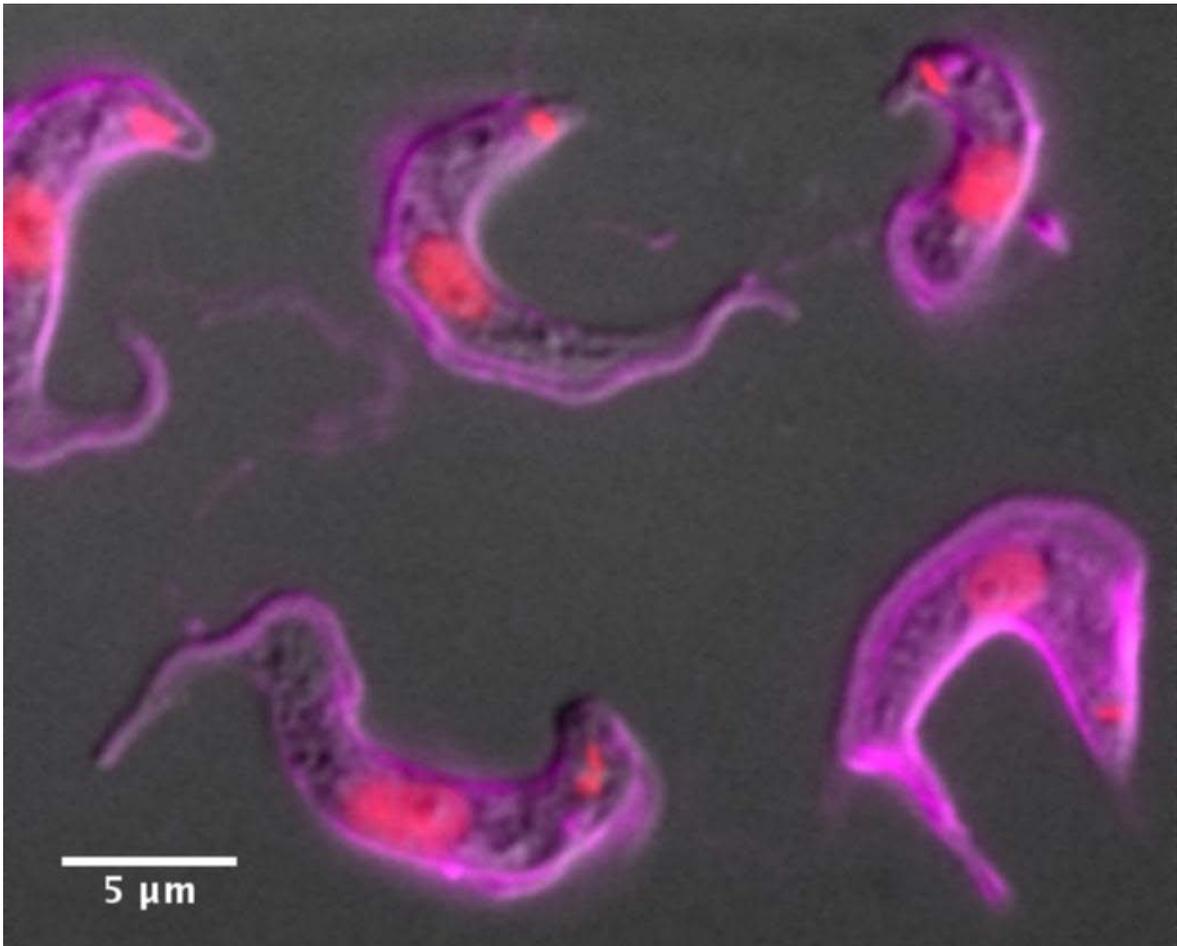


Estudio de la velocidad de evolución molecular en tripanosomátidos.



Forma del parásito *Trypanosoma brucei* en la corriente sanguínea teñido con marcadores para el ADN (rojo) y proteínas de superficie (magenta) y sobreesituado con la imagen DIC (Correlación de Imagen Digital).¹

Pasante: Uri Milstein
Licenciatura en Ciencias Biológicas
Tutor: Dr. Fernando Álvarez-Valín
Sección Biomatemática

TABLA DE CONTENIDO

INTRODUCCIÓN	3
Sistemática y biología molecular de los tripanosomátidos.....	3
Ciclos de vida y morfología de los tripanosomátidos.....	5
Evolución molecular en tripanosomátidos.....	11
Genómica comparativa de los tripanosomátidos.....	11
Evolución molecular, mutaciones, sustituciones y el reloj molecular.....	13
Filogenia de tripanosomátidos.....	20
OBJETIVOS	22
Objetivos generales.....	22
Objetivos específicos.....	22
MATERIALES Y MÉTODOS	23
Datos utilizados.....	23
Scripts.....	24
Árboles filogenéticos.....	28
Test de las tasas relativas.....	31
Ontología génica y búsqueda de causas funcionales de la aceleración.....	36
RESULTADOS	37
Grupos de genes ortólogos utilizados en este análisis.....	37
Test de tasas relativas de Tajima.....	38
Inferencias filogenéticas.....	40
Búsqueda de causas funcionales de la aceleración.....	43
DISCUSIÓN	46
Resumen y discusión general.....	46
La filogenia de los tripanosomas, el TTR de Tajima, y los métodos de reconstrucción de árboles filogenéticos.....	47
El test de tasa relativa de Tajima y la hipótesis del reloj molecular.....	47
Los genes imparálogos.....	47
Búsqueda de causas funcionales de la aceleración.....	48
BIBLIOGRAFÍA	48
GLOSARIO	52
APÉNDICE	54

INTRODUCCIÓN

Sistemática y biología molecular de los tripanosomátidos

La familia Trypanosomatidae está formada por protozoarios parásitos pertenecientes al Dominio Eucariota, Clase Euglenozoa, Orden Kinetoplástida (ver tabla 1)².

Tabla 1: Cuadro taxonómico de la familia Trypanosomatidae

Sistemática
Órden: Kinetoplastida
Subórden: Bodonina (con 2 flagelos)
Familia: Bodonidae
Familia: Cryptobiidae (parásitos de peces, caracoles)
Género: <i>Trypanoplasma</i>
Género: <i>Cryptobia</i>
Subórden: Trypanosomatina (con un simple flagelo)
Familia: Trypanosomatidae
Género: <i>Leptomonas</i>
Género: <i>Leishmania</i>
Género: <i>Phytomonas</i>
Género: <i>Crithidia</i>
Género: <i>Blastocrithidia</i>
Género: <i>Trypanosoma</i>
– Stercoraria (los tripanosomas son transmitidos por insectos vectores del grupo hemíptera, vía fecal)
– Salivaria (los tripanosomas son transmitidos por la saliva y/o o vía los órganos de alimentación de los dípteros, su vector)

Dentro de dicha familia encontramos a los géneros *Leishmania* y *Trypanosoma*, ambos taxones, con organismos eucariotas unicelulares, con una estructura celular muy peculiar, llamada kinetoplasto (ver figura 1).

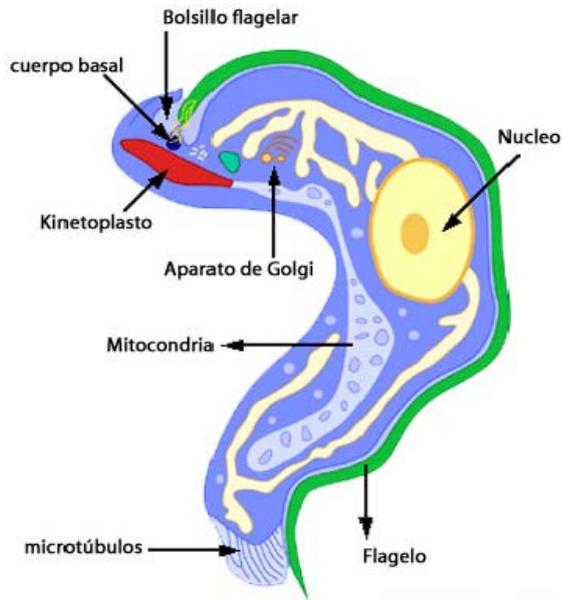


Figura 1: *Trypanosoma*. Note el kinetoplasto que presenta ³

Desde el punto de vista de su biología molecular, el kinetoplasto es una estructura que consiste en una malla de ADN localizada en una zona específica de la mitocondria (única en estos casos). Esta red de ADN representa una proporción importante del ADN total celular, que, dependiendo de la especie, puede contener del 10 al 20% del ADN total en la célula.⁴

Otras características importantes que tiene dicha familia son las siguientes: la no condensación de sus cromosomas durante la división mitótica, es decir, la cromatina se mantiene relativamente laxa durante todo el ciclo celular, el escaso intercambio sexual, (que en el caso de *Trypanosoma cruzi* llega a ser un comportamiento de población clonal)⁵ la variación antigénica, mecanismo único en la naturaleza de evasión de sistema inmune del mamífero hospedero (en los tripanosomas africanos)⁶ y los complejos mecanismos de regulación génica que presentan, los cuales son fundamentalmente post-transcripcionales.^{7, 8}

Ciclos de vida y morfología de los tripanosomátidos

Estos protozoarios presentan dos modalidades de ciclo parasitario, aquellos que requieren un solo hospedero en su ciclo de vida, llamados monogenéticos, y aquellos que tienen más de un hospedero en su ciclo vital, denominados heterogenéticos.

Entre los primeros, en general, el hospedero único es un insecto, como en los géneros *Crithidia* y *Leptomonas*. Entre los parásitos heterogenéticos se encuentra el género *Phytomonas*, que durante su ciclo parasita plantas e invertebrados. Y los parásitos digenéticos tripanosomátidos pertenecientes a los géneros *Trypanosoma* y *Leishmania*, que en general tienen como vector un insecto y parasitan una gran gama de hospederos definitivos, tales como aves, reptiles y mamíferos, incluido al hombre, razón por la cual, ambos géneros han sido ampliamente estudiados.

Durante su ciclo vital, los miembros pertenecientes a la familia Trypanosomatidae, adquieren diferentes morfologías, que varían en tamaño, localización del flagelo y ubicación del kinetoplasto (según el entorno en el cual el organismo se encuentre), a su vez, éstos, pueden ser extra o intracelulares (ver figura 2)

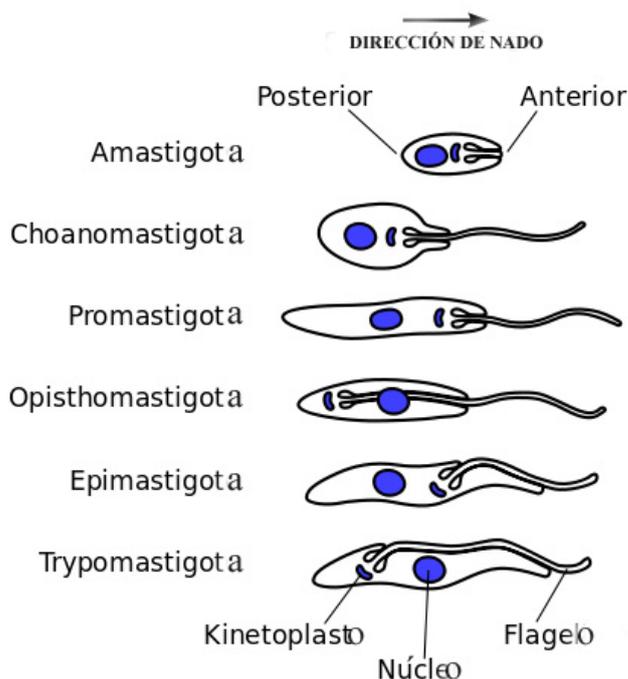


Figura 2

Las 6 principales morfologías que pueden adoptar los tripanosomátidos.⁹

El género *Trypanosoma* contiene 2 grandes grupos de importancia para la salud humana y veterinaria: Stercoraria y Salivaria.

El primero está integrado por especies cuya forma infectiva es desarrollada en la parte posterior del tracto digestivo del insecto vector, liberándose a través de las heces del insecto¹⁰. El conocido parásito humano característico de América, *Trypanosoma cruzi*, pertenece a este subgénero. Su hospedero intermediario son los hemípteros hematófagos triatomínios, conocidos popularmente como vinchucas y el definitivo es un mamífero, por ejemplo el hombre.

En la figura 3 se presenta el ciclo de vida de *T. cruzi*. Se puede observar como el parásito pasa durante su ciclo de vida por distintas etapas con marcadas diferencias morfológicas y bioquímicas.

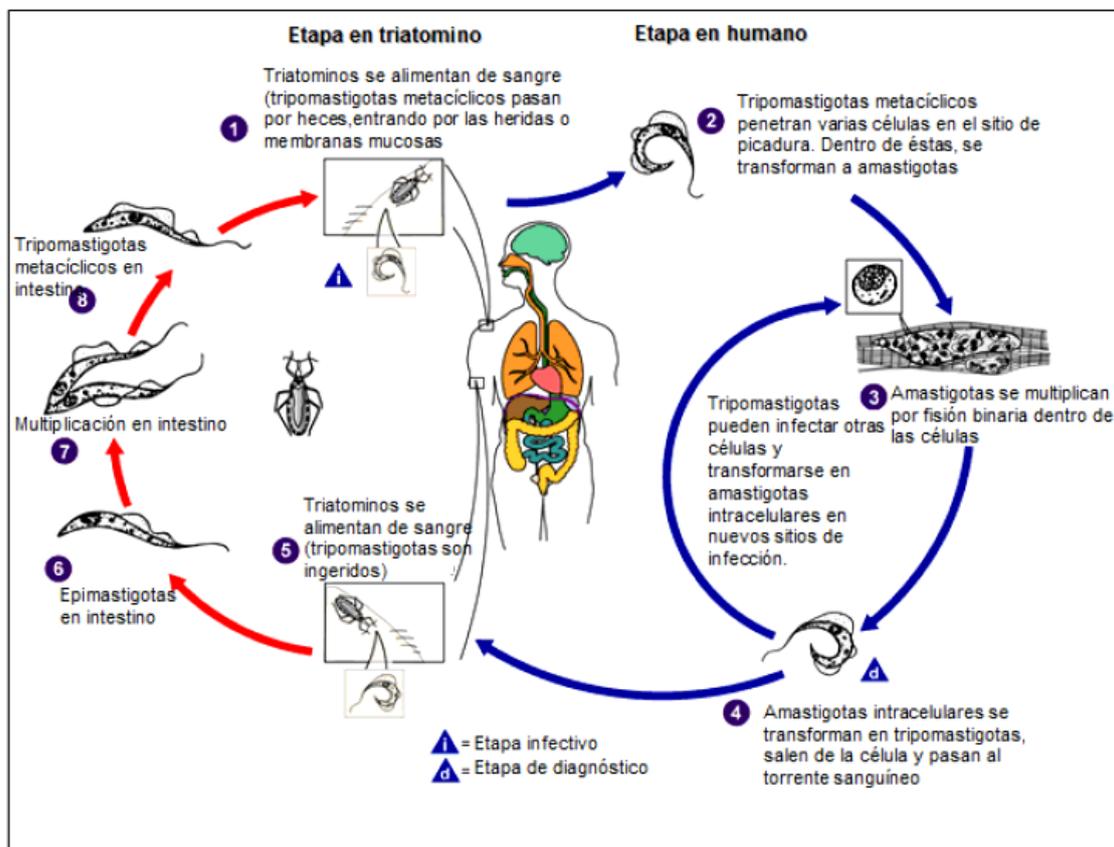


Figura 3: Ciclo de vida de *T. cruzi*¹¹

Un insecto vector triatomínio infectado, vinchuca, se alimenta de un hospedero, tomando sangre del mismo y liberando tripomastigotas en sus heces, cerca del sitio de mordida. Los tripomastigotas entran al huésped, a través de la mordida o las mucosas, como la conjuntiva. Dentro del huésped, los tripomastigotas invaden las células cerca

del sitio de inoculación, donde se diferencian a la forma amastigota intracelular. Los amastigotas se multiplican por fisión binaria y diferencian a tripomastigotas y luego son liberados a la circulación como tripomastigotas circulantes. Los tripomastigotas infectan células de una variedad de tejidos y se diferencian a amastigotas intracelulares en el nuevo sitio infectivo. La forma tripomastigota circulante de *T. cruzi* no se replica (a diferencia de la africana que sí lo hace). La replicación ocurre sólo cuando los parásitos entran a otra célula o son ingeridas por otro vector. La mordida del insecto se vuelve infectiva al alimentarse de sangre contaminada humana o animal que contiene parásitos circulantes.

Los tripomastigotas ingeridos se transforman a epimastigotas en el tracto digestivo medio del vector, forma morfológica caracterizada por tener el flagelo anterior al núcleo y conectado al cuerpo de la célula por una membrana ondulada corta (ver figura 2).

Luego el tripanosoma se multiplica y se ubica en la parte posterior del intestino del insecto diferenciándose a la forma infectiva tripomastigota metacíclica.

Por otro lado encontramos a los tripanosomátidos originarios de África, en especial al subgénero llamado *Salivaria*. Éstos se desarrollan en la parte anterior del tracto digestivo del insecto vector y se transmiten a través de la saliva (por eso la denominación de *Salivaria*) al alimentarse éste último de la sangre de los mamíferos¹⁰.

