuentra repetido en tándem en el genoma de forma de permitir suministrar el ARN SL en grandes cantidades que son requeridas por la célula, y que es posteriormente procesado. La adición del miniexón se produce en una secuencia consenso formada por dinucleótidos AG corriente arriba del codón de iniciación del gen (Daniels et al., 2010). Con respecto al resto de eucariotas superiores, la estructura de CAP presenta un mayor número de modificaciones, que consiste en una 7-metilguanosina además de grupos 2'O-metilo en los cuatro primeros nucleótidos (Bangs et al., 1992). En cuanto a la

poliadenilación del extremo 3', se sabe que dicho proceso está acoplado al *trans-splicing*, pero no está descrita una secuencia consenso que actúe como señal (**Figura 3**).

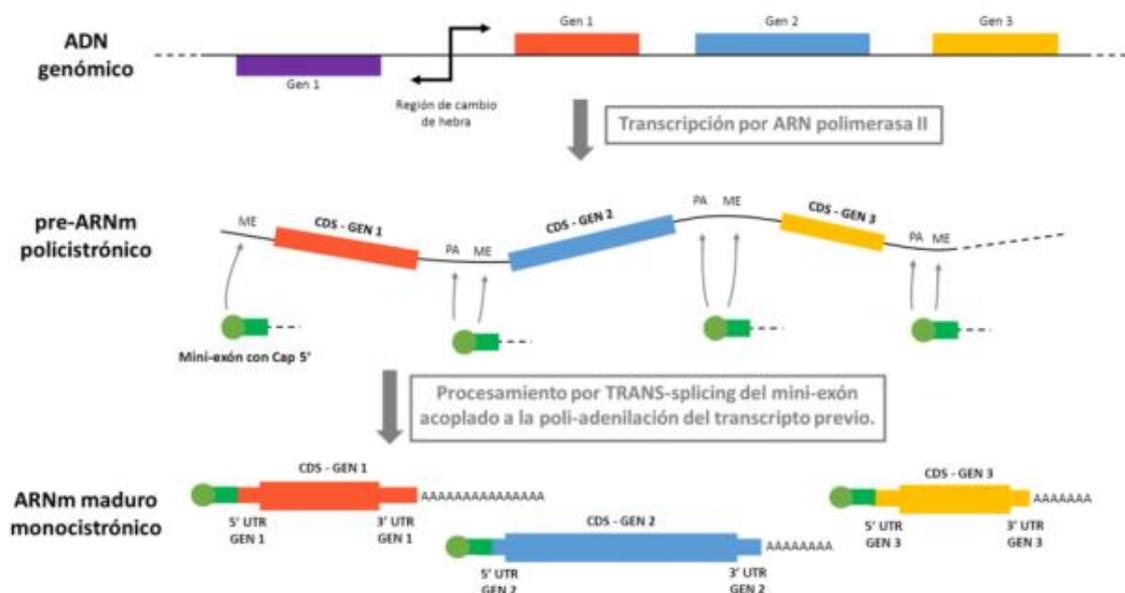


Figura 3. Esquema de la transcripción génica de *T. cruzi*. Extraído de la maestría de Santiago Chávez, 2016.

Las moléculas de ARNm maduro son uno de los principales blancos para la regulación génica, que se da mediante diferentes mecanismos. Uno de ellos es la regulación de los niveles de estado estacionario de los ARNm: la estabilización o degradación modula la vida media de los mismos en los diferentes estadios del ciclo de vida del parásito o en las diferentes condiciones en las que vive, y está principalmente determinada por secuencias presentes en las regiones no-traducidas 5' y 3' (UTR) del ARNm y proteínas que interactúan con las mismas, principalmente a nivel del 3' UTR (Coughlin et al., 2000; Di Noia et al., 2000; Vanhamme & Pays, 1995).

Otro mecanismo altamente conservado en eucariotas es la degradación de los ARNm mediante la acción de exonucleasas que actúan removiendo el CAP 5' y la cola poliA. Se ha descrito que la maquinaria de la vía de degradación de la CAP, junto con ARNm y otras proteínas de unión al ARN, se concentran formando gránulos en el citoplasma (*P-bodies*), que han sido reportado en *T. cruzi* (Barbieri Holetz et al., 2007). Este tipo de estructuras funcionarían como reservorios de ARNm que permiten modular su degradación o devolverlos para ser traducidos, en función de las condiciones a las que esté expuesta la célula.

La traducción parece ser otro punto clave de regulación de la expresión génica en tripanosomátidos, sin embargo, aún no están del todo dilucidados los mecanismos asociados. En organismos que presentan una fuerte regulación traduccional, los niveles de transcritos, cuantificados mediante microarreglos o RNA-seq, no reflejan de forma adecuada la cantidad de proteína presente en la célula. Estudios transcriptómicos y traductómicos realizados por nuestro grupo mediante *RNA-seq* y *Ribosome-profiling* indican que existe una mayor correlación entre el traductoma y el proteoma de *T. cruzi*, con respecto al transcriptoma (Smircich et al., 2015), lo que sugiere la presencia de mecanismos de regulación actuando a nivel traduccional. A su vez, este estudio demuestra la presencia de grandes diferencias en la eficiencia traduccional de transcritos presentes en el mismo estadio del ciclo de vida del parásito, y de iguales transcritos presentes en diferentes estadios del ciclo de vida (específicamente epimastigota y tripomastigota metacíclico).

La regulación de la traducción estaría dada principalmente a nivel de la formación del complejo de iniciación de la traducción y la posterior elongación. En eucariotas, el CAP 5' del ARNm se une a un complejo de exportación nuclear y, una vez fuera del núcleo, el factor de inicio de la traducción eIF4E se une al CAP 5' quien forma parte del complejo eIF4F. El extremo 5' UTR del ARNm es escaneado por el ribosoma hasta el primer codón AUG, donde comienza la síntesis del polipéptido. Interacciones entre eIF4F y la cola poliA resulta en la circularización del ARNm, acercando los extremos 3' y 5' UTR, que pueden presentar unidas proteínas represoras de la síntesis proteica (Clayton & Shapira, 2007). En *T. cruzi* se han reportado factores de inicio de la traducción homólogos a los mencionados (Zinoviev & Shapira, 2012). A su vez, la tasa traduccional puede estar influida por el uso diferencial de codones sinónimos de cada gen: genes de alta expresión estarían optimizados para mejorar la eficiencia y fidelidad traduccional (Hershberg & Petrov, 2008). Este fenómeno se conoce en tripanosomátidos desde hace tiempo, con estudios previos a la publicación de los genomas completos y por lo tanto realizados en un set reducido de genes (Alvarez et al., 1994).

1.4 USO DIFERENCIAL DE CODONES

Sesenta y un triplete de bases alternativos (codones) en el ADN y el ARNm codifican para veinte aminoácidos diferentes, de manera que varios aminoácidos están codificados por dos o hasta seis codones distintos o “sinónimos”.

Se ha reconocido hace varias décadas las diferencias en el uso de codones entre genes y entre especies (es decir, diferencias en la frecuencia de ocurrencia de codones sinónimos). Debido a la aparente naturaleza silenciosa de las sustituciones sinónimas de los codones, por mucho tiempo se creyó que no tenían impacto alguno en la expresión de los genes. Esto ha sido refutado en los últimos años a partir de estudios que evidencian que incluso una única sustitución sinónima puede tener impactos significativos en los niveles de expresión génica, plegamiento de proteínas y función celular de las mismas (Angov, 2011; Jeacock et al., 2018; Plotkin & Kudla, 2011).

Las principales hipótesis que explican los patrones del uso diferencial de codones caen dentro de dos categorías, asociadas principalmente a dos fuerzas evolutivas independientes: mutación y selección natural (Plotkin & Kudla, 2011). Por un lado, la explicación asociada a las mutaciones postula que el uso de codones surge como consecuencia de fenómenos de mutaciones neutrales al azar, que no necesariamente implican una ventaja o desventaja a nivel del *fitness* biológico del organismo. Este tipo de mecanismo usualmente se utiliza para explicar la variación inter-específica en el uso de codones (Plotkin & Kudla, 2011).

Por otro lado, la explicación asociada a la selección natural postula que las mutaciones sinónimas influyen en el *fitness* biológico del organismo, por lo que modularían los patrones en el uso diferencial de codones. Este mecanismo explica generalmente la variación en el uso de codones entre los genes de un organismo (Plotkin & Kudla, 2011).

El uso diferencial de codones influye en la tasa de traducción de los ARNm, el procesamiento co-traducciona del péptido nascente facilitado por el enlentecimiento de la traducción mediado por el uso de codones “raros”, la estabilidad y decaimiento de los ARNm, y más. Está reportado que el uso de codones y la cantidad de copias de ARNt correspondientes juega un rol importante en el control de la expresión génica de tripanosomátidos (Horn, 2008). A su vez, el grupo de Horn demostró experimentalmente en *Trypanosoma brucei* la influencia del uso de codones en el control de la abundancia

relativa de proteína, estudio en el que lograron también predecir la abundancia relativa de ARNm y proteínas basándose únicamente en el uso de codones de las secuencias codificantes (Jeacock et al., 2018).

El sesgo en el uso de codones influye también la tasa de traducción de los ARNm, fenómeno que estaría ligado a los niveles de ARNt presentes en la célula; codones menos frecuentes están asociados a ARNt menos frecuentes en las poblaciones citoplasmáticas. La hipótesis establece que en una región que presente codones “raros”, el ribosoma deberá esperar más tiempo a que llegue el ARNt cargado correcto, produciendo una “pausa ribosomal”, o como mínimo un enlentecimiento de su movimiento (Lesnik et al., 2000). Este fenómeno podría estar asociado a la optimización de los tiempos de síntesis proteica con el fin de permitir el correcto plegamiento de la proteína sintetizada (Thanaraj & Argos, 1996), así como a la estabilización de los ARNm (Collart & Weiss, 2020; Presnyak et al., 2015; Radhakrishnan et al., 2016). En este caso, se ha reportado en levaduras la asociación entre el uso no óptimo de codones y el enriquecimiento de Dhh1, proteína asociada al decaimiento de ARNm, y que ésta se une preferencialmente a ribosomas enlentecidos en el transcripto (Radhakrishnan et al., 2016).

En este trabajo se buscará caracterizar la influencia del uso de codones en la regulación de la expresión génica a nivel global, así como caracterizar el uso de codones sobre regiones donde la traducción presente un enlentecimiento, como un posible mecanismo de regulación post-transcripcional de la expresión génica en *T. cruzi*. Esto será llevado a cabo a través del estudio de perfiles de huellas ribosomales y su vinculación con los niveles esperados de ARNt en el organismo.

2 MATERIALES Y MÉTODOS

2.1 MEDIOS DE CULTIVO

Para el cultivo de parásitos de *Trypanosoma cruzi* se utilizó medio *Brain-Heart Infusion-Tryptose* (BHI): 33 g BHI, 3 g triptosa, 4 g Na₂HPO₄, 0.4 g KCl, 0.3 g glucosa, 10 mL hemina, 990 mL H₂O. Se realizó la esterilización de la solución por autoclavado durante 20 minutos a 121°C. Posteriormente se le incorporó 1% de penicilina (100 U/mL)/estreptomicina(100ug/mL) y 10% de suero fetal bovino descomplementado (tratado 15 minutos a 65°C para la inactivación del complemento).

2.2 CULTIVO *IN-VITRO* DE EPIMASTIGOTAS

Epimastigotas de *T. cruzi* de la cepa Dm28c (Contreras et al., 1988) fueron mantenidos en fase exponencial de crecimiento (2×10^7 parásitos/mL). Los cultivos fueron diluidos 1:10 con BHI fresco (detallado en 2.1) cada 3 días y cultivados a 28°C.

2.3 DIFERENCIACIÓN *IN-VITRO* DE EPIMASTIGOTAS A TRIPOMASTIGOTAS METACÍCLICOS

Para la diferenciación *in-vitro* se dejaron envejecer cultivos de epimastigotas durante 21 días a 28°C, sometidos a estrés nutricional. Pasado este tiempo se colocaron los cultivos en tubos Falcon de 50 mL y se centrifugaron a 3000xg por 15 minutos. Se repitió este procedimiento 2 veces, lavando con 30 mL de PBS 1X (0.145 M NaCl, 0.0027 M KCl, 0.0081 M Na₂HPO₄, 0.0015 M KH₂PO₄ pH 7.4) entre cada centrifugación. Finalmente, se resuspendieron los pellets de células en 40 mL de PBS 1X, y se cultivó a 28°C por 3-5 días, observando diariamente mediante microscopía óptica hasta observar el alargamiento característico de los parásitos en este estadio.

2.4 EXTRACCIÓN Y CUANTIFICACIÓN DE ARN TOTAL

Para la extracción de ARN total se realizó el protocolo TRIzol™ Reagent (Invitrogen) indicado por el fabricante. Brevemente, se tomaron los cultivos (3 réplicas biológicas para el experimento de qPCR y 3 réplicas biológicas para la cuantificación de ARNt), se colocaron en tubos Falcon de 50 mL y se centrifugaron a 3000xg por 15 minutos. Se repitió este procedimiento 3 veces, lavando con 30 mL de PBS 1X entre cada centrifugación.

Para cada réplica, se tomó el pellet celular y se solubilizó en 1 mL de Trizol, provocando la lisis celular y liberación del contenido citosólico. Se agregó cloroformo para separar la mezcla en dos fases, una orgánica que contiene las proteínas y otra acuosa conteniendo el ARN, y una interfase que contiene el ADN. Se tomó la fase acuosa y se precipitó con isopropanol. Finalmente, el ARN se lavó con etanol 75% y se solubilizó en H₂O libre de nucleasas.

