



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**DETECCIÓN Y DATACIÓN DE EVENTOS DE
TRANSFERENCIA HORIZONTAL DE GENES
DE AMIBAS DEL GÉNERO *ENTAMOEB*A CON
BACTERIAS Y ARQUEAS**

T E S I S

QUE PARA OBTENER EL TÍTULO DE

B I Ó L O G O

P R E S E N T A:

MIGUEL FERNANDO ROMERO GUTIÉRREZ



**DIRECTOR DE TESIS:
DOCTOR RENÉ CERRITOS FLORES**

Ciudad Universitaria, México D.F. 2014



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Hoja de Datos del Jurado

1. Datos del alumno

Romero
Gutiérrez
Miguel Fernando
(044)5516915668
Universidad Nacional Autónoma de México
Facultad de Ciencias
Biología
307591494

2. Datos del tutor

Doctor
René
Cerritos
Flores

3. Datos del sinodal 1

Doctora
Valeria Francisca Eugenia Leopoldina de María de Guadalupe
Souza
Zaldivar

4. Datos del sinodal 2

Doctora
Cecilia Teresa
Ximénez
García

5. Datos del sinodal 3

Doctor
José Hugo
Aguilar
Díaz

6. Datos del sinodal 4

Doctor
Luis David
Alcaraz
Peraza

7. Datos del trabajo escrito

Detección y datación de eventos de transferencia horizontal de genes de amibas del género *Entamoeba* con bacterias y arqueas
pp. 45
2014

Agradecimientos

A la Dra Cecilia Ximénez por darme la oportunidad de llevar a cabo este trabajo en su laboratorio e incorporarme dentro del mismo, en donde sin duda di mis primeros pasos como biólogo.

A mi tutor, el doctor René Cerritos, por su gran apoyo, confianza, sabiduría y amistad, por hacerme parte de su grupo y dirigir esta tesis. A mis compañeros del laboratorio, especialmente a Valeria Zermeño, a quien podría llamar hermana mayor en la biología, por compartir conmigo sus conocimientos, darme su amistad y los comentarios que me dio tras revisar este escrito. A mi asesor de servicio social, el Dr. Ángel Rodríguez, por su apoyo y paciencia.

A los sinodales que revisaron este trabajo: la Dra. Valeria Souza, de nuevo a la Dra. Cecilia Ximénez, al Dr. David Alcaraz y al Dr. Hugo Aguilar por tomarse el tiempo de evaluar y discutir este trabajo.

A cada profesor que me compartió sus conocimientos en la carrera, especialmente a Morena Avitia, Lev Jardón, Enrique Scheinvar, de nuevo a Hugo Aguilar, Rodolfo Salas, Xitlali Aguirre, Francisco Sánchez, Laura Vargas y Ana Flisser. Quiero agradecer sobre todo a Germán Bonilla, quien desde primer semestre es un ejemplo a seguir y se tomó el tiempo para mostrarme el camino en el mundo de la bioinformática.

A mi “familia” de la facultad: Alejandra, Anabell, Citlalli, Dana, Daniela, Fabiola, Fabián, Haydee y Mayra; con quienes desde primer semestre he compartido no solo horas de estudio, sino también experiencias que han marcado mi vida y nunca olvidaré.

A quienes en la facultad pude (y aún puedo) nombrarlos como mis amigos y colorearon este viaje: Memo, Nallely, Lalo, Gaby, Ariel, Alejandra, Angélica (los panchos), Ia, Marco, Emiliano, Paola, Toño, Annette, Gaby Gutiérrez, Lore, Ale, Tona (sr. coordinador), Ana Karen (Lehrerin), Daniel, Jessica, Mariana, Isaí, Iliana, Erendira, Luis, Yocelyn, Rodrigo, Silvana, Erick, Cecilia y David.

A quienes conocí afuera de la facultad, me dieron su amistad, influyeron en mi formación y consciente o inconscientemente aportaron a la realización de esta tesis: Cesar (mi maestro y guía en programación), Adolfo, Sofia, Tobías y Rodrigo.

A Adriana Franco, unsere Liebe, compañera de vida, por sus palabras de aliento, por su apoyo incondicional, por hacer que salga el sol cada vez que siento que todo se oscurece y por dejarme ver el

mundo a través de sus ojos. Gracias por mostrarme mis errores y alentarme a ser mejor, por revisar y corregir este escrito, por lo que hemos vivido y aún nos falta por compartir.

A mi padrino Mario Alberto González que es un verdadero ejemplo a seguir como persona y como académico. A mi primo Mario (colega) con quien puedo pasar horas platicando sobre cualquier tema, en especial sobre ciencias naturales y me motiva a seguir investigando.

A mi madre Norma Gutiérrez, quien no sólo me enseñó a pedir por favor y dar las gracias, sino también a luchar día a día con perseverancia y disciplina para lograr las metas que nos fijamos.

A mi padre Rubén Romero, por darnos el pan de cada día y más, pero sobre todo, porque gracias a él tengo la oportunidad de dedicar el resto de mi vida a hacer lo que me gusta.

A mi hermano Rubén Romero, con quien me he dedicado a descubrir el mundo desde que tengo memoria y me motiva a reinventarme e ir más allá sin importar el reto.

A todos ellos y a quienes por falta de espacio no fue posible incluir:

No tengo forma de agradecer lo que me han enseñado.

Dedicada a:

Mis padres, Rubén Romero y Norma Gutiérrez

y a mi hermano Rubén Romero.

Todo lo que soy y llegaré a ser se lo debo a ustedes.

Índice

I. Resumen.....	3
II. Abstract.....	4
1. INTRODUCCIÓN.....	5
1.1 El género <i>Entamoeba</i>	5
1.2 Transferencia horizontal de genes.....	7
1.3 Determinación de genes xenólogos.....	8
1.4 Fuentes de error en análisis filogenéticos.....	10
1.5 Antecedentes.....	10
2. JUSTIFICACIÓN.....	12
3. OBJETIVOS.....	13
4. MÉTODOS.....	14
4.1 Obtención de secuencias candidato.....	14
4.2 Selección de genes para datación y obtención de secuencias.....	14
4.3 Búsqueda de homólogos.....	15
4.4 Alineamiento de aminoácidos.....	15
4.5 Construcción de alineamientos de cDNA.....	15
4.6 Construcción de matrices de distancia evolutiva.....	16
4.7 Pruebas de saturación mutacional.....	16
4.8 Construcción de árboles filogenéticos por máxima verosimilitud.....	17
4.9 Construcción de árboles filogenéticos por probabilidad bayesiana.....	17
4.10 Datación de eventos de divergencia por probabilidad bayesiana.....	18
4.11 Análisis del uso preferencial de codones.....	20
5. RESULTADOS Y DISCUSIÓN.....	21
5.1 Búsqueda de homólogos.....	21
5.2 Matrices de distancia evolutiva.....	22
5.3 Análisis de saturación mutacional.....	22
5.4 Árboles filogenéticos.....	24
5.5 Estimación de tiempos de divergencia.....	31
5.6 Análisis del uso preferencial de codones.....	37
6. CONCLUSIONES.....	40
7. LITERATURA CITADA.....	41

III. ANEXOS.....	46
Anexo 1. Tabla de genes candidato.....	46
Anexo 2. Código en Python 2.7.....	49
2.1 Verificación de marcos de lectura.....	49
2.2 Construcción de alineamientos de cDNA.....	56
Anexo 3. Pruebas de saturación mutacional.....	60
Anexo 4. Filogenias a partir de alineamientos cortos.....	63
Anexo 5. Árboles consenso.....	69
Anexo 6. Tabla de grupos donadores asignados por gen.....	101

I. RESUMEN

El género *Entamoeba* está compuesto principalmente por amibas que colonizan el colon de diferentes especies de vertebrados, de entre las cuales destaca *E. histolytica* por ser uno de los parásitos protozoarios que infectan a humanos más prevalentes en países en vías de desarrollo. Estos organismos, al tratarse de protozoarios que se alimentan principalmente de bacterias, son propensos a obtener genes por transferencia lateral con mayor frecuencia que otros grupos de eucariontes. Se sabe que por medio de la transferencia horizontal de genes es posible obtener adaptaciones a diferentes ecosistemas y en organismos infecciosos su importancia aumenta porque es un mecanismo de propagación de factores de virulencia. En la primera anotación del genoma de *E. histolytica* se reportó una lista de 98 genes probablemente obtenidos por transferencia horizontal, actualmente se considera que 68 de ellos reflejan pruebas consistentes para considerarlos xenólogos.

Los objetivos del presente trabajo fueron identificar a los grupos donadores que han transferido genes a las especies relacionadas con *E. histolytica*, verificar si se ha tratado de transferencias masivas de factores adaptativos al microecosistema del colon y/o factores de virulencia e identificar genes xenólogos falsos positivos o que ya han perdido la señal filogenética dada la saturación de sustituciones.

Para ello se llevaron a cabo búsquedas de homólogos en la base de datos RefSeq del NCBI y se construyeron alineamientos de codones; estos últimos fueron sometidos a pruebas de saturación mutacional, así como a análisis filogenéticos por máxima verosimilitud y probabilidad bayesiana para asignar los posibles grupos donadores de cada gen. Posteriormente se calcularon los tiempos de divergencia de varias especies del género *Entamoeba* y con base en ello se dataron algunos eventos de transferencia horizontal. Finalmente se analizó la desviación del uso preferencial de codones de los genes candidatos. Los resultados de este trabajo revelan que este género ha adquirido mecanismos alternos de obtención de energía de distintos grupos de organismos procariontes. Es posible que la transferencia horizontal de genes haya sido relevante para la radiación de este grupo como parte de la microbiota intestinal.

II. ABSTRACT

The *Entamoeba* genus is formed by morphological similar amoebae that colonize the gut of several species of vertebrates, *E. histolytica* is specially relevant because it is one of the most important protozoan parasites endemic in developing countries. These organisms, which nourish in bacteria, are prone to get genes by horizontal gene transfer (HGT) more frequently than other groups of eukaryotes. It is well known that horizontal gene transfer is a source of evolutionary adaptations to different ecosystems and its importance grows in infectious organisms because virulence factors may spread through this mechanism. In the first annotation of the genome of *E. histolytica*, the authors proposed a list of 98 HGT candidates. Horizontal gene transfer remains the strongest hypothesis to explain 68 of the 96 original topologies.

The objective of this study was to identify the donor groups that have transferred genes to *E. histolytica* and other closely related species and verify if massive transfers of adaptation or virulence factors have occurred, as well as the identification of false positive xenologs and those gene candidates which have lost phylogenetic signal.

To study HGT within *Entamoeba*, homologs for each HGT candidates were collected from the NCBI's RefSeq database, and were aligned as codons. The codon alignments were analyzed for substitution saturation and then phylogenetic trees were built by posterior probability and maximum likelihood, in order to find the putative donor group of each candidate. Afterwards, the divergence time of several species of the *Entamoeba* genus were estimated and accordingly, some gene transfer events were dated. Finally the codon usage bias of the HGT candidates was analyzed.

The results show that this genus has acquired alternate energy metabolism pathways from different prokaryotic donors. In fact, it is likely that horizontal gene transfer has been crucial for the radiation of this genus as a member of the gut microbiota.

1. INTRODUCCIÓN

1.1 El género *Entamoeba*

El género *Entamoeba* pertenece al filum Amoebozoa (amibas *sensu stricto*) y está compuesto por amibas unicelulares cuyas diferencias morfológicas más evidentes son el tipo de motilidad y el tamaño del trofozoíto, además del número de núcleos presentes por cada quiste. La mayoría de las especies que integran este género son parásitos intestinales de organismos vertebrados, otras infectan la cavidad bucal de los mismos, entre ellos el humano y otras incluso son de vida libre (figura 1). Algunas especies de *Entamoeba* pueden infectar más de una especie de hospedero (1).

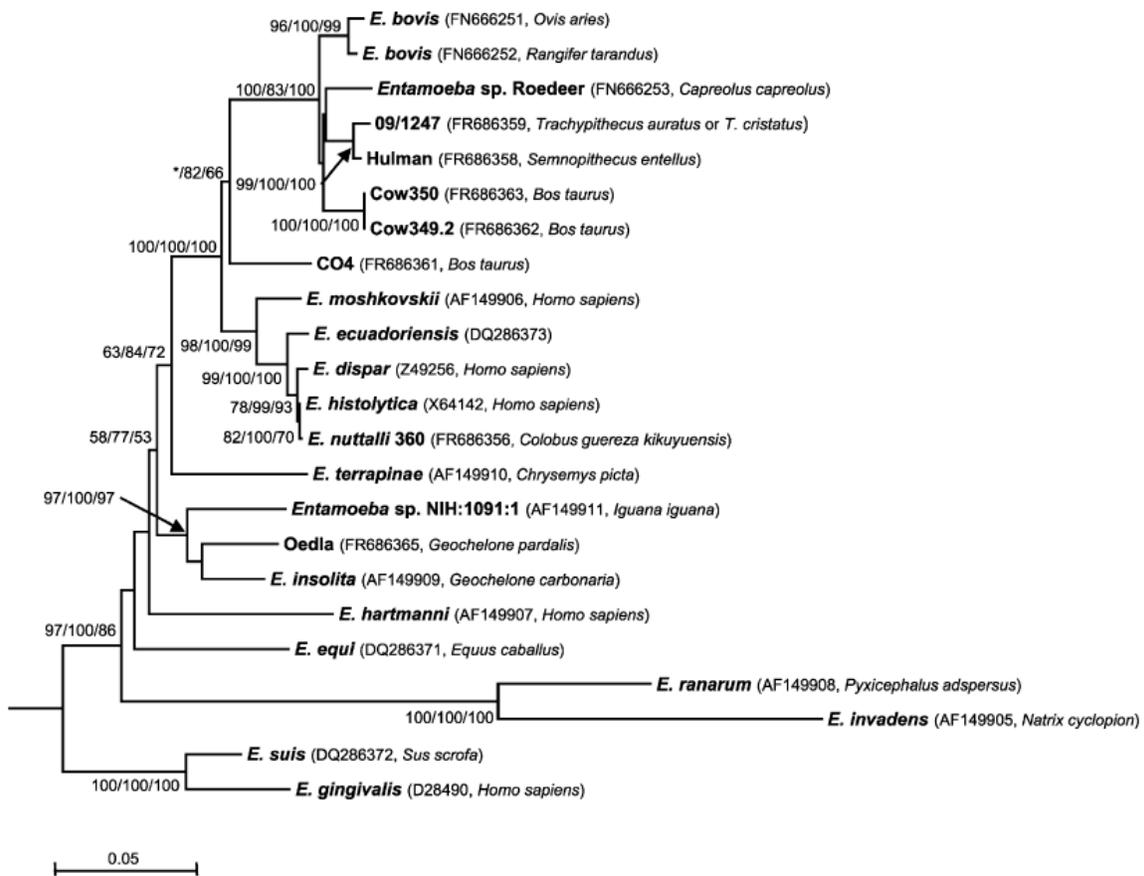


Figura 1. Árbol filogenético construido a partir de secuencias de la subunidad pequeña de RNA ribosomal de distintas especies de *Entamoeba*. El árbol mostrado es la topología obtenida por medio de Neighbor-Joining. Los valores en cada nodo pertenecen a valor bootstrap de Phyml/ probabilidad posterior de MrBayes/ Bootstrap de Neighbor-Joining. A un lado del nombre de la especie de *Entamoeba* se presenta el nombre binomial del hospedero común. La barra muestra el número de sustituciones esperadas por sitio. Modificado de Stensvold et al 2011.

Entamoeba histolytica es uno de los parásitos protozoarios más importantes a nivel mundial e infecta el intestino del humano y de otras especies de primates (2-5). La presencia de *E. histolytica* en el intestino puede permanecer asintomática o desencadenar colitis amébrica. En algunas ocasiones, esta amiba puede salir del intestino por la vena porta y colonizar el hígado produciendo abscesos, en casos extremos estos pueden aparecer en otros órganos como los pulmones o el cerebro, frecuentemente ocasionando la muerte del hospedero. Se estima que actualmente *E. histolytica* causa 70,000 decesos al año en todo el mundo (6), pero cabe mencionar que muchos autores consideran que esta cifra es imprecisa y que es necesario hacer un estudio detallado al respecto. La infección de este parásito es endémica en países en desarrollo, especialmente en regiones con altos índices de pobreza e higiene deficiente porque estas condiciones propician que se complete el ciclo de vida del parásito, provocando que una proporción importante de la población se encuentre infectada permanentemente. La transmisión de esta amiba sigue la vía fecal-oral y su ciclo de vida está compuesto por dos fases: la de quiste y la de trofozoíto. El quiste es la fase infectiva del parásito, se encuentra en cuerpos de agua contaminados por aguas residuales y por consiguiente en vegetales mal lavados, presenta cuatro núcleos y su pared celular está compuesta por quitina y diferentes proteínas. El trofozoíto, en contraste con otras amibas, se caracteriza por sus trofozoitos unicucleados o binucleados y su motilidad limitada a una sola región de la célula.

