



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**Estudio *in silico* de la estructura metabólica de
“*Candidatus* Midichloria mitochondrii”, posibles
consecuencias de su modo de vida
endosimbionte.**

TESIS

QUE PARA OBTENER EL TÍTULO DE:

BIÓLOGO

PRESENTA:

RICARDO AUGUSTO MERCADO JUÁREZ



**DIRECTOR DE TESIS:
M. EN C. DANIEL CERQUEDA GARCÍA
Ciudad Universitaria, CD. MX., 2016**



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

Apellido paterno: Mercado
Apellido materno: Juárez
Nombre(s): Ricardo Augusto
Teléfono: 5551969571
Universidad Nacional Autónoma de México
Facultad de Ciencias
Carrera: Biología
Número de cuenta: 300296569

2. Datos del tutor:

Grado M. en C.
Nombre(s): Daniel
Apellido paterno: Cerqueda
Apellido materno: García

3. Datos del sinodal Propietario:

Grado: Dr
Nombre(s): Luis David
Apellido paterno: Alcaraz
Apellido materno: Peraza

4. Datos del sinodal Propietario:

Grado: M. en C.
Nombre(s): Alfonso José
Apellido paterno: Vilchis
Apellido materno Peluyera

5. Datos del sinodal Propietario tutor:

Grado: M. en C.
Nombre(s): Daniel
Apellido paterno: Cerqueda
Apellido materno: García

6. Datos del sinodal Suplente:

Grado: Dr.
Nombre(s). Arturo Carlos II
Apellido paterno: Becerra
Apellido materno: Bracho

7. Datos del sinodal Suplente:

Grado: Dra.
Nombre(s). Luisa Isaura
Apellido paterno: Falcón
Apellido materno: Álvarez

8. Datos del trabajo escrito.

Título: Estudio *in silico* de la estructura metabólica de “*Candidatus* Midichloria mitochondrii”, posibles consecuencias de su modo de vida endosimbionte.
64 p.
Año 2016

AGRADECIMIENTOS

A la Universidad Nacional Autónoma de México por haberme otorgado la oportunidad de formarme como un ser humano en plenitud.

A la Facultad de Ciencias por brindarme la oportunidad de formarme como Biólogo.

Al Laboratorio de Ecología Bacteriana, en especial a la Dra. Luisa Falcón por el apoyo brindado y las facilidades para la realización de este trabajo.

Al M. en C. Daniel Cerqueda por toda la oportunidad, ayuda, tolerancia, espera, apoyo, crítica constructiva, recomendación y motivación para realizar esta tesis.

Al grupo de sinodales que revisaron de forma puntual este trabajo:

Dr. Luis David Alcaraz Peraza

M. en C. Alfonso José Vilchis Peluyera

Dr. Arturo Carlos II Becerra Bracho

Dra. Luisa Isaura Falcón Álvarez

DEDICATORIA

A mi madre María Guadalupe Juárez Tejeda, por todo el amor, cariño, cuidados, consejos y demás.

Mis abuelos Agustín Juárez, Clara Tejeda, Aurelio Mercado, Rosa García por su cariño y atenciones de su parte.

Tía Rocío, Rocío (Piolina), Alex, Tío Ricardo...

...Al resto de mis familiares que me han apoyado y han estado conmigo siempre.

A Rocío Yanel por tu amor, comprensión, ayuda y motivación para seguir adelante.

“...Quero ser feliz também, navegar nas águas do teu mar...”

A todos los compañeros y amigos que han convivido, compartido y han estado conmigo durante todos estos años.

CONTENIDO

AGRADECIMIENTOS.....	IV
DEDICATORIA	V
CONTENIDO	VI
Lista de Figuras.....	VII
Lista de tablas	IX
RESUMEN.....	1
INTRODUCCIÓN.....	2
Evolución y simbiogénesis	2
Evolución reductiva del genoma	3
Transferencia horizontal de genes	5
“ <i>Candidatus</i> Midichloria mitochondrii”	10
HIPÓTESIS	12
OBJETIVOS	12
MATERIALES Y MÉTODOS	13
Genomas utilizados.....	13
Comparación de capacidades metabólicas y genomas	14
Análisis de transferencia horizontal de genes	15
Análisis filogenético.....	16
RESULTADOS Y DISCUSIÓN.....	17
Capacidades metabólicas y genomas.....	17
Transferencia horizontal.....	23
Análisis filogenético.....	29
CONCLUSIONES.....	34
REFERENCIAS	35
APÉNDICES.....	41
Apéndice A Categorías metabólicas.....	41
Apéndice B Resumen de genomas.....	43
Apéndice C Resultados ‘DarkHorse’	44
Apéndice D BLAST	46