Para eliminar las trazas de ADN que podrían permanecer contaminando las muestras de ARN extraído, se trató cada una de ellas con el kit DNAFree (Thermo Fisher Scientific), según el protocolo establecido por el fabricante. Brevemente, se incubaron las muestras de ARN con buffer, la ADNasa, y H₂O, se incubó a 37°C por 30 minutos. Se agregó un buffer inactivador de la enzima y se incubó a temperatura ambiente por 5 minutos. Pasado este tiempo se centrifugó 1,5 minutos a 10.000xg, y se recolectó la fase acuosa que contenía el ARN purificado.

Todas las muestras de ARN se cuantificaron midiendo su absorbancia a 260 nm utilizando NanoDrop™ Lite Spectrophotometer de Thermo Fisher Scientific. Para determinar la pureza de las muestras se tuvo en cuenta el cociente Abs₂₆₀/Abs₂₈₀, considerando como pura aquella muestra cuya relación fuese mayor o igual a 2. En el caso de las muestras destinadas a la cuantificación de ARNt mediante microarreglos, se cuantificó y determinó su calidad mediante Agilent 2100 Bioanalyzer.

2.5 OBTENCIÓN DE DATOS GENÓMICOS, TRANSCRIPTÓMICOS Y TRADUCTÓMICOS

Los datos de genomas, transcriptomas y traductomas para los diferentes análisis bioinformáticos fueron obtenidos de la base de datos especializada en tripanosomátidos TriTrypDB (Aslett et al., 2010) y de la base de datos del SRA, según lo reportado en la publicación correspondiente (Smircich et al., 2015).

2.6 ESTUDIO DE COMPOSICIÓN DE CODONES

El estudio de la composición de codones en el transcriptoma y traductoma de *T. cruzi* se realizó mediante diferentes scripts en R (Ihaka & Gentleman, 1996), utilizando los paquetes *seqinr*, *coRdon*, *stringr* y *ggplot*.

Por un lado, se calculó la composición nucleotídica (% A, % T, % G, % C) de los codones en las posiciones 1, 2 y 3, así como su contenido GC (GC1, GC2, GC3) para cada uno de los genes del CDS.

A su vez, se calculó el valor ENC (*Effective Number of Codons*) (Wright, 1990), que estima la magnitud de la desviación del uso de codones sinónimos de los genes, reflejando la preferencia de ciertos codones, variando de un valor de 20 (uso de codones más sesgado, donde se utiliza solamente un codón sinónimo por aminoácido) a 61 (uso de codones menos sesgado). Utilizando los valores ENC y GC3 de cada uno de los genes, se graficó el ENC vs GC3 *plot*, y la curva esperada bajo la presunción de ausencia de selección.

Por otro lado, utilizando los valores GC12 (que representa el promedio de GC1 y GC2) y GC3 de cada uno de los genes, se graficó un *Neutrality plot*, que permite estimar el balance entre los mecanismos de selección y mutaciones al azar que determinan la composición de los codones en los genes. A su vez, se graficó la curva esperada bajo la presunción de ausencia de selección, donde actuarían únicamente mecanismos mutacionales.

2.7 CONSTRUCCIÓN Y ANÁLISIS DE PERFILES DE EXPRESIÓN DE GENES EN TRANSCRIPTOMA Y TRADUCTOMA

Para la construcción de los perfiles de expresión génica se realizó en primer lugar un reporte de calidad de las lecturas de *RNAseq* y *Ribosome profiling* con la herramienta FastQC (Andrews, 2010). Dado los resultados obtenidos se decidió filtrar las lecturas por calidad y largo con las herramientas incluidas en el paquete CLC Genomics Workbench v10 de QIAGEN, versión que permite el análisis de lecturas de formato *colospace*. Este mismo software fue utilizado para mapear las lecturas filtradas contra el genoma de referencia. Para el caso de los datos de *RNA-seq*, se conservaron aquellas lecturas que tuvieran un largo de entre 18 y 50 nucleótidos, y un *phred score* mayor a 13, mientras que para el caso de los datos de *Ribo-seq*, se conservaron las lecturas que tuvieran un largo de entre 25 y 40 nucleótidos, y un *phred score* mayor a 13.

Las lecturas filtradas se mapearon al genoma de referencia de la cepa CLBrenner obtenida de la base de datos TriTrypDB (Aslett et al., 2010), utilizando el software CLC Genomics Workbench que, como se mencionó anteriormente, permite el manejo de lecturas en

colorspace y además realiza un tratamiento específico de las lecturas de múltiple mapeo (asignando las mismas de forma proporcional al perfil de las lecturas de mapeo único).

Posteriormente, se realizó el conteo de *features* utilizando *featureCounts* (Liao et al., 2014) para construir una tabla que permitió cuantificar la cantidad de lecturas por gen. Se utilizó el paquete de R DEseq2 para normalizar los niveles de expresión de genes de acuerdo con la profundidad de secuenciado de cada librería y luego este valor se normalizó por el largo de cada gen, obteniéndose una tabla con los valores nRPK (“RPK normalizado”), reflejo de su nivel de expresión.

Por último, se procedió a la determinación de *subsets* de genes de alta y baja expresión tomando los 300 genes más expresados (no se tomaron en cuenta genes ribosomales) y los 300 menos expresados (estableciendo un umbral para considerar únicamente genes que tuvieran un mínimo de expresión), respectivamente. A su vez, se determinó el *subset* de genes ribosomales. Las secuencias de los genes de estos *subsets* fueron obtenidas a través de scripts en bash y Python para el posterior estudio de su composición de codones.

Los datos fueron visualizados en programas de interfaz gráfica tales como IGV (Ttir et al., 2012) y Artemis (Carver et al., 2005).

2.8 VERIFICACIÓN DE ESTIMACIÓN DE NIVELES DE EXPRESIÓN MEDIANTE QPCR DE GENES SELECCIONADOS

Con el fin de verificar la estimación de los niveles de expresión obtenidos previamente mediante análisis bioinformáticos, se realizó un ensayo de qPCR de un grupo de genes derivados del mismo: **Alta Expresión:** Mucina - TcSMUGL (TcCLB.506533.142), **Baja Expresión:** MASP (TcCLB.511875.80), Transialidasa TSIII (TcCLB.511129.40), Amastina (TcCLB.511071.40)

A partir del ARN extraído se procedió a su retrotranscripción a ADNc utilizando hexámeros al azar como cebadores y la SuperScript IV (Thermo Fisher Scientific) como retrotranscriptasa. Brevemente, se tomó 1 ug de ARN, 50 ng de cebadores, 1 uL de dNTPs 10 mM y H₂O hasta completar un volumen de 13 uL, y se incubó la mezcla durante 5 minutos a 65°C. Se agregó 4 uL de buffer de reacción 5x First-Strand Buffer, 1 uL de DTT 0,1 M y 1 uL de RNaseOUT 40 U/uL. Se añadió 1 uL de SuperScript IV RT 200

U/uL, se incubó durante 5 minutos a 25°C y luego 45 minutos a 50°C. Finalmente se inactivó la enzima incubando 15 minutos a 70°C.

Para la qPCR, se tomaron las muestras de ADNc y se procedió a su amplificación utilizando cebadores específicos para cada uno de los genes, realizando réplicas técnicas de cada una de las réplicas biológicas. Se utilizó el kit SensiFAST SYBR Hi-Rox Kit (Bioline), que contiene la sonda fluorescente SYBR Green, hot start ADN polimerasa, buffer específico y dNTPs necesarios para la reacción. Las reacciones se llevaron a cabo en un volumen final de 10 uL, con una concentración de molde de 50 ng/uL y de cebadores de 0,4 uM. La reacción se llevó a cabo mediante una primera incubación a 95°C durante 2 minutos para activar la ADN polimerasa, luego 40 ciclos a 95°C por 5 segundos para desnaturalizar las hebras seguido de 30 segundos a 60°C para la hibridación de los cebadores. Las reacciones se llevaron a cabo en un equipo BIORAD CFX96™ Real-Time System y los datos fueron procesados utilizando el software CFX Manager.

Para cuantificar los valores relativos de ARN, se normalizaron los valores de cada gen relativos al gen de expresión constitutiva de la gliceraldehído-3-fosfato deshidrogenasa (GAPDH) ($CT_{gen} - CT_{goi}$), y la magnitud del cambio se expresó como $2^{-(CT_{gen}-CT_{goi})}$

La eficiencia de los cebadores fue establecida previamente en estudios llevados adelante por el grupo (Tesina de grado, Lucía Bilbao), todas mayores a 90%.

2.9 ANÁLISIS DE CORRESPONDENCIA DEL USO DE CODONES Y NIVELES DE EXPRESIÓN

El estudio de la relación entre la composición de codones y los niveles de expresión génica se realizó utilizando análisis estadísticos multivariados, mediante análisis de correspondencia, utilizando el paquete de R ade4 (Charif et al., 2005; Thioulouse et al., 1997). Brevemente, la lógica del script consistió primero en filtrar genes que tuvieran codones stop prematuros y menos de 100 codones en su secuencia. Luego, se construyó para cada *subset* una matriz con el conteo de codones de cada gen, se concatenaron las matrices, se definió un factor que determinaba para qué aminoácido codificaba cada codón, y se realizó el análisis de correspondencia analizando los dos primeros componentes (Charif et al., 2005).

2.10 IDENTIFICACIÓN DE ARNt EN EL GENOMA DE *T. CRUZI* Y *T. BRUCEI*

Las secuencias actualmente anotadas de ARNt de *T. cruzi* y *T. brucei* se descargaron de la base de datos especializada en tripanosomátidos TriTrypDB (Aslett et al., 2009) y se concatenaron en un único archivo multifasta para cada especie. Posteriormente, se filtraron teniendo en cuenta criterios de longitud de secuencia de entre 60 y 100 nucleótidos, y de estructura secundaria utilizando el software ARAGORN (Laslett & Canback, 2004) evaluando que la misma correspondiera a la de ARNt, descartando aquellas secuencias que no cumpliera con alguno de estos criterios. A continuación, se realizó un BLASTn contra el genoma de Dm28c obtenido mediante la tecnología de secuenciación de tercera generación PacBio por (Berná et al., 2018) en el caso de las secuencias de ARNt de *T. cruzi*, y contra el genoma de Lister 427 en el caso de las secuencias de *T. brucei*, también obtenido mediante PacBio por (Müller et al., 2018), con el fin de buscar posibles genes de ARNt no anotados y obtener así el número de copias exacto de cada uno.

2.11 ALINEAMIENTO DE SECUENCIAS DE ARNt DE *T. CRUZI* Y *T. BRUCEI*

Para la comparación de las secuencias de ARNt de *T. cruzi* y *T. brucei* (ver Resultados sección 4.3.2) se realizó un alineamiento utilizando el *software* MUSCLE (Edgar, 2004) y se evaluó la cantidad de *mismatches* en función del largo total de las secuencias.

2.12 DETERMINACIÓN DE SITIOS DE ENLENTECIMIENTO RIBOSOMAL EN *T. CRUZI*

Para la determinación de los sitios de enlentecimiento ribosomal en *T. cruzi* se implementó un flujo de trabajo bioinformático similar al realizado por (Smircich, 2014). La lógica del procedimiento fue encontrar regiones de los ARNm donde las huellas ribosomales (secuencias de fragmentos de ARNm protegidos por ribosomas, obtenidos a partir de ensayos de *ribosome profiling*) se encuentren sobrerrepresentadas con respecto al resto del gen. Para ello, se mapearon las lecturas a los transcritos anotados en el genoma de Dm28c obtenido mediante PacBio (ver sección 2.10), al cual previamente se le filtraron y descartaron gran parte de los genes de familias multigénicas mediante *scripts* realizados en bash y Python. Posteriormente se calculó la cobertura sobre cada nucleótido

utilizando *genomecov* de BEDTools (Quinlan & Hall, 2010) para luego evaluar la cobertura de una ventana de 15 nucleótidos solapante. Se descartaron los genes en donde exista una ventana o más en las que el promedio de los valores de cobertura en no fuese mayor a 2, de modo de continuar el estudio con genes que tuvieran un mínimo de cobertura homogéneo. Para obtener las regiones de pausa se extrajeron las ventanas que superaran 10 veces la mediana de cobertura del gen, descartando las que presentaran también sobrerrepresentación en el transcriptoma. Para estudiar la representación diferencial de codones en los sitios de enlentecimiento ribosomal, se tomaron los conteos de codones sinónimos en las regiones de pausa ribosomal y se comparó, mediante un *test* chi-cuadrado de contingencia, con el conteo de codones sinónimos en los genes que presentaban pausas ribosomales. Una vez obtenidos los codones que tienen una diferencia significativa en su representación en los distintos *sets* de genes, se calculó la frecuencia relativa de cada codón en las regiones de pausa ribosomal, en los genes con pausas ribosomales, en todos los genes del CDS y en los genes de alta expresión.

Por otro lado, a modo comparativo, se utilizó el software *Scikit-ribo* (Fang et al., 2018), que mediante metodologías de aprendizaje automático supervisado permite, entre otras cosas, predecir tiempos relativos de permanencia del ribosoma en los codones (*dwell-time*) a nivel global, a partir de datos de *RNA-seq* y *Ribo-seq*.

3 OBJETIVOS

OBJETIVO GENERAL

Estudiar la influencia del uso diferencial de codones en la dinámica traduccional en *Trypanosoma cruzi*.