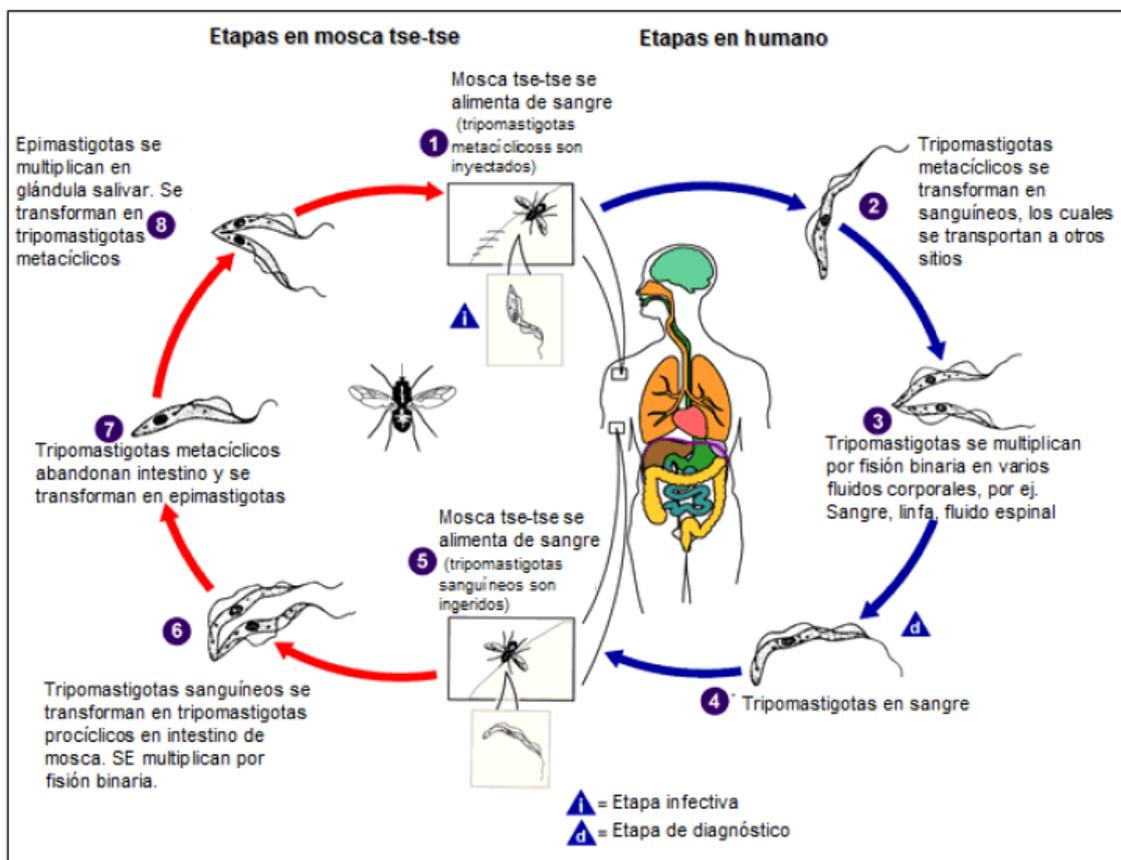


Figura 4: Ciclo de vida de *T. brucei*¹¹

La figura 4 muestra el ciclo de vida de *Trypanosoma brucei*, un integrante del grupo africano. Una mosca tsetse del género *Glossina*, infectada por *T. brucei*, inyecta tripanosomas en su forma tripomastigotas metacíclicos a un huésped mamífero al alimentarse de la sangre de este último.

Dentro del huésped, los parásitos tripanosomátidos, se transforman en tripomastigotas circulantes, son llevados a otros sitios del cuerpo, alcanzando los fluidos sanguíneos, la linfa y el fluido espinal y continuando su replicación por fisión binaria.

Todo el ciclo de vida de los tripanosomas africanos es representado por las etapas extracelulares. La mosca tsetse se vuelve infectada con tripomastigotas circulantes, al alimentarse de sangre de un huésped mamífero ya infectado.

En el tracto digestivo medio de la mosca, el parásito se transforma en tripomastigotas procíclicos, se multiplica por fisión binaria, deja el tracto digestivo, y se transforma a la forma epimastigota. Los epimastigotas dejan las glándulas salivares de la mosca y continúan su multiplicación por fisión binaria.

El ciclo en la mosca toma 3 semanas aproximadamente. Los humanos somos el principal reservorio para *T. brucei gambiense* aunque se encuentra también en los animales.

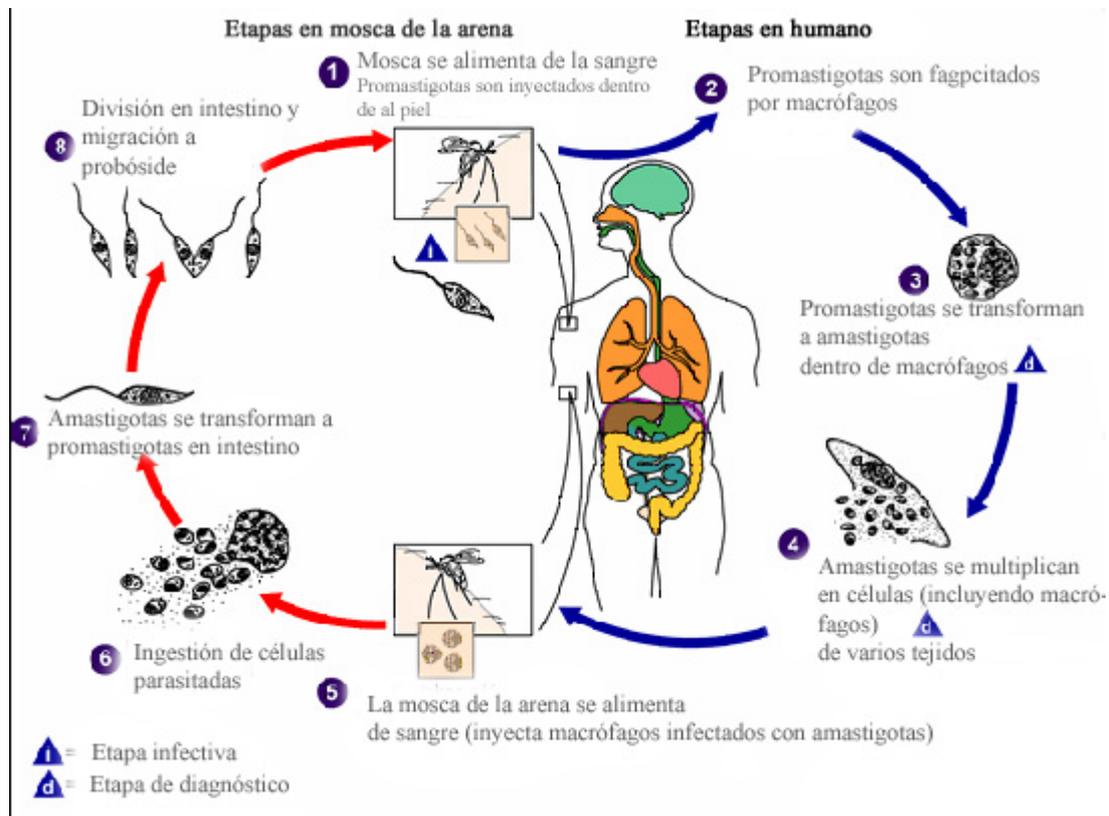


Figura 5: Ciclo de vida de *Leishmania*¹¹

La figura 5 muestra el ciclo de vida de *Leishmania*. La mosca de la arena inyecta la etapa infectiva, promastigota desde su probóscide durante su alimentación con sangre. Los promastigotas que alcanzan la punta de la herida son fagocitados por macrófagos, y otros tipos de células fagocíticas mononucleares. Dentro de estas células los promastigotas se transforman en amastigota, y se multiplican por simple división. La mosca se vuelve infectiva ingiriendo células infectadas al comer sangre de un hospedero. En la mosca de la arena, los amastigotas se transforman en promastigotas, y se desarrollan en la parte media del tracto digestivo, y migran a la probóscide. En la tabla 2 se resume diferentes enfermedades en humanos y animales relacionadas con estos parásitos. (Se hace mención únicamente de las especies tratadas en el trabajo)

Tabla 2. Algunas enfermedades causadas por parásitos tripanosomátidos

Enfermedad	Agente causante	Zona geográfica	Vector	Hospedero
Leishmaniasis visceral	<i>Leishmania infatum</i>	África Europa Sudamérica	Mosca de la arena	Humanos Perros
Leishmaniasis cutánea (forma húmeda)	<i>Leishmania major</i>	África del Norte Europa del Sur Oriente Medio Asia Central	Mosca de la arena (géneros <i>Lutzomyia</i> y <i>Phlebotomus</i>)	Humanos Roedores
Leishmaniasis mucocutánea Espundia	<i>Leishmania braziliensis</i>	América del Sur América Central	Mosca de la arena	Humanos Roedores
Nagana	<i>Trypanosoma brucei brucei</i>	África	Mosca tsetse (género <i>Glossina</i>)	Ganado Caballos
Tripanosomiasis Nagana	<i>Trypanosoma congolense</i>	África	Mosca tsetse (género <i>Glossina</i>)	Ganado
Tripanosomiasis Nagana	<i>Trypanosoma vivax</i>	África	Mosca tsetse	
Tripanosomiasis Africana del Oeste	<i>Trypanosoma brucei gambiense</i>	Parte Oeste de África	Mosca tsetse (<i>Glossina palpalis</i>)	Humanos (Enfermedad del sueño)
Tripanosomiasis americana (enfermedad de Chagas)	<i>Trypanosoma cruzi</i>	Sudamérica Y América Central	Triatomínidos o vinchucas (<i>Rhodnius</i> spp)	Humanos Animales salvajes

Tanto *T. cruzi*, *T. brucei*, como *L. major*, son transmitidos por diferentes insectos y cada uno posee su ciclo de vida característico (ver figura 6), así como afectan a diferentes tejidos y provocan distintas enfermedades en sus huéspedes mamíferos correspondientes (El-Sayed et al.).

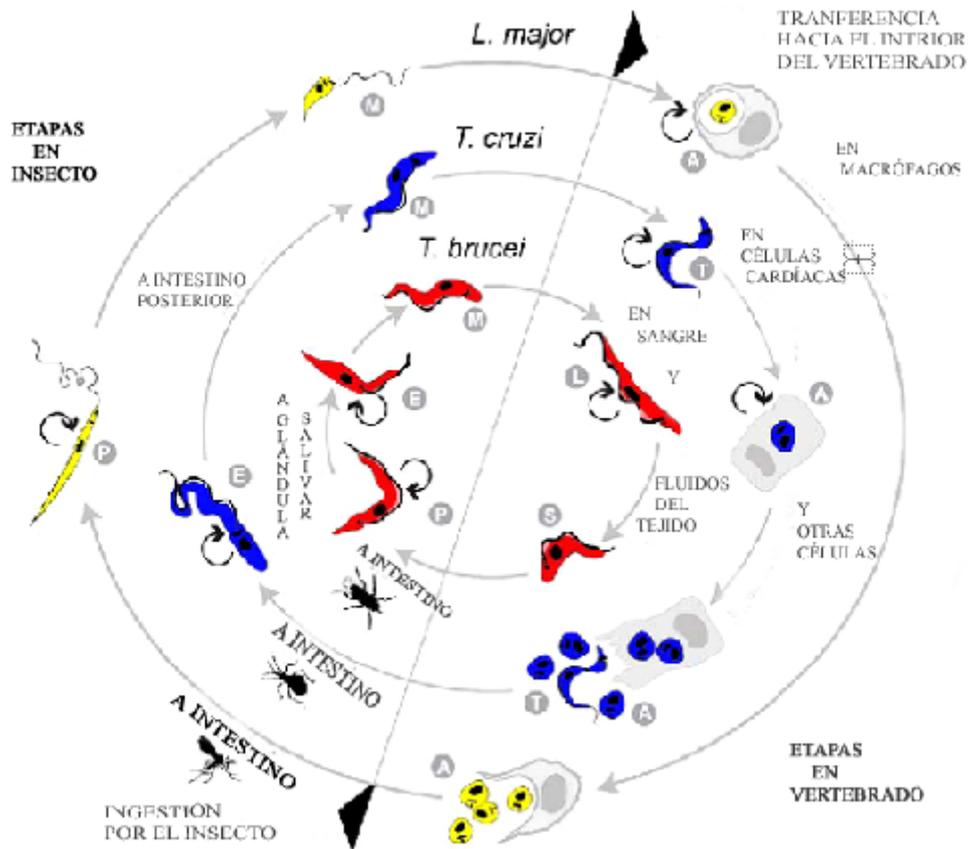


Figura 6: Comparación de los ciclos de vida de *L. major*, *T. cruzi* y *T. brucei*.¹¹

Evolución molecular en tripanosomátidos

Genómica comparativa de los tripanosomátidos

La disponibilidad de las secuencias de genomas completos de varias especies hace posible la comparación entre genomas y de esta manera nos empieza a permitir responder preguntas tales como qué genes son compartidos entre 2 especies, qué proteínas homólogas comparten diferentes phyla y cuáles son únicas de un phylum determinado¹². También permite comparar la localización de genes homólogos entre especies y si dicha localización se conserva (conservación de sintenia) entre otras posibilidades. Esta área de investigación, conocida como genómica comparativa, comenzó su desarrollo producto de los proyectos de secuenciación genómica iniciados hace relativamente poco tiempo, como por ejemplo el genoma de *Haemophilus influenzae*, en 1995, el genoma humano completo en el 2003 entre otros proyectos que le siguieron. El advenimiento de nuevas tecnologías de secuenciación de ultra-rendimiento a partir del 2005 ha hecho de esta área de la biología una de las más dinámicas.

Entre estos proyectos genómicos, se encontraron, los que permitieron realizar este trabajo de pasantía, nos referimos a los tres tripanosomátidos con mayor importancia para la salud humana: *Trypanosoma brucei*, *Trypanosoma cruzi*, y *Leishmania major* publicados en el año 2005.^{6,13, 14}

Los genomas de las 3 especies recién mencionadas, se caracterizan por contener entre 25 a 55 Mpb distribuidos sobre 11 a 36 cromosomas, los cuales codifican para entre 8300 a 12.000 genes, aproximadamente. Las características genómicas generales de estos organismos tales como tamaño, contenido GC, composición génica, conservación de sintenia, etc., se presentan en la tabla 3. Un aspecto a remarcar es que estas tres especies comparten un grupo de 6158 genes ortólogos, 5812 ubicados en la región sinténica y 346 en la no sinténica¹⁵ (ver también figura 12 más adelante). Estos genes ortólogos compartidos por los tres grupos de tripanosomátidos, de importancia para la salud humana, han permitido extraer varias conclusiones acerca de las funciones básicas comunes de ellos, tales como, vías metabólicas, aspectos de estructura celular general, composición de membranas, etc..

Tabla 3: Características generales de los genomas de *T. brucei*, *T. cruzi* y *L. major*

	<i>T. brucei</i>	<i>T. cruzi</i>	<i>L. major</i>
GENOMA			
Tamaño del genoma (pb)	^{&} 26.075.393	[*] 60.372.279	32.816.678
Contenido G+C(%)	46,4	51	59,7
Nº de cromosomas (por genoma haploide)	11	28	36
Secuencia scaffolds		838	
Secuencias contigs	30	4.008	36
Porcentaje codificante	50,5	58,9	47,9
GENES CODIFICANTES DE PROTEÍNAS			
Nº de genes (por genoma haploide)	9.068	12.000	8272
Pseudogenes	904	3.590	39
Promedio del largo de CDS (pb)	1.592	1.513	1901
Media del largo de CDS (pb)	1.241	1.152	1407
Contenido G+C(%)	50,9	53,4	62,5
Densidad génica (genes/Mb)	317	385	252
REGIONES INTERGÉNICAS (entre los CDSs codificantes de proteínas)			
Promedio de largo (pb)	1.279	1.024	2.045
Contenido G+C (%)	41	47	57,3
COMPARACIÓN			
Región total con bloques sinténicos (Mpb)	19,9	NC	30,7
Media en tamaño CDS (pb) en 3-COGs sinténicos	1.511	1.457	1.731
Media tamaño entre CDS (pb) entre 3-COGs sinténicos	721	561	1.431
GENES codificantes de RNA			
tRNA	65	115	83
rRNA	56	219	63
siRNA	>28	192	63
snRNA pequeños nucleares	5	19	6
RNA pequeños nucleolares	353	1.447	695
SrpRNA		2	1

Referencia: CDS, secuencias codificadoras de proteínas.

NC, no computado

[&] excluye 100 cromosomas de tamaño pequeño e intermedio

^{*} incluye todos los scaffolds y contigs >5 kb de ambos haplotipos

Evolución molecular, mutaciones, sustituciones y el reloj molecular.

Una simple definición de evolución molecular sería: estudio de los cambios en la secuencia de nucleótidos del ADN ocurridos en la historia de las especies, que permite reconstruir las secuencias de sus ancestros¹⁶ y estudiar los patrones y velocidades de evolución. Mientras que la genética de poblaciones trata con las frecuencias alélicas dentro de las poblaciones y por lo tanto la historia evolutiva reciente, la evolución molecular trata con las secuencias de las moléculas informativas (ADN, proteínas) que evolucionan a través de las sustituciones considerando una escala evolutiva de más largo aliento (ver más adelante).

A continuación definiremos dos términos que serán usados recurrentemente en esta tesis: la mutación y la sustitución.

Una **mutación**, es un cambio en el ADN (no causado por recombinación genética). Puede darse a nivel del gen, llamándose mutación génica (o genética, ver tabla 4) a nivel del cromosoma, mutación cromosómica o a nivel de un set completo de cromosomas siendo ésta una mutación genómica.¹⁷ Se clasifican según su origen en espontáneas, si ocurren al azar, o inducidas si son causadas por agentes físicos, químicos y biológicos. La causa más común que genera las mutaciones espontáneas, son los errores en la síntesis y/o reparación del ADN, participando en dicho proceso la ADN polimerasa. También se clasifican en gaméticas, si son transmitidas a la descendencia o somáticas si no lo son.