Lista de Figuras

- Figura 1. Representación de las diferentes formas de transferencia horizontal en bacterias y arqueas. En la conjugación se transfiere material genético entre dos bacterias, en este caso por medio de un plásmido. En la transducción una partícula viral (fago) incorpora DNA a la célula receptora. En la transformación la célula receptora incorpora DNA externo. Modificado de Stewart (2013). 5
- Figura 2. Representación de transferencia intracelular o endosimbótica de genes donde los organismos endosimbiontes transfieren material genético al genoma del organismo hospedero debido a la estrecha cercanía de ambas células, modificado de Soucy et al., (2015)..... 6
- Figura 3. El esquema general del proceso del algoritmo de 'DarkHorse'. Comienza con una búsqueda no rigurosa (BLAST) de las secuencias de proteínas del genoma contra una base de datos, después se forman grupos de las secuencias de proteínas coincidentes en linajes distintos, se hace un cálculo de probabilidad de linaje, usando información taxonómica de las secuencias, basado en los mejores valores de los grupos formados de las secuencias de proteínas se eligen las mejores probabilidades y se calcula de nueva cuenta la probabilidad de linaje 'LPI' para formar una tabla con los resultados finales. 9
- Figura 4. 1) Mitocondria de un oocito de *I. ricinus* "infectada" por "Ca. *M. mitochondrii*" modificada de Pistone et al., (2012). 2) Imagen de microscopia electrónica de transmisión que muestra en (A) inicio de la infección de la bacteria [b] entre las membranas de la mitocondria [m] y (B) dos bacterias dentro de la mitocondria modificada de Sasser et al., (2011). 3) Micrografías electrónicas de células de ovario de *I. ricinus* con "Ca. *M. mitochondrii*" modificada de Sasser et al., (2006)..... 11
- Figura 5. 'Heatmap' que muestra el número de rutas en distintas categorías metabólicas afines al metabolismo central de diferentes bacterias, la intensidad del tono, en las celdas del 'heatmap', está relacionada con la cantidad de rutas en cada categoría. Los grupos principales en los dendrogramas (izquierda y arriba) están resaltados. BcofGpoPrAcEl = Biosíntesis Cofactores Grupos Prostéticos y Acarreadores de Electrones, DAcGrsLip = Degradación Ácidos Grasos y Lípidos, BMetSec = Biosíntesis Metabolitos Secundarios, DAmiPol = Degradación Aminas y Poliaminas, BAmiPol = Biosíntesis Aminas y Poliaminas, BCarb = Biosíntesis Carbohidratos, MNutInor = Metabolismo Nutrientes Inorgánicos, DNucNucl = Degradación Nucleótidos Nucleósidos, DMetSec = Degradación Metalitos Secundarios, DCarb = Degradación Carbohidratos, DCarbx = Degradación Carboxilatos, BAcGrsLip = Biosíntesis Ácidos Grasos Lípidos, GenMetPreEne = Generación Metabolitos Precursores Energía, BAA = Biosíntesis Aminoácidos, BNuclNucl = Biosíntesis Nucleótidos Nucleósidos. 18
- Figura 6. Gráficas comparativas entre la cantidad de genes/proteínas respecto al tamaño del genoma por cada organismo (arriba) y por forma de vida (abajo). Datos para análisis tomados de 'RefSeq' (Pruitt et al. 2002; Tatusova et al. 2014) Apéndice B Tabla 6. 20
- Figura 7. Gráficas comparativas del tamaño de genoma en relación con el porcentaje GC por cada organismo (arriba) y respecto a la forma de vida (abajo). Datos para análisis gráfico tomados de (Pruitt et al. 2002; Tatusova et al. 2014) Apéndice B Tabla 6. 22
- Figura 8. Resultado de la búsqueda de transferencia horizontal en 'DarkHorse'. Arriba, Frecuencias 'LPI' ('lineage probability index') respecto a las coincidencias del análisis de transferencia horizontal en 'DarkHorse', el color rojo indica el número de coincidencias con alta posibilidad de transferencia horizontal, las de color azul se descartan o se analizan individualmente. Abajo, Histograma de las mayores coincidencias encontradas por organismo. 24
- Figura 9. Resumen gráfico de las alineaciones de la búsqueda en la plataforma web de BLAST, de las secuencias de nucleótidos de los genes *mmn:midl_00391* (A, D), *mmn:midl_00900* (B, E) y *mmn:midl_00973* (C) de "*Candidatus Midichloria mitochondrii*" contra la base de datos de las secuencias genómicas (SRX821226) <ABC> y las secuencias del transcriptoma (SRX821392) del intestino medio <DE> de *I. ricinus*..... 27
- Figura 10. Resultados de la búsqueda de homólogos contra de la base de datos 'UniProtKB' en la plataforma web de HMMER (<http://hmmer.org/>) de las secuencias de aminoácidos identificadas en los 'contigs' de *I. ricinus*. A) *mmn:midl_00391*; B) *mmn:midl_00900* y C) *mmn:midl_00973*. Muestra