OBJETIVOS ESPECÍFICOS

- 1) Estudiar la composición de codones a nivel genómico en *T. cruzi*.
- 2) Estimar los niveles de expresión génica a nivel global a partir de datos de transcriptómica y traductómica.
- 3) Verificar los niveles de expresión mediante una técnica alternativa (qPCR), para un grupo de genes seleccionados.
- 4) Correlacionar el uso de codones diferencial con la expresión génica a nivel transcriptómico y traductómico de *T. cruzi*.
- 5) Identificar y curar la anotación de los genes de ARNt determinando su número de copias en el genoma de *T. cruzi*.
- 6) Determinar la influencia del uso de codones en la dinámica del movimiento del ribosoma a partir de datos de *Ribo-seq*.

4 RESULTADOS Y DISCUSIÓN

4.1 USO GENERAL DE CODONES EN *TRYPANOSOMA CRUZI*

4.1.1 Estudio de selección en el sesgo de uso de codones en transcriptoma y traductoma

Como consecuencia de la degeneración del código genético, la mayoría de los aminoácidos están codificados por más de un codón. El sesgo en el uso de codones es un fenómeno que se da en los genomas, donde los codones sinónimos no son usados con las mismas frecuencias que son esperadas por azar. Se propone entonces que el uso de codones es el resultado de un equilibrio entre las mutaciones producto del azar y la selección natural (Plotkin & Kudla, 2011).

Se propuso, por tanto, analizar cuál de estos mecanismos impera en la modulación del uso de codones de los genes de *T. cruzi*. En primer lugar, se realizó un estudio de la variación de la composición GC entre los primeros dos nucleótidos (GC12) y el tercer nucleótido (GC3) de sus codones. Para ello, se construyó un *Neutrality plot*, donde se graficó para cada gen, la frecuencia de GC12 en función de la frecuencia de GC3 (**Figura 4**). Bajo la hipótesis de que la composición nucleotídica de los codones está afectada únicamente por mutaciones al azar, la relación entre la frecuencia de GC12 y la de GC3 debería ser prácticamente igual, y gráficamente se visualizaría este fenómeno como una recta de pendiente 1. Esto no es lo que se observa en nuestro caso, siendo el contenido GC es mucho más variable en la tercera posición de los codones con respecto a la primera y segunda, posición que determina en mayor medida el uso diferencial de codones sinónimos sin alterar el aminoácido codificado.

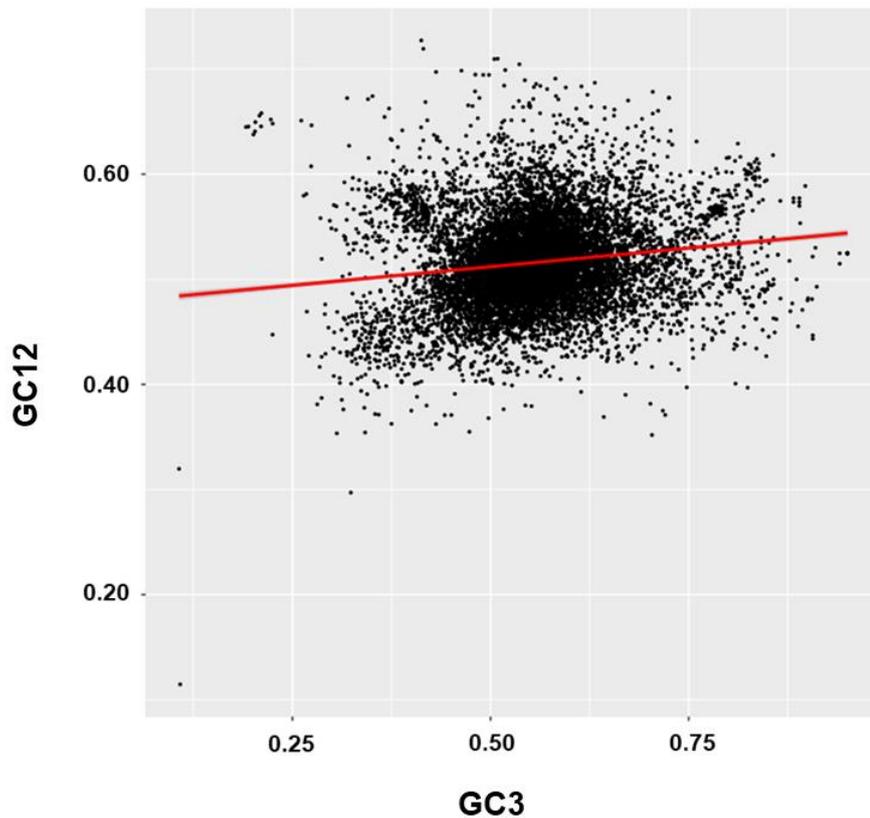


Figura 4. *Neutrality plot.* Se grafica para cada gen su contenido GC en la posición 1 y 2 de sus codones (GC12) contra su contenido GC en la posición 3 (GC3). En rojo se muestra la línea de regresión. En el caso de que la composición nucleotídica de los codones estuviera determinada únicamente por mutaciones al azar, la línea de regresión esperada tendría pendiente 1.

Por otro lado, se analizó la preferencia que tienen los genes por determinados codones sinónimos, en función de su contenido GC3. Para ello, se calculó para cada gen el valor ENC (número efectivo de codones) que estima la magnitud de la desviación del uso de codones sinónimos. Este valor varía entre 20 y 61, siendo 20 un valor ENC que refleja un gran sesgo en el uso de codones, donde solamente se utiliza un codón sinónimo por aminoácido, y 61 siendo un valor ENC donde se utilizan todos los codones sinónimos de forma indistinta, reflejando la ausencia de sesgo en el uso de codones. Utilizando los valores ENC y GC3 de cada uno de los genes, se graficó el ENC vs GC3 *plot* (**Figura 5**). Bajo la hipótesis de que el uso de codones está determinado únicamente por mutaciones al azar, se espera que los genes se sitúen sobre la curva roja. En nuestro caso, se observa que la mayor parte de los genes se sitúa por debajo de esta curva, reflejo de que estarían actuando otros mecanismos evolutivos en el modulado del uso de codones.

Estos análisis permiten concluir que existe un claro sesgo en el uso de codones de los genes de *T. cruzi*, y que este no está determinado únicamente por cambios mutacionales azarosos, sino que hay otros factores evolutivos actuando, que modulan la conformación de los codones en los genes.

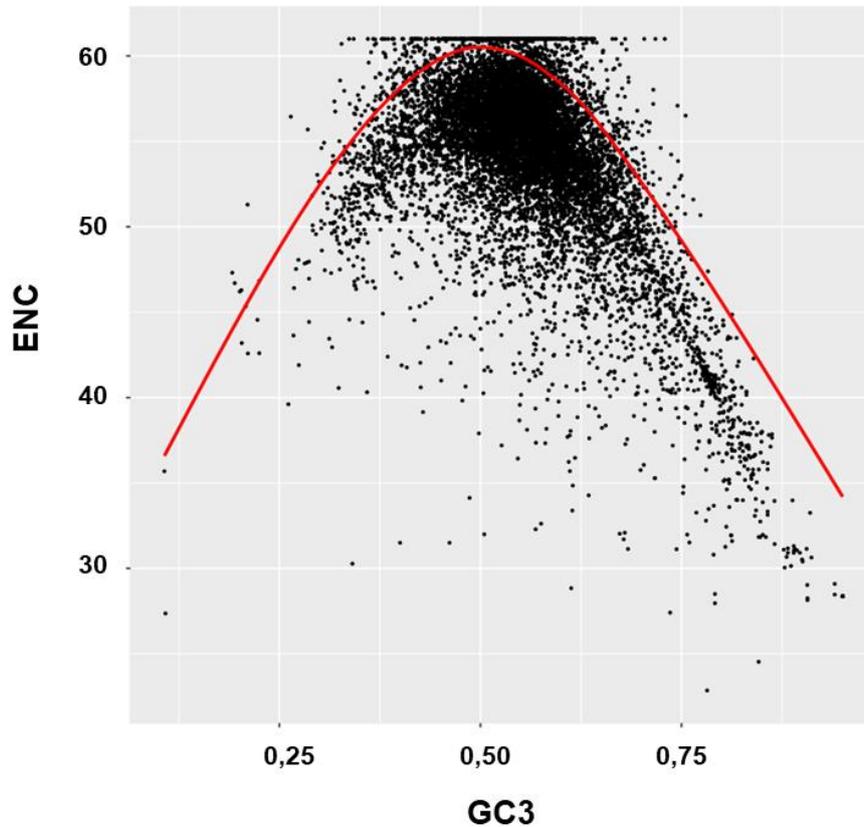


Figura 5. *ENC vs GC3 plot.* Se grafica para cada gen, el número de codones efectivo (ENC) contra el contenido GC en la posición 3 de los codones (GC3). En rojo se muestra la curva esperada si el uso de codones estuviera determinado únicamente por mutaciones al azar.

4.1.2 Estimación de niveles de expresión génica mediante RNAseq y Ribosome profiling

Se reconstruyeron los perfiles de expresión génica utilizando datos de *RNA-seq* y *Ribo-seq* del estadio epimastigota del parásito, generados por el grupo (Smircich et al., 2015). Cabe destacar que estos datos de secuenciación masiva fueron obtenidos mediante la tecnología de SOLiD, que produce un formato de lecturas basados en espacio de color (*colorspace*). Actualmente esta tecnología ha sido descontinuada, por lo que la mayoría de los *softwares* más extensamente utilizados para el alineamiento de lecturas ya no incorporan métodos para el alineamiento de este tipo de lecturas. Para este estudio se

decidió, por tanto, replicar la metodología utilizada por los autores de la publicación mencionada, y utilizar el *software* de interfaz gráfica CLCGenomics Workbench v10.

En primer lugar, se realizó un control de calidad de las lecturas utilizando el software FastQC, que brinda estadísticas sobre la calidad de las lecturas, tales como número total de lecturas, scores de calidad por base, contenido AGCT promedio de las lecturas, contenido GC, distribución del largo de las lecturas, secuencias sobrerrepresentadas (que pueden ser contaminación de ARNr, presencia de adaptadores, etc). Previo a mapear las lecturas contra el genoma para la cuantificación de los niveles de expresión génica, se realizó un filtrado de las lecturas por largo y calidad, utilizando el software CLCGenomics Workbench v10. Los resultados del filtrado de calidad se resumen en la **Tabla 1**.

Tabla 1. Resultados del filtrado de lecturas de *RNA-seq* y *Ribo-seq* utilizando CLCGenomics Workbench v10, realizado a partir de los reportes de calidad obtenidos mediante FastQC.

Datos	N° de lecturas	Promedio de largo de lecturas	N° de lecturas conservadas	Porcentaje de lecturas conservadas	Promedio de largo de lecturas conservadas
Epi1 Ribo-seq	182.299.881	49,1	21.656.335	11,88%	33,1
Epi2 Ribo-seq	182.381.480	49,1	34.044.609	18,67%	33,0
Epi3 Ribo-seq	225.425.090	49,7	36.152.929	16,04%	32,3

Datos	N° de lecturas	Promedio de largo de lecturas	N° de lecturas conservadas	Procentaje de lecturas conservadas	Promedio de largo de lecturas conservadas
Epi1 RNA-seq	52.475.038	49,8	24.368.404	46,44%	34,2
Epi2 RNA-seq	60.519.383	49,8	30.073.924	49,69%	34,4
Epi3 RNA-seq	66.463.404	49,8	29.981.613	45,11%	33,7

Luego del filtrado de las lecturas, se procedió al mapeo de las lecturas de buena calidad, alineando las réplicas al genoma de referencia utilizando el mismo software que el utilizado durante el filtrado de lecturas. Los resultados de este alineamiento se resumen en la (**Tabla 2**).

Tabla 2. Resultados del alineamiento (mapeo) de lecturas de *RNA-seq* y *Ribo-seq* utilizando CLCGenomics Workbench v10.

EPI Ribo-seq	Nº de secuencias	%
Lecturas mapeadas	9.172.114	9,99
Lecturas no mapeadas	82.681.759	90,01
Total	91.853.873	100,0

EPI RNA-seq	Nº de secuencias	%
Lecturas mapeadas	34.564.121	40,94
Lecturas no mapeadas	49.859.820	59,06
Total	84.423.941	100,0

Una vez obtenidos los valores de expresión normalizados para cada gen (ver Métodos), se generaron grupos de IDs de genes: por un lado, los 150 genes ribosomales, por otro lado, se generaron dos grupos en función de sus valores de nRPK: genes de alta expresión que incluyó los 300 genes con valores más altos (excluyendo genes ribosomales), y genes de baja expresión que incluyó los 300 genes con valores más bajos pero que superaron el límite de detección (nRPK mayor a 2).

4.1.3 Validación de niveles de expresión de genes seleccionados mediante qPCR

Con el objetivo de validar los niveles de expresión obtenidos anteriormente mediante secuenciación masiva, se realizaron ensayos de qPCR de genes representantes de los análisis antes descritos: TcSMUGL (TcCLB.506533.142) con un RPK = 21341,9, MASP (TcCLB.511875.80) con un RPK = 13,2, Transialidasa TSIII (TcCLB.511129.40) con un RPK = 21,6 y Amastina (TcCLB.511071.40) con un RPK = 391,9.