Muchas características de *E. histolytica* son un misterio: aún se ignora su ploidía y número de cromosomas. Sin embargo, se ha reportado que el contenido de DNA de los trofozoitos es 10 veces mayor en cultivos axénicos que en cultivos xénicos (7). Investigaciones a nivel de genoma han identificado los factores clave de virulencia del parásito: lectinas Gal/GalNac, proteasas de cisteína, peroxirredoxina, NADH:flavin oxidoreductasa, superóxido dismutasa, el amebaporo A, entre otros (8). En diversos estudios se ha encontrado que el grado de virulencia de *E. histolytica* puede ser afectado por otros microorganismos con los que se encuentre (9, 10). La infección por *E. histolytica* y otros patógenos microaerofílicos es tratada alrededor del mundo por metronidazol (1-(2-hidroxietil)-2-metil-5-nitroimidazol), y aunque la resistencia a este fármaco está reportada (11), su uso es efectivo en la mayoría de los casos (12).

En 2013 Tachibana y colaboradores reportaron un linaje de *Entamoeba* que infecta exclusivamente a macacos rhesus (*Macaca mulatta*) y debido a sus peculiaridades genéticas se propuso que se le considerara una especie distinta y se reviviera el nombre de *Entamoeba nuttalli* para este grupo (13). Se encontró que *E. nuttalli* infecta tanto a macacos salvajes como a especímenes en cautiverio y que es capaz de generar abscesos en hígados de hamster (14). *E. dispar* es muy similar a *E. histolytica*, a tal grado que hasta apenas algunas décadas a ambos linajes eran considerados miembros de la misma

especie. Asimismo, hasta hace poco se creía que *E. dispar* era una especie comensal del intestino humano, que se alimentaba de bacterias sin causar síntomas a los portadores. A pesar de ello, recientemente *E. dispar* se ha encontrado asociada con casos de colitis amibiana y abscesos hepáticos (15). Aunque se ha sugerido a partir de la comparación de secuencias ribosomales que la divergencia entre *E. histolytica* y *E. dispar* ocurrió hace decenas de millones de años (16-18); no obstante, hasta el día de hoy no aparecen estudios detallados que discutan este tema en la literatura. *Entamoeba invadens* presenta cuatro núcleos por quiste; infecta reptiles pudiendo ocasionar colitis amébrica, abscesos amibianos en el hígado u otros órganos e incluso puede provocar la muerte. Esta especie ha sido ampliamente estudiada debido a que es la única especie de *Entamoeba* que tiene la capacidad de formar quistes *in vitro* (1).

1.2 Transferencia horizontal de genes

La transferencia horizontal de genes o transferencia lateral de genes (HGT o LGT por sus siglas en inglés) es un fenómeno biológico en el cual un organismo incorpora a su genoma una o más secuencias de DNA de manera alterna a la transferencia vertical, que es la transmisión de genes de padre a hijo a través de la reproducción. A estos genes se les conoce como genes xenólogos. Gracias a estudios de biología molecular y de genómica comparada se sabe que este mecanismo es un proceso fundamental en la evolución genómica procarionte (19-20) y de hecho fue determinante para la formación de los tres dominios de la vida que conocemos en la actualidad (21). Por otro lado la transferencia horizontal de genes ocurre todo el tiempo en la naturaleza: se observa desde la infección viral que incorpora el provirus al genoma de la célula hospedera, hasta los procesos que otorgan la posibilidad de recombinación a las células bacterianas: conjugación, transducción y transformación. A pesar de que los procesos antes mencionados se conocen desde hace relativamente mucho tiempo y se tienen modelos detallados sobre su funcionamiento, la forma en la cual los organismos eucariontes obtienen genes por HGT es aún desconocida. En 2009, Whitaker y colaboradores propusieron que este fenómeno debería ser más frecuente en organismos eucariontes unicelulares que se alimentan de bacterias por fagocitosis (22). Aun así se ha detectado genes obtenidos por LGT en eucariontes multicelulares como los rotíferos de la clase Bdelloidea (23) e incluso en el ser humano (24).

En el caso de organismos infecciosos, la transferencia horizontal tiene una importancia especial, ya que es una fuente de factores de virulencia o de resistencia a las defensas del hospedero ya sean naturales o artificiales, como los antibióticos. Gracias al experimento de Griffith, que no sólo proporcionó pruebas de que el DNA es el portador de la información genética, desde el siglo pasado se tienen evidencias de que los factores de virulencia pueden ser transmitidos horizontalmente. Otros casos son las llamadas islas de patogenicidad que son módulos de genes relacionados con la virulencia de ciertos organismos cuya

presencia en diversos grupos se debe a la LGT (25). De igual forma la obtención de estos factores de virulencia por transferencia horizontal también ha sido descrita en organismos eucariontes, un ejemplo asombroso son los organismos fungoides (Oomycota) que han recibido genes de lisis celular y resistencia a las defensas del hospedero a partir de donadores fungales y por lo tanto han facilitado la evolución hacia la vida parasitaria (26). Por otro lado, varios ensayos de laboratorio demuestran que las bacterias son capaces de obtener factores de resistencia a antibióticos por medio de conjugación o transformación (27, 28). En el marco de la evolución de las relaciones de parasitismo, existen modelos que sugieren una estrecha relación entre las características de la resistencia del hospedero y el aumento o disminución de la virulencia del parásito (29).

Independientemente de que este fenómeno ocurra entre organismos virales, procariontes o eucariontes, la transferencia horizontal de genes puede ser una fuente de innovaciones evolutivas para el grupo receptor, ya que al incorporarse esta información a su genoma, ésta comenzará a propagarse por transferencia vertical a toda su descendencia (en organismos multicelulares la transferencia debe ocurrir en la línea celular germinal). Hay que tomar en cuenta que una vez incorporado el nuevo material al pool genético de la población, su perpetuación o desaparición está sujeta al efecto de las fuerzas evolutivas que actúan sobre la misma, principalmente la selección y la deriva génica. Por lo tanto no es posible observar todos los eventos de HGT que ha experimentado una especie, así como tampoco darle un valor adaptativo a todo gen obtenido de esta manera.

1.3 Determinación de genes xenólogos

La localización de genes obtenidos lateralmente es una tarea primordialmente bioinformática que requiere sin duda sólidos conocimientos de evolución molecular. La forma más sencilla de detectar genes xenólogos es buscar genes homólogos al gen problema en las grandes bases de datos disponibles en internet y comparar la historia evolutiva que reflejan estos genes con la historia evolutiva de las especies a las cuales pertenecen.

La falta de concordancia entre las relaciones de homología de los genes problema y las especies correspondientes, es un indicio de que haya ocurrido un evento de transferencia lateral. Para llevar a cabo esta comparación es importante tomar en cuenta que las relaciones genealógicas de un grupo de genes no necesariamente reflejan la historia evolutiva de las especies estudiadas (figura 2). Por eso, para definir la historia evolutiva de las especies comúnmente se utiliza marcadores moleculares específicos que cumplen ciertas características: deben estar presentes en todas las especies, su tasa de cambio debe ser proporcional al nivel taxonómico estudiado (por ejemplo, usar el gen *rpoB* para organismos de diferentes fila o el gen

RAG1 para comparar géneros de aves) y no deben ser genes obtenidos por transferencia horizontal.

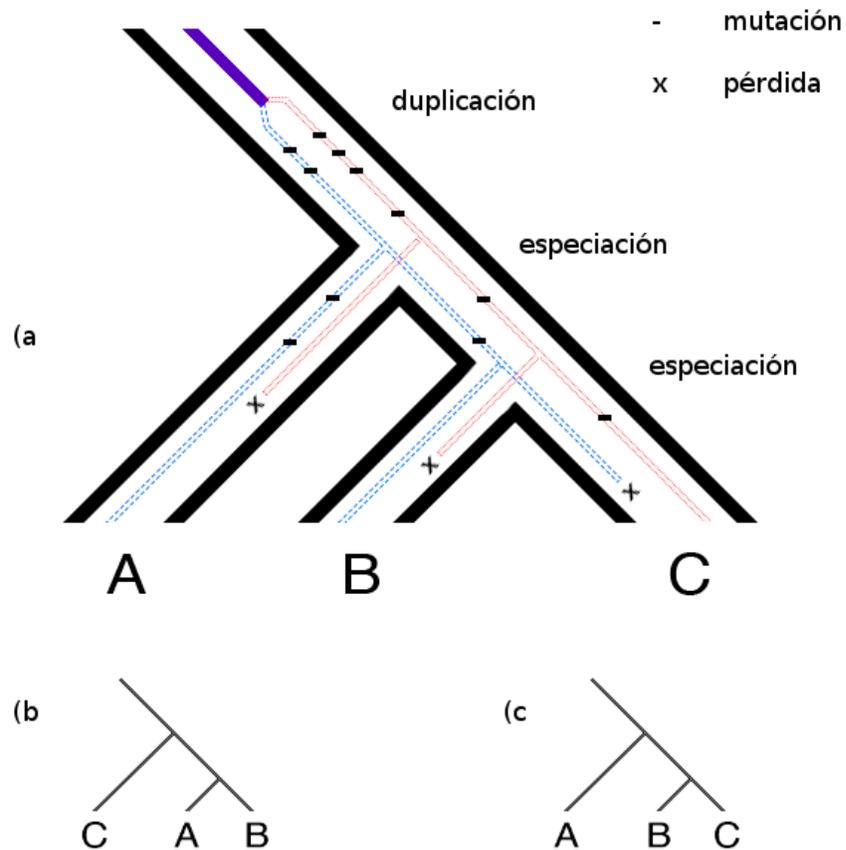


Figura 2. Un ejemplo que ilustra un caso en el cual un árbol de genes no coincide con el árbol de especies. En la figura se muestra la filogenia de tres especies hipotéticas junto con la filogenia de la familia de un gen cualquiera. Este gen pudo haber sufrido una duplicación y posteriormente mutaciones en cada uno de los linajes de los dos genes hijos. En caso de que se pierdan ciertos genes en el genoma de las especies (a), es posible que en la actualidad se observe un árbol construido a partir de los de genes (b) que no corresponde con el árbol de especies (c).

En la actual era postgenómica es común encontrar genes xenólogos durante el ensamblaje y anotación de genomas recién secuenciados cuando no existen genes ortólogos en el genoma de referencia o en organismos filogenéticamente cercanos. Existen diversos métodos que validan estos descubrimientos que parten de la premisa de que todo gen obtenido horizontalmente debe mantener ciertas características del grupo donador como el uso preferencial de codones (30) o el contenido de G-C (19). Sin embargo, la prueba que demuestra un evento de LGT por excelencia es la inferencia filogenética. Hoy en día es común encontrar en la literatura científica estudios donde se construye árboles filogenéticos, principalmente por probabilidad bayesiana o máxima verosimilitud, para verificar la verdadera xenología

de estos genes argumentando que estas secuencias se agrupan con secuencias que pertenecen a taxones no relacionados en sus filogenias. No obstante se han dado casos en los cuales estas suposiciones están débilmente fundamentadas y son desmentidas en artículos publicados tiempo después, como en el caso de varios genes en el genoma del humano, a los cuales se les atribuía un origen procarionte (31).

1.4 Fuentes de error en análisis filogenéticos

Cuando se construyen árboles filogenéticos a partir de secuencias de nucleótidos o aminoácidos se deben considerar una serie de posibles fuentes de error que de otra forma llevarían a conclusiones erróneas (32), por ejemplo la atracción de ramas largas. La atracción de ramas largas es un artificio que se observa cuando varias secuencias se agrupan en un dendrograma, no porque tengan características en común, sino porque tienen diferencias considerables con respecto a todas las demás. Otra fuente de equívoco que podría llevar a concluir erróneamente que un gen es xenólogo a partir de estudios filogenéticos es la saturación mutacional. Dado que el DNA está constituido por únicamente cuatro estados de carácter, en todo sitio de un alineamiento de dos secuencias existe un 25% de probabilidad de que se observe el mismo estado de carácter debido al azar. Esta observación es resultado de una tasa alta de sustitución de nucleótidos. Las secuencias que están sujetas a cambiar de esta manera son aquellas bajo selección positiva y aquellas que evolucionan de manera neutral, especialmente los pseudogenes. Por lo tanto es posible que dos secuencias sin una historia evolutiva en común presenten una distancia genética lo suficientemente pequeña para agruparse en una filogenia, lo que daría lugar a una interpretación errónea (24). Existen diferentes pruebas que verifican que un alineamiento tenga un nivel aceptable de saturación y sea por consiguiente “filogenéticamente útil”, estas pruebas deben tomarse en cuenta en todo análisis filogenético y en este estudio se le da un valor especial para rechazar casos falsos positivos de transferencia horizontal.

1.5 Antecedentes

En la primera anotación del genoma de *E. histolytica* en 2005, Loftus y colaboradores hicieron una lista de 96 genes posiblemente obtenidos por transferencia horizontal (33). Posteriormente se llevó a cabo una segunda anotación del genoma de *E. histolytica* en la cual se clasificaron estos 96 genes en subcategorías basándose en filogenias construidas por probabilidad bayesiana y máxima verosimilitud. Del total de 96 genes, 41 tienen pruebas sólidas de xenología, en el caso de 27 los valores de bootstrap y/o de probabilidad posterior disminuyeron en las filogenias, ampliando el muestreo taxonómico la xenología de 14 genes perdió consistencia, nueve candidatos se encontraron en otros eucariontes unicelulares y para cinco genes la transferencia vertical se convirtió en la explicación más parsimoniosa de la topología obtenida (34).

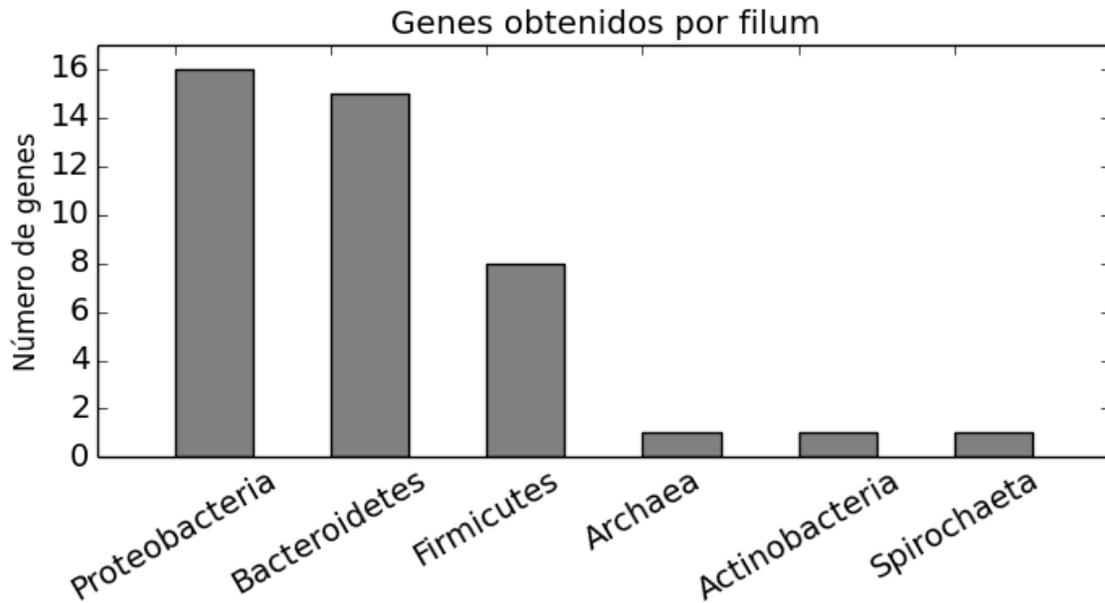


Figura 3. Número de genes obtenidos por cada filum bacteriano. Modificada de Alsmark et al 2009.