parte de los resultados de las secuencias, las características, una gráfica de barras y un listado de los resultados más significativos de acuerdo al 'e-value'. 30

Figura 11. Árbol filogenético de los homólogos de la proteína codificada por el gen *midi_00900* FliX de "Candidatus Midichloria mitochondrii". Se muestran los valores de 'bootstrap' en los nodos y se colapsaron ramas con un corte de ≤ 50 32

Figura 12. Árbol filogenético de los homólogos de la proteína codificada por el gen *midi_00973* de "Candidatus Midichloria mitochondrii". Se muestran los valores de 'bootstrap' en los nodos y se colapsaron ramas con un corte de ≤ 50 33

Lista de tablas

Tabla 1. Descripción taxonómica de “Ca. M. mitochondrii”, tabla basada de Sassera, et al. (2006).	11
Tabla 2. Organismos utilizados para los análisis con los códigos de acceso de las bases de datos a las que pertenecen.....	13
Tabla 3. Posibles proteínas transferidas de “Ca. M. mitochondrii” a <i>I. ricinus</i>	27
Tabla 4. El total de rutas de Biosíntesis y Generación de Metabolitos Precursores y Energía recuperadas de la base de datos de ‘BioCyc’.....	41
Tabla 5. El total de rutas de degradación recuperadas de las categorías de la base de datos de ‘BioCyc’.	42
Tabla 6. El resumen de las características de cada uno de los genomas utilizados.	43
Tabla 7. Resumen de resultados de la búsqueda de transferencia horizontal ‘DarkHorse’.....	44
Tabla 8. Resultado BLAST de la búsqueda de la secuencia mmn:mid_00391 contra la anotación del genoma de <i>I. ricinus</i>	46
Tabla 9. Resultado BLAST de la búsqueda de la secuencia mmn:mid_00900 contra la anotación del genoma de <i>I. ricinus</i>	48
Tabla 10. Resultado BLAST de la búsqueda de la secuencia mmn:mid_00973 contra la anotación del genoma de <i>I. ricinus</i>	51
Tabla 11. Resultado BLAST de la búsqueda de la secuencia mmn:mid_00391 en la anotación del transcriptoma del intestino medio de <i>I. ricinus</i>	55
Tabla 12. Resultado BLAST de la búsqueda de la secuencia mmn:mid_00900 en la anotación del transcriptoma del intestino medio de <i>I. ricinus</i>	55

RESUMEN

En este proyecto se estudió a “*Candidatus Midichloria mitochondrii*”, un organismo que pertenece a la clase Alphaproteobacteria, muy cercana a la familia Rickettsiaceae, organismos relacionados con el ancestro de la mitocondria. Se caracterizan por ser bacterias simbiotes que complementan la dieta de sus hospederos, muchas de ellas son simbiotes de hospederos causantes de enfermedades, tienen genomas reducidos y muestran evidencia de transferencia horizontal de genes, particularmente la transferencia de genes por endosimbiosis. “*Ca. M. mitochondrii*” es la primera bacteria intramitocondrial descrita, tiene un genoma reducido, producto de evolución reductiva, presenta una disminución de sus rutas metabólicas. “*Ca. M. mitochondrii* es endosimbionte de *Ixodes ricinus*, una garrapata de importancia médica por ser vector de la enfermedad de Lyme. Todas estas características la hacen un modelo biológico importante para poder afrontar varias hipótesis sobre evolución, ciencias genómicas, salud y genética bacteriana. Sin embargo, es una bacteria no cultivable, esto dificulta su análisis experimental, a pesar de esto, se puede analizar por métodos *in silico*, por medio de la comparación, modelaje y predicción a partir de su genoma, mediante algoritmos computacionales que permitan resolver hipótesis sobre las características biológicas de esta bacteria. En este estudio se compararon las características metabólicas y genómicas de “*Ca. M. mitochondrii*” con especies cercanas filogenéticamente y especies de estilos de vida diferentes, se reafirma que su capacidad metabólica y características genómicas son producto del tipo de vida endosimbionte. Se encontraron indicios de transferencia horizontal de genes, bacteria-eucarionte, al comparar el genoma de *I. ricinus* con el genoma de “*Ca. M. mitochondrii*”, se descubrieron secuencias de genes de origen bacteriano en el genoma parcial de *I. ricinus*.