Nos concentraremos en las mutaciones genéticas. Existen 3 tipos: inserción de nucleótidos, delección de nucleótidos o remplazo de un nucleótido por otro.

Tanto las inserciones como las delecciones pueden ser de una o más bases. Cuando estas mutaciones ocurren dentro de los genes codificantes de proteínas pueden cambiar el marco de lectura (frameshift) pues trastocan el sentido de lectura del ARNm, con la consiguiente alteración total de la secuencia del polipéptido correspondiente. Si la inserción o delección es de 3 nucleótidos (o múltiplo de 3) no cambia el sentido de lectura aunque agrega o quita aminoácidos, siendo menos perjudicial que si sí lo cambiara.

Tabla 4: Clasificación de mutaciones génicas según varios criterios

CRITERIO QUÍMICO	
Transiciones	Sustitución de purina por purina o de pirimidina por pirimidina
Transversiones	Sustitución de purina por pirimidina o de pirimidina por purina
CRITERIO CODIFICANTE	
Sin sentido	Sustitución de una base dando un codón de terminación (UAA UAG UGA) interrumpiendo la formación del polipéptido
De reemplazo No sinónimas (con sentido)	Sustitución de un aminoácido por otro con función diferente
De reemplazo Neutrales	Sustitución de un aminoácido por otro con función similar
Sinónimas (silenciosas)	Sustitución de base pero no de aminoácido por haber más de un codón para un mismo aminoácido (redundancia del código genético)
Frameshift	Cambio del marco de lectura
CRITERIO AMBIENTAL / FUNCIONAL	
Ventajosa	Mutación que da ventaja selectiva a quien la porta, dependiendo del ambiente donde se encuentre
Neutral	Mutación que no es ni más ventajosa ni más deletérea que la variante dominante en la población (el tipo salvaje)
Desventajosa	Mutación desventajosa que afecta la funcionalidad de la proteína codificada, siendo en general válido independiente del ambiente en que se encuentre dicha mutación. Estas mutaciones no contribuyen a la evolución , por lo general no se transforman en sustituciones

Existe una gran variación en las diferentes tasas de mutación. Éstas pueden variar 500 veces entre genes de una misma especie (la tasa de mutación en el maíz es de aproximadamente 10^{-5} por gameto¹⁸) y 100.000 veces entre especies. En los humanos se observa una frecuencia entre 10^{-5} a 10^{-6} mutaciones por gameto.¹⁹

Las tasas de mutación son muy bajas considerando como base el gen. Pero hay tantos loci en los organismos que quizá haya gametos que lleven una mutación detectable fenotípicamente. Además debemos tener en cuenta que estas tasas están subestimadas por no tener presente a los nuevos alelos surgidos por sustituciones sinónimas (ver más adelante la definición de sustitución). Por lo tanto se vuelve muy probable que la mayoría de los descendientes lleven al menos un nuevo alelo en su genoma.

Una **sustitución**, por otro lado, es la fijación de una mutación en una determinada población. El proceso puede dividirse en 2 partes. La primera, es el surgimiento de la mutación. La segunda es el proceso de fijación, en el cual, la frecuencia del alelo va en aumento hasta que luego de varias generaciones, el 100% de la población lo porta.

Cada región del genoma será más o menos proclive a recibir una sustitución en función de su importancia funcional, es decir una proteína, región o sitio de mayor importancia funcional, recibirá menor número de sustituciones que otra con menor importancia.(ver tabla 5)

Tabla 5: Distintas tasas de sustitución ordenadas por región, valor, importancia funcional y algunos ejemplos proteicos (tabla modificada ²⁰).

IMOPORTANCIA FUNCIONAL (+ → -)	TASAS DE SUSTITUCIÓN			
	Sustituciones en 10 ⁹ años			
	REGIÓN (nucleotídicas)		Proteínas (aminoacídicas)	
	Intrón	4.4	Fibrinopéptidos	8.3
	Flanqueante 3´	4.3	Ribonucleasa Pancreática	2.1
			Lisozima	2.0
	No traducida	2.3	α – globina	1.2
	Flanqueante 5´	2.2	Mioglobina	0.89
			Insulina	0.44
			Citocromo c	0.3
		Histona H4	0.01	

Si la mutación es neutral, es decir, si no es ni más ni menos ventajosa que el alelo preexistente, puede llegar a fijarse, siendo la deriva genética quien la fije. Por el contrario si la mutación es desventajosa, la misma será eliminada por la selección natural negativa, por lo que en general, no llegará a fijarse. Por último, si la mutación es

beneficiosa, entonces también podrá llegar a ser una sustitución, siendo la misma fijada a través de la deriva genética o la selección positiva, principalmente.

Una mutación sinónima, en general, no afecta al fenotipo del organismo, porque ésta, no modifica la secuencia aminoacídica, entonces, en general, no se vería afectada por la selección natural.

Los alelos que no tienen efecto sobre la eficacia se dice que son neutros.

Por otro lado las sustituciones no sinónimas pueden dar lugar tanto efectos deletéreos como beneficiosos sobre la eficacia, efectos que en el alelo resultante dependerán en última instancia del ambiente en el cual se encuentre.

Es de esperar que una mutación no sinónima sea seleccionada cuando haya un cambio drástico ambiental.

La **tasa de sustitución K de alelos neutros**, se calcula de la siguiente manera:

K es igual al nº de mutantes nuevos por la probabilidad de fijación de dicha mutación $2N\mu$ es el número de mutaciones que entran a la población cada generación (siendo $2N$ el tamaño poblacional y μ la tasa de mutación de un gen en particular).

El modelo de alelos infinitos por otra parte, sostiene que cada mutación puede considerarse como un evento único, donde su probabilidad de fijación es $1/2N$, por lo tanto tendríamos que:

$$K = 2N\mu \times 1/2N = \mu \quad \text{eq.1}$$

De esta ecuación se desprende que la tasa de sustitución de alelos neutrales, K, es igual a la tasa de mutación de alelos neutrales, μ y que la misma es independiente del tamaño de la población. Al excluir el tamaño poblacional de la eq.1, se puede observar una de las predicciones que la hipótesis del **reloj molecular** plantea, y es que los genes evolucionan a una tasa constante.²¹

Por otro lado:

$$\mu = v \times f \quad \text{eq.2}$$

Esto significa, que a la tasa de mutación del gen por generación (μ), la descomponemos tanto en v , tasa de mutación por gen como en f , porción de sitios de la proteína que puede aceptar mutaciones sin alterar su función.

De las ecuaciones anteriores fácilmente se deduce:

$$K = v \times f \quad \text{eq.3}$$

Por lo tanto, K será constante para cada gen, pero 2 genes distintos, tendrán diferentes valores de K por diferir en su f (siendo $1-f$ la proporción de aminoácidos que no acepta cambios por alterar la función o estructura de la proteína).

Esto implicaría que si bien no hay un único reloj, cada gen sí estaría actuando como un reloj molecular distinto, lo que permitiría la comparación entre las distintas especies utilizando un gen determinado (ver figura 7).

DIFERENTES RELOJES TIENEN DIFERENTES TIEMPOS

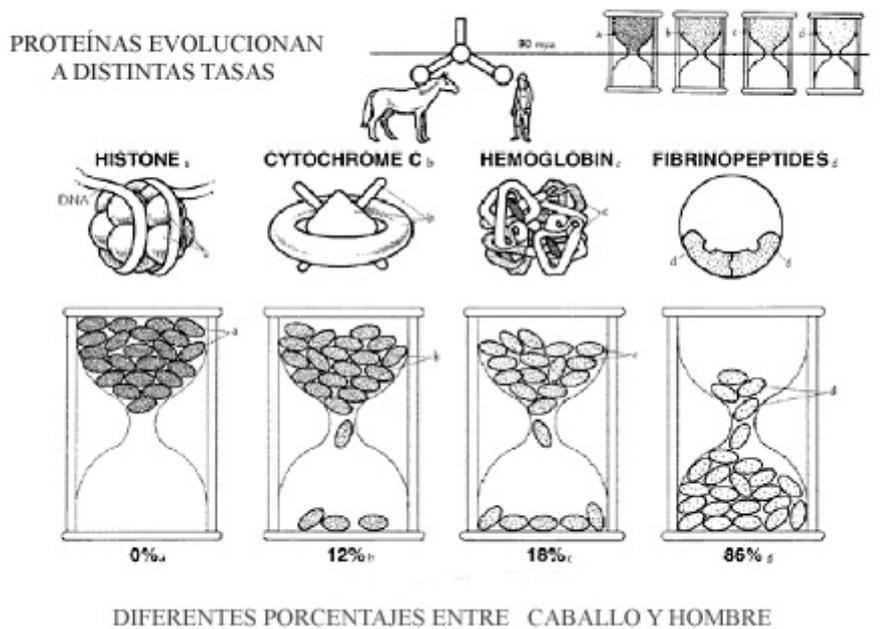


Figura 7: Representación del reloj molecular para diferentes tiempos en diferentes proteínas (modificado ²²).

El reloj molecular por lo tanto puede establecer una escala de tiempo para la evolución (ver figura 8).

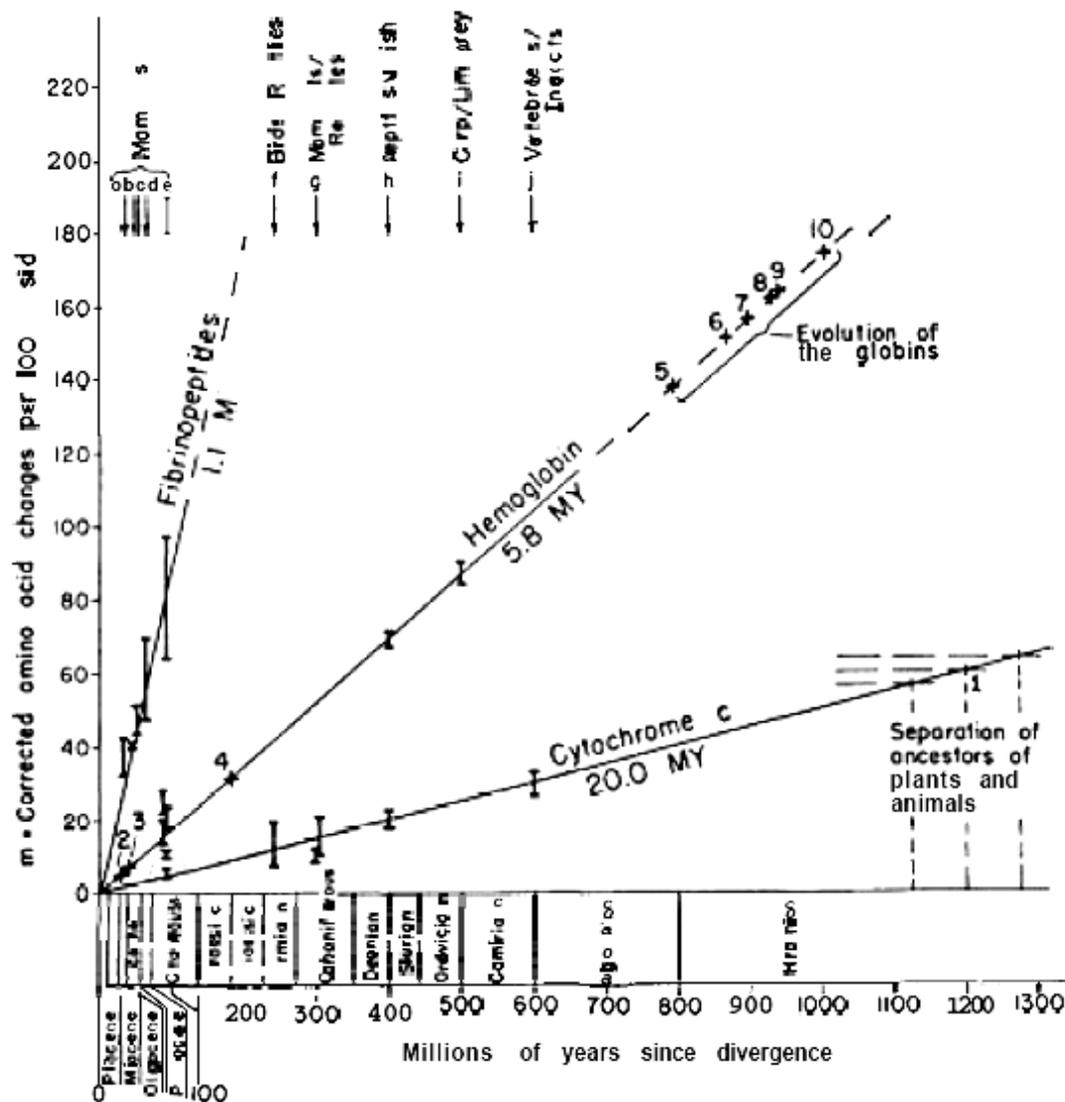


Figura 8: Tasas de sustituciones aminoacídicas en fibrinopéptidos, hemoglobina, y citocromo c.

El punto 1 muestra una fecha de 1200 ± 75 millones de años, tiempo de la separación de plantas y animales, basado en una extrapolación lineal del citocromo c. Los puntos 2-10 se refieren a los eventos en la evolución de la familia de las globinas. Con g se indicó la separación entre mamíferos y reptiles; i entre teleosteos y condricteos; j entre vertebrados e insectos²³.

Para la comparación entre genes iguales en diferentes especies, es importante resaltar, que dichos genes son homólogos por descendencia, **genes ortólogos**; para diferenciarlos de los **genes parálogos** (ver glosario).

Una aproximación para establecer cuando 2 genes homólogos son ortólogos, se basa en la búsqueda de los llamados mejores aciertos bidireccionales (Best Bidirectional Hits, BBH). Según esta aproximación, dos proteínas P1 y P2 en dos organismos O1 y O2 se definen como ortólogas o BBH, si al realizar el alineamiento de P1 contra todas las proteínas del organismo O2, el alineamiento con mejor resultado se produce entre P1 y P2 y viceversa. Adicionalmente se imponen las siguientes restricciones: que la similitud entre ambas secuencias supere un cierto e-value umbral y que la región en la cual dichas secuencias son similares supere, por ejemplo, un 75% de la longitud de las mismas.²⁴ Entonces, en la comparación en una misma proteína de varias secuencias ortólogas de distintas especies, podremos inferir que existen diferencias de uno a varios aminoácidos entre las mismas. De esta forma podemos calcular la tasa de sustitución de dicha proteína empíricamente, y esto fue lo que hicieron Pauling y Zukerland en el año 1964. Tomando la diferencia aminoacídica existente en el citocromo c por ejemplo, entre varias especies, se ve que especies que se separaron hace menos tiempo poseen menor número de cambios que aquellas con tiempo de separación mayor.

Una vez calibrado un reloj molecular para una determinada proteína, es decir, tomando diferencias aminoacídicas entre 2 especies por un lado y diferencias aminoacídicas para otro par de especie por otro lado, y teniendo el registro fósil para saber el tiempo de separación en estos 2 casos, podemos calcular un nuevo tiempo de separación para otras 2 especies en ausencia de registro fósil.

Filogenia de tripanosomátidos

El género *Trypanosoma* posee una gran cantidad de especies, que forman un grupo monofilético, es decir, un grupo con un antecesor común a todas ellas no compartido. Estudios filogenéticos iniciales basados en genes ARNr 18S propusieron que el género *Trypanosoma* era un grupo parafilético, análisis posteriores basados en genes codificantes de proteínas o incluyendo mayor cantidad de especies de tripanosomátidos, mostraron que el género *Trypanosoma* es monofilético.^{25,26}

Estos resultados contradictorios indicarían que tal vez pudieran existir algunos problemas a la hora de la construcción de las filogenias moleculares relacionados con los métodos con que se llegan a la misma, que es preciso tener en cuenta y estudiar con mayor precisión.

Hay que aclarar también que la filogenia de las especies no tiene porque coincidir con la del gen o genes de las especies correspondientes.

Por ejemplo si existiese gran variabilidad en las tasas de evolución molecular de las diferentes especies de tripanosomátidos, esto habría que tenerlo en cuenta a la hora de utilizar un determinado método de reconstrucción de árboles filogenéticos.

Por ejemplo el método **UPGMA** (**U**nweighted **P**air **G**roup **M**ethod with **A**rithmetic **M**ean) u otros dependen de la constancia en las tasas de evolución molecular de las secuencias seleccionadas, surgiendo problemas en la utilización de los mismos en ausencia de reloj molecular (ver figuras 9 y 10).