Para realizar este experimento fue necesario establecer 3 cultivos (réplicas biológicas) de epimastigotas (ver sección 2.2 de Materiales y Métodos). Se purificaron los parásitos mediante sucesivos pasos de centrifugación y lavado del pellet con PBS 1X, y se extrajo el ARN total utilizando TRIzol. Una vez obtenidas las muestras de ARN, se cuantificaron mediante NanoDrop y se realizó un tratamiento de las muestras con ADNasa, de forma de eliminar las trazas de ADN que pudieran haber quedado al momento de tomar la fracción de ARN durante el protocolo de extracción. El objetivo de este tratamiento con ADNasa fue evitar la amplificación de ADN genómico durante el posterior experimento de qPCR, que de otro modo no permitiría extraer conclusiones certeras sobre los niveles

de expresión relativa de los genes, más aún al tratarse de genes de familias multigénicas con numerosas copias en el genoma.

Una vez obtenidas las muestras de ARN purificado, se procedió a generar el ADNc mediante retrotranscripción *in vitro*, que funcionaría como molde durante la qPCR para cuantificar y analizar la expresión de los genes seleccionados.

Para el ensayo de qPCR se utilizaron cebadores específicos para cada gen, que se resumen en la **Tabla 3**. Si bien los cebadores utilizados fueron dirigidos a regiones únicas y particulares de cada transcripto, dado que se trata de familias multigénicas que pueden presentar variantes y que puede haber errores en las secuencias obtenidas de las bases de datos, se analizaron las curvas de *melting* con el fin de determinar si efectivamente durante la reacción se obtuvo un único producto de amplificación para cada par de

Tabla 3. Cebadores utilizados para verificar la expresión del grupo de genes seleccionados en *T. cruzi*. Se muestra la secuencia codificante a amplificar, el ID del gen, y las secuencias de los cebadores *forward* (FWD) y *reverse* (REV).

	Familia	Grupo	ID
	Transialidasa	TSIII	TcCLB.511129.40
Secuencia	CCGCCATCGTGTTTGATCATTACGACGTGAAAATTGAC CGCCTGTTGAGCCCAACTGCCATTGTGGATGAAAGGG ATAGTGAACAAATGCTCTTGTCGGGG		
Cebadores	CCGCCATCGTGTTTGATCATT		FWD
	CCCCGACAAGAGCATTGTTC		REV
	Amastina	Amastina 40	TcCLB.511071.40
Secuencia	CTTCCTCCATCGTTTCCATTGCCCTGCTGCTTGTGCCACCATC CTTGCGTGTTGTACACCTCTACAAAAGTCTCTCAGGACCT TGACGAATGTGCTGTTGGTGTCTGGCAT		
Cebadores	CTTCCTCCATCGTTTCCATTGC		FWD
	ATGCCAGACACCAACAGCACAT		REV
	Mucina	TcSMUGL	TcCLB.506533.142
Secuencia	GTGCTTTTGTTTGCCCTCTGCTGTGCGTGCGCCACGG CCCAGGAGGAGGGCAGTACGATGCTGCTTTTTGAAGGC		
Cebadores	GTGCTTTTGTTTGCCCTCTGCT		FWD
	GCCTTCAAACAGCAGCATCGT		REV
	Masp	MASP	TcCLB.511875.80
Secuencia	GGCATTGAGCCAACAACAGATGCAAGGCAGAATGCAATAACT GAGGGTCAGGCAGAAACAACCTCACCATCTTCTTCTGCGG		
Cebadores	GGCATTGAGCCAACAACAGATG		FWD
	CCGCAGAAGAAGATGGTGAGG		REV

cebadores (**Figura 6**). Se observa que para el caso de la Transialidasa hay un único pico en las diferentes curvas a determinada temperatura de *melting*, lo que está asociado a un único producto de amplificación. Por otro lado, para los casos de SMUGL, MASP y Amastina, se puede observar que la curva presenta pequeños hombros en T_m menores con respecto al pico principal. Esto indica la presencia de productos inespecíficos que podrían estar dados por variantes génicas de las familias. La alta similitud entre estos genes duplicados hace que el diseño de cebadores específicos para un único miembro de la familia sea extremadamente complejo.

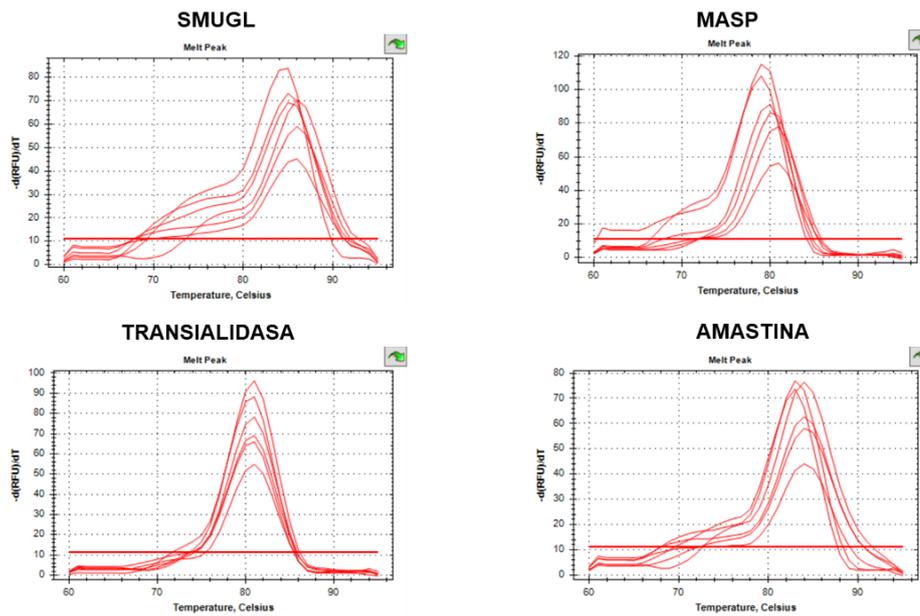


Figura 6. Resultados de curvas de *melting* para los genes estudiados.

A continuación, se cuantificaron los transcritos de los genes de interés relativo a la cantidad del transcrito del gen de expresión constitutiva gliceraldehído-3-fosfato deshidrogenada (GAPDH) (**Figura 7**).

Se observa que, del grupo de genes seleccionados, el más expresado es SMUGL, el menos expresado es MASP, y la Amastina y Transialidasa tienen niveles de expresión que se sitúan entre los niveles de los otros dos genes. Este resultado está en concordancia con lo observado durante el análisis de expresión a partir de datos de secuenciación masiva, por lo que se considera validado este último análisis.

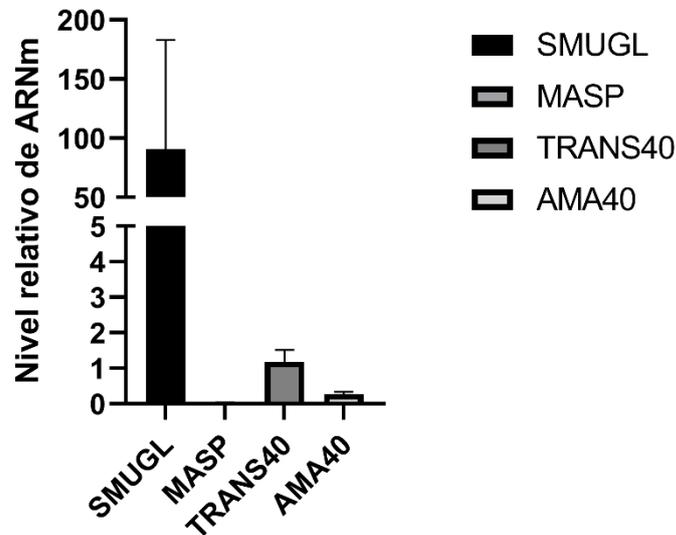


Figura 7. Cuantificación de la expresión relativa a GAPDH del grupo de genes seleccionados.

4.1.4 Asociación entre el uso diferencial de codones y niveles de expresión génica

Con el fin de establecer la existencia de una asociación entre la optimización (o no) en el uso de codones de los genes y sus niveles de expresión, se realizó un análisis de correspondencia con el paquete de R *ade4*, a partir de los datos traductómicos mencionados anteriormente.

Para este análisis, tomamos 4 *subsets* de genes obtenidos a partir de los análisis previos de estimación de los niveles de expresión: genes de alta expresión (*High-expressed genes*), genes de baja expresión (*Low-expressed genes*), genes ribosomales (*Ribosomal genes*) y todos los genes (*All genes*). Se realizó un análisis de correspondencia a partir del conteo de codones para cada gen de cada uno de los *subsets*, y se graficaron las dos componentes en espacio bidimensional (**Figura 8**). Se visualiza que los genes se

distribuyen de una forma particular: los genes ribosomales agrupados por un lado, los genes de alta expresión parecen tener un perfil de uso de codones que tiende a asemejarse al de los genes ribosomales, mientras que los genes de baja expresión lo contrario. Partiendo de la hipótesis de que los genes ribosomales tienen una expresión alta y constitutiva, y un uso de codones optimizado para ello, y observando el comportamiento de los perfiles de uso de codones de los genes de alta y baja expresión, se puede concluir la existencia de una correlación entre el uso diferencial de codones y los niveles de expresión génica. Estudios experimentales realizados en *T. brucei*, demostraron la disminución de la expresión de genes reporteros, alterando su composición de codones utilizando aquellos que se encuentran subrepresentados en genes de alta expresión (Alvarez et al., 1994; Jeacock et al., 2018; Nascimento et al., 2018), resultados que respaldarían nuestros estudios.

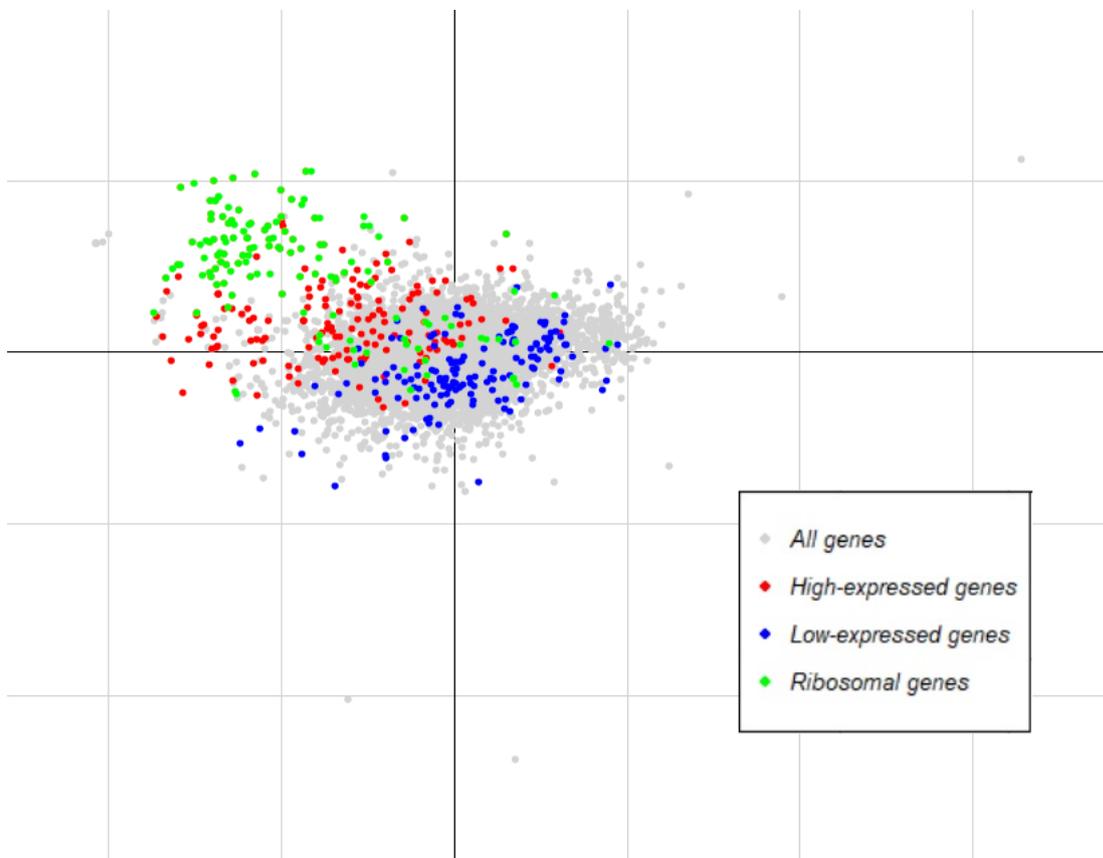


Figura 8. Análisis de correspondencia (CoA) muestra diferencias en los patrones de uso de codones en función de los niveles de expresión génica

4.2 DETERMINACIÓN DE SITIOS DE ENLENTECIMIENTO RIBOSOMAL A PARTIR DE DATOS DE *RIBOSOME-PROFILING*

En el estudio llevado adelante por el grupo mencionado anteriormente (Smircich et al., 2015), utilizando la técnica de *Ribosome profiling* (*Ribo-seq*), se detectaron diferencias significativas en los niveles de expresión y las eficiencias traduccionales de los transcriptos en los estadios epimastigota y tripomastigota metacíclico. Esta metodología, desarrollada por (Ingolia et al., 2009), permite estimar el traductoma, es decir, aquellos transcriptos que están siendo traducidos en la célula, mediante la detención de la traducción utilizando cicloheximida, la purificación de la fracción de polisomas, la digestión controlada mediante ensayo de protección a nucleasas, y la secuenciación de los fragmentos de ARNm protegidos por los ribosomas de aproximadamente 30 nucleótidos (denominados huellas ribosomales). Una vez secuenciados los fragmentos, las lecturas son mapeadas contra los CDS, y se obtiene un perfil de huellas ribosomales sobre cada mensajero, reflejo de la cantidad de ribosomas presentes en los ARNm de ese gen y particularmente en esa región de los mismos.