INTRODUCCIÓN

Evolución y simbiogénesis

Hace aproximadamente 50 años que se descubrió 'DNA' dentro de cloroplastos y mitocondrias. Esto fue realizado por Hans Ris, que retomó la teoría del científico ruso Konstantin Mereschkowsky, donde explicaba que los cloroplastos se originaron a partir de la fusión de una cianobacteria y un protozoo. Mereschkowsky inventó el término simbiogénesis para referirse a fusiones entre organismos filogenéticamente distintos, en un solo organismo. Su idea se sumó a la de Ivan Wallin, que explicó el origen de la mitocondria a partir de un origen bacteriano, ambas ideas fueron tomadas con relevancia científica, por lo tanto se tiene la idea que la simbiogénesis tiene un papel clave en la evolución (Margulis 1990; Cavalier-Smith 2003).

En la naturaleza están presentes asociaciones simbióticas entre distintos organismos, que en su conjunto forman consorcios, esto implica un nuevo conocimiento del mundo natural, con la simbiosis como clave para la adquisición de genomas y capacidades metabólicas, esto impulsa la evolución y la generación de biodiversidad. Entonces "Simbiosis es simplemente la convivencia de organismos que son diferentes entre sí" (Margulis y Sagan 2002) y la simbiogénesis puede interpretarse como "Origen de una novedad evolutiva a través de simbiosis" (Margulis 1990), (Carrapiço 2015).

El término simbiosis fue acuñado por Heinrich Anton de Bary para describir la convivencia de "organismos de distinto nombre". La simbiosis es una asociación física a largo plazo de dos o más organismos. Simbiogénesis se refiere a la aparición de nuevas morfologías, tejidos, vías metabólicas, comportamientos, u otras novedades evolutivas reconocibles en holobiontes. La simbiogénesis es el resultado de la coexistencia permanente de diversos biontes para formar el holobionte, es decir, el organismo hospedero y su microbioma (Guerrero et al. 2013). El hologenoma es la suma total de los genomas de los componentes en un organismo eucariota; que comprende el genoma de un miembro individual de un

determinado taxón (el genoma del huésped) y todos los genomas que componen el microbioma (bacterias, arqueas, hongos y/o virus) (Rosenberg et al. 2009; Cerqueda-García y Falcón 2016).

Conforme a la localización de los organismos en la relación simbiótica, ésta puede ser ectosimbiosis o endosimbiosis. Entre las simbiosis intracelulares hay diferencias que dependen del tiempo de la asociación entre el huésped y el simbiote, por lo que existen endosimbiontes obligados primarios y simbiotes facultativos secundarios, los endosimbiontes primarios generalmente han tenido una larga historia evolutiva junto a sus hospederos, mientras que los endosimbiontes secundarios han establecido su relación de forma más reciente, por lo que estos mantienen su capacidad de regresar a la condición de vida libre (Moya et al. 2008).

La mayoría de las asociaciones simbióticas tienen como base el intercambio bioquímico (Moran et al. 2008). Las asociaciones de insectos con bacterias son de las más estudiadas y el establecimiento de esta asociación simbiótica se calcula que surgió hace ~300 Ma., estas asociaciones permiten visualizar historias evolutivas y las adaptaciones específicas de bacterias hacia la simbiosis (Gil et al. 2004; Moya et al. 2008).

Evolución reductiva del genoma

Muchas especies bacterianas intracelulares ya sean patógenas o mutualistas, tienen genomas reducidos comparados con especies de vida libre con las que están filogenéticamente relacionadas. Esta reducción se relaciona con la pérdida de un gran número de genes como una adaptación a las nuevas condiciones, donde muchas moléculas pueden ser obtenidas del hospedero sin requerir ser sintetizadas. La mayoría de los genomas con tamaños de ~500 kb se han encontrado en simbiotes obligados de varios insectos que han desarrollado asociaciones con diversos linajes bacterianos, principalmente con Proteobacteria y Bacteroidetes; estas bacterias proveen nutrimentos a sus hospederos a pesar de tener genomas reducidos (Moran y Bennett 2014).