Clark y colaboradores concluyeron que la transferencia lateral de genes sigue siendo la explicación más probable para 68 de los 96 candidatos de Loftus et al. En una publicación posterior, Alsmark y colaboradores sugieren que estos resultados no están sujetos a cambiar a pesar del crecimiento de las bases de datos y la ampliación del muestreo taxonómico (35). En este estudio se reporta que el filum bacteriano que ha donado el mayor número de genes al genoma de *Entamoeba histolytica* es Proteobacteria con 16 genes, seguido de Bacteroidetes con 15 genes y en tercer lugar Firmicutes con 8 genes (figura 3). Estos resultados reflejan la alta frecuencia y dominancia de estos grupos en el intestino del ser humano (36, 37).

2. JUSTIFICACIÓN

Los estudios de evolución son parte de las investigaciones de ciencia básica que permiten entender a cada uno de los agentes que interactúan a lo largo de una infección y eventualmente posibilitan el desarrollo de estrategias de prevención y/o tratamiento en el área médica. Hoy en día no está claro bajo qué condiciones *Entamoeba histolytica* desencadena la patogénesis, a pesar de ello existen pruebas de que la virulencia está fuertemente ligada a la interacción con procariontes. Por lo tanto es necesario estudiar a fondo toda interacción de *E. histolytica* con procariontes, en este caso la transferencia horizontal de genes que pudo otorgar mecanismos de infección o de resistencia a antibióticos a las amibas. De la misma manera puede que uno de estos genes de virulencia obtenidos por HGT se haya perdido en *E. dispar* lo que explicaría por qué esta última rara vez es patogénica. O es posible que *E. histolytica* haya obtenido alguna isla de patogenicidad de bacterias que colonizan el intestino y utiliza estos genes a lo largo de la infección, siendo esto verdadero esperamos encontrar transferencias horizontales de factores de virulencia simultáneas. Dado el caso de encontrar factores de virulencia o genes involucrados en el metabolismo de origen procarionte en *E. histolytica*, se podrían implementar estrategias de tratamiento a los síntomas de la amibiasis similares a los que hoy en día se usan contra infecciones bacterianas. Por el otro lado, tomando en cuenta que el género *Entamoeba* está compuesto por amibas que se alimentan principalmente de bacterias, este género es el modelo perfecto para estudiar la transferencia horizontal de genes en eucariontes y por lo tanto se cuenta con una oportunidad privilegiada para observar sus diferentes resultados. Además el presente trabajo incluye la datación de los linajes cercanos a *E. histolytica*, estudio que hasta ahora no se había llevado a cabo de manera concluyente y que sirve de marco de referencia para explicar muchas características particulares de cada una de las especies en cualquier contexto.

En un marco más técnico, dada la susceptibilidad de encontrar falsos positivos en estudios filogenéticos causados por la saturación mutacional, es necesario validar que los candidatos de HGT de *Entamoeba* compartan una historia evolutiva con secuencias procariontes y las topologías obtenidas en trabajos anteriores no se deben a artificios metodológicos. Tanto en los estudios de Loftus et al. como en los de Clark y colaboradores se utilizan secuencias de aminoácidos, en cambio en el presente trabajo se presentan resultados con una resolución diferente en cuanto a la variación observada en las secuencias de nucleótidos. Es importante considerar que dado el vertiginoso crecimiento de secuencias genómicas y metagenómicas en las bases de datos públicas, se espera que el muestreo taxonómico de linajes procariontes se haya ampliado. Así, este trabajo da seguimiento a las publicaciones de Loftus y colaboradores y Clark y colaboradores para verificar si el tiempo es determinante en la asignación de grupos donadores.

3. OBJETIVOS

- Localizar las secuencias de *Entamoeba* que se agrupan con genes procariontes por causa de saturación mutacional.
- Inferir eventos de transferencia horizontal a partir de secuencias de nucleótidos.
- Asignar un grupo donador procarionte a cada candidato siempre que sea posible.
- Verificar que el crecimiento de las bases de datos no repercute en la asignación de grupos donadores como predijo Alsmark y colaboradores.
- Datar los tiempos de divergencia de las especies de *Entamoeba* que se incluyen en el análisis.
- Datar los eventos de transferencia lateral de genes siempre que sea posible.
- Encontrar casos de transferencias horizontales simultáneas y/o la pérdida de estos genes en la historia evolutiva de las especies de *Entamoeba* que se incluyen en el análisis.

4. MÉTODOS

Durante la ejecución de este estudio fue necesario apoyar los programas publicados en la red con scripts y/o programas escritos por el autor ya sea para automatizar procesos, extraer datos o resolver problemas bioinformáticos. El lenguaje empleado en la mayoría de los casos fue python 2.7. Python es un lenguaje de programación interpretado de código abierto cuya filosofía sostiene que haciendo fácil la tarea de programar se obtienen beneficios como escritura rápida e integración eficiente de sistemas. Corre en Windows, Linux/Unix y Mac OS X. Actualmente se encuentra en la versión 2.7 y 3.3 simultáneamente (www.python.org). Existen librerías para llevar a cabo fines específicos y el cómputo científico no es la excepción: numpy (Numerical Python) para el manejo eficiente de datos numéricos, scipy (Scientific Python) para analizar datos científicos, matplotlib (Mathematical Plotting Library) para la construcción de gráficas con calidad de publicación de manera similar a MATLAB® y Biopython para llevar a cabo análisis bioinformáticos. En este trabajo se emplearon todas las librerías mencionadas.

4.1 Obtención de secuencias candidato

Los números de acceso de la base de datos refseq del Centro Nacional de Información Biotecnológica (NCBI por sus siglas en inglés) de Estados Unidos de los 68 candidatos consistentes se obtuvieron de la publicación original de Clark y colaboradores (34). Las 68 secuencias de aminoácidos fueron descargadas del NCBI y después fueron usadas como búsqueda para el algoritmo de búsqueda básica por alineamiento local (BLAST por sus siglas en inglés) en la base de datos Amoebadb, <http://amoebadb.org/amoeba/>, y de esta manera no sólo se obtuvieron las secuencias codificantes sin intrones (cDNA) y secuencias de aminoácidos, sino también anotaciones actualizadas, comentarios, información de sintenia y expresión disponible en esta última base de datos (38).

4.2 Selección de genes para datación y obtención de secuencias

Para llevar a cabo la datación de tiempo de divergencia de los linajes de *Entamoeba* se buscaron genes que aprobaron los siguientes criterios de inclusión: estar presentes en el genoma de todos los linajes incluidos en el estudio (*E. histolytica*, *E. nuttallii*, *E. dispar* y *E. invadens*), ser codificantes para proteínas, ser genes constitutivos y que no sean xenólogos. Los genes elegidos fueron los siguientes (entre paréntesis se muestra el número de acceso de la base de datos amoebadb): subunidad RPA2 de la RNA polimerasa DNA-dependiente I (EHI_186020), factor de elongación 2 (EHI_189490), actina (EHI_131230), cadena gamma de tubulina(EHI_008240) y cadena pesada de la clatrina (EHI_201510). Se obtuvieron las secuencias de cDNA y aminoácidos de cada uno de estos genes de cada especie de *Entamoeba* en la base de datos amoebadb (38). Posteriormente se obtuvieron las secuencias de

aminoácidos y cDNA correspondientes a estos genes de *Dictyostelium discoideum* en la base de datos dictybase (<http://dictybase.org/>, 39).

4.3 Búsqueda de homólogos

La secuencia de aminoácidos de cada uno de los 68 candidatos originales de la lista de Clark y colaboradores se usó como entrada para el algoritmo BLASTp, con parámetros estándar (40), y se buscó contra la base de datos de aminoácidos RefSeq del NCBI (41). Los primeros 50 resultados de cada ejecución del BLASTp se muestrearon para evaluar la hipótesis de transferencia lateral. La recolección de los hits de BLAST se llevó a cabo a través del sistema Entrez.

4.4 Alineamiento de aminoácidos

Para los genes candidato de HGT, cada secuencia de aminoácidos se alineó con los 50 primeros resultados de la búsqueda por BLASTp y con sus homólogos de *E. dispar* y *E. invadens*, siempre que estuvieran presentes. Para los estudios de datación de *Entamoeba*, se generó un alineamiento de aminoácidos por separado por gen incluyendo las secuencias de *E. histolytica*, *E. nuttalli*, *E. dispar*, *E. invadens* y *D. discoideum*. Todos los alineamientos de aminoácidos se ejecutaron con el programa Clustalw (42) con parámetros estándar.

4.5 Construcción de alineamientos de cDNA

Como las secuencias de cDNA descargadas de Entrez eran secuencias de origen genómico fue necesario verificar y homologar el marco de lectura de estas secuencias. Para ello, se escribió un programa en lenguaje python 2.7, el cual obtiene el marco de lectura +1 de una secuencia de cDNA basándose en su secuencia de aminoácidos (Anexo 2.1). Posteriormente, los alineamientos de cDNA se construyeron a partir de los alineamientos de aminoácidos usando un programa diseñado específicamente para esta investigación escrito en python 2.7 el cual requiere como entrada el alineamiento de aminoácidos y un archivo con las secuencias codificantes, ambos en formato FASTA (anexo 2.2). Este programa toma cada secuencia del alineamiento y su secuencia de cDNA correspondiente y genera una secuencia nueva de cDNA, al principio vacía. Posteriormente revisa cada sitio de la secuencia de aminoácidos. Siempre que encuentre un guión o “gap” (-), inserta tres gaps (---) en la secuencia nueva de cDNA, por el contrario, siempre que encuentre una letra mayúscula (aminoácido), corta el primer codón que encuentre en la secuencia de cDNA original y lo inserta al final de la secuencia de cDNA nueva. De esta manera se obtiene el alineamiento de cDNA correspondiente a cada alineamiento de aminoácidos sin importar el código genético que emplean las diferentes especies.

4.6 Construcción de matrices de distancia evolutiva

Se construyó una matriz de distancia por alineamiento de nucleótidos de cada uno de los genes xenólogos, para ello se utilizó el modelo de sustitución de nucleótidos de Máxima verosimilitud compuesta tomando en cuenta la distribución de variación de tasas a lo largo del alineamiento (parámetro alfa). Se analizaron las tres posiciones de cada codón y los sitios ambiguos (sitios que presentaran indels) fueron eliminados del análisis. Estas matrices se llevaron a cabo con paquete MEGA (43). De esta forma se construyeron alineamientos (que denominaremos alineamientos cortos) que incluyeran las secuencias con menor distancia evolutiva con respecto a las secuencias de *Entamoeba*.

4.7 Pruebas de saturación mutacional

Se obtuvo el índice Iss e Iss.c descrito por Xia y colaboradores para cada alineamiento corto de cDNA tomando en cuenta las tres posiciones de cada codón (32). Estos análisis se llevaron a cabo con el programa DAMBE (44, 45). El índice Iss. es una medida de entropía (según la definición de entropía en el ámbito de la teoría de la información) dado un alineamiento de nucleótidos: un sitio altamente conservado es un sitio con baja entropía y por el otro lado un sitio en donde la frecuencia relativa de los cuatro nucleótidos es igual o cercana a 0.25 es un sitio con alta entropía y por lo tanto con alta saturación mutacional. El índice Iss. es la sumatoria de los valores de entropía de todos los sitios de un alineamiento. Como el valor de Iss. crece en relación con el número de sitios del alineamiento, para decidir si el alineamiento es útil para llevar a cabo inferencia filogenética, se simula un alineamiento con las mismas dimensiones que el alineamiento problema, pero con una distribución de nucleótidos que refleja alta entropía. El valor de Iss. para esta secuencia simulada es el índice Iss.c. Siempre que el valor de Iss. resulte significativamente menor que el valor de Iss.c se asume que las secuencias tienen poca saturación. En cambio, si las diferencias no alcanzan a ser significativas, se dice que el alineamiento presenta saturación substancial y es necesario verificar la topología de un árbol construido a partir de este alineamiento, en este caso una topología asimétrica confirmaría una alta saturación mutacional. Dado el caso que el valor de Iss. resulte mayor que el de Iss.c y las diferencias sean significativas, el alineamiento es considerado como inútil para estudios de filogenética. Cuando el índice Iss. es mayor pero las diferencias no son significativas nuevamente es recomendable verificar la topología de un árbol construido con el alineamiento problema. Siempre que un alineamiento presentaba saturación considerable (substantial saturation) el análisis se repitió pero ignorando el tercer sitio de cada codón para aminorar la saturación mutacional observada dada la degeneración del código genético. Para cada alineamiento se llevó a cabo una prueba de t de Student para comparar el valor de Iss. con los de Iss.c, la significancia estadística se estableció en $p < 0.5$. Para los alineamientos que mostraron saturación considerable tras excluir la tercera posición de cada codón, fue necesario el análisis de la topología

resultante de un árbol filogenético para concluir si el alineamiento contenía o no información evolutiva, esperando ver un árbol de topología asimétrica como resultado de un alineamiento con alta saturación y por lo tanto inútil filogenéticamente.

4.8 Construcción de árboles filogenéticos por máxima verosimilitud

Antes de construir los dendrogramas como tal, cada alineamiento de cDNA por separado fue introducido al programa jModeltest 2 para encontrar el modelo de sustitución de nucleótidos que mejor se ajustara a cada alineamiento (46). Se revisaron los modelos construidos a partir de 11 esquemas de sustitución, agregando combinaciones de los distintos parámetros que se mencionan a continuación: frecuencias relativas de cada base (+F), proporción de sitios invariables (+I) y variación de las tasas de sustitución a lo largo del alineamiento (+G). Estas combinaciones producen un total de 88 modelos probados por cada alineamiento, el árbol base para llevar a cabo los cálculos se construyó usando el algoritmo BioNJ y el algoritmo usado para buscar topologías fue el NNI. Posteriormente se usó el criterio de selección de modelos AICc para evitar la sobreparametrización que produciría un modelo demasiado complicado.

A continuación se utilizó el programa PhyML 3.0 para obtener el mejor árbol filogenético de cada alineamiento (47), se empleó la mejor topología resultante de la comparación entre las obtenidas por los diferentes algoritmos NNI y SPR, el árbol base se construyó con BioNJ, se usó el modelo de sustitución de nucleótidos elegido por el criterio AICc del programa jModeltest 2 para cada alineamiento y finalmente se llevaron a cabo 100 pruebas de bootstrap por alineamiento. El número relativamente bajo de bootstraps fue una consecuencia de la mejor topología resultante entre la comparación de los algoritmos NNI y SPR. Tanto para la búsqueda del mejor modelo como para la construcción de filogenias, se eliminó la tercera posición de cada codón siempre que de esta manera el alineamiento haya mostrado niveles de saturación mutacional aceptables.

4.9 Construcción de árboles filogenéticos por probabilidad bayesiana

Los análisis filogenéticos por probabilidad bayesiana se ejecutaron con el programa MrBayes 3.2 (48). Para los candidatos que presentaron niveles aceptables de saturación mutacional tomando en cuenta los tres sitios de cada codón, se corrió un millón de generaciones de cadenas de Markov Monte Carlo (MCMC) muestreando un árbol cada doscientas generaciones. De la misma manera para los alineamientos que pasaron la prueba de saturación mutacional tomando en cuenta solamente los primeros dos sitios de cada codón, también se corrieron un millón de generaciones de MCMC muestreando un árbol cada doscientas generaciones pero ignorando la tercera posición de cada codón. En ambos casos el modelo de sustitución GTR+I+G fue utilizado, además se corrieron dos cadenas “frías” y dos cadenas

“calientes” con una temperatura igual a 0.1. El 25% de los árboles muestreados (mil doscientos cincuenta árboles) fueron eliminados en calidad de “burn-in” y a partir de las muestras restantes se construyó un árbol consenso.

Los árboles de ambos análisis fueron revisados y editados con el programa Dendroscope 3 (49). Los árboles presentados como resultados son los árboles consenso arrojados por los análisis de probabilidad bayesiana y muestran en cada nodo la probabilidad posterior y el valor de bootstrap de dicho nodo siempre que haya estado presente en el árbol generado por máxima verosimilitud y su valor haya sido mayor o igual a 50.