Cuando se observan los perfiles de huellas ribosomales utilizando un visualizador genómico, se pueden observar regiones que presentan picos de cobertura muy grandes. La hipótesis es que esos picos de cobertura podrían estar dados por ribosomas que se encuentran detenidos o enlentecidos en codones “raros”, cuyos niveles de ARNt asociado serían bajos en la célula, por lo que el ribosoma debe esperar más tiempo a llegue que el ARNt cargado correcto, afectando de esa forma su dinámica de traducción. Se cree que este fenómeno podría estar asociado a la optimización de los tiempos de síntesis proteica con el fin de permitir el correcto plegamiento de la proteína sintetizada (Thanaraj & Argos, 1996), así como a la estabilización de los ARNm (Radhakrishnan et al., 2016).

Con el fin de evaluar esta hipótesis en *T. cruzi*, se propuso identificar regiones de pausa o enlentecimiento ribosomal donde las huellas ribosomales se encontraran sobrerrepresentadas con respecto al resto del gen, utilizando los datos de *RNA-seq* y *Ribo-seq* generados por el grupo (Smircich et al., 2015). Para ello, se utilizó un flujo de trabajo

bioinformático detallado en la (Figura 9). En primer lugar, se realizó el filtrado de genes de familias multigénicas del CDS (ver sección 2.12 de Materiales y Métodos), con el fin de evitar que la gran cantidad de lecturas que mapearían a esos genes (que tienen un uso

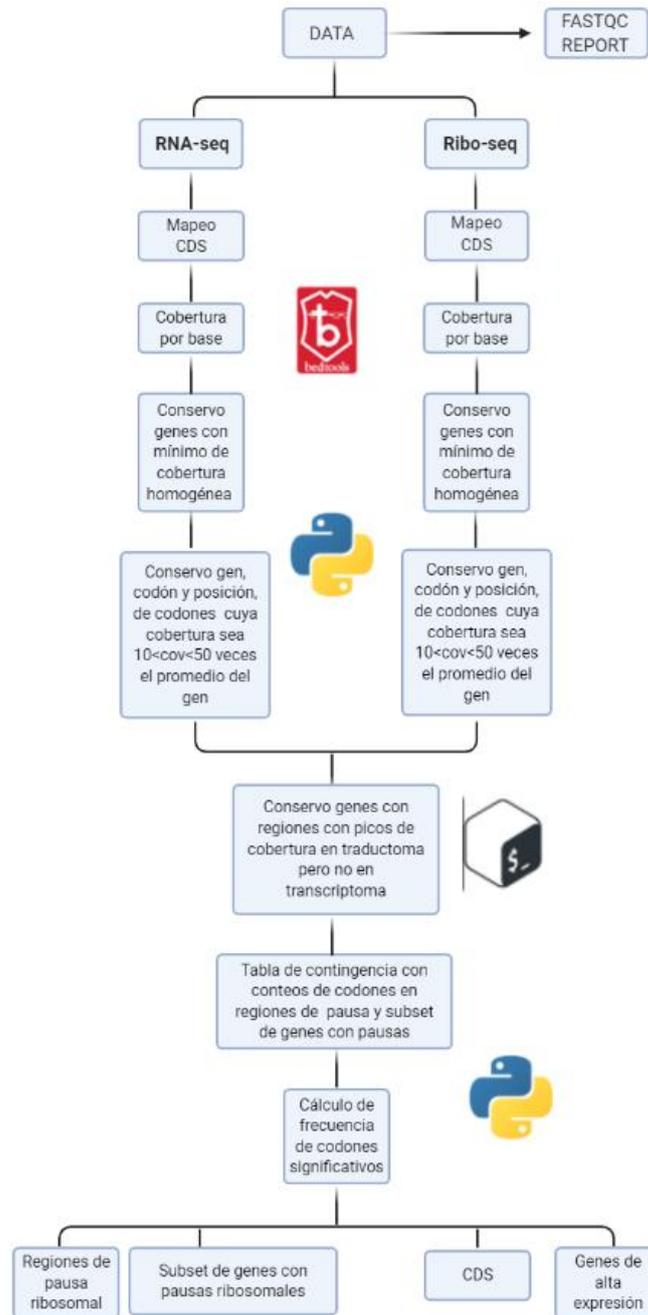


Figura 9. Flujo de trabajo bioinformático utilizado para el estudio de representación diferencial de codones en sitios de enlentecimiento (pausa) ribosomal

de codones similar dentro de las familias), sesgaran el estudio para el análisis de codones posterior, pues podría darse que haya una sobrerrepresentación de determinados codones simplemente por el hecho de que esos genes se expresan mucho y utilizan mucho esos

codones. Una vez filtrado el CDS, se realizó un control de calidad de las lecturas, un filtrado de estas y finalmente el mapeo al los CDSs filtrados. A continuación, se identificaron regiones de los transcriptos que tuvieran picos de cobertura en el traductoma, descartando aquellos que presentaran también sobrerrepresentación de cobertura en el transcriptoma, ya que estos picos presentes en ambos podrían estar asociados a sesgos en los procesos de generación de las librerías y/o secuenciación, y no realmente a un sitio de pausa ribosomal (ver sección 2.12 de Materiales y Métodos).

Una vez obtenidas las regiones de pausa ribosomal, se propuso estudiar su composición de codones evaluando la representación diferencial de los mismos mediante un *test* chi-cuadrado de contingencia, a partir del cual se obtuvieron ciertos codones que tenían una diferencia significativa en su representación en las regiones de pausa ribosomal con respecto al resto del gen.

Con el fin de descartar la posibilidad de que los genes que presentaban regiones de pausa ribosomal tuvieran de por sí un sesgo en su composición de codones, se realizó una comparación entre las frecuencias de los codones sinónimos de ese subset de genes y las

frecuencias de todos los genes del CDS, resumido en la **Tabla 4**. A partir de este resultado no halló evidencia de la existencia de tal sesgo, por lo que se continuó con el análisis.

Aminoácido	Codón	Genes con Pausas	Genes del CDS
A	GCG	0,29	0,3
A	Resto	0,71	0,7
C	TGT	0,45	0,43
C	TGC	0,55	0,57
F	TTC	0,46	0,47
F	TTT	0,54	0,53
G	GGC	0,32	0,34
G	Resto	0,68	0,66
K	AAA	0,48	0,47
K	AAG	0,52	0,53
L	CTA	0,03	0,03
L	Resto	0,97	0,97
L	CTC	0,29	0,29
L	Resto	0,71	0,71
P	CCA	0,27	0,26
P	Resto	0,73	0,74
P	CCT	0,19	0,2
P	Resto	0,81	0,8
R	CGC	0,24	0,26
R	Resto	0,76	0,74
V	GTG	0,31	0,32
V	Resto	0,69	0,68
V	GTC	0,26	0,26
V	Resto	0,74	0,74
Y	TAT	0,42	0,42
Y	TAC	0,58	0,58

Tabla 4. Comparación de frecuencias de codones significativos en los Genes con Pausas y todos los Genes del CDS.

Una vez obtenidos estos codones con representación diferencial significativa en las regiones de pausa ribosomal, se calculó la frecuencia relativa de cada uno de ellos en estas regiones, en todos los genes con pausas, y en los genes de alta expresión (**Figura 10**).

Estos resultados muestran que para la mayor parte de los codones sinónimos (44 de 61), no hay diferencias significativas en su frecuencia entre los genes con pausas ribosomales y las regiones donde éstas se producen (**Tabla 1 Anexo**). Sin embargo, en el caso de Cisteína (C), Lisina (K), Tirosina (Y), Valina (V), y Prolina (P) los codones más frecuentemente utilizados se invierten en las regiones de pausa. En el caso de la cisteína, el codón más frecuentemente usado es TGC, mientras que en las regiones de pausa el más utilizado es TGT. Para la lisina, el codón más usado es AAG, pero en las regiones de pausa el más utilizado pasa a ser AAA. En el caso de la tirosina, TAT está enriquecido

en las regiones de pausa con respecto a TAC. Para la valina, GTC se enriquece en las regiones de pausa por sobre el resto de los codones sinónimos. En el caso de la prolina, el codón CCA está enriquecido en las regiones de pausa con respecto al resto de los codones sinónimos. Por último, para la fenilalanina, el codón más frecuentemente usado no se invierte, pero sí está enriquecido en las regiones de pausa. Para todos los casos, se observa un comportamiento opuesto en los genes de alta expresión.

Por otro lado, en el caso de la Alanina (A), Glicina (G), Arginina (R) y Valina (V), se observa la subrepresentación de determinados codones sinónimos en las regiones de pausa. En el caso de la alanina, GCG es el codón más usado a nivel general, mientras que en pausas se encuentra disminuido; lo mismo sucede para el codón de la glicina GGC, el de la arginina CGC, y el de la valina GTG. Sin embargo, para todos los casos, se observa una sobrerrepresentación de estos codones en los genes de alta expresión.

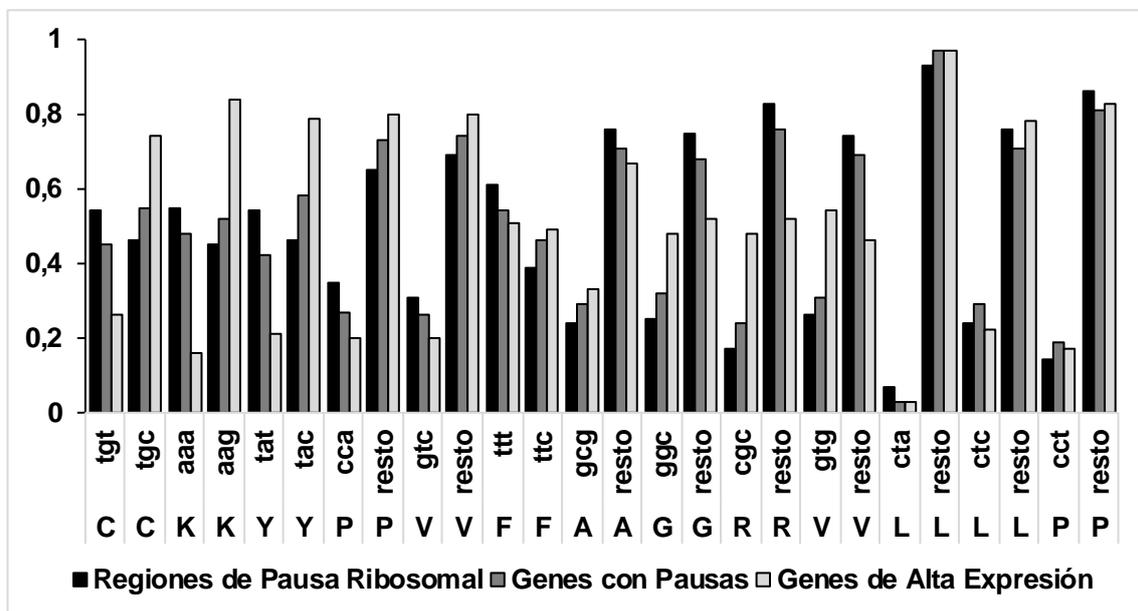


Figura 10. Frecuencias de codones significativos en Regiones de Pausa Ribosomal, Genes con Pausas y Genes de Alta Expresión. En el caso de aminoácidos que presentan más de dos sinónimos, se comparó la frecuencia del codón significativo con la suma de las frecuencias del resto de sinónimos.

Por otro lado, utilizamos el *software Scikit-ribo*, que permite estimar el tiempo de permanencia de los ribosomas sobre los codones en regiones de pausa ribosomal mediante métodos de aprendizaje automático, utilizando datos de *RNA-seq* y *Ribo-seq*.

Particularmente, este *software* evalúa y elimina posibles sesgos en la identificación de pausas dados por la abundancia diferencial de ARNm y estructuras secundarias de los mismos.

Utilizando este *software*, se observa que algunos de los codones con representación diferencial significativa en regiones de pausa obtenidos mediante la metodología desarrollada por nosotros, siguen la misma tendencia cuando se evalúa el tiempo de permanencia del ribosoma sobre los mismos (**Tabla 2 Anexo**). En el caso de la cisteína (C), el codón TGT presenta una mayor frecuencia en las regiones de pausa y un mayor tiempo de permanencia del ribosoma mientras que el codón TGC lo opuesto. Lo mismo sucede cuando observamos el comportamiento de los codones de la fenilalanina (F) y la tirosina (Y). Por otro lado, en el caso de la glicina (G), se observa que el codón GGC, que dentro de los sinónimos es uno de los menos frecuentes en las regiones de pausa, también posee uno de los menores tiempos de permanencia del ribosoma.

4.3 ANÁLISIS DE ARNT

4.3.1 Análisis de secuencias de ARNt en el genoma de *T. cruzi*

Como mencionamos anteriormente, el uso de codones óptimo (es decir, qué tan eficiente y rápido un codón de un ARNm es traducido por los ribosomas) se cree que está principalmente determinado por las poblaciones de ARNt de la célula. Genes más expresados tienden a tener un uso de codones particular y óptimo, y a su vez codifican proteínas de gran abundancia. Estos codones óptimos presentan escasa representación en genes de baja expresión, y habría una asociación entre el uso de codones no-óptimos en este tipo de transcritos y su estabilidad, como demostraron independientemente (Jeacock et al., 2018; Nascimento et al., 2018) de forma experimental en algunos tripanosomátidos.

Luego de obtener ciertos codones sub o sobrerrepresentados en los sitios de enlentecimiento ribosomal, y bajo la hipótesis de que estos podrían estar asociados a los niveles de ARNt, se propuso estudiar los niveles esperados de estas moléculas a partir del número de copias de cada isoceptor en el genoma de *T. cruzi*.