Varios estudios han corroborado que la capacidad metabólica entre los miembros del holobionte, principalmente en insectos - consorcios bacterianos, se ve complementada. Por ejemplo, hospederos fitófagos tienen una dieta rica en azúcares, compuestos inorgánicos y minerales, pero deficiente o pobre en compuestos nitrogenados como aminoácidos, vitaminas y/o cofactores. Otros hospederos de dietas basadas en hematofagia y omnivoría presentan deficiencia en vitaminas, cofactores, grupos prostéticos y acarreadores. Estos compuestos son suministrados por los endosimbiontes; a cambio los endosimbiontes reciben compuestos que no pueden sintetizar debido a la reducción de sus capacidades metabólicas, por ejemplo compuestos orgánicos sulfurados, compuestos intermediarios de rutas metabólicas centrales, aminoácidos no esenciales, compuestos para la síntesis de envoltura celular, entre otros (Gosalbes et al. 2010).

Las posibles causas de la reducción del genoma de las bacterias endosimbióticas son: la transición entre el estilo de vida simbiote facultativa a endosimbiote obligada, la ausencia de recombinación, mayor tasa de sustitución de nucleótidos, alto contenido de A+T, acumulación de mutaciones deletéreas por deriva génica, la preferencia codónica hacia A o T y la rápida evolución en sus secuencias (Moran 2002; Gil et al. 2004; Moran et al. 2008). Esto se debe al cambio de un ambiente de vida libre a uno intracelular y protegido, lo que implica que muchos genes se vuelven innecesarios y otros se vuelven redundantes, debido a que sus funciones pueden ser suministradas por el hospedero (Moya et al. 2008).

Durante el proceso, el endosimbiote es obligado a depender irreversiblemente de su huésped para la supervivencia, debido a que se han perdido muchos genes (Gil et al. 2004). Incluso genes aparentemente fundamentales (reparación de ADN, regulación de la transcripción y los mecanismos de replicación de ADN). Por lo que la reducción del genoma se dirige específicamente a los genes de menor importancia selectiva o menos esenciales para la supervivencia de la asociación hospedero/simbiote (Delmotte et al. 2006).

Hay dos momentos importantes en la reducción del genoma, el primero cuando el genoma se erosiona y pierde genes de forma masiva, debido a la restricción del nuevo hábitat, es decir cuando se establece la simbiosis. El segundo momento es donde la reducción del genoma es más gradual y ya se ha establecido un intercambio metabólico complementario entre el simbiote y el hospedero (Moya et al. 2008; Moran y Bennett 2014).

Transferencia horizontal de genes

La transferencia horizontal de genes ('HGT'-*Horizontal gene transfer*) es la compartición de material genético entre organismos que no tienen una relación de descendencia directa. Es ampliamente reconocida como un mecanismo para la adaptación de bacterias y arqueas; se da con mayor frecuencia entre especies que están relacionadas con características genómicas similares. Los tres mecanismos más reconocidos para la transferencia horizontal de genes en bacterias y arqueas son la conjugación, transformación y la transducción (Figura 1).

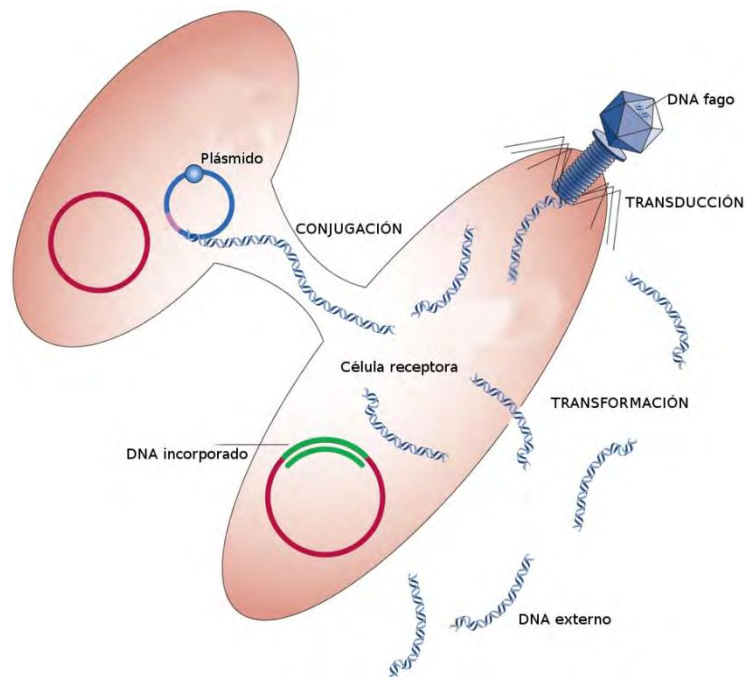


Figura 1. Representación de las diferentes formas de transferencia horizontal en bacterias y arqueas. En la conjugación se transfiere material genético entre dos bacterias, en este caso por medio de un plásmido. En la transducción una partícula viral (fago) incorpora DNA a la célula receptora. En la transformación la célula receptora incorpora DNA externo. Modificado de Stewart (2013).