4.10 Datación de eventos de divergencia por probabilidad bayesiana

Se llevaron a cabo dos series distintas de estimaciones de tiempos de divergencia, la primera se ejecutó sobre los cinco genes constitutivos (solamente de *E. histolytica*, *E. nuttallii*, *E. dispar*, *E. invadens* y *D. discoideum*): subunidad RPA2 de la RNA polimerasa DNA-dependiente I, factor de elongación 2, actina, cadena gamma de tubulina y cadena pesada de la clatrina. El árbol de entrada utilizado para llevar a cabo los análisis de divergencia fue el siguiente: ((((*E. nuttallii*, *E. histolytica*), *E. dispar*), *E. invadens*), *D. discoideum*). Las estimaciones se ejecutaron usando los programas Estbranches y Multidivtime (50, 51). El nodo que une a *E. histolytica* y a *E. nuttallii* fue el elegido para calibrar el cálculo de los tiempos de divergencia con un intervalo de 5.5 a 6.5 millones de años antes del presente, siguiendo el razonamiento explicado en la sección de resultados. De manera independiente por cada gen se empleó el programa Baseml dentro del programa Paml 4.7 (52) con el modelo de sustitución F84+G para calcular la frecuencia nucleotídica, el cociente de tasas transición/transversión (valor del parámetro κ) y la variación de tasas de sustitución a lo largo del alineamiento (valor del parámetro α).

A continuación la longitud de las ramas de los árboles de entrada, junto con su máxima verosimilitud y la matriz de covarianza fueron estimadas por el programa Estbranches. Finalmente se ejecutó el análisis bayesiano de Cadenas de Markov Monte Carlo del programa Multidivtime con el fin de estimar las distribuciones posteriores de las tasas de sustitución y los tiempos de divergencia. Se corrieron un total de cinco millones cien mil generaciones MCMC, de las cuales cien mil fueron descartadas en calidad de “burn-in” y posteriormente se tomó una muestra cada cien generaciones dando un total de cincuenta mil muestras. La unidad de tiempo elegida fue millones de años y serán referidas como millones de años antes del presente (Map). *A priori*, se eligió el valor de cien Map para la raíz del árbol de entrada, una desviación estándar de cincuenta Map y un tiempo máximo de trescientos millones de años antes del presente. Todos los genes constitutivos fueron ingresados simultáneamente en el programa Multidivtime.

La media y la desviación estándar de la distribución de tasas de evolución molecular del nodo que une las secuencias de *Entamoeba* se asignaron como la mediana de los valores de las tasas de evolución molecular arrojadas por Estbranches de todos los genes constitutivos en conjunto. Los resultados se muestran como el tiempo de divergencia en Map +/- la desviación estándar arrojada por Multidivtime en las mismas unidades.

Por otro lado, la segunda serie de estimaciones se realizó con base en los resultados de los análisis filogenéticos. De los sesenta y ocho candidatos de transferencia horizontal de genes solamente se tomaron en cuenta aquellos que cumplieran ciertos requisitos: agrupación bien soportada en los árboles filogenéticos, grupo donador conocido por lo menos hasta filum y finalmente que se encuentre anotado un ortólogo en el genoma de *E. nuttallii*. Las secuencias utilizadas para esta serie incluyó las mismas secuencias introducidas para los análisis filogenéticos pero solamente de ciertos organismos: *E. histolytica*, *E. dispar* y *E. invadens*, hasta cuatro especies incluidas dentro del grupo hermano de los candidatos de HGT resultante en las filogenias (a los cuales nos vamos a referir en general como “a”, “b”, “c” y “d”) y hasta tres especies ubicadas basalmente con respecto a la secuencia ancestral compartida entre las secuencias de *Entamoeba* y su grupo hermano “a”, “b”, “c” y “d” (a los cuales nos vamos a referir como “x”, “y” y “z”). Siempre que se encontró un homólogo anotado en el genoma de *E. nuttallii*, este se alineó a ojo con el resto de las secuencias utilizando el programa Seaview (53). Un árbol de entrada común sería el siguiente: ((((*E. nuttalli*, *E. histolytica*), *E. dispar*), *E. invadens*), ((a, b), (c, d))), ((x, y), z)), aunque la topología dentro del grupo hermano puede variar, así como el número de secuencias en el grupo hermano y/o en el grupo externo. En este caso se utilizaron los tiempos de divergencia estimados para los genes constitutivos para calibrar las divergencias de las transferencias horizontales. El nodo que une a *E. histolytica* con *E. nuttallii*, el nodo que une al ancestro *E. histolytica*-*E. nuttallii* con *E. dispar*, el nodo que une al ancestro de *E. dispar* con su grupo hermano y el nodo que una a *E. invadens* con el resto de las especies de *Entamoeba* fueron usados para calibrar las divergencias de las transferencias horizontales.

Nuevamente se utilizó el programa Paml 4.7 (52) para calcular la frecuencia nucleotídica, el cociente de tasas transición/transversión (valor del parámetro κ) y la variación de tasas de sustitución a lo largo del alineamiento (valor del parámetro α) ingresando los mismos parámetros que se emplearon para los genes constitutivos. Después se usó el programa Estbranches para obtener la máxima verosimilitud y los valores de longitud de las ramas de los árboles de entrada y la matriz de covarianza. Por último se calcularon los tiempos de divergencia con el programa Multidivtime. Se corrieron un total de cinco millones cien mil generaciones MCMC, cien mil fueron descartadas en calidad de “burn-in” y posteriormente se tomó una

muestra cada cien generaciones dando un total de cincuenta mil muestras. *A priori*, se eligió el valor de cien Map para la raíz del árbol de entrada, una desviación estándar de cincuenta Map y un tiempo máximo de trescientos millones de años antes del presente. Tanto la media como la desviación estándar de la distribución de tasas de evolución molecular del nodo que une las secuencias de *Entamoeba* con su grupo hermano se estableció como la mediana de las tasas de evolución molecular arrojadas por Estbranches. Cada gen candidato fue analizado individualmente. Los resultados se muestran como el tiempo de divergencia en Map +/- la desviación estándar arrojada por Multidivtime (50) en las mismas unidades.

4.11 Análisis del uso preferencial de codones

Los análisis de uso preferencial de codones se llevaron a cabo en todas las secuencias candidato de *E. histolytica*, así como en los genes constitutivos elegidos para la datación de los linajes dentro del género y finalmente en todas las secuencias de cDNA anotadas hasta ahora en el genoma de *E. histolytica* (descargadas del portal amoebadb.org, 38). Posteriormente se obtuvieron la media y la desviación estándar de la frecuencia relativa por codón. Finalmente se obtuvieron la frecuencia relativa por codón y la frecuencia absoluta por codón y por aminoácido tanto de las secuencias candidato como de los genes constitutivos independientemente. Para estos cálculos se utilizó el código genético estándar sin excepción.

A continuación se calculó el índice de medida independiente de longitud y composición (MILC por sus siglas en inglés, 54, 55), desarrollado por Supek y Vlahovicek, sobre las frecuencias relativas de los genes xenólogos y los genes constitutivos. Siempre que el valor de MILC resultara mayor a 0.5 el uso de codones se consideró significativamente diferente de la distribución esperada.

5. RESULTADOS Y DISCUSIÓN

5.1 Búsqueda de homólogos

Cada uno de los candidatos para transferencia horizontal de genes fue usado como entrada para el programa BLASTp (40) contra la base de datos de referencia de proteínas (RefSeq) con parámetros estándar. Se eligió la base de datos RefSeq porque, además de ser una base de datos curada manual y automáticamente, tiene una “no redundancia” real frente a otras bases de datos, como por ejemplo la base de datos nr del NCBI, en la cual es posible que el algoritmo arroje distintos segmentos de una misma secuencia como diferentes hits (56). En las 68 búsquedas, sin excepción, el hit con el número e más bajo (0.0) fue la misma secuencia que fue introducida al algoritmo como búsqueda, ésta fue omitida cuando se recuperaron las secuencias encontradas por el BLASTp por medio de Entrez. En la mayoría de los casos el programa devolvió secuencias pertenecientes a especies procariontes, sobre todo bacterianas. En el caso de los genes RNA metiltransferasa (EHI_00670), fosfoesterasa (EHI_013150), Orn/Arg/Lys descarboxilasa (EHI_098380), cinasa (EHI_178490), isopentenil-difosfato delta-isomerasa (EHI_194410), proteína hipotética (EHI_198800) y citidiltransferasa (EHI_202040) el hit más alto correspondió a secuencias de arqueas, todas con un e-valor menor a $3.99e^{-21}$. En algunos casos el programa arrojó péptidos de especies eucariontes, de estos, el género más frecuente fue *Dictyostelium*, cuyas secuencias se alinearon con los genes ser/thr proteína fosfatasa (EHI_035210), proteína hipotética (EHI_046680) y deoxicitidina trifosfato deaminasa (EHI_140240), para este último, el BLAST también arrojó una secuencia perteneciente al género *Naegleria*, que es un género de protozoos del filum Percolozoa, de entre los cuales, la especie *Naegleria fowleri* causa infecciones accidentales en humanos. Además, entre las secuencias putativamente homólogas para el gen candidato de la flavoproteína tipo A (EHI_096710), se encuentra una secuencia perteneciente al género *Trichomonas*, un protozoario parásito del filum Metamonada que infecta vertebrados.

Como se encuentra reportado en el sitio amoebadb.org, casi todos los candidatos de *E. histolytica* presentan genes ortólogos en las otras especies del mismo género (38). Los genes de la hidrolasa metal dependiente (EHI_05469) y la proteína hipotética (EHI_085050) se encuentran exclusivamente en el genoma de *E. histolytica* y de *E. nuttallii*. Además, el gen de la proteína hipotética (EHI_198610) está ausente en *E. dispar* y las secuencias para la proteína de resistencia al 5-nitroimidazol (EHI_068430), la proteína hipotética (EHI_07264) y la proteína inducida por daño/competencia (EHI_090260) carecen de homólogo en el genoma de *E. invadens*.

Por el otro lado, no fue posible encontrar ortólogos para ocho candidatos en la base de datos de

secuencias de aminoácidos anotadas de *E. nuttallii*. La mayoría de estos genes tienen pocos homólogos en el genoma de *E. histolytica*: proteína hipotética (EHI_017590) y serin/treonin fosfatasa (EHI_035210). Por el contrario la proteína hipotética (EHI_104900) y la aldehído-alcohol deshidrogenasa (EHI_160940) no se encontraron en *E. nuttallii* a pesar de que en *E. histolytica* se observe cuatro o más parálogos para cada uno de estos genes.

5.2 Matrices de distancia evolutiva

A partir de los alineamientos de cDNA se construyeron matrices de distancia evolutiva. El modelo de sustitución empleado fue el de máxima verosimilitud compuesta tomando en cuenta el parámetro α de la distribución γ , las tres posiciones de cada codón fueron incluidas y por cada par de secuencias las posiciones ambiguas (con indels) fueron ignoradas. Los valores de sustituciones por base estimadas arrojadas por el análisis resultaron polarizadas. En general se encontraron grupos de secuencias con baja distancia entre sí y un valor mucho más alto en comparación con las demás. Las secuencias con la menor distancia evolutiva en relación con las secuencias de las diferentes especies de *Entamoeba* fueron seleccionadas para formar alineamientos más cortos para llevar a cabo las pruebas de saturación mutacional. Esto último con el fin de analizar a las secuencias de *Entamoeba* solamente junto con los grupos donadores más cercanos.

5.3 Análisis de saturación mutacional

Los índices Iss. e Iss.c (32) fueron calculados por cada alineamiento corto, en primera instancia, tomando en cuenta los tres sitios de cada codón. En el primer análisis, treinta y ocho alineamientos devolvieron valores de Iss. significativamente menores que los valores de Iss.c correspondientes ($p < 0.05$). Por el otro lado en catorce casos el valor de Iss. resultó menor que el de Iss.c pero sin diferencias significativas, de igual forma en diez alineamientos se observó una entropía mayor que la de los alineamientos simulados correspondientes aunque las diferencias no fueron significativas (figura 4). Finalmente, los alineamientos de los seis candidatos restantes resultaron con valores de Iss. mayores a los valores de Iss.c correspondientes y por lo tanto inútiles para llevar a cabo inferencia filogenética (Anexo 3). Dado que el código genético es degenerado para muchos aminoácidos, la tercer posición de cada codón cambia a una tasa mayor en comparación con los primeros dos sitios, dado que la secuencia de aminoácidos no se modifica y por tal motivo no hay efecto a nivel funcional. Es posible que este fenómeno aumente la saturación de sustituciones observada en los alineamientos, generando resultados falsos negativos.

Con esta observación en mente se realizó una segunda serie de pruebas de saturación mutacional pero ignorando el tercer sitio de cada codón para mitigar la saturación observada dada la degeneración del

código genético. En este segundo ensayo, los treinta y ocho candidatos que resultaron con saturación aceptable en la primer serie de análisis fueron excluidos. Como resultado, quince alineamientos tuvieron un valor de Iss. significativamente menor y cuatro resultaron con valores de Iss significativamente mayores a los valores de Iss.c correspondientes. Estos últimos fueron los genes para la fosfoesterasa (EHI_013150), proteína hipotética (EHI_085050), proteína de competencia ComEC (EHI_156240) y la aldehído-alcohol deshidrogenasa 2 (EHI_160940).

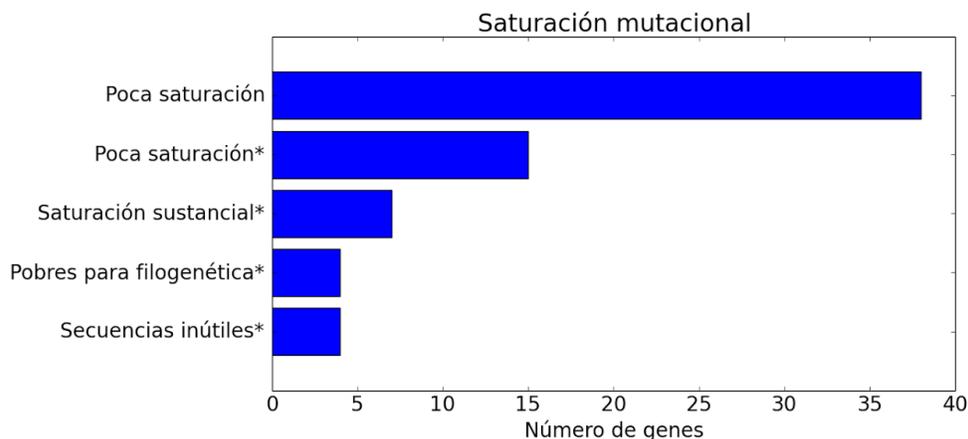


Figura 4. Grado de saturación según el índice Iss. e Iss.c de los sesenta y ocho candidatos originales. Poca saturación: Iss. significativamente menor. Saturación sustancial: Iss. menor pero sin diferencias significativas. Pobres para filogenética: Iss. mayor pero sin diferencias significativas. Secuencias inútiles: Iss. significativamente mayor. * Resultados excluyendo el tercer sitio de cada codón.

Los once candidatos restantes no reflejaron diferencias significativas en los análisis y en consecuencia, fue necesario construir árboles filogenéticos a partir de cada uno de estos alineamientos para definir el grado de saturación a partir de la topología resultante. De esta manera se construyeron once árboles con el programa MrBayes 3.2 (48), de los cuales 8 tuvieron una topología simétrica (Anexo 4). Por el contrario, tres alineamientos dieron lugar a filogenias con topologías asimétricas, estos fueron los alineamientos de la monoglicérido lipasa (EHI_045330), acetiltransferasa (EHI_065290) y el de la proteína hipotética (EHI_104900). Por consiguiente, tanto los cuatro candidatos que presentaron valores de Iss. significativamente mayores que los de Iss.c, como los tres candidatos cuyas filogenias construidas con los alineamientos cortos que resultaron con topologías asimétricas, fueron descartados de los análisis posteriores.

Los resultados del programa BLASTp lograron dar un acercamiento a los resultados de los análisis posteriores. Los candidatos cuyo hit más alto presentaron valores-e altos (EHI_085050 y EHI_156240)

finalmente fueron descartados en los análisis de saturación mutacional. Es probable que estos últimos candidatos de transferencia horizontal ya hayan perdido la señal filogenética dadas las sustituciones que han ocurrido a lo largo del tiempo desde que haya ocurrido la transferencia horizontal. Como ya no es posible asegurar que las secuencias que el BLAST encontró sean homólogas a estos candidatos, no se construyeron los árboles filogenéticos para no asignar grupos donadores falsos positivos. Por otro lado es posible que estas secuencias no sean xenólogas al resto de los genes de *Entamoeba*, sino que tras un largo tiempo de evolucionar de manera neutral, hoy en día se alineen parcialmente con secuencias procariontes por convergencia evolutiva, lo que explica los altos valores e.