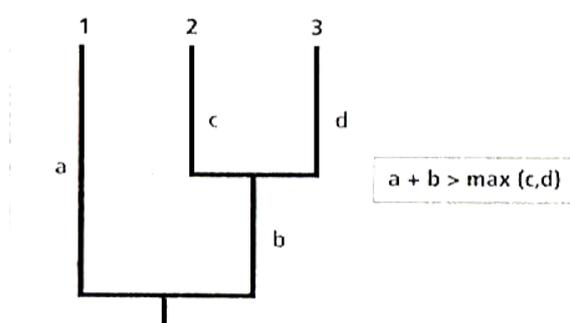


Figura 9: La condición requerida por UPGMA para reconstruir satisfactoriamente el verdadero árbol es que las sumas de los bordes que llevan a la secuencia 1 ($a+b$) debe ser mayor que el más largo de c y d .²⁷

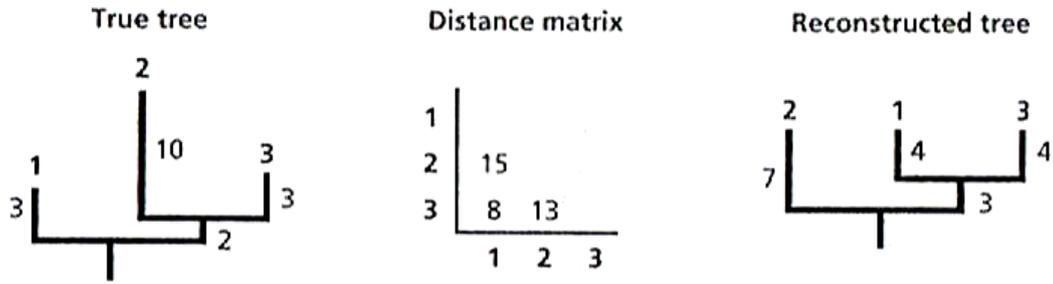


Figura 10: Un ejemplo donde UPGMA reconstruirá el árbol mal.

El borde largo en el árbol verdadero viola la condición mostrada en la figura anterior, ya que $a+b = 3+2 < \max(c,d) = 10$. La secuencia 2 ha evolucionado más rápidamente que las otras dos secuencias, por lo tanto la secuencia 1 y 3 serían más similares entre sí que cualquiera con la 2.²⁷

Por lo tanto a la hora de llevar a cabo reconstrucciones de relaciones evolutivas al menos mediante estos métodos, así como otros análisis filogenéticos, un aspecto fundamental es determinar si son o no constantes las tasas de evolución molecular por un lado, y qué especie o especies evolucionan comparativamente más rápido dentro de una familia y en lo que respecta a este trabajo, dentro del grupo de tripanosomas africanos de nuestro interés.

El objetivo de este trabajo es comparar las velocidades de evolución molecular relativas en los tripanosomátidos, es decir, analizar si las distintas especies de tripanosomátidos evolucionan a la misma velocidad, o si unas especies evolucionan más rápido que otras, haciendo especial énfasis en los tripanosomas africanos. Para este fin se utilizarán todos los genes ortólogos disponibles en 8 especies de tripanosomátidos representativas de la diversidad filogenética del grupo.

OBJETIVOS

Se intentará analizar si todos los tripanosomátidos evolucionan a nivel molecular a la misma velocidad.

En la figura 11 se observa la filogenia del género *Trypanosoma*.

Objetivo general:

Realizar un análisis de la velocidad de evolución molecular de los tripanosomátidos.

Para este fin se utilizará el test estadístico de Tajima de tasas relativas.

Objetivos específicos:

-Analizar la velocidad de evolución molecular de los tripanosomas africanos intentado determinar si alguno de ellos evoluciona rápidamente.

-Investigar las posibles causas funcionales de una eventual aceleración mediante una asociación funcional de los genes que presenten alteraciones en sus tasas evolutivas.

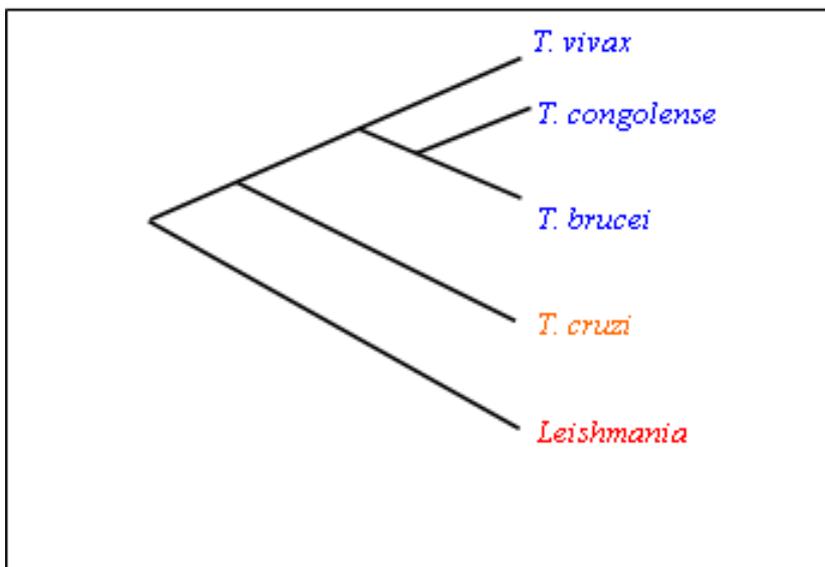


Figura 11

Filogenia del género *Trypanosoma*.

En azul los tripanosomas africanos.

En naranja el grupo externo de estos tres (*T. cruzi*) y en rojo

el género *Leishmania*, grupo externo del género *Trypanosoma*.

MATERIALES Y MÉTODOS

Datos utilizados

Se utilizaron las secuencias génicas de los genomas de los siguientes organismos: *Leshmania infatum*, *L. major*, *L. braziliensis*, *Trypanosoma brucei brucei*, *T. cruzi*, *T. brucei gambiense*, *T. vivax* y *T. congolense* (los genomas de estas 2 últimas especies utilizadas en este trabajo fueron obtenidas mediante un pipeline de anotación por cortesía de Miguel Ponce de León y Guillermo Lamolle).

En la tabla 6 se presenta la cantidad de genes utilizados en este trabajo, de una base datos disponibles a la fecha.

Es importante aclarar que la misma difiere en algún grado a la publicada en el 2005 por El-Sayed, Berriman y colaboradores y a la actual publicada en el Genbank, debido a que *T. vivax* y *T. congolense* fueron obtenidos indirectamente, mediante métodos predictivos.

Tabla 6: Las 8 especies utilizadas en este trabajo con sus abreviaciones respectivas y cantidad de genes de cada una de las mismas.

<u>TABLA 6</u>		
Especie	Abreviación (usada en Este trabajo)	Número de Genes
<i>Leshmania infatum</i>	Li	8173
<i>Leishmania major</i>	Lm	8331
<i>Leishmania braziliensis</i>	Lb	8408
<i>Trypanosoma brucei brucei</i>	Tbb	9079
<i>Trypanosoma congolense</i>	Tco	10546
<i>Trypanosoma cruzi</i>	Tcr	19607
<i>Trypanosoma brucei gambiense</i>	Tbg	8623
<i>Trypanosoma vivax</i>	Tvi	8321

Scripts

Se utilizaron varios scripts informáticos cedidos por la cátedra de Biomatemática de la Facultad de Ciencias, para la realización de este trabajo.

Un script, es un fichero de texto que contiene una serie de instrucciones que se pueden ejecutar en una línea de órdenes seguidas.²⁸

El primero de ellos fue **fasta_nuc2pep**, script que obtiene la secuencia aminoacídica a partir de una nucleotídica realizando una traducción virtual.

El script **ORTOLOGADOR** utilizado en este trabajo, obtiene de 2 proteomas de 2 especies distintas, una lista con los nombres de las proteínas ortólogas comunes a ambas.

El script toma la secuencia aminoacídica de cada uno de los genes de una especie X determinada y realiza alineamientos entre cada gen y cada uno de los genes de la otra especie, especie Y, utilizando el programa **BLAST**²⁹.

El script **ORTOLOGADOR**, cuenta con la opción **BRH -best reciprocal hit-** para realizar lo mencionado anteriormente pero en ambos sentidos, es decir tomando también cada uno de los genes de la especie Y, y alineándolos con todos los genes de la especie X, así se evitan proteínas que pudieran estar repetidas en la lista y eliminar grupos de genes cuyas relaciones son de paralogía.

Para obtener una lista con los 3-COG, se utilizó el script **cog_m.pl** que permite, teniendo 3 listas de genes ortólogos de 2 especies (ortólogos entre especie X y especie Y; ortólogos entre especie Y y especie Z; ortólogos entre especie X y especie Z), obtener una sola lista con genes ortólogos de 3 especies, es decir, ortólogos entre especie X, especie Y y especie Z.

La figura 12 nos ilustra un ejemplo de COGs y 3-COG para los datos de El-Sayed et al. Allí puede verse que *T.brucei* y *T.cruzi* poseen un COG de 6616 proteínas, *T. brucei* y *L. major* poseen un COG de 6232 (6158 + 74) y *L. major* y *T. cruzi* un COG de 6640 proteínas (6158+482). En este ejemplo se ve que las 3 especies, *T. brucei*, *T.cruzi* y *L. major*, poseen un 3-COG, de 6158 proteínas.

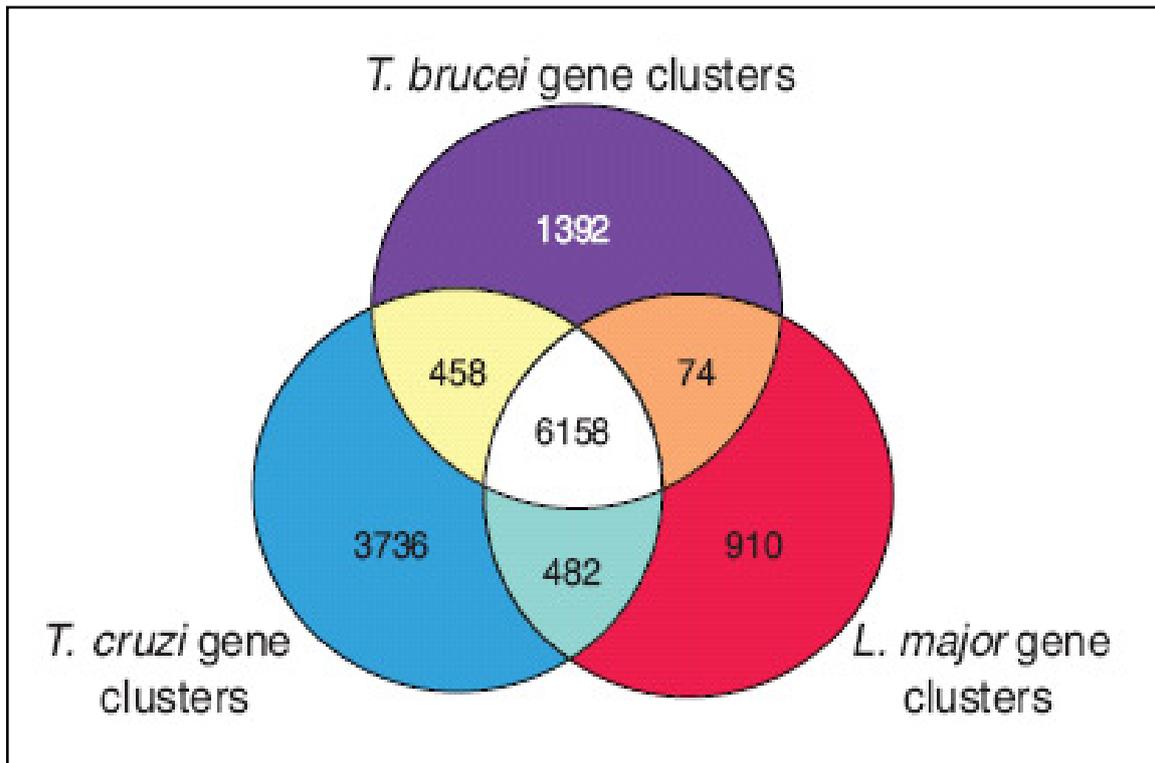


Figura 12

Distribución de genes para cada una de las especies de TriTryps, genes ortólogos para pares de especies y número de ortólogos para las tres especies.¹⁵

Siguiendo el ejemplo anterior con esos 6158 grupos de 3 proteínas ortólogas tendríamos la siguiente lista:

<i>T. brucei</i>	<i>T. cruzi</i>	<i>L. major</i>
Proteína A ₁	Proteína B ₁	Proteína C ₁
Proteína A ₂	Proteína B ₂	Proteína C ₂
Proteína A ₃	Proteína B ₃	Proteína C ₃
Proteína A ₄	Proteína B ₄	Proteína C ₄
Proteína A ₅	Proteína B ₅	Proteína C ₅
.	.	.
Proteína A ₆₁₅₈	Proteína B ₆₁₅₈	Proteína C ₆₁₅₈

El script **mysplit**, se utilizó para obtener de esa lista de nombres de genes sus respectivas secuencias.

Con el Programa **CLUSTAL 2.0.10 multiple sequence alignment**³⁰, se alinearon las secuencias, y con la opción -PIR-del mismo se obtuvieron las secuencias en el orden preferido.

ALINEAMIENTO 1	
<i>T. brucei</i>	Proteína A ₁
<i>T. cruzi</i>	Proteína B ₁
<i>L. major</i>	Proteína C ₁

ALINEAMIENTO 2	
<i>T. brucei</i>	Proteína A ₂
<i>T. cruzi</i>	Proteína B ₂
<i>L. major</i>	Proteína C ₂

ALINEAMIENTO 3	
<i>T. brucei</i>	Proteína A ₃
<i>T. cruzi</i>	Proteína B ₃
<i>L. major</i>	Proteína C ₃

ALINEAMIENTO 4	
<i>T. brucei</i>	Proteína A ₄
<i>T. cruzi</i>	Proteína B ₄
<i>L. major</i>	Proteína C ₄

ALINEAMIENTO 6158	
<i>T. brucei</i>	Proteína A ₆₁₅₈
<i>T. cruzi</i>	Proteína B ₆₁₅₈
<i>L. major</i>	Proteína C ₆₁₅₈

Mediante el script **tatarel**, se realizó el test de Tajima de tasas relativas (ver más adelante la aplicación del mismo), para dos secuencias y un outgroup, grupo externo conocido

Por otro lado, otra forma de comparar la velocidad de evolución molecular de diferentes especies, es la realización de un árbol filogenético con la totalidad de las secuencias disponibles de dichas especies.

Para lograr esto se utilizó primero el script **agrega1or**. El mismo permite agregar a un 3-COG específico, los genes ortólogos de una 4^a especie, luego una 5^a, y así recursivamente hasta la N especie, obteniendo de esta forma, los genes ortólogos de N especies de interés, formando un N-COG, como se verá a continuación:

<i>Especie 1</i>	<i>Especie 2</i>	<i>Especie 3</i>	<i>Especie 4</i>	...	<i>Especie N</i>
Proteína A ₁	Proteína B ₁	Proteína C ₁	Proteína D ₁	...	Proteína N ₁
Proteína A ₂	Proteína B ₂	Proteína C ₂	Proteína D ₂	...	Proteína N ₂
Proteína A ₃	Proteína B ₃	Proteína C ₃	Proteína D ₃	...	Proteína N ₃
Proteína A ₄	Proteína B ₄	Proteína C ₄	Proteína D ₄	...	Proteína N ₄
Proteína A ₅	Proteína B ₅	Proteína C ₅	Proteína D ₅	...	Proteína N ₅
.
Proteína A _n	Proteína B _n	Proteína C _n	Proteína D _n	...	Proteína N _n

Una vez obtenido el N-COG, con los genes correspondiente a las N especies de interés, se necesita, tener sus secuencias. El script **mysplit**, se utilizó para obtener de esta lista de nombres de genes sus respectivas secuencias aminoacídicas.

Luego necesitamos que se encuentren alineadas todas las secuencias en un único alineamiento.

<i>Especie 1</i>	Proteína A ₁ -- Proteína A ₂ --- Proteína A ₃ – Proteína A ₄ --.... -- Proteína A _n
<i>Especie 2</i>	Proteína B ₁ -- Proteína B ₂ -- Proteína B ₃ -- Proteína B ₄ -- ... -- Proteína B _n
<i>Especie 3</i>	Proteína C ₁ -- Proteína C ₂ -- Proteína C ₃ -- Proteína C ₄ -- ...-- Proteína C _n
<i>Especie 4</i>	Proteína D ₁ -- Proteína D ₂ -- Proteína D ₃ -- Proteína D ₄ -- ...--Proteína D _n
...	...
<i>Especie N</i>	Proteína N ₁ -- Proteína N ₂ -- Proteína N ₃ -- Proteína N ₄ -- ... -- Proteína N _n

Una vez alineadas las secuencias de dichas especies, se utiliza el programa **RAxML-7.0.4**³¹, que infiere árboles filogenéticos basados en Máxima Verosimilitud secuencial y paralela, que permite además, procesar gran cantidad de datos como lo requiere este estudio.

Por último, el programa **Figtree.v1.2.2.**, permite la visualización del árbol filogenético resultante.

Árboles filogenéticos

Éstos nos permiten mediante un algoritmo determinado inferir una hipótesis evolutiva, es decir, inferir cómo fue la historia evolutiva actual de un grupo de secuencias de organismos.

Los árboles pueden tener raíz o no tenerla. Diferentes tipos de árboles pueden ser usados para describir diferentes aspectos de la historia evolutiva de los distintos organismos en cuestión.

Un cladograma, muestra sólo el orden relativo de las especies. Por otra parte, un **filograma**, es un cladograma con el largo de las ramas proporcional a la cantidad de evolución que ocurrió en dichas ramas y por último un dendograma es un tipo especial de filograma en la cual los extremos del árbol son todos equidistantes a la raíz, dejando de lado una eventual disparidad en las tasas de evolución entre las diferentes OTUs y mostrando por otro lado el tiempo de divergencia entre los mismos.²⁷

En este trabajo se realizó un filograma, pues, la cantidad de evolución fue lo que se quiso calcular.