Se sabe que organismos eucariontes pueden evolucionar a través de la adquisición de genes por parte de sus microbiomas, en este caso hay transferencia de genes endosimbióticos. Este tipo de transferencia genética se da cuando el material genético del endosimbionte u organelo se incorpora al genoma del hospedero (Figura 2). La transferencia horizontal de genes permite la explotación de nuevos nichos, lo que permite la explotación de recursos sin competencia alguna (Soucy et al. 2015).

La transferencia horizontal de genes ('HGT') de bacteria a eucarionte se divide en dos posibilidades; las originadas por bacterias de vida libre y las de bacterias endosimbiontes. La 'HGT'-endosimbiontes a eucariontes son probablemente más frecuentes, debido a la proximidad entre las células de ambos organismos. Estas transferencias podrían ser aún más comunes para endosimbiontes que viven en células germinales susceptibles a pasar a las siguientes generaciones. Las mitocondrias y los cloroplastos son ejemplos de estos escenarios, la mayoría de los casos de 'HGT' endosimbionte a eucariontes se han descritos en bacterias del género *Wolbachia* (Dunning Hotopp 2011).

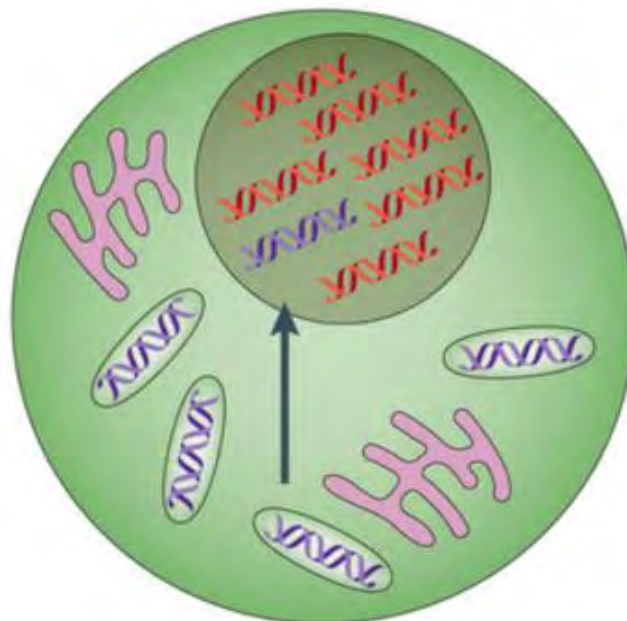


Figura 2. Representación de transferencia intracelular o endosimbiótica de genes donde los organismos endosimbiontes transfieren material genético al genoma del organismo hospedero debido a la estrecha cercanía de ambas células, modificado de Soucy et al., (2015).

Existen mecanismos como T4SS o sistema de secreción tipo IV, que son cruciales para la transferencia horizontal de genes (transformación y conjugación), donde las células son capaces de transportar diversos sustratos macromoleculares, proteínas, factores de virulencia y secuencias de DNA, por lo general están asociados a bacterias patógenas simbioses. Sin embargo, se desconoce el mecanismo por el cual ocurre la transferencia horizontal de bacterias a eucariontes (Le et al. 2014; Juhas 2015; Soucy et al. 2015).

La identificación de la transferencia horizontal de genes se basa en la composición de sus secuencias o de su historia evolutiva. Los métodos basados en la composición de la secuencia ("paramétricos") buscan diferencias en el promedio genómico (GC%, uso de codones, 'ORF'), mientras que los enfoques evolutivos basados en la historia evolutiva ("filogenéticos") identifican genes cuya historia evolutiva difiere significativamente del genoma que los contiene (Ravenhall et al. 2015).

El análisis de transferencia horizontal por medio de evidencia filogenética es poco práctico debido a la cantidad de información genómica a analizar y al tiempo de consumo de cómputo (Podell et al. 2008). Actualmente existe una herramienta bioinformática capaz de identificar sistemáticamente todos los genes adquiridos lateralmente (THG) en un genoma completo ('DarkHorse', Podell y Gaasterland 2007), la cual utiliza un enfoque heurístico, es decir encontrar la solución a un problema por medio de 'atajos'. Una definición formal sería "Un heurístico, es un algoritmo que es capaz de producir una solución aceptable a un problema en muchos escenarios prácticos, pero para los que no hay una prueba formal que sea la correcta" (Simpson 2009).