5.4 Árboles filogenéticos

Se utilizó el programa jModeltest 2 (46) para elegir el modelo de sustitución que mejor describiera la evolución de cada alineamiento; la decisión final se tomó con el criterio AICc para evitar usar un modelo sobreparametrizado. El modelo de sustitución más común fue el GTR+I+G, siendo elegido para 39 de los alineamientos; el segundo modelo más frecuente fue el TVM+I+G con 12 alineamientos.

61 árboles fueron construidos tanto en Phyml 3 (47) como en MrBayes 3.2 (48) los cuales, en la mayoría de los casos, las topologías obtenidas coincidieron. Las diferencias más comunes consisten en la presencia de politomías en los árboles de probabilidad bayesiana en donde los árboles de máxima verosimilitud si mostraron dicotomías definidas. Sin embargo los nodos de estas últimas dicotomías presentan valores bajos de bootstrap (menores a 50), lo que se interpreta como posibles politomías, sugiriendo una congruencia aún mayor con los árboles de probabilidad bayesiana (Anexo 5).

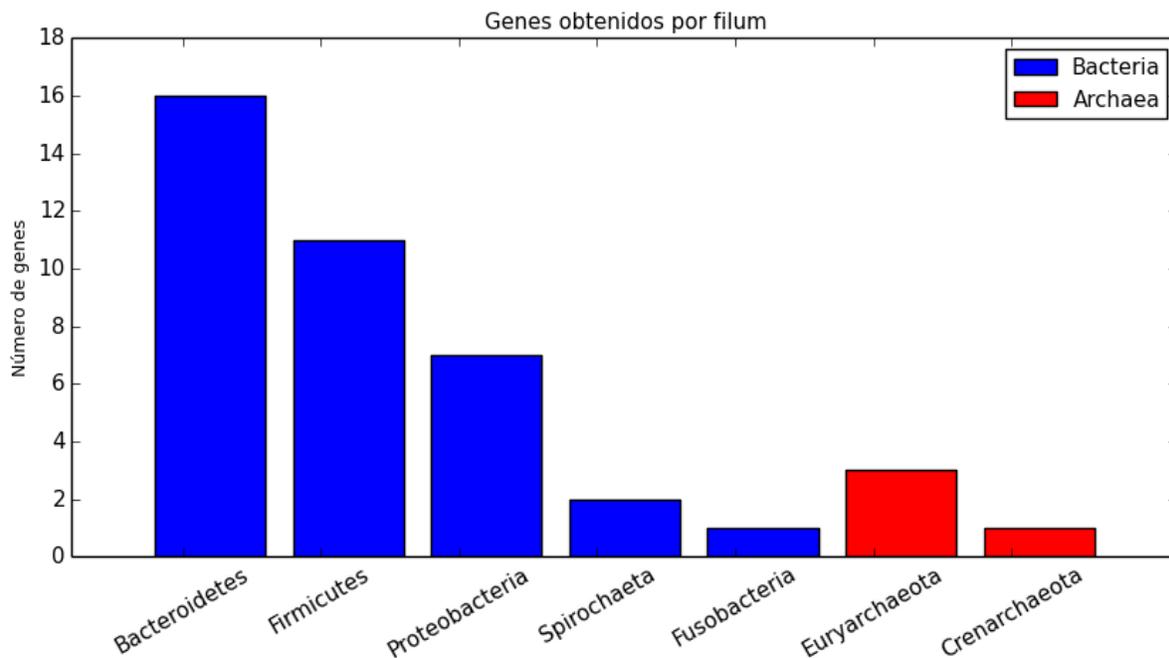


Figura 5. Número de genes obtenido por cada filum donador identificado en los árboles filogenéticos.

Para un total de 54 árboles fue posible localizar el grupo donador por lo menos en el nivel de filum, 35 de ellos construidos con los codones completos y el resto ignorando el tercer sitio de cada codón. Por el otro lado, en únicamente 12 casos distintos fue posible asignar el dominio donador dado que el grupo hermano del grupo de secuencias amebianas estaba compuesto por secuencias de organismos del mismo dominio pero de fila distintos como en el caso de los genes para la fructocinasa pirofosfato-dependiente (EHI_00730) y la 2-fosfosulfolactato fosfatasa (EHI_199080) para el dominio Bacteria.

Finalmente, en los últimos ocho casos no fue posible asignar un grupo donador, ya que el racimo de secuencias amebianas se agrupó con secuencias de organismos del dominio Bacteria, Archaea y en algunos casos, de otros eucariontes. Como se puede observar en la figura 5, la mayor parte de las transferencias fueron asignadas al dominio Bacteria, al cual se le asignó un total de 50 transferencias. De estos casos, 12 no reflejaron una asociación clara con algún filum en específico. Comparando la figura 3, que muestra los resultados de Alsmark y colaboradores en 2009 (35), con la figura 5, que ilustra los resultados del presente trabajo, se puede afirmar que el crecimiento de las bases de datos es un factor que modifica los resultados de este tipo de estudios a largo plazo (Anexo 6).

La mayoría de los posibles grupos donadores pertenecen a taxones que habitan comúnmente el intestino

humano. Bacteroidetes fue el filum con el mayor número de transferencias, con 16 de ellas. El orden Bacteroidales fue el donador asignado para 10 topologías, el cual incluye organismos anaerobios facultativos y anaerobios, al igual que *E. histolytica*. Se tienen pruebas de que el filum Bacteroidetes es un componente fundamental de la microbiota intestinal en cuanto a abundancia (57), hecho que se ve reflejado en la gran proporción de genes que el género *Entamoeba* ha obtenido de este filum. Aunque dos de estos genes tengan una anotación funcional relacionada con la resistencia a antibióticos, el gen de la proteína de la superfamilia beta-lactamasa (EHI_115720, 58) y el gen de resistencia al 5-nitroimidazol (EHI_068430), no se puede asegurar que estén llevando a cabo esta función (Figura 6).

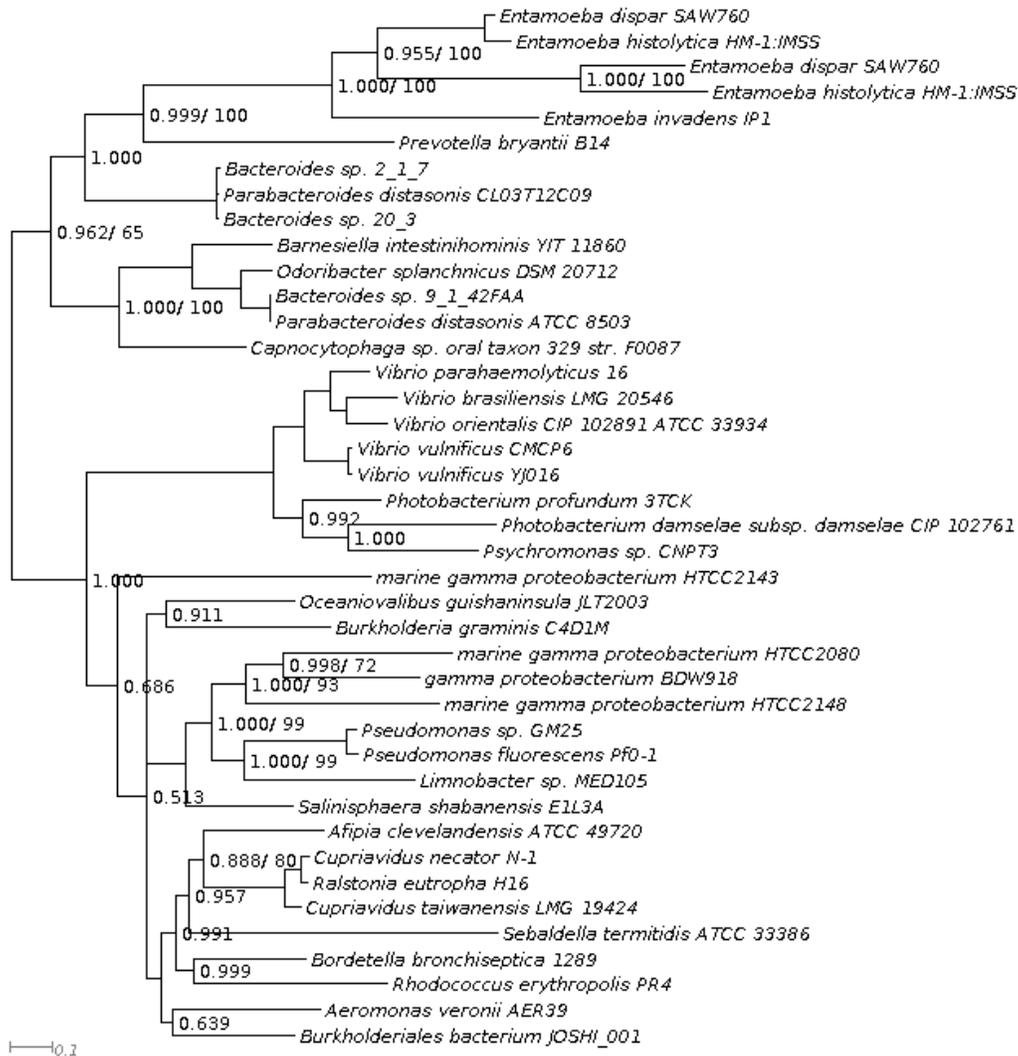


Figura 6. Metalo-beta-lactamasa (EHI_115720). La topología presentada fue arrojada por el programa MrBayes 3.2, se muestra en algunos nodos la probabilidad posterior / bootstrap generada por MrBayes y PhymI respectivamente. Las barras indican el número de sustituciones esperadas por sitio.

El resto de los genes donados por el filum Bacteroidetes, que no están marcados como hipotéticos, están involucrados en el metabolismo de carbohidratos, como el gen de la beta-galactosidasa (EHI_170020) y el de la fructocinasa (EHI_054510), pero ninguno pertenece a vías metabólicas anaerobias o anaerobias facultativas exclusivamente.

Por otra parte el segundo filum que transfirió más genes al género *Entamoeba* fue el filum Firmicutes, con 11 transferencias en total. Al igual que el filum Bacteroidetes, este grupo es de crucial importancia en el colon dada su abundancia y diversidad (59), de hecho la razón Firmicutes: Bacteroidetes cambia a lo largo del desarrollo del ser humano (60). Fue posible determinar el grupo donador en siete topologías, seis de Clostridiales y una de Bacillales. Ambos Órdenes comprenden tanto a organismos de vida libre, que se encuentran comúnmente en el suelo o en cuerpos de agua, como a organismos patógenos y oportunistas.

Algunos genes donados por el filum Firmicutes codifican para proteínas que intervienen en el metabolismo de aminoácidos como el que codifica para la fosfoglicerato mutasa (EHI_129820). Sin embargo, una proporción importante de los genes donados por este filum están involucrados en metabolismo de nucleótidos como la desoxirribosa-fosfato aldolasa (EHI_121800) o la pirimidina nucleótido-disulfuro oxidoreductasa (EHI_153000). Uno de los genes obtenidos de bacterias del filum Firmicutes por el género *Entamoeba* tiene una antación funcional relacionada con la resistencia a antibióticos (61), que es el que codifica para la RNA metiltransferasa (EHI_047740) (Figura 7).

El tercer filum bacteriano con mayor número de donaciones fue el filum Proteobacteria, siendo a su vez, el filum con la mayor diversidad de órdenes (Tabla 1). Una secuencia fue asignada para cada uno de los órdenes siguientes: Campylobacterales, Pseudomonadales, Thiotrichales y Enterobacterales. Este filum no sólo incluye especies comensales como la ampliamente estudiada *Escherichia coli*, sino también con especies patógenas como *Salmonella typhimurium*. Dos de las siete secuencias obtenidas de este filum son homólogas a proteínas de la familia de las beta-lactamasas, los genes EHI_068560 y EHI_169860, las cuales están relacionadas con la resistencia a antibióticos.

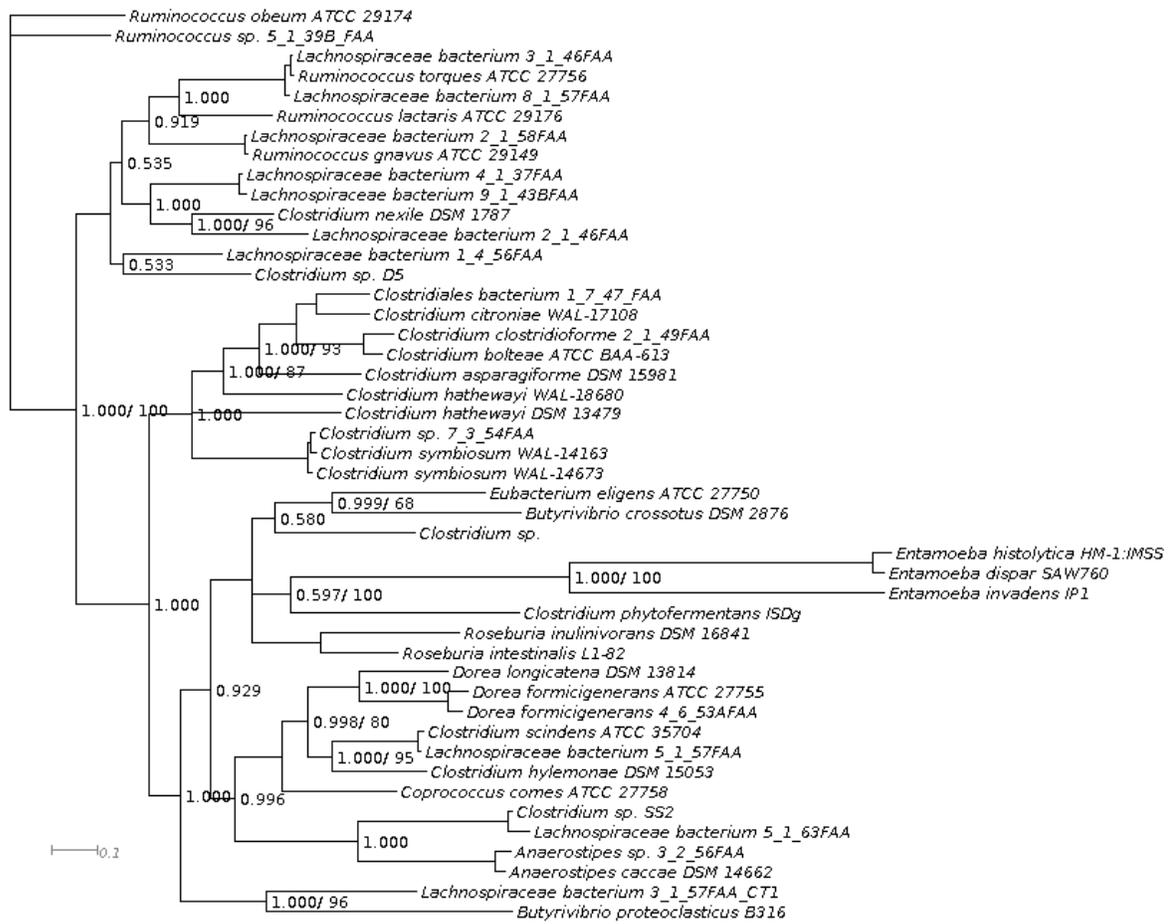


Figura 7. RNA metiltransferasa (EHI_047740). La topología presentada fue arrojada por el programa MrBayes 3.2, se muestra en algunos nodos la probabilidad posterior / bootstrap generada por MrBayes y PhymI respectivamente. Las barras indican el número de sustituciones esperadas por sitio.

Tabla 1. Número de genes por taxón donador.

Dominio	Filum	Orden	Número de genes	
Bacteria	Bacteroidetes	Bacteroidales	10	
		Sin asignar	6	
	Firmicutes	Clostridiales	4	
		Bacillales	1	
		Sin asignar	6	
		Campylobacterales	2	
	Proteobacteria	Enterobacteriales	1	
		Thiotrichales	1	
		Pseudomonadales	1	
		Sin asignar	2	
		Spirochaeta	Sin asignar	2
		Fusobacteria	Fusobacteriales	1
	Sin asignar		10	
	Archaea	Euryarchaeota	Methanococcales	1
Sin asignar			2	
Crenarchaeota		Sin asignar	1	

Tanto la alcohol deshidrogenasa NADP-dependiente (EHI_042260) como la enzima málica (EHI_044970) fueron obtenidas de organismos del filum Spirochaetes. Este filum está compuesto por bacterias anaerobias, comúnmente de forma larga y torcida (de ahí el nombre del filum). El gen de la deshidrogenasa NADP-dependiente de *E. histolytica* participa en la fermentación alcohólica anaerobia de *E. histolytica* (62), al igual que el gen de la enzima málica. En un trabajo del año 2000, Field y colaboradores ya habían probado la hipótesis de que *E. histolytica* usa genes xenólogos para llevar a cabo la fermentación alcohólica (63), pero no incluyeron un análisis de saturación mutacional a sus análisis ni les fue posible asignar un grupo donador.