Existen varias formas de construir un árbol filogenético. Los métodos de construcción filogenética se clasifican según Page & Holmes (1998) de acuerdo a:

- a) Como utilizan los datos de las secuencias (distancias o caracteres discretos)
- b) Como construyen el árbol (agrupamiento o búsqueda)

Distancias o caracteres discretos:

Los métodos de distancias convierten las secuencias alineadas en una matriz de distancias (el número inferido de sustituciones entre dos secuencias) y luego utilizan esta matriz para construir el árbol. Los métodos de caracteres discretos utilizan cada sitio de la secuencia (o una función de cada sitio) directamente.

Agrupamiento o búsqueda:

Otra forma de clasificar los métodos de reconstrucción filogenética es por la forma en que construyen los árboles. Los métodos de búsqueda utilizan un criterio de optimización para decidir, entre todas las posibles hipótesis (árboles evolutivos), cual es el que mejor se comporta de acuerdo a cierto criterio de optimización.

De lo dicho anteriormente y de acuerdo a los criterios anteriores, podemos ver la siguiente clasificación:

		Tipos de datos	
		Matriz de distancia	Sitios nucleotídicos
Métodos de construcción del árbol	Algoritmo de agrupamiento	<p>Método de Grupos de Pares No Ponderada con Media Aritmética (UPGMA)</p> <p>Unión de Vecinos (Neighbour Joining)</p>	
	Criterio de optimización	<p>Evolución Mínima</p>	<p>Máxima Parsimonia</p> <p>Inferencia Bayesiana</p> <p>Máxima Verosimilitud</p>

El método utilizado en este trabajo fue el de máxima verosimilitud, a través del ya mencionado programa **RAxML-7.0.4**.

El método de inferencia filogenética bayesiano (ver clasificación), por otro lado, genera la probabilidad posterior de un parámetro, un árbol filogenético y/o un modelo evolutivo, basada en la probabilidad anterior de ese parámetro y la función de verosimilitud de los datos. La aplicación del análisis bayesiano en la inferencia filogenética presenta varias ventajas en comparación con otros métodos de inferencia, como la fácil interpretación de los resultados, la posibilidad de usar información *a priori* además de otras ventajas computacionales³² pero de todas maneras fue utilizada el Test de Tajima, comentado más adelante, debido a su fácil aplicación.

En la figura 13 puede verse un árbol filogenético de toda la familia Trypanosomatidae (elipse en rojo) donde es posible apreciar sus relaciones de parentesco.

Tal como ya se indicó, el género *Trypanosoma* es monofilético, y además dentro de este género, los tripanosomas africanos (Salivaria) también forman un grupo monofilético (elipse en azul).

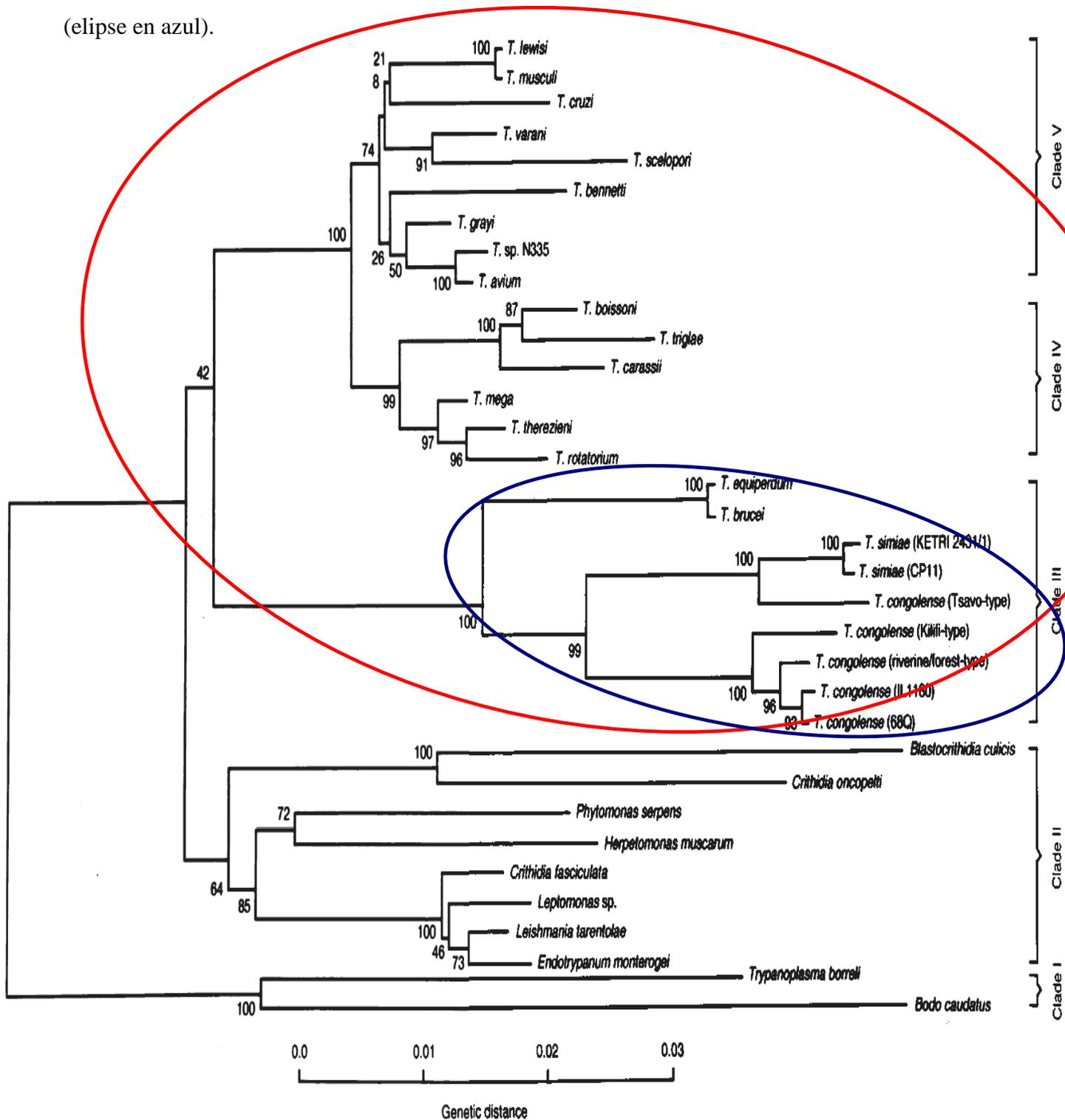


Figura 13 Árbol filogenético de la familia Trypanosomatidae.

Utilizando el método de Neighbor-joining y secuencias de rRNA y DNA de cepas de tripanosomas.³³

Test de las tasas relativas

Como mencionamos anteriormente, una forma de evaluar la velocidad de evolución del género *Trypanosoma*, es utilizando el test estadístico de Tajima de tasas relativas.

En términos generales un test de tasas relativas consiste en comparar la cantidad de evolución de dos especies desde que se separaron de un ancestro común. Para este fin se utiliza un grupo externo a ambas especies. Bajo la hipótesis nula de paridad en la velocidad de evolución, la distancia de ambas especies al grupo externo, debería ser la misma. Nótese, que dos especies tengan la misma distancia al grupo externo, es equivalente a decir que tienen la misma distancia al ancestro común, pues si difieren en su distancia al grupo externo esta diferencia sólo podría haber surgido luego que se separaran de este ancestro, pues antes eran un mismo linaje.

El test de Tajima, posee la ventaja al igual que otros test de tasas relativas, de no necesitar de ningún conocimiento del tiempo de divergencia de los taxa en cuestión. A diferencia de otros test no necesita asumir algún modelo particular de sustitución para estimar el número de sustituciones nucleotídicas por sitio (o aminoacídicas en nuestro caso), ni saber la variación de las tasas de sustitución entre diferentes sitios.

La único que precisamos tener es una determinada filogenia y un grupo externo lo más cercano posible a las especies en comparación.³⁴

El test de Tajima de tasas relativas toma las 3 secuencias alineadas de las respectivas especies, siendo n_{ijk} el número observado de sitios donde las secuencias 1,2 y 3 tienen los aminoácidos i , j y k respectivamente. El subconjunto i , j y k puede tomar los valores 1, 2, 3, 4, 5 y así sucesivamente hasta el 20, correspondiendo a los 20 aminoácidos existentes.

Teniendo la filogenia de las 3 especies de estudio, asumimos que la secuencia número 3 es el grupo externo conocido. Nuestra hipótesis nula es que deberían existir igual número de sustituciones en las secuencias 1 y 2 cuando las comparamos con el grupo externo. La diferencia entre las sustituciones que ocurrieron en la rama que conduce a la especie 1 y la que conduce a la especie 2 sigue una distribución de chi cuadrado con un grado de libertad.³⁴

Entonces en términos formales podemos plantear que el valor esperado de n_{ijk} será igual al de n_{jik} , es decir,

$$E(n_{ijk}) = E(n_{jik}) \quad (1)$$

Y esta igualdad debería cumplirse, cualquiera sea el modelo de sustitución e incluso si las tasas de sustitución varían entre los diferentes sitios. Si la igualdad de la ecuación (1) no se cumple, la tasa evolutiva cambió en alguna o ambas ramas que están siendo comparadas.³⁴

Para ejemplificar este test, tomemos la filogenia de la figura 14

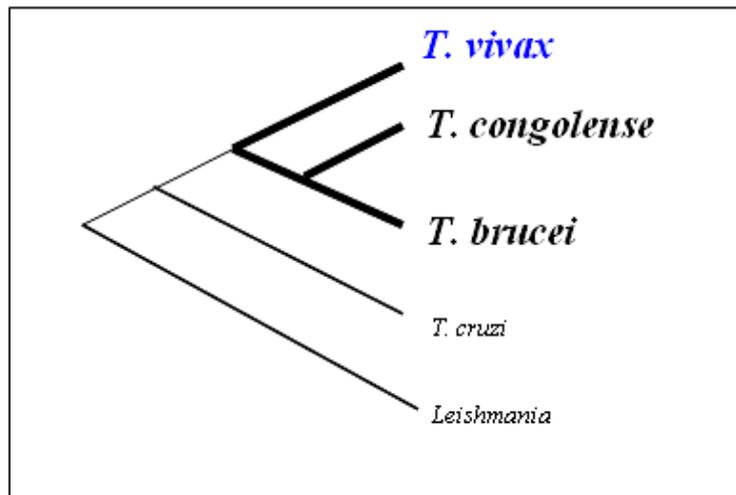


Figura 14

Aplicación del test de Tajima de tasas relativas.

Filogenia elegida en la cual se testaron las velocidades de evolución entre *T. congolense* y *T. brucei* tomando *T. vivax* como grupo externo.

Las 3 especies *T. congolense*, *T. brucei* y *T. vivax* comparten 4705 genes (ver resultados tabla 7 columna 7).

En la figura 14 se resaltan 2 especies de tripanosomas, *T. congolense* y *T. brucei* comparada con *T. vivax*, grupo externo más cercano de ambas. Un tramo de un

alineamiento (entre los 4705 existentes) entre estas 3 especies es mostrado en la figura 15 con los sitios conservados en amarillo y los variables en blanco.

Para la realización del test se utilizan sólo algunos de los sitios variables, aquellos sitios que cumplen con la condición donde hay 2 tipos diferentes de aminoácidos en las 3 secuencias.

Se puede definir el número observado de sitios en los cuales los aminoácidos de la secuencia 1 son diferentes a los aminoácidos de la secuencia 2 y 3 (mientras que la secuencia 2 y 3 tienen el mismo aminoácido en este sitio) a través de m_1 :

$$m_1 = \sum \sum n_{ijj} \quad (2a)$$

Por otra parte definimos a m_2 como el número observado de sitios en los cuales los aminoácidos de la secuencia 2 son diferentes a los aminoácidos de la secuencia 1 y 3 (siendo los aminoácidos de 1 y 3 iguales en este sitio) a través de m_2 :

$$m_2 = \sum \sum n_{jij} \quad (2b)$$

Y m_3 el número observado de sitios en los cuales los aminoácidos de la secuencia 3 difieren de los de la 1 y 2:

$$m_3 = \sum \sum n_{jji} \quad (2c)$$

Cuando la secuencia 3 es el grupo externo, y utilizando la eq (1) vista anteriormente, se llega a que el valor esperado de m_1 debe ser igual al de m_2

$$E(m_1) = E(m_2) \quad (3)$$

Esta igualdad puede ser testeada usando el chi-cuadrado

$$\chi^2 = (m_1 - m_2)^2 / (m_1 + m_2) \quad (4)$$

La eq 4 sigue aproximadamente una distribución chi-cuadrado con un grado de libertad.

Y este es el método el que Tajima denomina Método con un grado de libertad para el caso donde el grupo externo es conocido.

En la figura 15, se observa un alineamiento múltiple entre los genes ortólogos Tb07.11L3.90 (*T. brucei*), Scaff.1431.14 (*T. congolense*) y Tviv896g02.q1kgen16 (*T. vivax*, grupo externo).

En dicha figura se observa un fragmento del alineamiento desde el aminoácido 137 al 168 que representa sólo un fragmento de un alineamiento de 396 aminoácidos (no mostrado).

De estos 396 sitios, 29 son gaps, 152 conservados (igual aminoácido en las 3 especies) y 233 variables.

De estos sitios variables, hay 78 sitios con aminoácidos distintos en las 3 especies, los cuales no son informativos pues representan cambios que ocurrieron tanto en la rama que conduce a *T. congolense* como a *T. brucei* y 86 son iguales en las 2 especies de estudio. Por último los sitios de interés para este análisis son los que son iguales entre *T. congolense* y el grupo externo, que llegan a 22 sitios, los cuales representan sustituciones en la rama que lleva a *T. brucei* (m_1); y aquellos que son similares entre *T. brucei* y el grupo externo (sustituciones de *T. congolense*) que llegan a 35 (m_2).

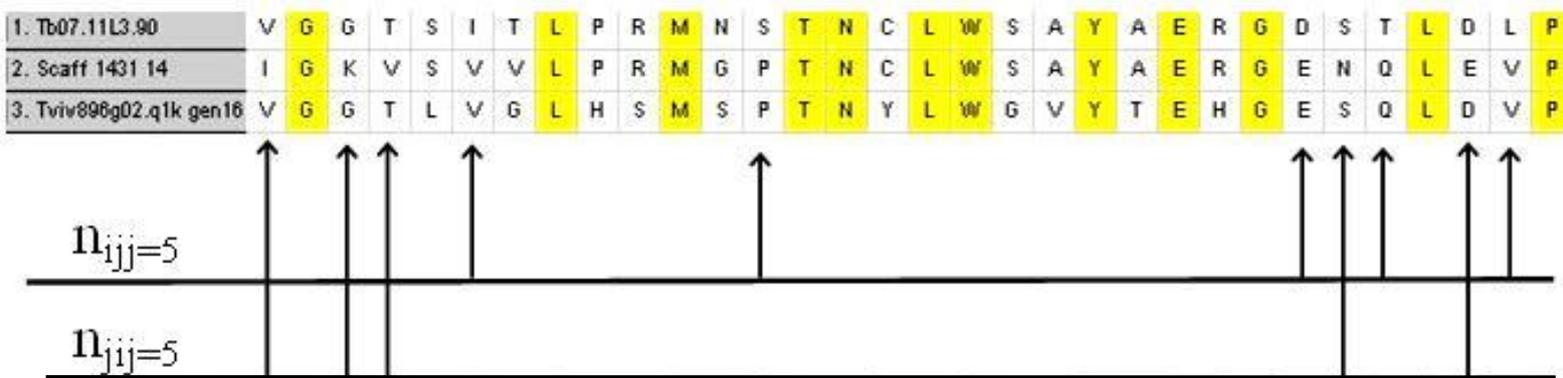


Figura 15

Aplicación del test de Tajima de tasas relativas.

Parte del alineamiento de un gen ortólogo determinado entre las 3 especies de la figura 15 (*T. congolense*, *T. brucei* y *T. vivax*) visualizado con el **MEGA 4.0**.³⁵

Ya que lo matemáticamente esperado es que $E(m_1)=E(m_2)$, entonces, esperaríamos que existieran igual número de singletons en cada una de las dos secuencias ortológicas. Como en la práctica, no siempre nos encontramos con observaciones exactamente iguales, a veces existen diferencias, y necesitamos evaluar que tan determinante es la misma, como para aceptar o no un determinado resultado, a tal efecto contamos con la distribución de chi cuadrado, la cual podemos usar si las desviaciones de la esperanza son mayores de lo que esperaríamos por azar. En la figura 16 se presenta esta distribución para un grado de libertad y los valores críticos de alfa (probabilidad de error de tipo I) para 5% y 1%.

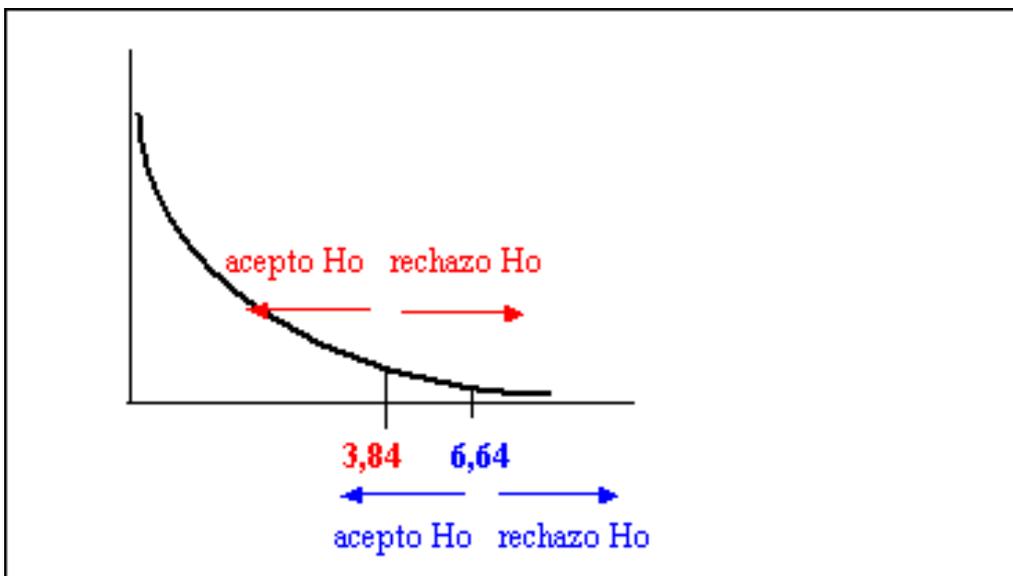


Figura 16: Distribución de χ^2 para un grado de libertad con niveles significativos de aceptación de la H_0 al 95% (3,84) y 99% (6,64) respectivamente.