En primer lugar, se descargaron de la base de datos TriTrypDB todas las secuencias anotadas como ARNt, un total de 165. Se realizaron diferentes filtrados con el fin de eliminar secuencias asignadas incorrectamente como ARNt (ver sección 2.10 de Materiales y Métodos). Posteriormente, se realizó un BLASTn con las secuencias conservadas luego del filtrado contra el genoma de *T. cruzi* de la cepa Dm28c reportado por (Berná et al., 2018), obtenido mediante PacBio. Este ensamblado, según lo reportado por los investigadores, permitió estimar de forma más precisa el número de copias de genes repetidos en tándem, como es el caso de genes codificantes para ARNt. Se generó un archivo con todas las secuencias obtenidas del BLASTn, al cual se le realizó nuevamente el filtrado descrito anteriormente, conservando finalmente 149 secuencias de ARNt.

Es interesante destacar que en análisis previos realizados por (Horn, 2008), donde realiza una comparación de codones enriquecidos en los genes y el número de copias de ARNt asociados en TriTryps, los investigadores reportan 115 secuencias de ARNt en el genoma de *T. cruzi*. En nuestro análisis se reporta un aumento de 34 secuencias, mejorando de esta forma la precisión para este tipo de estudios. La comparación entre el número de copias de cada isoceptor reportadas por Horn y las reportadas por nuestro análisis se resumen en la (Tabla 3 Anexo).

4.3.2 Estrategias de cuantificación de ARNt

Con el fin de estimar de forma más precisa los niveles de ARNt presentes en la célula, se propuso estudiar y llevar a cabo diferentes técnicas para su cuantificación. En este caso, se planteó realizarla en los estadios epimastigota y tripomastigota metacíclico. Por un lado, los datos generados asociados al estadio epimastigota servirán para este estudio, mientras que los datos asociados al estadio tripomastigota metacíclico servirán para futuros estudios de búsqueda de codones sobrerrepresentados en pausas ribosomales en este estadio y su asociación con los niveles de ARNt, y estudios comparativos de posible expresión diferencial de ARNt entre ambos estadios.

Dada la naturaleza intrínseca de estas moléculas (particularmente su fuerte estructura terciaria y múltiples modificaciones en sus bases nucleotídicas (Phizicky & Hopper, 2010)), los protocolos tradicionales de cuantificación de ARN pequeños, ya sea RT-qPCR, *RNA-seq*, microarreglos, no son del todo efectivos ni precisos. Es por esta razón

que se planteó evaluar distintas metodologías alternativas desarrolladas en los últimos años.

En primer lugar, se estableció una colaboración con la Dra. Zoya Ignatova, quien, junto a su grupo, desarrolló un protocolo de cuantificación de poblaciones de ARNt humanos mediante microarreglos (Kirchner et al., 2017). A su vez, la Dra. Ignatova disponía de chips de microarreglos diseñados para *Trypanosoma brucei*. Dado que estos chips funcionan mediante la hibridación de las muestras de ARNt marcados a sondas unidas a los chips, se realizó un análisis comparativo de las secuencias de ARNt de *T. cruzi* y *T. brucei*, mediante su alineamiento utilizando el *software* MUSCLE, con el fin de determinar si existirían diferencias que imposibilitarían el experimento. A modo de ejemplo, en la **Figura 11** se muestra el alineamiento de las secuencias del ARNt para Alanina y Treonina. Como se observa en la figura, las secuencias tienen un gran porcentaje de similitud entre ellas, por lo que se decidió proseguir con el estudio.

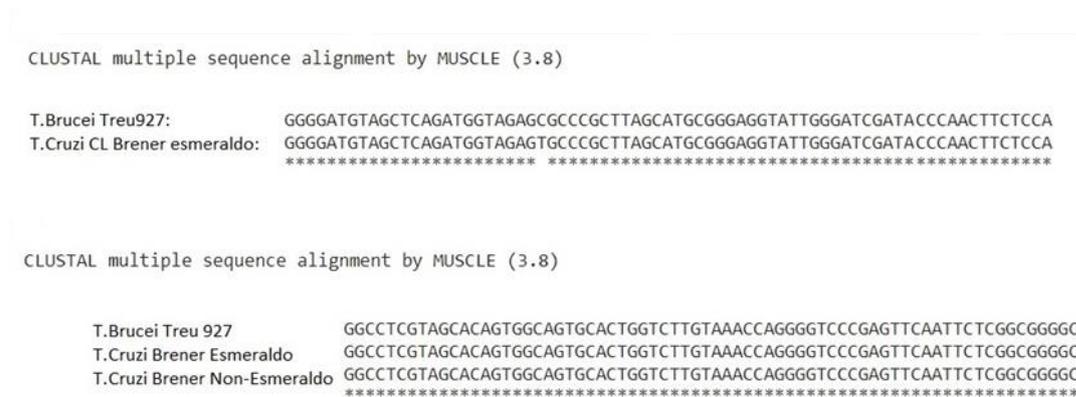


Figura 11. Ejemplos de alineamiento de secuencias de ARNt de *T. cruzi* y *T. brucei*, para Alanina (arriba) y Treonina (abajo). Se observa un muy alto porcentaje de similitud entre ellas.

Lo siguiente que se realizó fue el cultivo de parásitos epimastigotas y tripomastigotas metacíclicos con duplicados, y la extracción de ARN total. Una vez extraído el ARN, con el fin de garantizar la buena calidad de las muestras, se analizaron las mismas en un Agilent 2100 Bioanalyzer, cuyos resultados se resumen en la **Figura 12**. Como se aprecia

en la Figura 12, el ARN se encuentra íntegro, se observan las 3 bandas características correspondientes a los ARN ribosomales a los 2000 pb aproximadamente.

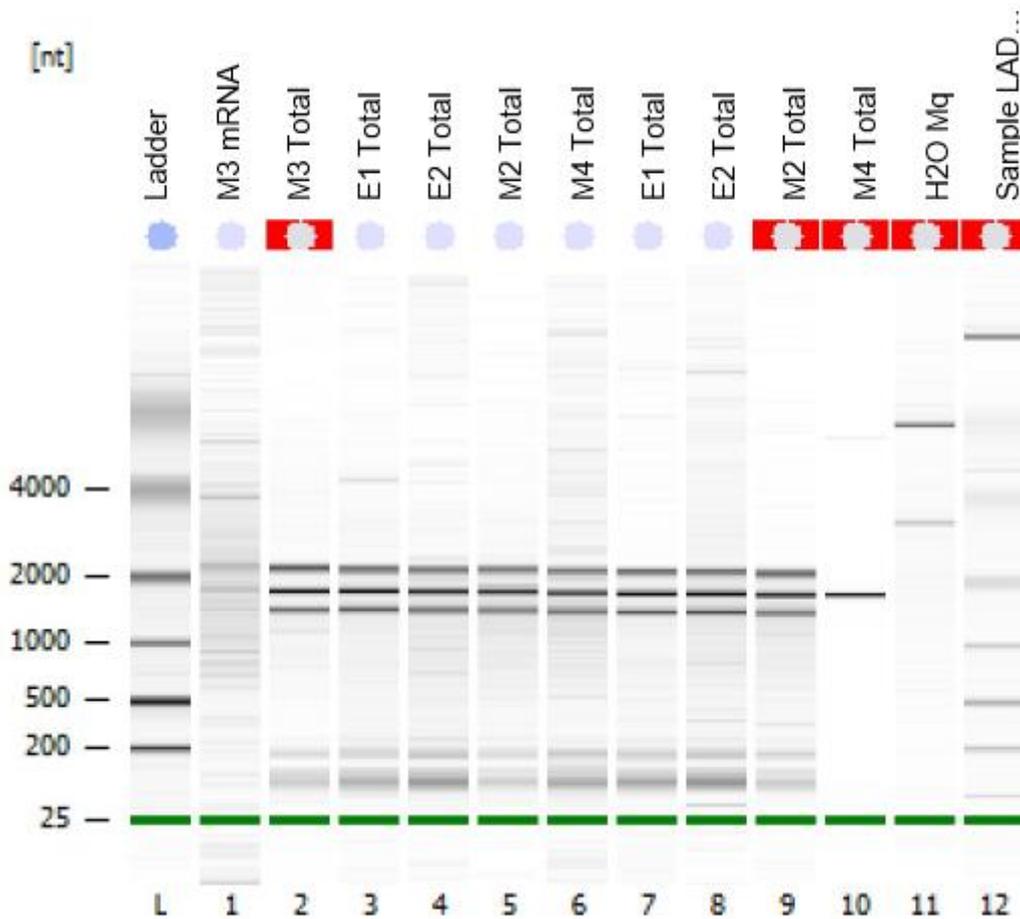


Figura 12. Resultados del control de calidad de las muestras de ARN realizado con un Agilent 2100 Bioanalyzer. Se muestran los duplicados de epimastigotas (E1 y E2, carriles 3 y 4) y sus réplicas técnicas (carriles 7 y 8), y los duplicados de tripomastigotas metacíclicos (M2 y M4, carriles 5 y 6) y sus réplicas técnicas (carriles 9 y 10).

Una vez verificada la calidad de las muestras, las mismas fueron enviadas al laboratorio de la Dra. Ignatova para la cuantificación de los ARNt. Lamentablemente, hubo inconvenientes durante el marcado de estas moléculas previo a la hibridización en el chip de microarreglos, por lo que no se obtuvieron resultados. Continuaremos con esta colaboración esperando obtener resultados satisfactorios en el futuro.

Por otro lado, se planteó utilizar estrategias alternativas a los microarreglos para la cuantificación de ARNt, específicamente mediante secuenciación masiva. Actualmente existen diversos protocolos específicos de secuenciación masiva de ARNt, cada uno con sus particularidades y limitaciones, tales como DM-tRNA-seq (Zheng et al., 2015),

YAMAT-seq (Shigematsu et al., 2017) y QuantM-tRNA-seq (Pinkard et al., 2020). Cada uno de ellos presenta distintas alternativas metodológicas para la construcción de las librerías, de modo de sortear las dificultades asociadas a la naturaleza de los ARNt, tales como el uso de demetilinas con el fin de remover este tipo de modificaciones de las bases nucleotídicas, el uso de retrotranscriptasa termoestables de alta procesividad durante la retrotranscripción a ADNc, o el uso de adaptadores particulares durante la construcción de la librería que hacen más eficiente su ligación y la posterior secuenciación.

Actualmente se están poniendo a punto nuevamente los cultivos celulares, con el fin de realizar nuevas extracciones de ARN y utilizar las dos metodologías propuestas para la cuantificación de las poblaciones de ARNt.

4.4 CORRELACIÓN ENTRE CODONES DIFERENCIALMENTE REPRESENTADOS EN SITIOS DE ENLENTECIMIENTO RIBOSOMAL Y NIVELES ESPERADOS DE ARNT

Habiendo obtenido determinados codones con una representación diferencial en los sitios de pausa ribosomal con respecto a su uso en el genoma y en los genes de alta expresión, y el número de copias de los genes de ARNt, se realizó una comparación entre ambos, con el fin de evaluar la hipótesis de que codones sobrerrepresentados en los sitios de pausa podrían estar asociados a ARNt menos frecuentes en la población celular de estas moléculas, que determinarían el enlentecimiento en el movimiento del ribosoma. Los resultados se resumen en la **Tabla 5**.

Se observa que todos los codones, a excepción de CTC codificante para L, cumplen con el criterio de que si están sobrerrepresentados en sitios de pausa ribosomal son menos usados en los genes de alta expresión, y viceversa. Por otro lado, se observa que, para el caso de C, V, Y, F, y L, los codones sobre y subrepresentados están vinculados con el número de copias de su ARNt correspondiente. Sin embargo, resulta imprescindible la cuantificación mediante microarreglos o secuenciación masiva, experimentos que permitirán establecer este vínculo de forma más precisa.

Tabla 5. Resumen de información sobre la frecuencia en pausa, frecuencia en genes de alta expresión y número de copias de ARNt asociados, de los codones con representación diferencial significativa en sitios de pausa ribosomal.

Aminoácido	Codón	Frecuencia en Pausas	Frecuencia en Genes de Alta Expresión	N° de copias de ARNt asociado	¿Cumple con todos los criterios?
C	TGT	↑	↓	Menor	✓
C	TGC	↓	↑	Mayor	✓
V	GTG	↓	↑	Mayor	✓
V	GTC	↑	↓	Menor	✓
Y	TAT	↑	↓	Menor	✓
Y	TAC	↓	↑	Mayor	✓
F	TTC	↓	↑	Mayor	✓
F	TTT	↑	↓	Menor	✓
G	GGC	↓	↑	Mayor	X
K	AAA	↑	↓	Mayor	X
K	AAG	↓	↑	Menor	X

R	CGC	↓	↑	Menor	X
A	GCG	↓	↑	Menor	X
P	CCA	↑	↓	Intermedio	X
P	CCT	↓	↑	Intermedio	X
L	CTA	↑	↓	Menor	✓
L	CTC	↓	↓	Menor	X

A su vez, se observa que los cambios nucleotídicos de los codones sinónimos son en la tercera posición, donde los codones más frecuentes en pausas son ricos en AT, mientras que los más frecuentes en genes de alta expresión son ricos en GC. Esto concuerda con reportes previos sobre las preferencias de codones en *T. cruzi* según sus niveles de expresión génica (Alvarez et al., 1994).