El único filum bacteriano que donó un solo gen fue Fusobacteria, que contiene bacterias anaerobias obligadas. El gen transferido al genoma de las especies de Entamoeba fue la proteína hipotética (EHI_017590) en la cual se agrupó con altos valores de probabilidad posterior y bootstrap con sus homólogos bacterianos (Anexo 5).

Por el otro lado, se encontró que el dominio Archaea también ha transferido genes al género *Entamoeba*:

cuatro genes de los sesenta y uno fueron donados por este dominio procarionte. Para 3 de ellos se encontró que el posible filum donador fue Euryarchaeota, el cual que incluye organismos con características ecológicas muy contrastantes. Tanto la isopentenil-difosfato delta-isomerasa (EHI_194410), obtenida del orden Methanococcales, como la citidililtransferasa (EHI_198800) de *E. histolytica*, están involucradas en la biosíntesis de membranas lipídicas.

Otro filum donador fue Crenarchaeota, que incluye tanto organismos extremófilos, como organismos marinos. La única proteína que se encontró asociada a este filum fue la RNA metiltransferasa (EHI_006720), cuya anotación funcional está relacionada con la resistencia a linezolid (61), que es un antibiótico sintético (Figura 10).

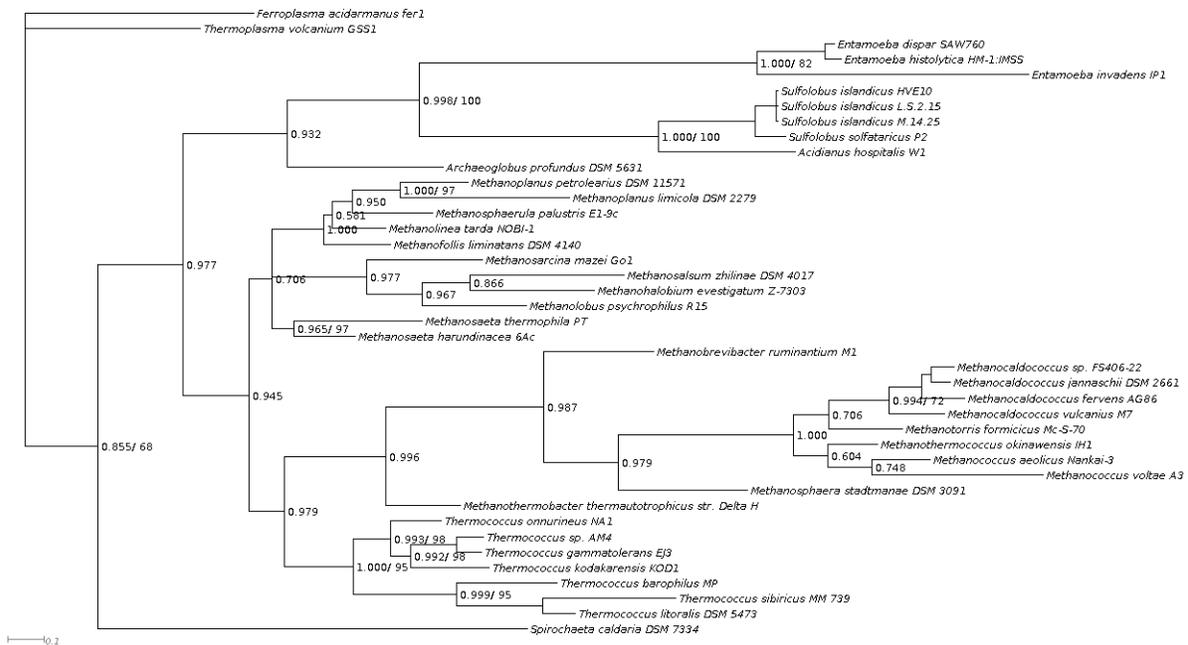


Figura 10. RNA metiltransferasa, familia TrmH (EHI_006720). La topología presentada fue arrojada por el programa MrBayes 3.2, se muestra en algunos nodos la probabilidad posterior / bootstrap generada por MrBayes y PhymI respectivamente. Las barras indican el número de sustituciones esperadas por sitio.

Varios de los genes obtenidos horizontalmente, tienen anotaciones funcionales relacionadas con la resistencia a antibióticos como la RNA metiltransferasa (EHI_006720) o las proteínas de la familia de las beta-lactamasas, los genes EHI_068560 y EHI_169860. Aunque los antibióticos que se utilizan hoy en día para tratar infecciones bacterianas sean derivados de compuestos encontrados en la naturaleza, como el thioestrepton, la clortetraciclina y la penicilina (64), y que de la misma manera los factores de resistencia a antibióticos también se encuentren expresados en la naturaleza (65), existen varias razones por las

cuales no es posible afirmar que los genes transmitidos al género *Entamoeba* via transferencia horizontal que tienen anotación funcional relacionada con la resistencia a antibióticos hayan fungido como factores de adaptación en la historia evolutiva del género. En primer lugar, no se sabe si estos antibióticos que encontramos hoy en día ya estaban presentes cuando ocurrieron estas transferencias. Asimismo, es posible que estos genes hayan cumplido una función determinada en el genoma donador y esta cambie por el contexto genómico que encuentra en el grupo receptor. Dicho de otra manera, un gen que fue transferido al genoma del ancestro del género *Entamoeba* puede resultar útil como factor de resistencia a antibióticos en la actualidad, después de haber evolucionado en el genoma amibiano. En este caso es posible hablar de la transferencia de un factor pre-adaptativo o exaptación (66). Por el otro lado es necesario llevar a cabo estudios bioquímicos que confirmen que la anotación funcional de la secuencia sea correcta, ya que el algoritmo bioinformático usado para buscar péptidos homólogos utilizado está basado en alineamientos locales y no en las funciones que realizan los productos de los genes o la predicción de estas funciones.

En algunos casos la secuencia de *E. invadens* tiene un grupo hermano distinto al de *E. histolytica* o tiene valores de distancia muy elevados en comparación, no sólo con las secuencias procariontes, sino con las demás secuencias del género de amibas. Esto sugiere que estas secuencias de *E. invadens* son en verdad homoplasias.

5.5 Estimación de tiempos de divergencia

Dado que no existe registro fósil de la historia evolutiva del linaje del género *Entamoeba*, fue necesario calibrar las estimaciones de tiempos de divergencia de manera relativa. El descubrimiento de *E. nuttallii* contribuyó de manera crucial a la calibración del nodo que representa al ancestro de *E. histolytica* y *E. nuttallii*: Como la especie *E. histolytica* ha sido encontrada en heces de primates del género *Papio*, entre otros, y asumiendo que *E. nuttallii* es un linaje aislado que infecta exclusivamente a macacos rhesus, es probable que *E. nuttallii* divergió al mismo tiempo o inmediatamente después de que el género *Macaca* se separara de su linaje ancestral. Por lo tanto la calibración se llevó a cabo con el tiempo de divergencia entre los géneros *Papio* y *Macaca*. En cuanto al registro fósil de estos primates, se tienen fósiles de macacos en África y Europa de hace 5.5 Map (67), y de papioninos de hace unos cuatro Map (68), asimismo Delson y colaboradores estimaron esta divergencia entre siete y ocho Map (67). Por lo tanto, se usó el intervalo de 5.5 y 6.5 Map, asumiendo que la especie *E. nuttallii* surgió después de la divergencia de los primates y que no hubo saltos a hábitos de vida libre en este proceso.

Cuando se calibra nodos para llevar a cabo estimaciones de tiempos de divergencia de organismos

microscópicos, es inevitable el uso de dataciones relativas. En este caso, se asumió una divergencia en espejo de *E. histolytica* y *E. nuttallii*, frente a los géneros *Papio* y *Macaca*, respectivamente. Sin embargo, no es posible asegurar que toda la historia evolutiva del género *Entamoeba* tenga esta característica, dado que las relaciones filogenéticas de la mayoría de los hospederos no sigue el mismo patrón, por ejemplo los hospederos de *E. invadens*, *E. equi* y *Entamoeba sp. NIH:1091:1* son *Natrix cyclopion*, *Equus caballus* e *Iguana iguana* respectivamente (figura 1, 1). En un trabajo reciente, Fiz-Palacios y colaboradores dataron la divergencia del filum Amoebozoa incluyendo un microfósil reconocido como ancestro del filum Amoebozoa, además de calibraciones paralelas usando fósiles de plantas, hongos y animales (69). Tanto el uso del microfósil ancestral, como el de calibraciones paralelas fue una estrategia coherente a sus objetivos ya que se buscaba indagar las relaciones de grupos en las jerarquías de clases y órdenes, probablemente usar estos datos resultaría en una sobreestimación de tiempos de divergencia para la comparación de especies, además de que Fiz-Palacios et al. usaron una clasificación distinta específicamente para la especie *E. histolytica* (69). En términos generales una buena estimación de tiempos de divergencia por probabilidad bayesiana depende de la longitud de las secuencias analizadas y de por lo menos la calibración de dos nodos con registros fósiles (70), por lo tanto es posible que las estimaciones de este trabajo resulten subestimadas ya que solamente se calibró un nodo.

Los tiempos de divergencia de las cuatro especies de *Entamoeba* incluidas en este análisis fueron calculados a partir de los siguientes genes nucleares: subunidad RPA2 de la RNA polimerasa DNA-dependiente I (EHI_186020), factor de elongación 2 (EHI_189490), actina (EHI_131230), cadena gamma de tubulina (EHI_008240) y cadena pesada de la clatrina (EHI_201510). En este análisis se incluyó homólogos de *E. histolytica*, *E. nuttallii*, *E. dispar*, *E. invadens* y *D. discoideum*. La tasa mediana de evolución molecular preliminar de los cinco genes constitutivos arrojada por el programa Estbranches fue 0.02364 sustituciones por sitio por millón de años. Este mismo valor fue utilizado como la distribución del prior de la media y desviación estándar de la tasa de evolución molecular en el nodo que agrupa únicamente a las secuencias de *Entamoeba* para el programa Multidivtime. Tomando en cuenta que este valor es de 0.0022 sustituciones por sitio por millón de años en eucariontes en general (71), la tasa de sustitución de estos genes es un orden de magnitud más alta, observación que puede ser explicada por el tiempo generacional corto y el gran tamaño poblacional del parásito.

El análisis llevado a cabo por el programa Multidivtime arrojó los siguientes valores para los tiempos de divergencia de las especies de *Entamoeba*: la divergencia de *E. histolytica* y *E. nuttallii* ocurrió hace 5.93+/-0.28 Map, mientras que los linajes de *E. histolytica* y *E. dispar* se separaron hace 9.97+/-1.37 Map y por último, la divergencia de los linajes de *E. histolytica* y *E. invadens*, se dió hace 68.18+/-16.04 Map.

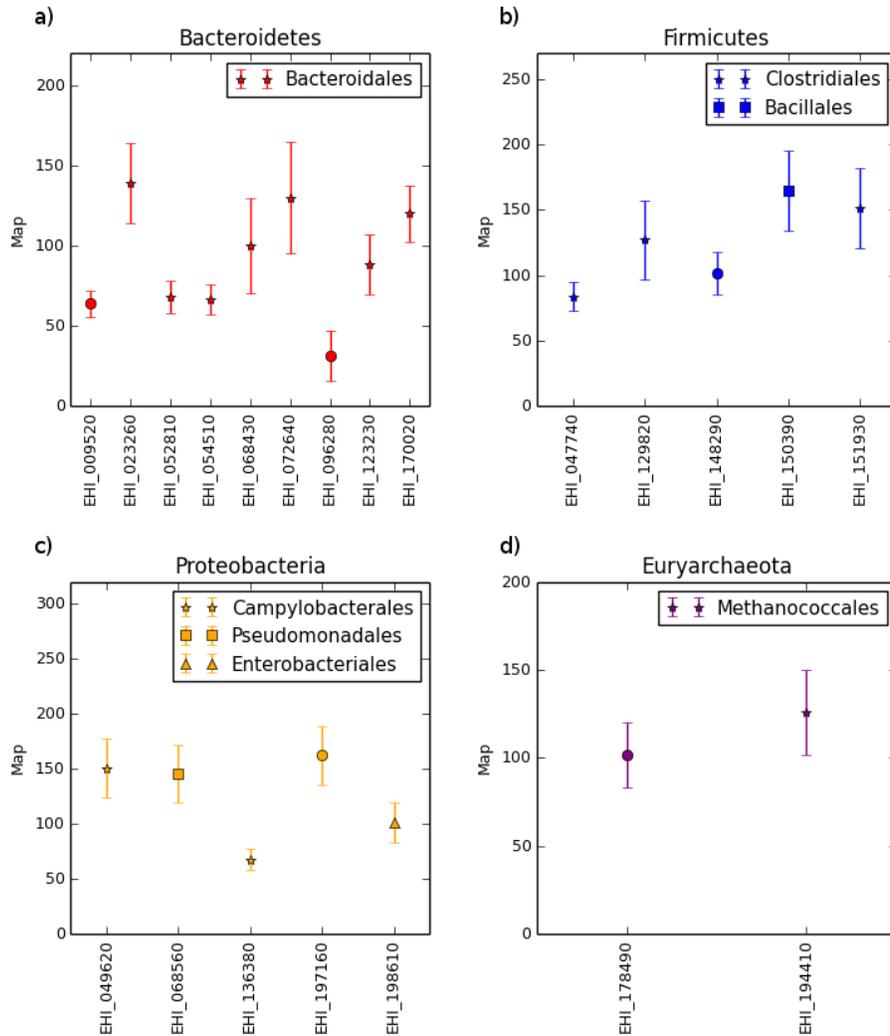


Figura 11. Línea de tiempo en la cual se muestra la media +/- desviación estándar de los tiempos de divergencia calculados por el programa Multidivtime para las secuencias obtenidas por el filum Bacteroidetes (a), Firmicutes(b), Proteobacteria (c) y Euryarchaeota(d). Las unidades mostradas en el eje de las abscisas son Map (millones de años antes del presente).

Con los resultados anteriores se calibraron los nodos para llevar a cabo la datación de las transferencias horizontales. De los sesenta y ocho candidatos de transferencia horizontal de genes solamente se tomaron en cuenta 21 genes debido a que fueron los únicos que cumplieron las características siguientes: agrupación bien soportada en los árboles filogenéticos, grupo donador conocido por lo menos hasta filum y la presencia de un gen ortólogo en el genoma de *E. nuttallii*. Según los resultados obtenidos, las transferencias acontecieron dentro del intervalo entre 15 Map y 285 Map, la más reciente, correspondiente al gen de la endo-1,4-beta-xilanasasa (EHI_096280), ocurrió hace 31.45+/-15.69 Map y fue transferida por

el filum Bacteroidetes. De hecho, ésta resultó más reciente que la divergencia que separó al ancestro de *E. histolytica*, *E. nuttallii* y *E. dispar* del linaje de *E. invadens*. Este hecho explica por que este gen está ausente en el genoma de *E. invadens*.

La media de los tiempos de divergencia es de 121.69 Map, lo que representa más del doble de tiempo que la divergencia entre *E. histolytica* y *E. invadens*. Dados estos tiempos tan profundos, es posible que se haya podido encontrar falsos negativos en las pruebas de saturación de nucleótidos. Es posible que un gen transferido horizontalmente haya sufrido tantos cambios, en este intervalo de tiempo tan grande, que haya perdido la identidad en la secuencia con sus homólogas procariontes y que hoy en día los algoritmos de búsqueda de homólogos fallen en encontrarlas. El evento de transferencia lateral más antiguo obtenido fue el de la tartrato deshidrogenasa (EHI_E143560) hace 253+/-28.91 Map, una datación tan profunda sólo pudo ser posible gracias a la tasa de evolución molecular, cuya mediana resultó ser de 0.00089 cambios por sitio por millón de años, este valor fue incluso menor a la mediana de evolución molecular final de los genes constitutivos que fue de 0.0014 cambios por sitio por millón de años. Por lo tanto, es probable que la señal de transferencia horizontal de genes se conserve dependiendo de dos factores: el tiempo que ha transcurrido y la tasa de sustitución del gen en el genoma del organismo receptor. Sin embargo, asignar el tiempo máximo que puede transcurrir antes de que una secuencia xenóloga pierda la señal filogenética para localizar a su grupo donador, no es trivial y depende tanto de la tasa de sustituciones, como de la proporción de sitios invariables y de la heterogeneidad de tasas de sustitución a lo largo del alineamiento. Describir esta relación en un lenguaje formal no se encuentra entre los objetivos de este trabajo.