Del ejemplo anterior:

	Observado	Esperado	obs-esp	$(\text{obs-esp})^2$	$(\text{obs-esp})^2/\text{esp}$
m_1	22	28,5	-6,5	42,25	1,48
m_2	35	28,5	6,5	42,25	1,48
					$\chi^2 = 2,96$

O lo que es lo mismo, $\chi^2 = (m_1 - m_2)^2 / (m_1 + m_2) = (22 - 35)^2 / (22 + 35) = 2,96$

Como $P(\chi^2 < 3,841) = 0,95$ para un grado de libertad (figura 16); por lo tanto para este gen ortólogo en particular, aceptamos la hipótesis nula que ambas especies evolucionan a la misma velocidad con un nivel significativo del 95%.

Resumiendo entonces tenemos que:

$H_0, E(m_1) = E(m_2)$

Si **ACEPTO** H_0 , es equivalente a decir que no tengo razones fuertes para afirmar que las 2 especies evolucionan a velocidades diferentes.

Si **RECHAZO** H_0 , afirmo que las 2 especies no evolucionan a la misma velocidad.

En este último caso, la diferencia de distancias aminoacídicas entre ambas 2 secuencias con respecto a un grupo externo cercano deberá ser tal (verificada a través de un test de chi cuadrado con un valor mayor a 3,84) como para rechazar la hipótesis nula (igualdad de evolución de un determinado gen en esas 2 especies) y de esta forma afirmar que ambas especies en ese gen ortólogo en particular no evolucionan de igual forma, evolucionando más la especie que haya incorporado mayor número de sustituciones.

Ontología génica y búsqueda de causas funcionales de la aceleración

La ontología génica podríamos definirla como el ordenamiento y jerarquización de los diferentes productos génicos y la relación entre éstos.

El GO Consortium, fue el grupo que desarrolló este lenguaje estructurado, el cual se representa por un grafo acíclico directo. El mismo permite que un término dado aparezca en varios lugares de la jerarquía. En cuanto a la jerarquización está formado por 3 grandes ontologías: 1) función molecular, 2) proceso biológico y 3) componente celular.

Con respecto a la relación entre los mismos, pueden relacionarse de 3 formas (A es un B), (A es parte de B), (A regula a B). [(A regula positivamente a B) o (A regula negativamente a B)]. Realizamos una anotación funcional de genes de tripanosomas mediante el software **Blast2GO**³⁶, para asignarle términos de ontología a los genes de *T. brucei*.

Este análisis fue realizado con el fin de explorar si el grupo de genes que presenta mayor tasa de evolución molecular cae preferentemente en algunas categorías funcionales concretas.

Para este fin se procedió de la siguiente forma:

Primero se buscaron a través del test de tasas relativas de Tajima los genes con mayor tasa de evolución de las especies de tripanosomas africanos.

Luego se eligió de forma arbitraria a *T. brucei*, como representativo del grupo de tripanosomas africanos para buscar la posible causa funcional de dicha aceleración.

Como ya se mencionó se utilizó el software **Blast2GO**, para asignarle ontología a los de genes de *T. brucei*

Por último se comparó al grupo de genes más evolucionados de *T. brucei* con respecto a la totalidad de los genes de esa especie con el fin de encontrar grupos funcionales acelerados por medio de un test exacto de Fisher con $P=0.01$.

RESULTADOS

Grupos de genes ortólogos utilizados en este análisis

Se utilizaron las secuencias génicas de las siguientes especies:

L. infatum, *L. major*, *L. braziliensis*, *T. brucei brucei*, *T. cruzi*, *T. brucei gambiense*, *T. vivax* y *T. congolense*

Con script casero (**fasta_nuc2pep**) se obtuvieron las secuencias aminoacídicas, a partir de las nucleotídicas, de cada uno de los genomas de las especies recién mencionadas, obteniendo de esta manera sus respectivos proteomas.

Una vez obtenidos estos 8 proteomas, se procedió a buscar las secuencias ortólogas de 2 especies a la vez (2-COG). Para este fin fue utilizado el script que se basó en los mejores hits recíprocos de BLAST usando un script en perl tal como fue explicado en la sección de métodos.

Repitiendo este procedimiento para todos los pares de especies, se obtuvieron un total de 28 2-COGs.

A partir de los 2-COGs obtenidos y utilizando el script **cog_m.pl** que va comparando los 3 pares de ortólogos, se obtuvieron 3-COG.

Estos contienen genes ortólogos de tres especies, una de las cuales será usada como grupo externo, a los efectos de realizar los test de tasas relativas. Mediante este procedimiento se obtuvieron 9 grupos de 3-COG, que se aprecian en la tabla 8.

Test de tasas relativas de Tajima

Se realizaron 9 grupos de test, filas número 1 al 9 para 2 especies determinadas y un grupo externo (ver tabla 7). Para cada uno de los grupos, se muestra, el número de genes que no presentaron diferencias significativas y aquellos que sí lo hicieron con su respectivo nivel de significación. Por una parte en rojo se aprecian las secuencias significativas al 5% y al 1%, para la especie 1, en cada uno de los 9 grupos; mientras que en verde se aprecia lo mismo pero para la especie 2 del grupo correspondiente.

La última columna, la de más a la derecha, indica la cantidad de secuencias analizadas en cada uno de los casos (Nº de sec).

A modo de ejemplo se explica el primer grupo de test de tasas relativas, el cual fue realizado en *L. infantum* y *L. major*, teniendo a *L. braziliensis* como grupo externo, tal como se presenta en la fila nº 1 de la tabla 8.

En este primer caso, se hicieron 6942 test de tasas relativas correspondientes a cada uno de los COGs de este trío de especies.

De los mismos, en 695 casos, dio un valor de 0 el χ^2 , por lo tanto, aceptamos la H_0 , es decir, afirmamos que en estos genes ortólogos, *L. infantum* y *L. major* evolucionan a la misma velocidad.

En 5497 casos, el valor de χ^2 fue menor a 3,84, por lo tanto, al igual que en el caso anterior, también aceptamos la hipótesis nula de igualdad de velocidad de evolución para estos genes ortólogos (téngase en mente el esquema de la figura 17).

Por otro lado, en los restantes 750 genes, sus respectivos test de tasas relativas presentaron valores de χ^2 mayores a 3,84. En 462 COGs, la significancia fue sólo al 5%, ($3,84 < \chi^2 < 6,64$) mientras que en los restantes 288 COGs, la significación estadística fue al 1% ($\chi^2 > 6,64$).

En estos 750 COGs se rechaza la hipótesis nula, pudiendo afirmar que estos genes evolucionan a distinta velocidad.

Veremos a continuación el detalle de las diferencias de evolución discriminando por especies. En 58 casos (tabla 8, fila 1, columna 9 en rojo) al encontrarse el valor de chi-cuadrado entre 3,84 y 6,64, rechazamos la H_0 con un nivel significativo del 5%, siendo *L. infantum* más evolucionado que *L. major*.

Por otro lado en 404 tests (tabla 8, fila 1, columna 11 en verde) rechazamos la H_0 al 5% siendo *L. major* más evolucionado que *L. infantum*.

Cuando consideramos los casos de significación al 1% (es decir chi-cuadrado mayor que 6,64) en 102 casos (tabla 7, fila 1, columna 13 en rojo) *L. infatum* ha evolucionado más que *L. major*, mientras que en 186 casos (tabla 7, fila 1, columna 15 en verde) *L. major* ha evolucionado más que *L. infatum*. En resumidas cuentas estos resultados nos permiten concluir que la tasa de evolución aminoacídica es más alta en *L. major* que en *L. infantum*.

Al ver la tendencia general para las especies del género *Trypanosoma* es posible extraer algunas observaciones de carácter general.

En primer lugar, vemos en la tabla 7, que en las comparaciones que involucran a los tripanosomas africanos y *Trypanosoma cruzi* (filas 3, 4 y 5), se usa siempre *L. major* como grupo externo. Esto último está justificado, por el hecho que el género *Trypanosoma* es un taxón monofilético. Podemos ver en todas estas comparaciones, que siempre existen más COGs donde evoluciona significativamente más rápido la especie africana que *T. cruzi*. Entonces, en las comparaciones que involucran a *T. brucei brucei*, *T. congolense*, y *T. vivax* con *T. cruzi*, los genes evolucionan más rápido en los africanos, en una relación de 4 a 1 (*T. brucei* - *T. cruzi*) a una de 50 a 1 (*T. vivax* - *T. cruzi*).

Además, se compararon las distintas especies de tripanosomas africanos entre sí. Se observa que *T. congolense* evolucionó más rápido que *T. brucei brucei* (filas números 2 y 7, de dicha tabla), y que *T. vivax* (fila número 9). Por otro lado al comparar estas 2 últimas entre sí, vemos que *T. vivax* evolucionan más rápido que *T. brucei brucei* (fila número 8).

Por último observamos que dentro de la especie *T. brucei*, la subespecie *T. b. gambiense* evolucionan más rápido que *T. b. brucei* (fila número 6).

En resumen, se desprende que los tripanosomas africanos poseen más cantidad de genes “rápidos” que el resto de los tripanosomas; por lo tanto, podemos decir que en términos generales evolucionan más rápido, y dentro de los africanos *T. congolense* está en primer lugar en su tasa evolutiva y *T. vivax* en segundo.

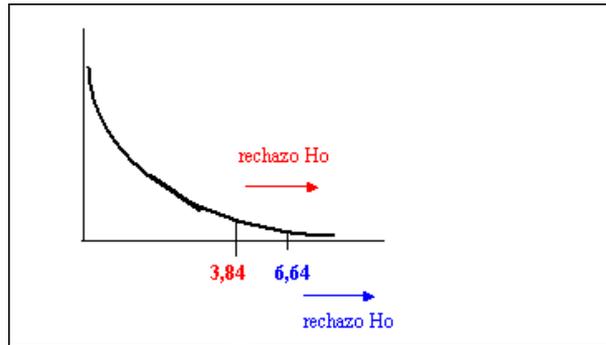


Tabla 7: Test de tasas relativas

Nro	Test de Tasas Relativas	Outgr.	spp 1	spp 2	NO sign.	dist. eq	significativo al 5%				significativo al 1%				TOTALES				Nro. de sec.
							spp 1	valor	spp2	valor	spp 1	valor	spp 2	valor	Li	Lm	Tbb	Tco	
1	Li Lm Lb	Lb	Li	Lm	5497	695	Li	58	Lm	404	Li	102	Lm	186	Li	160	Lm	590	6942
2	Tbb Tco Lm	Lm	Tbb	Tco	3864	302	Tbb	56	Tco	194	Tbb	22	Tco	114	Tbb	78	Tco	308	4552
3	Tbb Tcr Lm	Lm	Tbb	Tcr	2985	150	Tbb	337	Tcr	69	Tbb	277	Tcr	96	Tbb	614	Tcr	165	3864
4	Tco Tcr Lm	Lm	Tco	Tcr	2424	134	Tco	360	Tcr	31	Tco	491	Tcr	10	Tco	851	Tcr	41	3450
5	Tcr Tvi Lm	Lm	Tcr	Tvi	2316	124	Tcr	8	Tvi	333	Tcr	5	Tvi	394	Tcr	13	Tvi	727	3180
6	Tbb Tbg Tco	Tco	Tbb	Tbg	2734	2881	Tbb	21	Tbg	56	Tbb	3	Tbg	72	Tbb	24	Tbg	128	5767
7	Tbb Tco Tvi	Tvi	Tbb	Tco	3661	231	Tbb	66	Tco	389	Tbb	51	Tco	307	Tbb	117	Tco	696	4705
8	Tbb Tvi Tcr	Tcr	Tbb	Tvi	2877	141	Tbb	124	Tvi	292	Tbb	136	Tvi	347	Tbb	260	Tvi	639	3917
9	Tco Tvi Tcr	Tcr	Tco	Tvi	2549	152	Tco	241	Tvi	155	Tco	245	Tvi	154	Tco	486	Tvi	309	3416

Referencias: Li__*Leishmania infatum*
 Lm__*Leishmania major*
 Lb__*Leishmania braziliensis*
 Tbb__*Trypanosoma brucei brucei*
 Tco__*Trypanosoma congolense*
 Tcr__*Trypanosoma cruzi*
 Tbg__*Trypanosoma brucei gambiense*
 Tvi__*Trypanosoma vivax*

Inferencias filogenéticas

Una forma alternativa de tener una visión general de las tasas de evolución es la aproximación filogenómica, ésta básicamente consiste en reconstruir la historia evolutiva de las especies en cuestión, usando un número de genes suficientemente alto, de forma tal que pueda ser considerado representativo del genoma. En tal sentido se construyó un árbol filogenético con toda la información a nuestra disposición correspondiente a las 8 especies de estudio.

Para tal fin, se utilizó primero el script **agregalor**. El mismo nos permitió agregar al 3-COG de las especies, *L. infatum*, *L. major* y *L. braziliensis*, (fila 1 de la tabla 7) los genes ortólogos de una 4^a especie, luego una 5^a, y así recursivamente hasta la 8^a especie,

obteniendo de esta forma, los genes comunes de nuestras 8 especies de interés, y generando un 8-COG, como se puede ver en la figura 17.

```

6:alanina - default - SSH Secure Shell
File Edit View Window Help
Quick Connect Profiles
LbrM01_V2.0030 LinJ01_V3.0630 LmjF01.0610 Tc00.1047053511585.40 Tb09.160.2990 Tbgamb.22086 Scaff_860_1 Tviv114e09.plk_gen19
LbrM01_V2.0050 LinJ01_V3.0020 LmjF01.0020 Tc00.1047053508501.300 Tb09.160.2280 Tbgamb.21891 Scaff_696_5 Tviv199a07.qlk_gen2
LbrM01_V2.0060 LinJ01_V3.0030 LmjF01.0030 Tc00.1047053508501.290 Tb09.160.2260 Tbgamb.21886 Scaff_696_4 Tviv199a07.qlk_gen7
LbrM01_V2.0110 LinJ01_V3.0080 LmjF01.0080 Tc00.1047053506895.10 Tb09.160.2230 Tbgamb.21871 Scaff_1968_2 Tviv1334g02.plk_gen10
LbrM01_V2.0130 LinJ01_V3.0100 LmjF01.0100 Tc00.1047053508501.240 Tb09.160.2220 Tbgamb.21866 Scaff_695_5 Tviv1334g02.plk_gen5
LbrM01_V2.0140 LinJ01_V3.0110 LmjF01.0110 Tc00.1047053506893.80 Tb09.160.2210 Tbgamb.21861 Scaff_695_4 Tviv1334g02.plk_gen2
LbrM01_V2.0160 LinJ01_V3.0130 LmjF01.0130 Tc00.1047053508501.210 Tb09.160.2180 Tbgamb.21851 Scaff_1985_2 Tviv1263g01.plk_gen15
LbrM01_V2.0170 LinJ01_V3.0140 LmjF01.0140 Tc00.1047053506893.50 Tb09.160.2170 Tbgamb.21846 Scaff_1985_1 Tviv1263g01.plk_gen12
LbrM01_V2.0180 LinJ01_V3.0150 LmjF01.0150 Tc00.1047053506893.40 Tb09.160.2160 Tbgamb.21841 Scaff_1422_2 Tviv1263g01.plk_gen9
LbrM01_V2.0300 LinJ01_V3.0270 LmjF01.0270 Tc00.1047053504797.40 Tb09.160.2020 Tbgamb.21786 Scaff_2064_1 Tviv676c08.plk_gen35
LbrM01_V2.0310 LinJ01_V3.0280 LmjF01.0280 Tc00.1047053504797.30 Tb09.160.2000 Tbgamb.21776 Scaff_1504_3 Tviv676c08.plk_gen32
LbrM01_V2.0320 LinJ01_V3.0290 LmjF01.0290 Tc00.1047053504797.20 Tb09.160.1990 Tbgamb.21771 Scaff_1504_2 Tviv676c08.plk_gen31
LbrM01_V2.0340 LinJ01_V3.0310 LmjF01.0310 Tc00.1047053511577.110 Tb09.160.1950 Tbgamb.21756 Scaff_1274_10 Tviv676c08.plk_gen26
LbrM01_V2.0400 LinJ01_V3.0370 LmjF01.0350 Tc00.1047053511577.160 Tb09.160.2400 Tbgamb.21926 Scaff_1427_2 Tviv1721a02.plk_gen31
LbrM01_V2.0430 LinJ01_V3.0400 LmjF01.0380 Tc00.1047053504231.20 Tb09.160.2460 Tbgamb.21951 Scaff_1152_7 Tviv1721a02.plk_gen23
LbrM01_V2.0440 LinJ01_V3.0410 LmjF01.0390 Tc00.1047053504231.30 Tb09.160.2470 Tbgamb.21956 Scaff_1152_8 Tviv1721a02.plk_gen18
LbrM01_V2.0470 LinJ01_V3.0430 LmjF01.0420 Tc00.1047053506593.30 Tb09.160.2490 Tbgamb.21966 Scaff_1650_5 Tviv1721a02.plk_gen13
LbrM01_V2.0480 LinJ01_V3.0450 LmjF01.0430 Tc00.1047053506829.50 Tb09.160.2620 Tbgamb.21981 Scaff_1153_2 Tviv1721a02.plk_gen6
LbrM01_V2.0500 LinJ01_V3.0470 LmjF01.0450 Tc00.1047053506829.80 Tb09.160.2600 Tbgamb.21976 Scaff_1651_1 Tviv1808h09.qlk_gen5
LbrM01_V2.0510 LinJ01_V3.0480 LmjF01.0460 Tc00.1047053506829.90 Tb09.160.2650 Tbgamb.21991 Scaff_1153_3 Tviv1808h09.qlk_gen11
--MÁ;s--
Connected to alanina SSH2 - aes128-cbc - hmac-md5 - none 142x21

```

Figura 17: Captura de pantalla con los primeros 20 genes ortólogos de las 8 especies de tripanosomátidos (8-COG) de un total de 3371. Para ver la lista completa ir al apéndice B. Cada columna es una especie y cada fila es un grupo de 8 genes ortólogos

Una vez obtenido el 8-COG, con los genes correspondientes a las 8 especies de interés, se utilizó el script **myscplit**, para obtener de esta lista de nombres de genes sus respectivas secuencias aminoacídicas.