5 CONCLUSIONES

Como conclusiones finales, en este trabajo:

- Se realizó un estudio de la composición de los codones a nivel genómico en *T. cruzi* y se observó evidencia de selección en la determinación del uso de codones sinónimos en los genes.
- Se logró estimar los niveles de expresión génica utilizando datos de transcriptómica y traductómica.
- Se logró verificar los perfiles de expresión génica mediante qPCR, utilizando un grupo de genes seleccionados del análisis previo.
- Se correlacionó el uso diferencial de codones con los niveles de expresión génica de *T. cruzi*.
- Se logró mejorar la anotación de genes de ARNt de *T. cruzi* y *T. brucei*.
- Se identificaron sitios de enlentecimiento o pausa ribosomal a partir de datos de *RNA-seq* y *Ribo-seq*.
- Se determinaron codones diferencialmente representados en los sitios de enlentecimiento o pausa ribosomal.
- Para determinados codones se logró correlacionar el enriquecimiento en pausas ribosomales con los niveles esperados de ARNt.

6 PERSPECTIVAS

- Realizar un estudio de búsqueda de sitios de enlentecimiento ribosomal y codones asociados utilizando datos de *Ribo-seq* del estadio tripomastigota metacíclico.
- Cuantificar poblaciones de ARNt en los diferentes estadios del parásito mediante microarreglos y secuenciación masiva.
- Correlacionar codones sobre y subrepresentados en sitios de pausa ribosomal con los niveles de expresión de ARNt de *T. cruzi*.
- Estudiar la asociación entre el uso de codones y la estabilidad de los ARNm de *T. cruzi*, y correlacionarla con los codones sobrerrepresentados en sitios de pausa ribosomal.

7 BIBLIOGRAFÍA

- Alvarez, F., Robeilo, C., & Vignalp, M. (n.d.). *Evolution of Codon Usage and Base Contents in Kinetoplastid Protozoans*. Retrieved August 31, 2020, from <https://academic.oup.com/mbe/article/11/5/790/1008717>
- Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*.
- Angov, E. (2011). Codon usage: Nature's roadmap to expression and folding of proteins. In *Biotechnology Journal* (Vol. 6, Issue 6, pp. 650–659). Wiley-Blackwell. <https://doi.org/10.1002/biot.201000332>
- Araújo, P. R., & Teixeira, S. M. (2011). Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in trypanosoma cruzi - A review. In *Memorias do Instituto Oswaldo Cruz* (Vol. 106, Issue 3, pp. 257–266). Fundacao Oswaldo Cruz. <https://doi.org/10.1590/S0074-02762011000300002>
- Aslett, M., Aurrecochea, C., Berriman, M., Brestelli, J., Brunk, B. P., Carrington, M., Depledge, D. P., Fischer, S., Gajria, B., Gao, X., Gardner, M. J., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Hertz-Fowler, C., Houston, R., Innamorato, F., Iodice, J., ... Wang, H. (n.d.). *TriTrypDB: a functional genomic resource for the Trypanosomatidae*. <https://doi.org/10.1093/nar/gkp851>
- Aslett, M., Aurrecochea, C., Berriman, M., Brestelli, J., Brunk, B. P., Carrington, M., Depledge, D. P., Fischer, S., Gajria, B., Gao, X., Gardner, M. J., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Hertz-Fowler, C., Houston, R., Innamorato, F., Iodice, J., ... Wang, H. (2009). TriTrypDB: A functional genomic resource for the Trypanosomatidae. *Nucleic Acids Research*, 38(SUPPL.1), D457. <https://doi.org/10.1093/nar/gkp851>
- Bangs, J., Crain, P., Hashizume, T., McCloskey, J., & Boothroyd, J. (1992). Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. *Undefined*.
- Barbieri Holetz, F., Correa, A., Rodrigues, A., Vila, A. ´, Nakamura, V., Krieger, M. A., & Goldenberg, S. (2007). *Evidence of P-body-like structures in Trypanosoma cruzi*. <https://doi.org/10.1016/j.bbrc.2007.03.104>

- Berná, L., Rodriguez, 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>
- Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M.-A., Barrell, B. G., & Parkhill, J. (2005). ACT: the Artemis comparison tool. *BIOINFORMATICS APPLICATIONS NOTE*, 21(16), 3422–3423.
<https://doi.org/10.1093/bioinformatics/bti553>
- Castro, J. A., De Mecca, M. M., & Bartel, L. C. (2006). Toxic side effects of drugs used to treat Chagas' disease (American trypanosomiasis). *Human and Experimental Toxicology*, 25(8), 471–479. <https://doi.org/10.1191/0960327106het653oa>
- Chagas disease (American trypanosomiasis)*. (n.d.). Retrieved November 19, 2020, from https://www.who.int/health-topics/chagas-disease#tab=tab_1
- Charif, D., Thioulouse, J., Lobry, J. R., & Perrière, G. (2005). Online synonymous codon usage analyses with the ade4 and seqinR packages. *Bioinformatics*, 21(4), 545–547. <https://doi.org/10.1093/bioinformatics/bti037>
- Clayton, C., & Shapira, M. (2007). Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. In *Molecular and Biochemical Parasitology* (Vol. 156, Issue 2, pp. 93–101). <https://doi.org/10.1016/j.molbiopara.2007.07.007>
- Collart, M. A., & Weiss, B. (2020). Ribosome pausing, a dangerous necessity for co-translational events. In *Nucleic acids research* (Vol. 48, Issue 3, pp. 1043–1055). NLM (Medline). <https://doi.org/10.1093/nar/gkz763>
- Contreras, V. T., Araujo-Jorge, T. C., Bonaldo, M. C., Thomaz, N., Barbosa, H. S., Meirelles, M. N., & Goldenberg, S. (1988). Biological aspects of the Dm 28c clone of *Trypanosoma cruzi* after metacyclogenesis in chemically defined media. *Memórias Do Instituto Oswaldo Cruz*, 83(1), 123–133.
<https://doi.org/10.1590/S0074-02761988000100016>
- Coughlin, B. C., Teixeira, S. M. R., Kirchhoff, L. V., & Donelson, J. E. (2000). Amastin mRNA abundance in *Trypanosoma cruzi* is controlled by a 3'-

- untranslated region position-dependent cis-element and an untranslated region-binding protein. *Journal of Biological Chemistry*, 275(16), 12051–12060.
<https://doi.org/10.1074/jbc.275.16.12051>
- Daniels, J.-P., Gull, K., & Wickstead, B. (2010). Cell Biology of the Trypanosome Genome. *MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS*, 74(4), 1092–2172. <https://doi.org/10.1128/MMBR.00024-10>
- de Souza, W. (2009). Structural organization of *Trypanosoma cruzi*. *Memorias Do Instituto Oswaldo Cruz*, 104(SUPPL. 1), 89–100. <https://doi.org/10.1590/s0074-02762009000900014>
- De Souza, W. (1984). Cell Biology of *Trypanosoma cruzi*. *International Review of Cytology*, 86(C), 197–283. [https://doi.org/10.1016/S0074-7696\(08\)60180-1](https://doi.org/10.1016/S0074-7696(08)60180-1)
- Di Noia, J. M., D’Orso, I., Sánchez, D. O., & Frasch, A. C. C. (2000). AU-rich elements in the 3’-untranslated region of a new mucin-type gene family of *Trypanosoma cruzi* confers mRNA instability and modulates translation efficiency. *Journal of Biological Chemistry*, 275(14), 10218–10227.
<https://doi.org/10.1074/jbc.275.14.10218>
- Docampo, R., & Moreno, S. N. J. (2011). Acidocalcisomes. In *Cell Calcium* (Vol. 50, Issue 2, pp. 113–119). Elsevier Ltd. <https://doi.org/10.1016/j.ceca.2011.05.012>
- Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5(1), 113.
<https://doi.org/10.1186/1471-2105-5-113>
- 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>
- Fang, H., Huang, Y. F., Radhakrishnan, A., Siepel, A., Lyon, G. J., & Schatz, M. C. (2018). Scikit-ribo Enables Accurate Estimation and Robust Modeling of

- Translation Dynamics at Codon Resolution. *Cell Systems*, 6(2), 180-191.e4.
<https://doi.org/10.1016/j.cels.2017.12.007>
- Hershberg, R., & Petrov, D. (2008). Selection on Codon Bias Comparative population genomics of latitudinal variation in *Drosophila simulans* and *Drosophila melanogaster* View project Adaptive dynamics of cuticular hydrocarbons in *Drosophila* View project. *Article in Annual Review of Genetics*.
<https://doi.org/10.1146/annurev.genet.42.110807.091442>
- Horn, D. (2008a). Codon usage suggests that translational selection has a major impact on protein expression in trypanosomatids. *BMC Genomics*, 9, 1–11.
<https://doi.org/10.1186/1471-2164-9-2>
- Horn, D. (2008b). Codon usage suggests that translational selection has a major impact on protein expression in trypanosomatids. *BMC Genomics*, 9(1), 1–11.
<https://doi.org/10.1186/1471-2164-9-2>
- Ihaka, R., & Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299–314.
<https://doi.org/10.1080/10618600.1996.10474713>
- Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., & Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, 324(5924), 218–223.
<https://doi.org/10.1126/science.1168978>
- Jeacock, L., Faria, J., & Horn, D. (2018). Codon usage bias controls mRNA and protein abundance in trypanosomatids. *ELife*, 7, 1–20. <https://doi.org/10.7554/eLife.32496>
- Kirchner, S., Rauscher, R., & Czech, A. (2017). *Microarray-Based Quantification of Cellular tRNAs Epigenetic regulation of stage transitions and commitment to sexual development in Plasmodium* View project *Investigation into the role of silent mutations in CFTR biogenesis* View project.
<https://doi.org/10.17504/protocols.io.hfcb3iw>
- Kramer, S. (2012). Developmental regulation of gene expression in the absence of transcriptional control: The case of kinetoplastids. In *Molecular and Biochemical*

- Parasitology* (Vol. 181, Issue 2, pp. 61–72). Elsevier.
<https://doi.org/10.1016/j.molbiopara.2011.10.002>
- Laslett, D., & Canback, B. (n.d.). *ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences*. <https://doi.org/10.1093/nar/gkh152>
- Lesnik, T., Solomovici, J., Deana, A., Ehrlich, R., & Reiss, C. (2000). Ribosome traffic in *E. coli* and regulation of gene expression. *Journal of Theoretical Biology*, 202(2), 175–185. <https://doi.org/10.1006/jtbi.1999.1047>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). *Sequence analysis featureCounts: an efficient general purpose program for assigning sequence reads to genomic features*. 30(7), 923–930. <https://doi.org/10.1093/bioinformatics/btt656>
- Michels, P. A. M., Bringaud, F., Herman, M., & Hannaert, V. (2006). Metabolic functions of glycosomes in trypanosomatids. In *Biochimica et Biophysica Acta - Molecular Cell Research* (Vol. 1763, Issue 12, pp. 1463–1477). Elsevier.
<https://doi.org/10.1016/j.bbamcr.2006.08.019>
- Minning, T. A., Weatherly, D. B., Atwood, J., Orlando, R., & Tarleton, R. L. (2009). The steady-state transcriptome of the four major life-cycle stages of *Trypanosoma cruzi*. *BMC Genomics*, 10. <https://doi.org/10.1186/1471-2164-10-370>
- Miranda, K., Benchimol, M., Docampo, R., & De Souza, W. (2000). The fine structure of acidocalcisomes in *Trypanosoma cruzi*. *Parasitology Research*, 86(5), 373–384. <https://doi.org/10.1007/s004360050682>
- Müller, L. S. M., Cosentino, R. O., Förstner, K. U., Guizetti, J., Wedel, C., Kaplan, N., Janzen, C. J., Arampatzi, P., Vogel, J., Steinbiss, S., Otto, T. D., Saliba, A. E., Sebra, R. P., & Siegel, T. N. (2018). Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature*, 563(7729), 121–125. <https://doi.org/10.1038/s41586-018-0619-8>
- Nascimento, J. de F., Kelly, S., Sunter, J., & Carrington, M. (2018). Codon choice directs constitutive mRNA levels in trypanosomes. *ELife*, 7, 1–26. <https://doi.org/10.7554/eLife.32467>

- Palenchar, J. B., & Bellofatto, V. (2006). Gene transcription in trypanosomes. *Molecular & Biochemical Parasitology*, *146*, 135–141.
<https://doi.org/10.1016/j.molbiopara.2005.12.008>
- Phizicky, E. M., & Hopper, A. K. (2010). tRNA biology charges to the front. In *Genes and Development* (Vol. 24, Issue 17, pp. 1832–1860).
<https://doi.org/10.1101/gad.1956510>
- Pinkard, O., McFarland, S., Sweet, T., & Collier, J. (n.d.). *Quantitative tRNA-sequencing uncovers metazoan tissue-specific tRNA regulation*.
<https://doi.org/10.1038/s41467-020-17879-x>
- Plotkin, J. B., & Kudla, G. (2011a). Synonymous but not the same: The causes and consequences of codon bias. In *Nature Reviews Genetics* (Vol. 12, Issue 1, pp. 32–42). Nature Publishing Group. <https://doi.org/10.1038/nrg2899>
- Plotkin, J. B., & Kudla, G. (2011b). Synonymous but not the same: The causes and consequences of codon bias. In *Nature Reviews Genetics* (Vol. 12, Issue 1, pp. 32–42). Nature Publishing Group. <https://doi.org/10.1038/nrg2899>
- Presnyak, V., Alhusaini, N., Chen, Y. H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K. E., Graveley, B. R., & Collier, J. (2015). Codon optimality is a major determinant of mRNA stability. *Cell*, *160*(6), 1111–1124.
<https://doi.org/10.1016/j.cell.2015.02.029>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842.
<https://doi.org/10.1093/bioinformatics/btq033>
- Radhakrishnan, A., Chen, Y. H., Martin, S., Alhusaini, N., Green, R., & Collier, J. (2016). The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. *Cell*, *167*(1), 122–132.e9.
<https://doi.org/10.1016/j.cell.2016.08.053>
- Rassi, A., Rassi, A., & Marin-Neto, J. A. (2010). Chagas disease. In *The Lancet* (Vol. 375, Issue 9723, pp. 1388–1402). [https://doi.org/10.1016/S0140-6736\(10\)60061-X](https://doi.org/10.1016/S0140-6736(10)60061-X)