Se encontraron varias transferencias con tiempos de divergencia superpuestos (figura 11) que, además, provinieron del mismo grupo donador, lo que fue considerado como un indicio de un evento de transferencia horizontal simultánea. Los cinco genes obtenidos tempranamente del orden Bacteroidales comparten un intervalo de tiempo entre los 129.99 y 137.7 Map, por el otro lado los genes alfa-1,2-manosidasa (EHI_009520), manosa-1-fosfato guanililtransferasa (EHI_052810), fructocinasa (EHI_054510) y proteína hipotética (EHI_123230), las tres últimas también de Bacteroidales, comparten otro intervalo de tiempo entre 69.18 y 72.17 Map. Aunque, estos no son los únicos periodos de tiempo compartidos entre todas las estimaciones de Bacteroidetes, si son los que agrupan el mayor número de trasposos. De las cinco transferencias datadas del filum Firmicutes, tres de ellas comparten un espacio de tiempo entre 134.23 y 157.27 Map, aunque no se sabe a qué órdenes pertenecen dos de ellas. De igual forma, el filum Proteobacteria pudo haber transferido juntos un gen de la familia metalo betalactamasa (EHI_06856) y la proteína de clúster de Fe-S de ensamblaje (EHI_049620) entre 176.93 y 119.89 Map.

Por último, los dos eventos datados del filum Euryarchaeota, comparten el intervalo entre 101.91 y 119.85 Map.

Aunque muchos eventos de transferencia horizontal de genes tengan tiempos de divergencia superpuestos, e incluso varios provengan de un mismo orden conocido, es necesario tomar en cuenta otros factores para evitar resultados falsos positivos. Dado que es común que un genoma procarionte esté organizado en forma de operones, teniendo secuencias codificantes para proteínas funcionalmente relacionadas unas seguidas de otras, siendo reguladas por un solo operador. Tomando esto en cuenta, se esperaría que una transferencia simultánea debería dar como resultados genes xenólogos con funciones relacionadas insertados uno tras otro en el genoma del organismo receptor, siempre que estos genes hayan sido obtenidos a partir de una sola secuencia de ácido nucleico.

Si sólo tomamos en cuenta los genes a los que se les asignó un orden donador, y agrupamos las proteínas por función, tenemos únicamente dos genes que pudieron ser obtenidos simultáneamente, éstos del orden Bacteroidales. Según la información proveniente de la Enciclopedia de Kyoto de Genes y Genomas (KEGG, por sus siglas en inglés, 72), tanto el gen de la manosa-1-fosfato guanililtransferasa (EHI_052810), así como el de la fructocinasa (EHI_054510) están involucrados en el metabolismo de fructosa, manosa, amino azúcares y nucleótido azúcares, por lo tanto podemos agrupar a estos dos genes en una sola transferencia (figura 12). Aunque existen pruebas de que estos genes llegaron al mismo tiempo al genoma de *E. histolytica*, no se puede afirmar que hayan estado juntos en el mismo operón puesto que, según información de la Base de Datos de Operones (ODB, por sus siglas en inglés, 73) el operón de la fructocinasa contiene solamente un gen. En cuanto a la sintenia, la distancia física que separa estos genes en la secuencia genómica de *E. histolytica* se ignora, ya que el genoma de este organismo no se ha ensamblado en su totalidad (38). De hecho, cada gen, cuya transferencia fue datada, se encuentra en un scaffold diferente y cuando llega a encontrarse más de uno en un mismo scaffold, pertenecen a taxa donadores distintos.

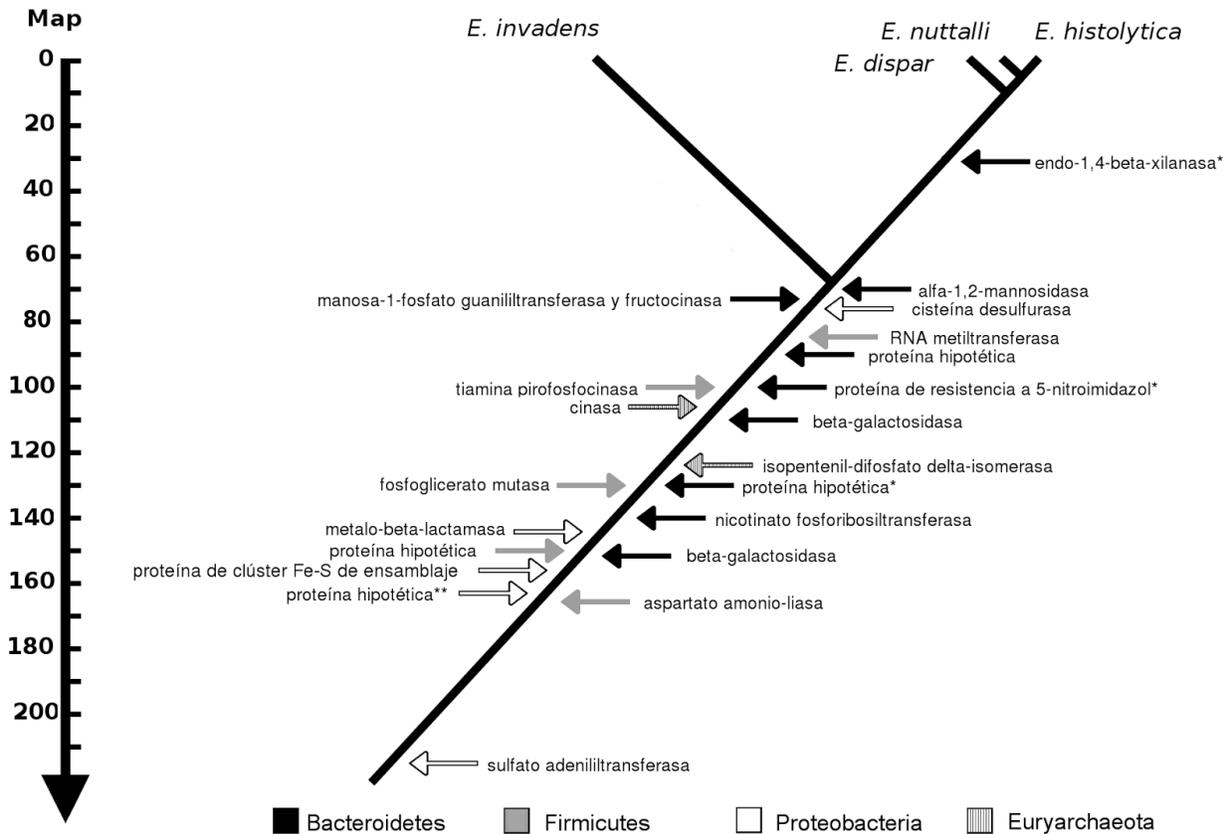


Figura 12. Esquema que ilustra la antigüedad, tanto de las divergencias de las especies de *Entamoeba*, así como de las transferencias horizontales. Las unidades empleadas son Map (millones de años antes del presente).

* genes sin homólogos en el genoma de *E. invadens*.

** genes sin homólogos en el genoma de *E. dispar*.

Como se muestra en la figura 12, la mayoría de estos genes fueron transferidos antes de la divergencia *E. histolytica*-*E. invadens*, lo que explica su presencia en cada uno de estos linajes, por otra parte, si se compara esta figura con la figura 1, es evidente que existe un número considerable de especies del género *Entamoeba* que no fueron incluidas en el análisis, por ejemplo *E. moshkovskii* y *E. bovis*. De la misma manera, como estos genes están presentes en *E. dispar*, es de esperarse que éstos también estén presentes en todas las especies que surgen del ancestro entre *E. equi* y su grupo hermano, a menos que sus genomas hayan sufrido pérdidas. Construyendo clusters de homólogos o realizando búsquedas con algoritmos bioinformáticos (por ejemplo el BLAST) es posible demostrar esta predicción, sin embargo, hoy en día es problemático llevar a cabo este procedimiento dado el estado actual de ensamblaje y anotación de los genomas disponibles de los integrantes de este género (38).

Tanto las especies del género *Entamoeba*, como las especies de su grupo hermano, el género *Endolimax* son especies de parásitos o comensales que colonizan principalmente el intestino de vertebrados (74). Al igual que en *E. histolytica*, se han observado mitosomas en amibas del género *Endolimax* (75), lo que sugiere que su metabolismo es igual, aunque no existen estudios experimentales que lo respalden.

Probablemente la transferencia de los genes de la alcohol deshidrogenasa NADP-dependiente (EHI_042260) y la enzima málica (EHI_044970) a partir de bacterias del filum Spirochaetes abrieron al ancestro *Entamoeba-Endolimax* la posibilidad de colonizar ambientes anaerobios y dar lugar a una radiación que hoy se ve reflejada en el amplio espectro de hospederos que aprovechan estas amibas. Se sabe que ambos genes son una parte fundamental del metabolismo de *E. histolytica* ya que por un lado, existen pruebas de que un agente antiamebico, el gosispol, actúa inhibiendo la enzima málica (76), y que la alcohol deshidrogenasa NADP-dependiente está involucrada directamente en el metabolismo de la amiba y se ha encontrado activa en los trofozoítos (77).

De hecho, la presencia de estos genes en el genoma de alguna especie de *Endolimax* podría dar fin a la controversia y determinar si estos dos géneros provienen de un mismo ancestro común (74). De ser así, es posible que esta radiación haya dado plasticidad al genoma de ambos géneros y de esta forma incorporar este gran número de genes bacterianos. Sin embargo, la imposibilidad de datar los genes de la alcohol deshidrogenasa NADP-dependiente y la enzima málica, y el hecho de que aún no se encuentra disponible el genoma de alguna especie de *Endolimax*, impide corroborar esta hipótesis.

5.6 Análisis del uso preferencial de codones

Se calculó el índice MILC (54) sobre las frecuencias relativas de los genes xenólogos y los genes constitutivos. Las frecuencias relativas observadas fueron consideradas significativamente diferentes de las esperadas siempre que la puntuación del índice MILC fuera igual o mayor a 0.5. Como se puede observar en la figura 12, son pocos los genes que tienen un uso preferencial de codones significativamente distinto al promedio de *E. histolytica*: la fosfofructocinasa pirofosfato-dependiente (EHI_000730), la xilulosa cinasa (EHI_012230) de Bacteroidetes, la enzima málica de Spirochaetes (EHI_044970), la proteína de clúster Fe-S de ensamblaje (EHI_049620) y la sulfato adenililtransferasa (EHI_197160) de Proteobacteria. Estos resultados confirman la homogeneización del código genético que experimentan los genes obtenidos lateralmente, que ha enmascarado a la mayoría de los genes que fueron transferidos de genomas procariontes al genoma del género *Entamoeba*, tras millones de años (figura 12). En la figura 13 también se muestra que todavía otro gen, el que codifica para la aldehído-alcohol deshidrogenasa 2 (EHI_160940), tiene un valor de MILC mayor a 0.5 y es de hecho la secuencia con el

valor más alto. Este gen fue descartado de los análisis por presentar alta saturación mutacional, por lo tanto estos resultados sugieren que esta proteína en específico no fue transferida horizontalmente y sus semejanza a secuencias procariontes son resultado de una evolución molecular relajada, aunque esta conclusión estaría rechazando parte de los resultados de Field y colaboradores del año 2002 (63).

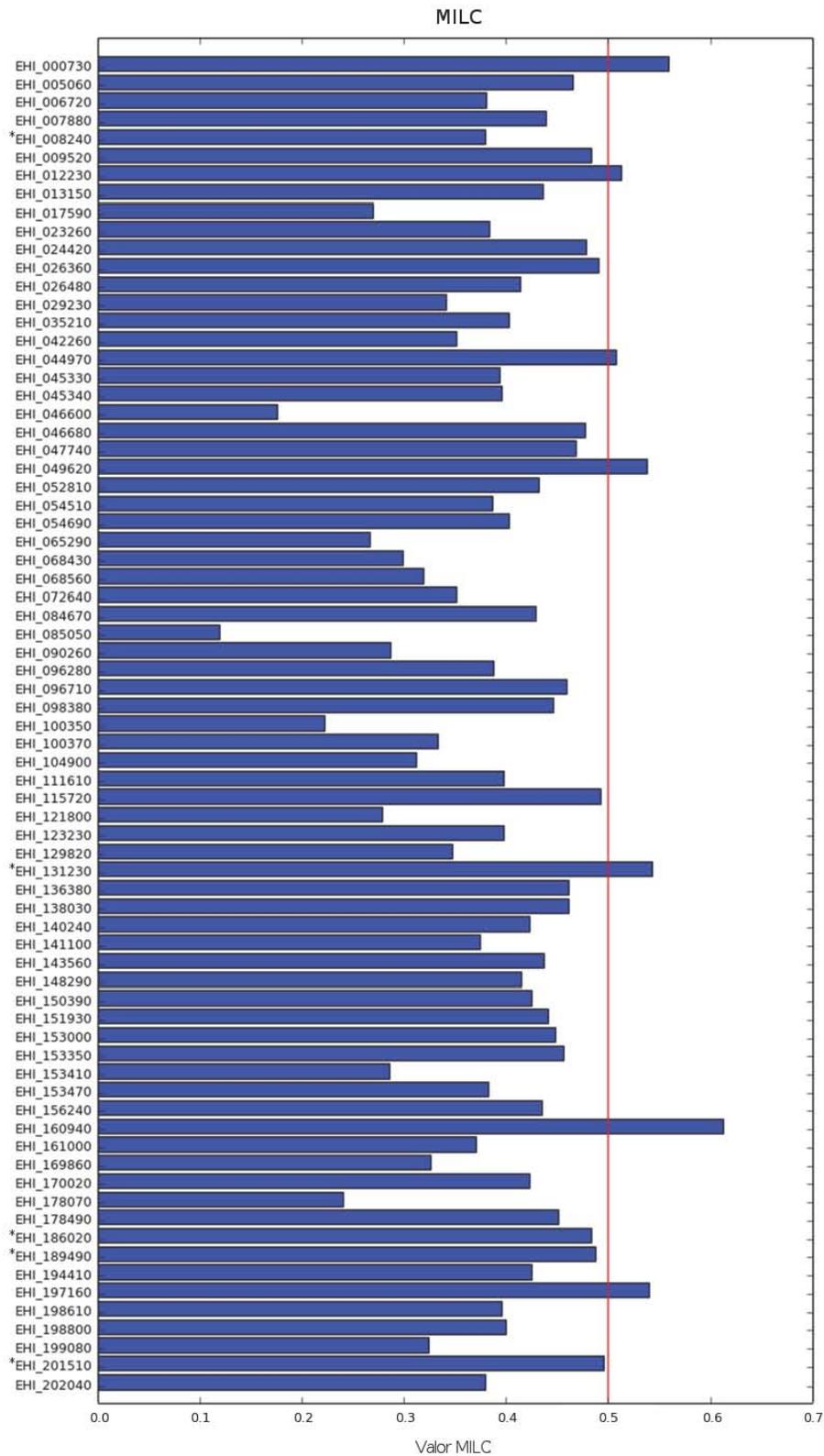


Figura 13. Valores de la medida independiente de longitud y composición para el uso de codones de los genes candidato y genes constitutivos(*).

6. CONCLUSIONES

El análisis de las secuencias candidato de transferencia horizontal de genes a partir de nucleótidos permitió evaluar el valor informativo de los alineamientos con los posibles homólogos encontrados en las bases de datos, de esta manera se encontraron siete alineamientos inútiles para llevar a cabo análisis filogenéticos. Al reconstruir filogenias con los 61 candidatos restantes se comprobó que efectivamente, los grupos donadores cambiaron con respecto a trabajos anteriores, lo que confirma que los resultados de este tipo de análisis están sujetos a cambiar a lo largo del tiempo, conforme las bases de datos públicas crecen. Además, gracias a la construcción de estas filogenias, también fue posible confirmar que el género *Entamoeba* ha obtenido factores adaptativos por medio de la transferencia horizontal de genes, específicamente genes de obtención de energía en hábitats anaeróbios. Además se encontró que algunos genes obtenidos vía transferencia horizontal pudieron funcionar como exaptaciones que hoy en día otorgan resistencia a antibióticos. Estos resultados acentúan la importancia de la interacción de este parásito intestinal con los otros organismos que habitan el colon, principalmente los procariontes. A través de análisis rigurosos para estimar los tiempos de divergencia entre especies de este género, fue posible datar algunos eventos de transferencias horizontales y predecir la presencia de estos genes en el genoma de especies del género *Endolimax*. De hecho es posible que la transferencia horizontal de genes haya sido relevante para la radiación de este grupo como parte de la microbiota intestinal gracias a la obtención de mecanismos alternos de obtención de energía de origen bacteriano. Finalmente, se determinó el acaecimiento una posible transferencia horizontal simultánea.