Luego, ante la necesidad de que se encuentren alineadas todas las secuencias en un único alineamiento, fue utilizado el programa **Clustal**, realizando los alineamientos para cada uno de los COGs por separado y concatenándolos posteriormente.

Una vez obtenido este alineamiento múltiple, se utilizó el **RAxML-7.0.4**, programa de inferencia filogenética, que permite obtener árboles filogenéticos basados en máxima verosimilitud secuencial y paralela, y con la ventaja, de procesar gran cantidad de datos, como lo requiere este tipo de estudio.

A pesar de esta ventaja, la información que poseíamos en el alineamiento era tal, que fue necesario dividirla en 2 sets de datos . Por lo que fueron obtenidos 2 árboles

filogenéticos, que luego se promediaron, obteniendo un único árbol filogenético visualizado a través del programa **Figtree.v1.2.2**. En la figura 18 se presenta el filograma resultante

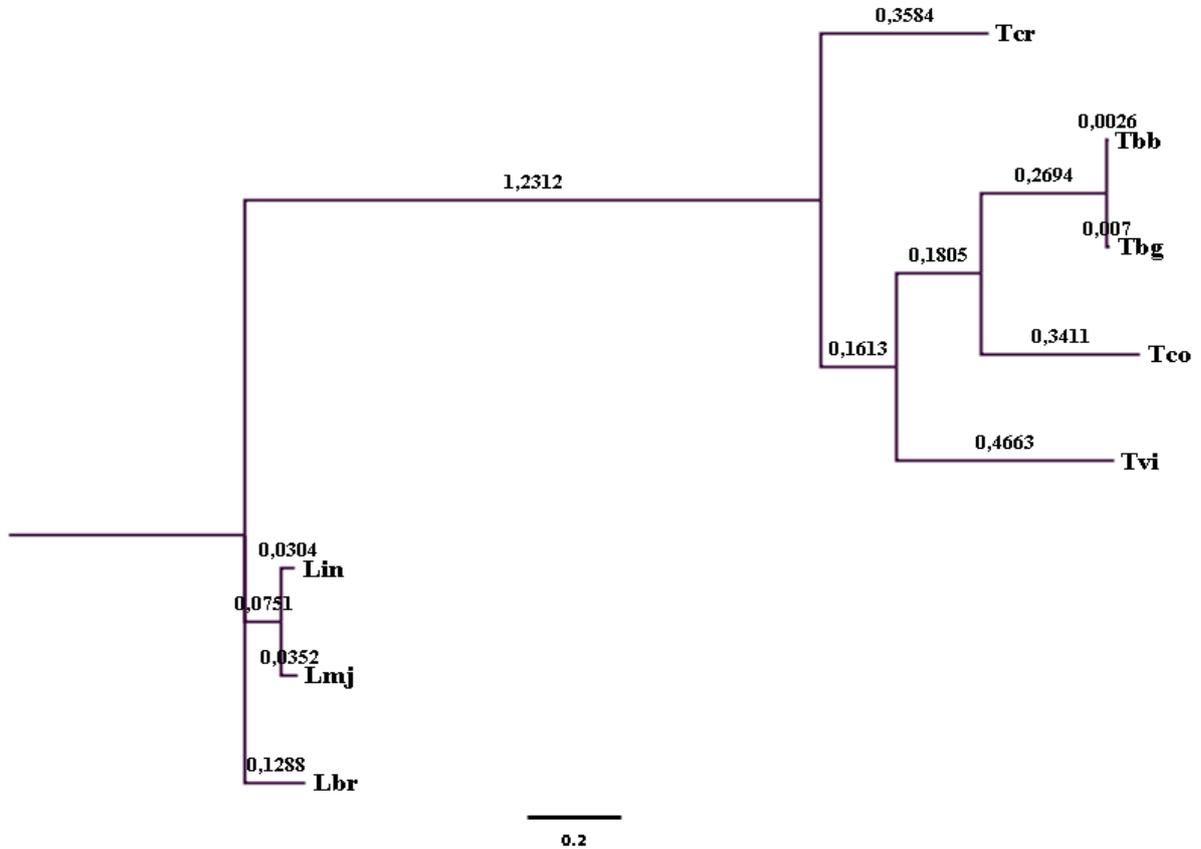


Figura 18

Filograma de la familia Trypanosomatidae, con la información de 3371 genes y obtenido por el método de máxima verosimilitud.

- Referencias: Li__*Leishmania infatum*
 Lm__*Leishmania major*
 Lb__*Leishmania braziliensis*
 Tbb__*Trypanosoma brucei brucei*
 Tco__*Trypanosoma congolense*
 Tcr__*Trypanosoma cruzi*
 Tbg__*Trypanosoma brucei gambiense*
 Tvi__*Trypanosoma vivax*

En dicho árbol filogenético, es posible tener una visión panorámica de la velocidad de evolución de los diferentes taxa observando el largo de las ramas. Dicho panorama es completamente compatible con el presentado en la sección anterior basado en el test de las tasas relativas. En concreto se puede apreciar que los tripanosomas africanos

evolucionan a una tasa más alta que *T. cruzi*, y que dentro de los africanos el orden en tasa evolutiva es el siguiente:

T. congolense, seguido de *T. vivax*, luego *T. brucei brucei*, *T. brucei gambiense*.

Lamentablemente por el momento no es posible realizar comparaciones entre las especies del género *Trypanosoma* y aquellas del género *Leishmania* debido a la ausencia de datos de un grupo externo adecuado

Búsqueda de causas funcionales de la aceleración

En las 2 secciones anteriores se presentaron resultados que muestran que la velocidad de evolución aminoacídica está lejos de ser homogénea en los tripanosomátidos, destacándose el hecho que los tripanosomas africanos son los que presentan la mayor velocidad de evolución y dentro de éstos se señala a *T. congolense* como el más veloz. En la presente sección intentaremos abordar el tema de las posibles causas funcionales de dicha aceleración. En particular nos concentraremos en los genes de los tripanosomas africanos con respecto al conjunto de genes ortólogos de todos los tripanosomas.

Debe tenerse en cuenta que determinar las causas funcionales del cambio de la tasa evolutiva no es trivial, pues esto puede deberse a una multiplicidad de factores, varios de ellos difíciles de determinar. En términos generales podemos decir que el cambio en la velocidad de evolución aminoacídica puede deberse a tres tipos de factores a saber: Incremento en la tasa mutacional, menor incidencia en la selección negativa o purificadora (ya sea porque esta es menos efectiva, debido por ejemplo a una disminución del tamaño poblacional, o debido a la relajación de las restricciones funcionales) o debido a la selección adaptativa (positiva).

Una diferencia sustancial entre los dos primeros factores y el último que éstos afectan a todo el genoma, es decir "empujan" como conjunto en una dirección, mientras que el último se espera que afecte a un grupo de genes, aquellos que están involucrados en la adaptación de los organismos a las nuevas condiciones ambientales. La idea subyacente en relación a este análisis es la siguiente. Si la aceleración o desaceleración obedece a causas funcionales, es decir es debida a la selección natural positiva, se espera entonces que afecte a grupos de genes con funciones específicas, es decir aquellos involucrados con la adaptación del organismo a las nuevas condiciones ambientales.³⁷ Esto en principio podría detectarse mediante un análisis de enriquecimiento de términos de ontología, es decir comparar los genes que presentan diferencias significativas en sus

tasas de evolución en relación al conjunto de genes del genoma y determinar qué términos GO están sobre-representados. Tal como fue mencionado, se eligió de forma arbitraria a *T. brucei* como representativo del grupo de tripanosomas africanos. A continuación se aprecian los 4 tipos de estudios realizados con sus explicaciones respectivas.

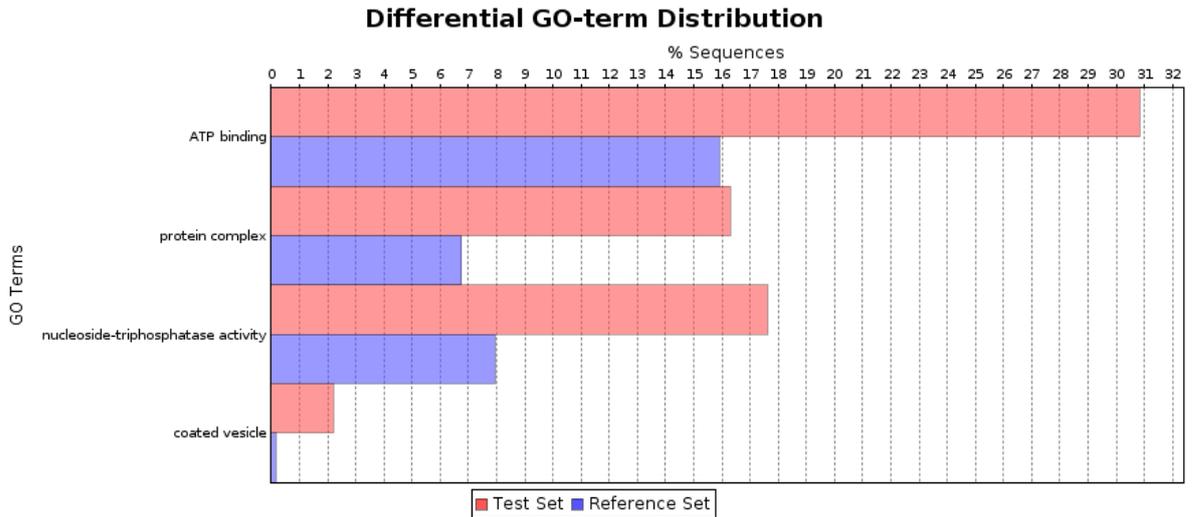


Figura 19_Distribución de los términos del GO para los genes más evolucionados de *T. brucei* sin tener en cuenta proteínas hipotéticas.

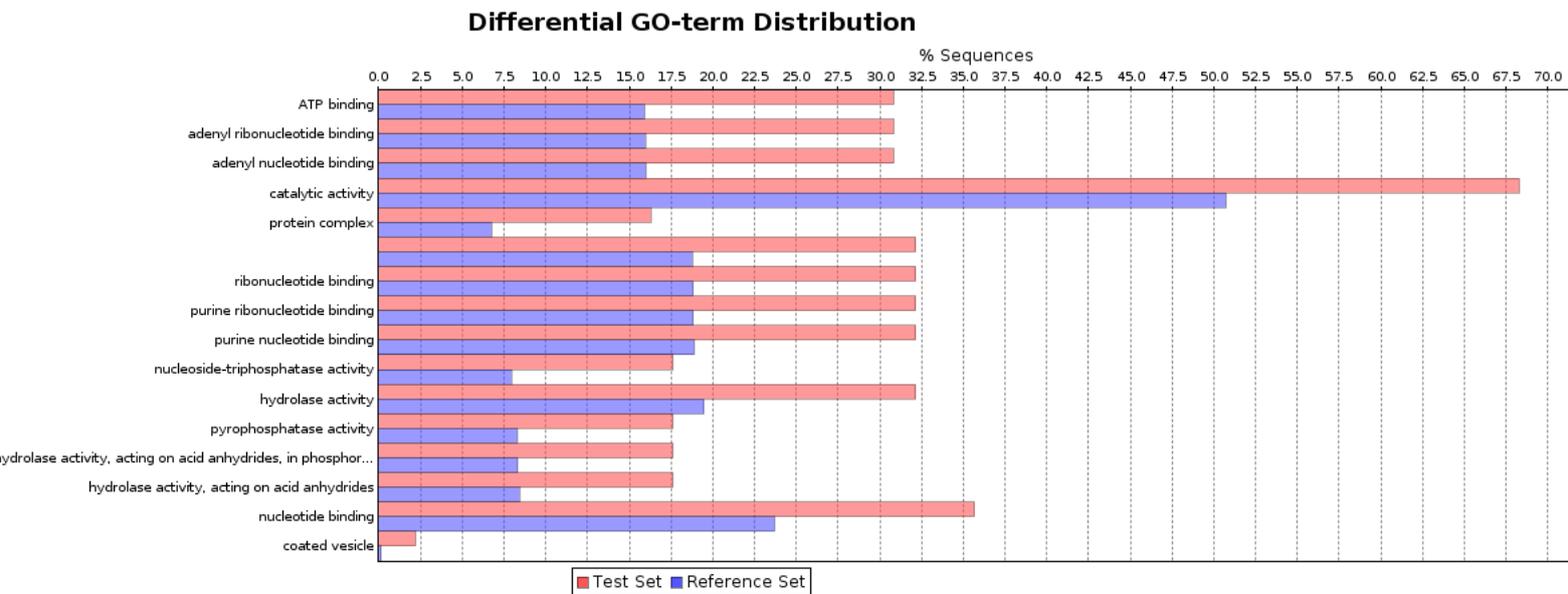


Figura 20_Distribución de los términos del GO para los genes más evolucionados de *T. brucei* sin incluir las proteínas hipotéticas y con un nivel mayor de abstracción dentro de la ontología.

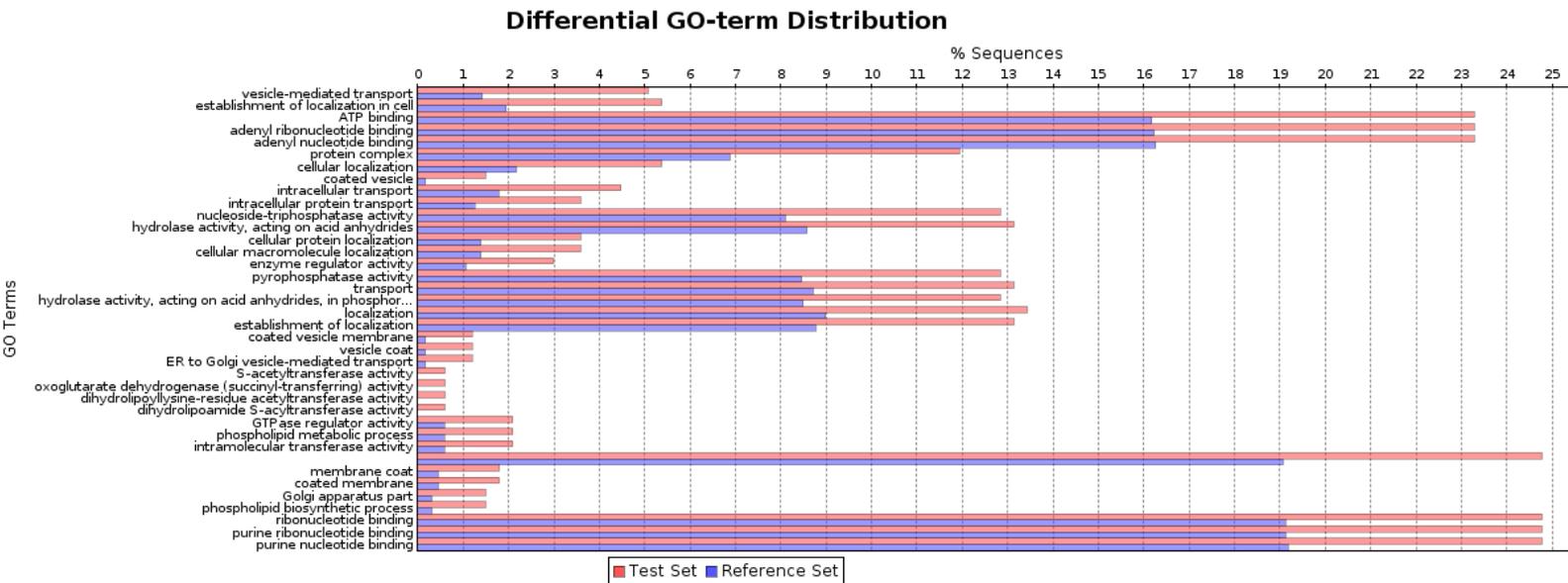


Figura 21_Distribución de los términos del GO para los genes más evolucionados de *T. brucei* realizado por medio de un test exacto de Fisher con p-valor=0.01, con los 614 genes