- Rohloff, P., Montalvetti, A., & Docampo, R. (2004). Acidocalcisomes and the contractile vacuole complex are involved in osmoregulation in *Trypanosoma cruzi*. *Journal of Biological Chemistry*, 279(50), 52270–52281. <https://doi.org/10.1074/jbc.M410372200>
- Schenkman, S., & Pascoalino, B. (2011). Nuclear Structure of *Trypanosoma cruzi* Toxoplasma epigenetics View project New Medicine for Trypanosomatidic Infections _ FP7 Research&Innovation project View project. *Article in Advances in Parasitology*. <https://doi.org/10.1016/B978-0-12-385863-4.00012-5>
- Shigematsu, M., Honda, S., Loher, P., Telonis, A. G., Rigoutsos, I., & Kirino, Y. (2017). YAMAT-seq: An efficient method for high-throughput sequencing of mature transfer RNAs. *Nucleic Acids Research*, 45(9), e70. <https://doi.org/10.1093/nar/gkx005>
- Simpson, A. G. B., Stevens, J. R., & Lukeš, J. (2006). The evolution and diversity of kinetoplastid flagellates. In *Trends in Parasitology* (Vol. 22, Issue 4, pp. 168–174). Elsevier Current Trends. <https://doi.org/10.1016/j.pt.2006.02.006>
- Smircich, P. (n.d.). *Aproximaciones de genómica estructural y funcional en tripanosomátidos*.
- Smircich, P., Eastman, G., Bispo, S., Duhagon, M. A., Guerra-Slompo, E. P., Garat, B., Goldenberg, S., Munroe, D. J., Dallagiovanna, B., Holetz, F., & Sotelo-Silveira, J. R. (2015). Ribosome profiling reveals translation control as a key mechanism generating differential gene expression in *Trypanosoma cruzi*. *BMC Genomics*, 16(1), 1–14. <https://doi.org/10.1186/s12864-015-1563-8>
- Smith, D. F., & Parsons, M. (1996). *Molecular biology of parasitic protozoa*. IRL Press at Oxford University Press. <https://agris.fao.org/agris-search/search.do?recordID=US201300300462>
- Teixeira, A. R. L., Calixto, M. A., & Teixeira, M. L. (1994). Chagas' disease: carcinogenic activity of the antitrypanosomal nitroarenes in mice. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, 305(2), 189–196. [https://doi.org/10.1016/0027-5107\(94\)90239-9](https://doi.org/10.1016/0027-5107(94)90239-9)

- Thanaraj, T. A., & Argos, P. (1996). Ribosome-mediated translational pause and protein domain organization. *Protein Science*, 5(8), 1594–1612.
<https://doi.org/10.1002/pro.5560050814>
- Thioulouse, J., Chessel, D., Dolédec, S., & Olivier, J. M. (1997). ADE-4: A multivariate analysis and graphical display software. *Statistics and Computing*, 7(1), 75–83.
<https://doi.org/10.1023/A:1018513530268>
- Ttir, H., Robinson, J. T., & Mesirov, J. P. (n.d.). *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration*.
<https://doi.org/10.1093/bib/bbs017>
- Vanhamme, L., & Pays, E. (1995). Control of gene expression in trypanosomes. In *Microbiological Reviews* (Vol. 59, Issue 2, pp. 223–240). American Society for Microbiology. <https://doi.org/10.1128/membr.59.2.223-240.1995>
- Wright, F. (1990). The “effective number of codons” used in a gene. *Gene*, 87(1), 23–29. [https://doi.org/10.1016/0378-1119\(90\)90491-9](https://doi.org/10.1016/0378-1119(90)90491-9)
- Zheng, G., Qin, Y., Clark, W. C., Dai, Q., Yi, C., He, C., Lambowitz, A. M., & Pan, T. (2015). Efficient and quantitative high-throughput tRNA sequencing. *Nature Methods*, 12(9), 835–837. <https://doi.org/10.1038/nmeth.3478>
- Zinoviev, A., & Shapira, M. (2012). Evolutionary conservation and diversification of the translation initiation apparatus in trypanosomatids. In *Comparative and Functional Genomics* (Vol. 2012). <https://doi.org/10.1155/2012/813718>

8 ANEXO

Tabla 1 Anexo. Frecuencias de todos los codones sinónimos en Regiones de Pausa Ribosomal, Genes con Pausa Ribosomal, todos los genes del CDS y Genes de Alta Expresión. Se muestra también el p-valor asociado al test de contingencia de chi-cuadrado realizado con el conteo de los codones.

	Regiones de Pausa	Genes con Pausas	CDS	Genes de Alta Expresión	p-valor
Ala_GCT	0,2333	0,1988	0,1978	0,1748	0.1022439838752262
Ala_GCC	0,2923	0,2813	0,2834	0,3367	0.6694793457029784
Ala_GCA	0,2385	0,2271	0,2236	0,1609	0.6372086515377068
Ala_GCG	0,2359	0,2928	0,2952	0,3276	0.01625201940069993
Cys_TGT	0,5429	0,4537	0,4341	0,2569	0.0120582953954763
Cys_TGC	0,4571	0,5463	0,5659	0,7431	0.0120582953954763
Asp_GAT	0,5051	0,5221	0,5098	0,3425	0.6020842235290582
Asp_GAC	0,4949	0,4779	0,4902	0,6575	0.6020842235290582
Glu_GAA	0,4749	0,4265	0,408	0,2517	0.1319816822905481
Glu_GAG	0,5251	0,5735	0,592	0,7483	0.1319816822905481
Phe_TTT	0,6057	0,5373	0,5299	0,5107	0.02630556503166444
Phe_TTC	0,3943	0,4627	0,4701	0,4893	0.02630556503166444
Gly_GGT	0,2766	0,2447	0,2294	0,2502	0.20198965257005913
Gly_GGC	0,2492	0,3214	0,3362	0,4826	0.006279040880476994
Gly_GGA	0,2584	0,2254	0,2255	0,1387	0.1751224746702874
Gly_GGG	0,2158	0,2085	0,2089	0,1285	0.79665281978322
His_CAT	0,4982	0,4676	0,4497	0,2689	0.3331026775501303
His_CAC	0,5018	0,5324	0,5503	0,7311	0.3331026775501303
Ile_ATT	0,4538	0,4237	0,4134	0,4556	0.25252488074499907

Ile_ATC	0,3974	0,3944	0,4023	0,4864	0.9451536397729858
Ile_ATA	0,1488	0,1818	0,1843	0,058	0.10508157377906242
Lys_AAA	0,5489	0,4821	0,4701	0,1574	0.015343908942609176
Lys_AAG	0,4511	0,5179	0,5299	0,8426	0.015343908942609176
Leu_TTA	0,052	0,0411	0,0409	0,0259	0.20870591080047624
Leu_TTG	0,2098	0,1811	0,1718	0,1331	0.0744382478592948
Leu_CTT	0,2163	0,2436	0,2449	0,2288	0.12744002608248528
Leu_CTC	0,239	0,2889	0,286	0,2249	0.007455568452310888
Leu_CTA	0,0699	0,0306	0,0291	0,0251	4,01E+07
Leu_CTG	0,213	0,2147	0,2273	0,3621	0.9581632511525889
Met_AT G	1	1	1	1	0
Asn_AAT	0,5586	0,5443	0,5273	0,2686	0.6939132816787764
Asn_AAC	0,4414	0,4557	0,4727	0,7314	0.6939132816787764
Pro_CCT	0,1428	0,1913	0,1999	0,1654	0.037340635416232504
Pro_CCC	0,2468	0,2568	0,2563	0,3042	0.7360850637179098
Pro_CCA	0,3474	0,2697	0,2628	0,1966	0.002778576132108294
Pro_CCG	0,263	0,2821	0,281	0,3338	0.496500780812808
Gln_CAA	0,4955	0,4684	0,4359	0,226	0.3554256896701006
Gln_CAG	0,5045	0,5316	0,5641	0,774	0.3554256896701006
Arg_CGT	0,2588	0,2239	0,2155	0,2587	0.06733206367806509
Arg_CG C	0,1727	0,2444	0,2561	0,4761	0.00020722420909496397
Arg_CG A	0,1686	0,1397	0,1343	0,0558	0.07075620302207075
Arg_CG G	0,1784	0,1663	0,1667	0,0975	0.5004779619545183
Arg_AG A	0,1176	0,1152	0,114	0,0327	0.9158942507972254
Arg_AG G	0,1039	0,1105	0,1133	0,0791	0.6868707244794279
Ser_TCT	0,1422	0,1487	0,1485	0,1325	0.7612031868562363
Ser_TCC	0,1735	0,204	0,2073	0,2227	0.1409444347001875

Ser_TCA	0,0795	0,0954	0,0936	0,0959	0.3100945043849527
Ser_TCG	0,1638	0,1705	0,1721	0,2139	0.7698183508351988
Ser_AGT	0,2	0,1665	0,1604	0,093	0.07955814527706341
Ser_AGC	0,241	0,2149	0,218	0,2421	0.2193472499421949
Thr_ACT	0,2027	0,1934	0,1891	0,122	0.6997415521375441
Thr_ACC	0,2432	0,2483	0,254	0,2651	0.8687182881342833
Thr_ACA	0,2649	0,2292	0,2268	0,187	0.11826744100515386
Thr_ACG	0,2892	0,3291	0,3301	0,428	0.11623802037032943
Val_GTT	0,276	0,2556	0,2453	0,2107	0.32192906474733624
Val_GTC	0,31	0,2587	0,2587	0,1974	0.010751172339093477
Val_GTA	0,156	0,1786	0,1762	0,0484	0.21070023739002727
Val_GTG	0,258	0,3072	0,3198	0,5435	0.020208668571886086
Trp_TGG	1	1	1	1	0
Tyr_TAT	0,5354	0,4195	0,4207	0,2123	0.010982232139118173
Tyr_TAC	0,4646	0,5805	0,5793	0,7877	0.010982232139118173

Tabla 2 Anexo. Frecuencias de codones significativos en Regiones de Pausa Ribosomal que correlacionan con el tiempo de permanencia del ribosoma (*Dwell-time*) obtenido utilizando el software *Scikit-ribo*.

Aminoácido_Codón	Regiones de Pausa	Dwell-time
Ala_GCT	0,2333	1,1331
Ala_GCC	0,2923	0,9355
Ala_GCA	0,2385	1,1582
Ala_GCG	0,2359	0,9672
Cys_TGT	0,5429	1,0407
Cys_TGC	0,4571	0,7208

Phe_TTT	0,6057	1,0207
Phe_TTC	0,3943	0,9456
Gly_GGT	0,2766	1,0173
Gly_GGC	0,2492	0,8917
Gly_GGA	0,2584	0,9904
Gly_GGG	0,2158	0,8708
Arg_CGT	0,2588	1,1273
Arg_CGC	0,1727	0,8909
Arg_CGA	0,1686	1
Arg_CGG	0,1784	0,9915
Arg_AGA	0,1176	1,1134
Arg_AGG	0,1039	0,8824
Tyr_TAT	0,5354	1,1352
Tyr_TAC	0,4646	0,9141

Tabla 3 Anexo. N° de copias de ARNt en el genoma de *T. cruzi* reportadas por este estudio utilizando un ensamblado de lecturas largas obtenido mediante PacBio comparadas con el n° de copias reportadas por (Horn, 2008).

Anticodón	N° de copias (PacBio)	N° de copias (Horn)
Ala CGC	2	2
Ala TGC	1	1
Ala AGC	4	2
Ala GGC	0	0
Arg CCT	1	2
Arg TCT	1	2
Arg CCG	2	2
Arg TCG	1	2
Arg GCG	0	0

Arg ACG	6	4
Asn GTT	6	4
Asn ATT	2	0
Asp GTC	4	2
Asp ATC	2	0
Cys GCA	1	2
Cys ACA	0	0
Glu CTC	4	4
Glu TTC	2	1
Gln CTG	6	4
Gln TTG	2	2
Gly CCC	1	2
Gly TCC	2	2
Gly GCC	6	4
Gly ACC	0	0
His GTG	8	4
His ATG	0	0
Ileu TAT	2	3
Ileu GAT	0	0
Ileu AAT	8	4
Leu CAG	2	2
Leu TAG	1	2
Leu GAG	0	0
Leu AAG	2	4
Leu CAA	4	2
Leu TAA	4	2
Lys TTT	5	2
Lys CTT	4	4
Met CAT	5	6
Phe GAA	4	4

Phe AAA	0	0
Pro CGG	6	2
Pro TGG	3	2
Pro AGG	3	2
Pro GGG	0	0
Ser GCT	4	2
Ser ACT	0	0
Ser CGA	8	2
Ser TGA	1	2
Ser GGA	0	0
Ser AGA	4	2
Thr CGT	4	2
Thr TGT	2	2
Thr GGT	0	0
Thr AGT	1	2
Trp CCA	2	2
Tyr GTA	1	2
Tyr ATA	0	0
Val CAC	2	4
Val TAC	2	2
Val GAC	0	0
Val AAC	1	2