7. REFERENCIAS

1. Stensvold, C. R., Lebbad, M., Victory, E. L., Verweij, J. J., Tannich, E., Alfellani, M., ... & Clark, C. G. (2011). Increased Sampling Reveals Novel Lineages of *Entamoeba*: Consequences of Genetic Diversity and Host Specificity for Taxonomy and Molecular Detection. *Protist*, 162(3), 525-541.
2. Jackson, T. F., Sargeant, P. G., Visser, P. S., Gathiram, V., Suparsad, S., & Anderson, C. B. (1989). *Entamoeba histolytica*: naturally occurring infections in baboons. *Archivos de investigacion medica*, 21, 153-156.
3. Ghandour, A. M., Zahid, N. Z., Banaja, A. A., Kamal, K. B., & Bouq, A. I. (1995). Zoonotic intestinal parasites of hamadryas baboons *Papio hamadryas* in the western and northern regions of Saudi Arabia. *The Journal of tropical medicine and hygiene*, 98(6), 431-439.
4. Muriuki, S. M. K., Murugu, R. K., Munene, E., Karere, G. M., & Chai, D. C. (1998). Some gastro-intestinal parasites of zoonotic (public health) importance commonly observed in old world non-human primates in Kenya. *Acta Tropica*, 71(1), 73-82.
5. Legesse, M., & Erko, B. (2004). Zoonotic intestinal parasites in *Papio anubis* (baboon) and *Cercopithecus aethiops* (vervet) from four localities in Ethiopia. *Acta Tropica*, 90(3), 231-236
6. WHO(1998). "The world health report 1998. Life in the 21st century: A vision for all." World Health Organization, Geneva, Switzerland.
7. Mukherjee, C., Clark, C. G., & Lohia, A. (2008). *Entamoeba* shows reversible variation in ploidy under different growth conditions and between life cycle phases. *PLoS neglected tropical diseases*, 2(8), e281.
8. Wilson, I. W., Weedall, G. D., & Hall, N. (2012). Host-Parasite interactions in *Entamoeba histolytica* and *Entamoeba dispar*: what have we learned from their genomes?. *Parasite immunology*, 34(2-3), 90-99.
9. Cheun, H. I., Cho, S. H., Lee, J. H., Lim, Y. Y., Jeon, J. H., Yu, J. R., ... & Kwon, M. H. (2010). Infection status of hospitalized diarrheal patients with gastrointestinal protozoa, bacteria, and viruses in the Republic of Korea. *The Korean journal of parasitology*, 48(2), 113-120.
10. Galván-Moroyoqui, J. M., Del Carmen Dominguez-Robles, M., Franco, E., & Meza, I. (2008). The interplay between *Entamoeba* and enteropathogenic bacteria modulates epithelial cell damage. *PLoS neglected tropical diseases*, 2(7), e266.
11. Wassmann, C., Hellberg, A., Tannich, E., & Bruchhaus, I. (1999). Metronidazole resistance in the protozoan parasite *Entamoeba histolytica* is associated with increased expression of iron-containing superoxide dismutase and peroxiredoxin and decreased expression of ferredoxin 1 and flavin reductase. *Journal of Biological Chemistry*, 274(37), 26051-26056.
12. Freeman, C. D., Klutman, N. E., & Lamp, K. C. (1997). Metronidazole. *Drugs*, 54(5), 679-708.
13. Tachibana, H., Yanagi, T., Lama, C., Pandey, K., Feng, M., Kobayashi, S., & Sherchand, J. B. (2013). Prevalence of *Entamoeba nuttalli* infection in wild rhesus macaques in Nepal and characterization of the parasite isolates. *Parasitology international*, 62(2), 230-235.
14. Tachibana, H., Yanagi, T., Akatsuka, A., Kobayashi, S., Kanbara, H., & Tsutsumi, V. (2009). Isolation and characterization of a potentially virulent species *Entamoeba nuttalli* from captive Japanese macaques. *Parasitology*, 136(10), 1169-1177.
15. Ximénez, C., Cerritos, R., Rojas, L., Dolabella, S., Morán, P., Shibayama, M., ... & Silva, E. F. (2010). Human amebiasis: breaking the paradigm?. *International journal of environmental research and public health*, 7(3), 1105-1120.

16. Clark, C. G., & Diamond, L. S. (1991). Ribosomal RNA genes of 'pathogenic' and 'nonpathogenic' *Entamoeba histolytica* are distinct. *Molecular and biochemical parasitology*, 49(2), 297-302.
17. Novati, S., Sironi, M., Granata, S., Bruno, A., Gatti, S., Scaglia, M., & Bandi, C. (1996). Direct sequencing of the PCR amplified SSU rRNA gene of *Entamoeba dispar* and the design of primers for rapid differentiation from *Entamoeba histolytica*. *Parasitology*, 112(04), 363-369.
18. Sehgal, D., Mittal, V., Ramachandran, S., Dhar, S. K., Bhattacharya, A., & Bhattacharya, S. (1994). Nucleotide sequence organisation and analysis of the nuclear ribosomal DNA circle of the protozoan parasite *Entamoeba histolytica*. *Molecular and biochemical parasitology*, 67(2), 205-214.
19. Koonin, E. V., Makarova, K. S., & Aravind, L. (2001). Horizontal gene transfer in prokaryotes: quantification and classification 1. *Annual Reviews in Microbiology*, 55(1), 709-742.
20. Marri, P. R., Hao, W., & Golding, G. B. (2007). The role of laterally transferred genes in adaptive evolution. *BMC evolutionary biology*, 7(Suppl 1), S8.
21. Mirkin, B. G., Fenner, T. I., Galperin, M. Y., & Koonin, E. V. (2003). Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC evolutionary biology*, 3(1), 2.
22. Whitaker, J., Mcconkey, G., & Westhead, D. (2009). Prediction of horizontal gene transfers in eukaryotes: approaches and challenges. *Biochemical Society Transactions*, 37(4), 792.
23. Gladyshev, E. A., Meselson, M., & Arhipova, I. R. (2008). Massive horizontal gene transfer in bdelloid rotifers. *science*, 320(5880), 1210-1213.
24. Andersson, J. O., Doolittle, W. F., & Nesbø, C. L. (2001). Are there bugs in our genome?. *Science (Washington)*, 292(5523), 1848-1850.
25. Groisman, E. A., & Ochman, H. (1996). Pathogenicity islands: bacterial evolution in quantum leaps. *Cell*, 87(5), 791-794.
26. Richards, T. A., Soanes, D. M., Jones, M. D., Vasieva, O., Leonard, G., Paszkiewicz, K., ... & Talbot, N. J. (2011). Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. *Proceedings of the National Academy of Sciences*, 108(37), 15258-15263.
27. Cohen, S. N., Chang, A. C., & Hsu, L. (1972). Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. *Proceedings of the National Academy of Sciences*, 69(8), 2110-2114.
28. Courvalin, P., & Carlier, C. (1986). Transposable multiple antibiotic resistance in *Streptococcus pneumoniae*. *Molecular and General Genetics MGG*, 205(2), 291-297.
29. Gandon, S., & Michalakis, Y. (2000). Evolution of parasite virulence against qualitative or quantitative host resistance. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1447), 985-990.
30. Lawrence, J. G., & Ochman, H. (1998). Molecular archaeology of the *Escherichia coli* genome. *Proceedings of the National Academy of Sciences*, 95(16), 9413-9417.
31. Salzberg, S. L., White, O., Peterson, J., & Eisen, J. A. (2001). Microbial genes in the human genome: lateral transfer or gene loss?. *Science*, 292(5523), 1903-1906.
32. Xia, X., Xie, Z., Salemi, M., Chen, L., & Wang, Y. (2003). An index of substitution saturation and its application. *Molecular phylogenetics and evolution*, 26(1), 1-7.

33. Loftus, B., Anderson, I., Davies, R., Alsmark, U. C. M., Samuelson, J., Amedeo, P., ... & Clark, C. G. (2005). The genome of the protist parasite *Entamoeba histolytica*. *Nature*, 433(7028), 865-868.
34. Clark, C. G., Alsmark, U. C. M., Tazreiter, M., Saito-Nakano, Y., Ali, V., Marion, S., ... & Hall, N. (2007). Structure and Content of the *Entamoeba histolytica* Genome. *Advances in parasitology*, 65, 51-190.
35. Alsmark, U. C., Sicheritz-Ponten, T., Foster, P. G., Hirt, R. P., & Embley, T. M. (2009). Horizontal gene transfer in eukaryotic parasites: a case study of *Entamoeba histolytica* and *Trichomonas vaginalis*. In *Horizontal Gene Transfer* (pp. 489-500). Humana Press.
36. Thomas, F., Hehemann, J. H., Rebuffet, E., Czjzek, M., & Michel, G. (2011). Environmental and gut Bacteroidetes: the food connection. *Frontiers in microbiology*, 2.
37. Marchesi, J. R. (2010). Prokaryotic and eukaryotic diversity of the human gut. *Advances in applied microbiology*, 72, 43-62.
38. Aurrecochea, C., Barreto, A., Brestelli, J., Brunk, B. P., Caler, E. V., Fischer, S., ... & Wang, H. (2011). AmoebaDB and MicrosporidiaDB: functional genomic resources for Amoebozoa and Microsporidia species. *Nucleic acids research*, 39(suppl 1), D612-D619.
39. Basu, S., Fey, P., Pandit, Y., Dodson, R., Kibbe, W. A., & Chisholm, R. L. (2013). dictyBase 2013: integrating multiple Dictyostelid species. *Nucleic acids research*, 41(D1), D676-D683.
40. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic acids research*, 36(suppl 2), W5-W9.
41. Pruitt, K. D., Tatusova, T., Brown, G. R., & Maglott, D. R. (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic acids research*, 40(D1), D130-D135.
42. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., ... & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), 2947-2948.
43. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*, 28(10), 2731-2739.
44. Xia, X., & Xie, Z. (2001). DAMBE: software package for data analysis in molecular biology and evolution. *Journal of Heredity*, 92(4), 371-373.
45. Xia, X., & Lemey, P. (2009). Assessing substitution saturation with DAMBE. *The phylogenetic handbook: a practical approach to DNA and protein phylogeny*, 2009, 615-630.
46. Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9(8), 772-772.
47. Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology*, 52(5), 696-704.
48. Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology*, 61(3), 539-542.
49. Huson, D. H., & Scornavacca, C. (2012). Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic biology*, 61(6), 1061-1067.
50. Thorne, J. L., & Kishino, H. (2003). Multidivtime. Available from the authors at <http://statgen.ncsu.edu/thorne/multidivtime.html>.

51. Rutschmann, F. (2005). Bayesian molecular dating using paml/multidivtime: a step-by-step manual. Version 1.5 (July 2005). Institute of Systematic Botany, University of Zurich, Zurich, Switzerland.
52. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*, 24(8), 1586-1591.
53. Gouy, M., Guindon, S., & Gascuel, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution*, 27(2), 221-224.
54. Supek, F., & Vlahoviček, K. (2005). Comparison of codon usage measures and their applicability in prediction of microbial gene expressivity. *BMC bioinformatics*, 6(1), 182.
55. Supek, F., & Vlahoviček, K. (2010). Correction: Comparison of codon usage measures and their applicability in prediction of microbial gene expressivity. *BMC Bioinformatics*, 11(1), 1-1.
56. Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research*, 35(suppl 1), D61-D65.
57. Thomas, F., Hehemann, J. H., Rebuffet, E., Czjzek, M., & Michel, G. (2011). Environmental and gut Bacteroidetes: the food connection. *Frontiers in microbiology*, 2.
58. Murray, B. E., & Mederski-Samaroj, B. (1983). Transferable beta-lactamase. A new mechanism for in vitro penicillin resistance in *Streptococcus faecalis*. *Journal of Clinical Investigation*, 72(3), 1168.
59. Marchesi, J. R. (2010). Prokaryotic and eukaryotic diversity of the human gut. *Advances in applied microbiology*, 72, 43-62.
60. Mariat, D., Firmesse, O., Levenez, F., Guimaraes, V. D., Sokol, H., Dore, J., ... & Furet, J. P. (2009). The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age. *BMC microbiology*, 9(1), 123.
61. Morales, G., Picazo, J. J., Baos, E., Candel, F. J., Arribi, A., Peláez, B., ... & Sánchez-García, M. (2010). Resistance to linezolid is mediated by the cfr gene in the first report of an outbreak of linezolid-resistant *Staphylococcus aureus*. *Clinical Infectious Diseases*, 50(6), 821-825.
62. Kumar, A., Shen, P. S., Descoteaux, S., Pohl, J., Bailey, G., & Samuelson, J. (1992). Cloning and expression of an NADP (+)-dependent alcohol dehydrogenase gene of *Entamoeba histolytica*. *Proceedings of the National Academy of Sciences*, 89(21), 10188-10192.
63. Field, J., Rosenthal, B., & Samuelson, J. (2000). Early lateral transfer of genes encoding malic enzyme, acetyl-CoA synthetase and alcohol dehydrogenases from anaerobic prokaryotes to *Entamoeba histolytica*. *Molecular microbiology*, 38(3), 446-455.
64. Nicolaou, K. E., Chen, J. S., Edmonds, D. J., & Estrada, A. A. (2009). Recent advances in the chemistry and biology of naturally occurring antibiotics. *Angewandte Chemie International Edition*, 48(4), 660-719.
65. Hawkey, P. M. (1998). The origins and molecular basis of antibiotic resistance. *Bmj*, 317(7159), 657-660.
66. Gould, S. J., & Vrba, E. S. (1982). Exaptation-a missing term in the science of form. *Paleobiology*, 4-15.
67. Delson, E., Tattersall, I., Van Couvering, J., & Brooks, A. S. (Eds.). (2004). *Encyclopaedia of Human Evolution and Prehistory*. Routledge.

68. Jablonski, N. G. (2002). Fossil Old World monkeys: the late Neogene radiation. *Cambridge Studies in Biological and Evolutionary Anthropology*, 255-299.
69. Fiz-Palacios, O., Romeralo, M., Ahmadzadeh, A., Weststrand, S., Ahlberg, P. E., & Baldauf, S. (2013). Did terrestrial diversification of amoebas (amoebzoa) occur in synchrony with land plants?. *PloS one*, 8(9), e74374.
70. Dos Reis, M., & Yang, Z. (2013). The unbearable uncertainty of Bayesian divergence time estimation. *Journal of Systematics and Evolution*, 51(1), 30-43.
71. Kumar, S., & Subramanian, S. (2002). Mutation rates in mammalian genomes. *Proceedings of the National Academy of Sciences*, 99(2), 803-808.
72. Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic acids research*, 42(D1), D199-D205.
73. Okuda, S., & Yoshizawa, A. C. (2011). ODB: a database for operon organizations, 2011 update. *Nucleic acids research*, 39(suppl 1), D552-D555.
74. Silberman, J. D., Clark, C. G., Diamond, L. S., & Sogin, M. L. (1999). Phylogeny of the genera *Entamoeba* and *Endolimax* as deduced from small-subunit ribosomal RNA sequences. *Molecular biology and evolution*, 16(12), 1740-1751.
75. Constenla, M., Padrós, F., & Palenzuela, O. (2014). *Endolimax piscium* sp. nov.(Amoebozoa), causative agent of systemic granulomatous disease of cultured sole, *Solea senegalensis* Kaup. *Journal of fish diseases*, 37(3), 229-240.
76. González-Garza, M. T., Matlin, S. A., Mata-Cárdenas, B. D., & Said-Fernández, S. (1992). *Entamoeba histolytica*: inhibition of malic enzyme and alcohol dehydrogenase by (+/-)-, (+)-, and (-)-gossypol. *Archives of medical research*, 24(2), 183-187.
77. Reeves, R. E. (1985). Metabolism of *Entamoeba histolytica* Schaudinn, 1903. *Advances in parasitology*, 23, 105-142.