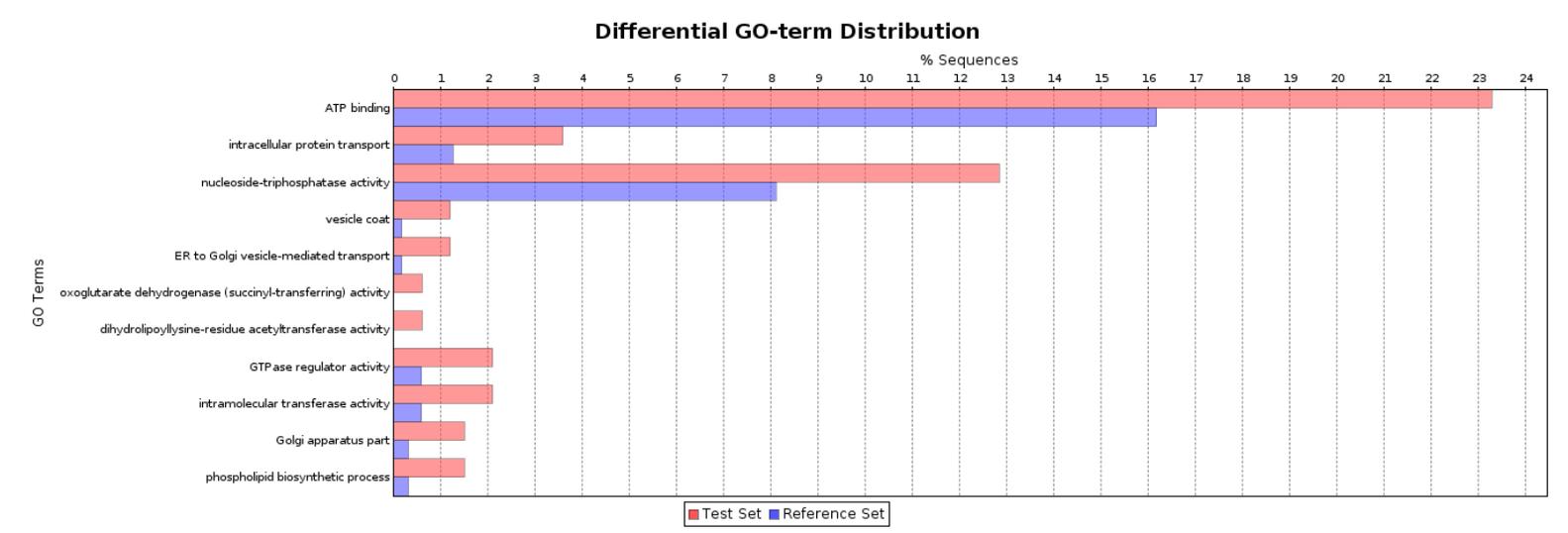


Figura 22_Distribución de los términos del GO para los genes más evolucionados de *T. brucei* realizado por medio de un Test exacto de Fisher con p-valor=0.01, con menor cantidad de genes

De la figura 19 principalmente (y figuras 20-22) se desprende que en el subgrupo de estudio predominan proteínas de unión a ATP, proteínas formadoras de complejos, actividad nucleósido trifosfato y proteínas que recubren vesículas.

De todas maneras, no se logra identificar claramente este enriquecimiento, con alguna causa funcional que lo esté generando, y/o que pueda asociarse en forma directa al estilo de vida de los tripanosomas africanos. Esto no quiere decir que la aceleración no se deba a causas de tipo adaptativo, pero el análisis aquí presentado no permite arrojar conclusiones claras.

En cuanto a los dos factores mencionados anteriormente (incremento en la tasa mutacional y debilitamiento de la selección purificadora), estos podrían testarse comparando la relación entre sustituciones sinónimas y no sinónimas (D_n/D_s), pues la mutación no afectaría esta relación mientras que la menor efectividad de la selección llevaría a un incremento de D_n pero no de D_s resultando en una disminución de los valores de este cocientes. Lamentablemente las distancias evolutivas entre tripanosomas africanos y americanos son demasiado grandes como para calcular de forma confiable los valores de D_s y por lo tanto impiden la aplicación de este test.

DISCUSIÓN

Resumen y discusión general

En este trabajo se estudió la velocidad de evolución molecular en los tripanosomátidos, utilizando el test estadístico de tasas relativas de Tajima.

Para tal fin, se analizaron 8 especies de interés tomando de a grupos constituidos por 3 especies, y alineando sus secuencias ortólogas, nos permitió comparar 2 especies cada vez con respecto a un grupo externo, con el fin de ver cuál de estas 2 especies evolucionaba más rápidamente.

Para esto se toman las sustituciones que ocurrieron en cada especie, comparándolas con el grupo externo, usando como hipótesis nula que habría la misma cantidad de cambio en las 2 especies, $E(m_1) = E(m_2)$. Usando la distribución de chi-cuadrado podemos, determinar si la diferencia observada entre $E(m_1)$ y $E(m_2)$ es mayor de lo que esperaría por azar, es decir se calcula que tan grande es esta diferencia, confrontándola con cierto p-valor arbitrario. Esto nos permite afirmar o refutar si las 2 especies evolucionan a la misma velocidad, y en caso de no hacerlo, poder ver cuál de ambas 2 evoluciona más.

Dentro de nuestras especies de estudio pudimos determinar que los tripanosomas africanos poseen más cantidad de genes “rápidos” y por lo tanto podemos decir que evolucionan más, y dentro de los africanos *T. congolense* en primer lugar seguido de *T. vivax*.

Esto lo pudimos probar tanto, a través de la realización de los test de tasas relativas, como por intermedio del filograma de secuencias concatenadas presentado en la figura 18.

La filogenia de los tripanosomas, el TTR de Tajima, y los métodos de reconstrucción de árboles filogenéticos

Como se dijo anteriormente, análisis anteriores indicaban al género *Trypanosoma* no monofilético, luego estudios posteriores mostraron que sí lo es.^{25. 26} Para la utilización del test de tasas relativas de Tajima, es crítico conocer con precisión la filogenia a utilizar, por ser ésta uno de los requerimientos para la utilización del test. Visto y considerando que dentro del género *Trypanosoma*, el grupo africano de tripanosomas evoluciona bastante más rápido que el resto, en este trabajo se pudo mostrar, porque el método de reconstrucción de árboles filogenéticos basado en distancia, UPGMA, podría fallar a la hora de reconstruir la filogenia del género *Trypanosoma*.

Además utilizando el método de máxima verosimilitud vimos que dentro del género *Trypanosoma*, los tripanosomas africanos forman un grupo monofilético.

El test de tasa relativa de Tajima y la hipótesis del reloj molecular

Al utilizar este método no afirmamos *a priori* la constancia del reloj molecular para un gen específico en particular, sino que asumimos que cada gen estaría funcionando como un “reloj local”, tomamos como hipótesis nula la equivalencia en la velocidad de evolución para 2 genes ortólogos particulares, al rechazar dicha hipótesis, afirmamos que 2 genes evolucionan a distinta velocidad, y a continuación observamos cual evoluciona más.

Los genes imparálogos

Es importante apuntar que al árbol de la figura 19 se le quitaron más de mil genes imparálogos, es decir, genes, con distinto nombre, pero que pueden representar duplicaciones que hayan ocurrido luego del proceso de especiación. El árbol obtenido luego de esa quita de genes y realizado por el mismo método, el de máxima verosimilitud, tuvo prácticamente la misma topología (árbol no mostrado). Es muy importante poder reconocerlos para quitarlos de cualquier estudio con el fin de evitar el ruido que puedan causar.

Búsqueda de causas funcionales de la aceleración

Los análisis de ontología de genes no permiten encontrar alguna causa funcional clara de la aceleración del subconjunto de genes que evolucionaron más en los tripanosomas africanos. Se requerirán análisis adicionales para clarificar este importante asunto, pero éstos están más allá de los objetivos de la presente tesina.

BIBLIOGRAFÍA

1. <http://www.wicksteadlab.co.uk/research.shtml>
2. Cevallos A.M & Hernández R. *Trypanosoma cruzi* y la enfermedad de Chagas (tripanosomiasis americana)
<http://biblioweb.unam.mx/libros/microbios/Cap15/capitulo.html>
3. Berná L (2010) Anotación y genómica comparativa de genes codificantes de VSGs de tripanosomas africanos. Tesis de Grado. Facultad de Ciencias. Universidad de la República Oriental del Uruguay
4. De Souza W (1999) A short review on the morphology of *Trypanosoma cruzi*: from 1909 to 1999. Mem. Inst. Oswaldo Cruz. 94,1:17-36.
5. De Souza W & Meyer H (1974). On the fine structure of the nucleus in *Trypanosoma cruzi* in tissue culture forms. Spindle fibers in the dividing nucleus. J Protozool. 21:48-52.
6. Berriman M et al. (2005) The Genome of the african trypanosome *Trypanosoma brucei*. Science. v. 309 p. 416-422.
7. Clayton C.E (2002) Life without transcriptional control? From fly to man and back again. EMBO Journal. v. 21, no. 8 p.1881–1888

8. Haile S & Papadopoulou B (2007) Developmental regulation of gene expression in trypanosomatid parasitic protozoa. *Current Opinion in Microbiology*. v. 10, no. 6, p. 569–577.
9. <http://es.wikipedia.org/wiki/Trypanosoma>
10. Donelson J.E & Rice-Fitch A.C (1985) Molecular biology of Trypanosome antigenic variation. *Microbiological Reviews* v. 49, n. 2, p. 107-125
11. <http://www.dpd.cdc.gov/dpdx/>
12. Lesk M. Arthur.(2002) *Introduction to Bioinformatics* Oxford University Press 283p
13. El-Sayed N.M. et al. (2005) The Genome Sequence of *Trypanosoma cruzi*, Etiologic agent of Chagas disease. *Science*. v. 309 p. 409-415.
14. Ivens A.C, et al. (2005) The genome of the kinetoplastid parasite, *Leishmania major*. *Science*. v. 309 p. 436-442.
15. El-Sayed N.M. et al. (2005) Comparative genomics of trypanosomatid parasitic protozoa. *Science*. v. 309 p. 404-409.
16. http://es.wikipedia.org/wiki/Evoluci%C3%B3n_molecular
17. Stickberger, M.W. (1985) *Genetics*. 3rd. Macmillan, New York
18. Stadler L. J (1939). Genetic studies with ultra-violet radiation. *Proc. 7th. Congr. Genet.* pp. 296-76
19. Nachman, Michael W. & Crowell, Susan L. 2000. Estimate of the Mutation Rate per Nucleotide in Humans. *Genetics* 156, 297-304.

20. Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge
21. Alvarez-Valín, F. 2000. Evolución molecular: neutralismo y seleccionismo. Capítulo 12 de *El prisma de la Evolución*. Editado por C. Altuna & M. Ubilla. Facultad de Ciencias, UdelaR.
22. PBS Evolution Library
<http://www.pbs.org/wgbh/evolution/library/>
23. Dickerson, R. E. (1971) The structure of cytochrome c and the rates of molecular evolution. *J. Molec. Evol.* 1, 26-45.
24. Ponce de León. M (2010) *Métodos computacionales aplicables a la anotación de genomas y reconstrucción de redes metabólicas*. Tesis de Grado. Facultad de Ciencias. Universidad de la República Oriental del Uruguay
25. Alvarez F, Cortinas MN, Musto H, (1996) The Analysis of Protein Coding Genes Suggests Monophyly of Trypanosoma. *Molecular Phylogenetics and Evolution*. v. 5, n. 2, p. 333-343
26. Stevens JR, Gibson W, (1999). The molecular evolution of trypanosomes. *Parasitol Today*. v. 15 p. 432-437.
27. Page, R. D.M. & Holmes, E. C. (1998). *Molecular Evolution. A Phylogenetic Approach*. Oxford, UK: Blackwell Science Ltd.
28. <http://www.demiurgo.org/doc/shell/shell-2.html>
29. Altschul, S.; Gish, W.; Miller, W.; Myers, E. Lipman, D. (1990). Basic local alignment search tool. *Journal of Molecular Biology*. v. 215; 403-410
30. Larkin M.A, Blackshields G, Brown N.P, Chenna R, McGettigan P.A, McWilliam H, Valentin F, Wallace I.M, Wilm A, Lopez R, Thompson J.D,

- Gibson T.J, Higgins D.G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, 23
31. Alexandros Stamakis (2006) RAxML-VI-HPC: Maximum Likelihood-based Phlogenetic Analyses with Thousands of Taxa and Mixed Models. *Bioinformatics* 22(21): 2688-2690
 32. Lemey, P. Salemi, M. Vandamme, A-M. (2008) *The Phylogenetic Handbook. A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, 2ed.
 33. Haag J, O’Huigin C & Overath P(1998).The molecular phylogeny of trypanosomes: evidence for an early divergence of the Salivaria. *Mol. Biochem.Parasitol.* 91:37–49.
 34. Tajima F (1993) Simple Methods for Testing the Molecular Evolution Clock Hypotesis. *Genetics* 135: 599-607
 35. Tamura K, Dudley J, Nei M & Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24:1596-1599.
 36. Conesa, S. Götz, J. M. Garcia-Gomez, J. Terol, M. Talon and M. Robles (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, v 21, September, 2005, pp. 3674-3676.
 37. Itoh, T., Martin, W., Nei, M. (2002). Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. *Proc. Natl. Acad. Sci.* 99, 12944–12948.

GLOSARIO

UPGMA (Unweighted Pair Group Method with Arithmetic Mean)

Es un simple método de agrupamiento jerárquico usado en bioinformática para la creación de fenogramas. UPGMA asume una tasa constante de evolución (hipótesis del reloj molecular), y no es considerado un buen método para inferir parentesco, a menos que esta asunción haya sido testada y justificada para los datos que se estén usando.

Neighbor joining

Es un método de agrupamiento para la creación de árboles fenéticos, fenogramas. Usualmente usado para árboles basados en datos de secuencias de ADN y proteicas. El algoritmo requiere conocimiento de las distancias entre cada par de taxa (es decir, especies o secuencias) para formar el árbol.

Pipeline

La arquitectura en pipeline (basada en filtros) consiste en ir transformando un flujo de datos en un proceso comprendido por varias fases secuenciales, siendo la entrada de cada una la salida de la anterior. Esta arquitectura es muy común en el desarrollo de programas para el intérprete de comandos, ya que se pueden concatenar comandos fácilmente con tuberías (pipe). También es una arquitectura muy natural en el paradigma de programación funcional, ya que equivale a la composición de funciones matemáticas.

Parsimonia

Es un método estadístico no paramétrico, comúnmente usado en filogenética computacional para estimar filogenias. Bajo la parsimonia, el árbol filogenético preferido es el árbol que requiere el mínimo cambio evolutivo para explicar los datos observados. La parsimonia es parte de una clase de métodos estimadores de árboles basados en caracteres, el cual usa una matriz de caracteres filogenéticos discretos, para inferir uno o más árboles filogenéticos óptimos para un conjunto de taxa. Estos métodos operan evaluando árboles filogenéticos candidatos de acuerdo a un criterio de optimización explícito; el árbol con el puntaje más favorable es tomado como el mejor estimador de las relaciones filogenéticas del taxa. Máxima parsimonia es usada con la mayoría de los tipos de datos filogenéticos; hasta hace poco tiempo era el único método estimador de árboles filogenéticos basado en caracteres, ampliamente utilizado por datos morfológicos.

Proteoma

Colección entera de proteínas codificadas por el genoma de un organismo.

BLAST (Basic Local Alignment Search Tool)

Es un programa informático de alineamiento de secuencias de tipo local, ya sea de ADN o de proteínas. El programa es capaz de comparar una secuencia problema (también denominada en la literatura secuencia *query*) contra una gran cantidad de secuencias que se encuentren en una base de datos. El algoritmo encuentra las secuencias de la base de datos que tienen mayor parecido a la secuencia problema. Es importante mencionar que BLAST usa un algoritmo heurístico por lo que no nos puede garantizar que ha encontrado la solución correcta. Sin embargo, BLAST es capaz de calcular la significación de sus resultados, por lo que nos provee de un parámetro para juzgar los resultados que se obtienen.

Hipótesis nula

Es una hipótesis en estadística construida para anular, o refutar, con el objetivo de apoyar una hipótesis alternativa. Cuando se la utiliza, la hipótesis nula se presume verdadera hasta que una evidencia estadística en la forma de una prueba empírica de la hipótesis indique lo contrario.

Ortólogo: La ortología es una relación de ancestría entre uno o más genes. Los genes ortólogos provienen de un mismo gen que se encontraba presente en la última especie ancestral común. Al producirse un evento de especiación ambas copias del gen divergen en dos nuevos linajes. En general se espera que los genes ortólogos continúen cumpliendo la misma función biológica en los nuevos linajes.

Parálogo

La paralogía es una relación de ancestría entre uno o más genes. Los genes parálogos, surgen como producto de la duplicación de un segmento cromosómico que resulta en dos copias de un mismo gen dentro del mismo genoma. Durante la evolución lo más común es que una sola de las copias mantenga su antigua función, mientras que la otra toma un rumbo diferente. Puede por ejemplo evolucionar hacia una nueva función, o acumular mutaciones hasta convertirse en un pseudogen.

Imparálogo

Parte de los genes parálogos cuya especie que la porta sufrió una duplicación luego de especiar.

APÉNDICE

En este apéndice se muestra la base de datos utilizada en este trabajo.

Se anexa:

- archivo **8-COGs**, con lista de 3371 grupo de ortólogos para las 8 especies
- archivo **8-COGs_2250genes**, con lista de 2250 grupo de ortólogos para las 8 especies (sin repetidos)