El enfoque heurístico es aplicable para reducir la complejidad de un problema, cuando esta complejidad rebasa la capacidad computacional para resolverlo, en este caso identificar los eventos de transferencia horizontal de todo un genoma.

Existe un método integrado en el algoritmo del programa 'DarkHorse' que es rápido, cuantitativo, reproducible y de baja demanda de cómputo. El método identifica posibles proteínas candidatas a transferencia horizontal en el genoma completo de un organismo; se ha demostrado en ejemplos de bacterias, arqueas y eucariontes. Los resultados pueden utilizarse para caracterizar el perfil histórico de actividad de transferencia horizontal de un organismo, la cobertura de la base de datos para especies cercanas filogenéticamente y las proteínas menos probables de haber sido heredadas verticalmente. El algoritmo 'DarkHorse' combina probabilidad, selección ponderada de linajes y un filtro de búsqueda que ajusta el número total de posibles candidatos a transferencia horizontal comparando las secuencias del genoma contra la base de datos que se utilice (Podell y Gaasterland 2007).

El algoritmo selecciona coincidencias homólogas y recupera la información taxonómica de la base de datos donde compara las secuencias, para formar el índice de probabilidad de linaje 'LPI'. El índice es un métrico que indica la posibilidad de transferencia horizontal, toma valores entre 0 y 1, clasifica las secuencias de aminoácidos coincidentes en la base de datos que pertenecen al genoma de búsqueda. Los valores 'LPI' son inversamente proporcionales a la distancia filogenética entre las secuencias que coinciden entre el genoma y la base de datos. El índice únicamente asigna valores de probabilidad donde se encuentran las secuencias coincidentes, los valores más altos se dan a los linajes donde las coincidencias entre secuencias ocurren con mayor frecuencia y valores más bajos a linajes que ocurren rara vez. La figura 3 muestra el esquema general del proceso del algoritmo.

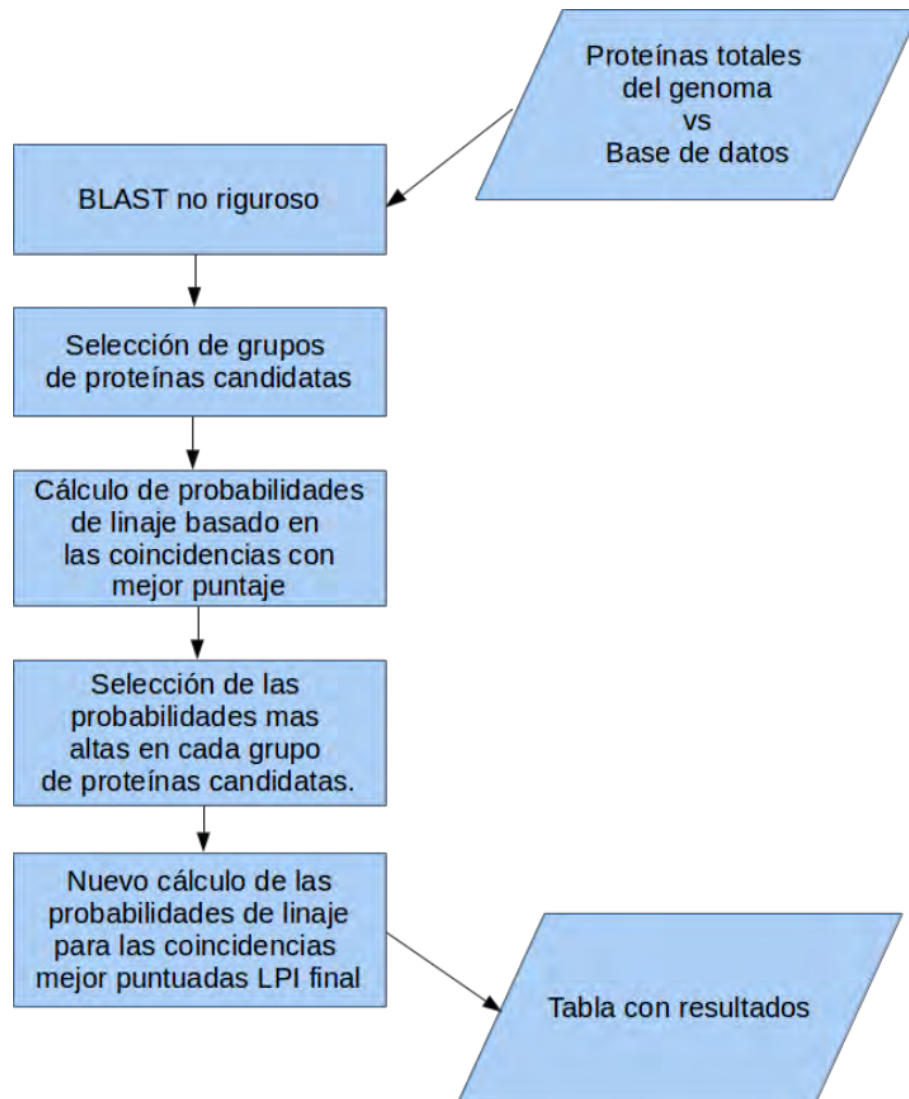


Figura 3. El esquema general del proceso del algoritmo de 'DarkHorse'. Comienza con una búsqueda no rigurosa (BLAST) de las secuencias de proteínas del genoma contra una base de datos, después se forman grupos de las secuencias de proteínas coincidentes en linajes distintos, se hace un cálculo de probabilidad de linaje, usando información taxonómica de las secuencias, basado en los mejores valores de los grupos formados de las secuencias de proteínas se eligen las mejores probabilidades y se calcula de nueva cuenta la probabilidad de linaje 'LPI' para formar una tabla con los resultados finales.

“*Candidatus* Midichloria mitochondrii”.

En este proyecto estudiamos a la bacteria “*Ca. M. mitochondrii*” como modelo endosimbionte. Esta bacteria es un bacilo Gram-negativo perteneciente al phylum Proteobacteria, clase Alphaproteobacteria, dentro del orden de las Rickettsiales. Tiene la característica única de ser endosimbionte de las mitocondrias del ovario de la garrapata *Ixodes ricinus* (Tabla 1, Figura 4) (Sassera et al. 2006).

I. ricinus es una especie de garrapata de importancia médica en Europa, debido a que es un vector de organismos patógenos. Además de estos patógenos, *I. ricinus* alberga a “*Ca. M. mitochondrii*” como la bacteria dominante en simbiosis, sin embargo aún se desconoce su papel biológico. La mayoría de las células de “*Ca. M. mitochondrii*” se localizan en los ovarios de la garrapata. Se calcula que hay alrededor de 15 millones de bacterias en un solo ejemplar de *I. ricinus*, principalmente en los ovarios y primordios, sin embargo también se han detectado en las glándulas salivales, en la saliva de la garrapata y en la sangre de vertebrados hospederos de *I. ricinus* (Di Venere et al. 2015). Aunque “*Ca. M. mitochondrii*” se descubrió primero en *I. ricinus*, también se ha detectado en otras garrapatas como *Ixodes holocyclus*, también en tabánidos (Tabanidae), chinches (Cimicidae) y ácaros (Acari) (Gofton et al. 2015).

El genoma de “*Ca. M. mitochondrii*” está secuenciado y consiste en un cromosoma circular de ~ 1.18Mb con un contenido de G+C de 36.6%. La mayoría del contenido del genoma es similar a miembros del grupo de las Rickettsiales, caracterizados por la escasez de genes que codifican las rutas de biosíntesis de aminoácidos y nucleótidos. Aunque “*Ca. M. mitochondrii*” tiene disminuida su capacidad de biosíntesis, posee genes para la producción de varios cofactores (coenzima A, biotina, ácido lipóico, tetrahidrofolato, pantotenato, heme y ubiquinona) que posiblemente suministra a las células de su hospedero (Sassera et al. 2011).

Tabla 1. Descripción taxonómica de “*Ca. M. mitochondrii*”, tabla basada de Sasser, et al. (2006)

Descripción	Característica
<i>Candidatus</i>	De acuerdo a los lineamientos del ‘ICSB’ las bacterias no cultivables se clasifican como <i>Candidatus</i>
Midichloria	mi.di.chlo’ria. N.L. fem. n. Nombre derivado de los ‘midiclorianos’, son entes ficticios de la saga ‘Star Wars’. Los ‘midiclorianos’ son endosimbiontes celulares que pueden comunicarse con ‘La Fuerza’.
<i>mitochondrii</i>	mi.to’chon.drii N.L. n. mitochondrium, <i>mitochondria</i> ; N.L. gen. n. mitochondrii, <i>de una mitochondria</i> Epíteto específico que se refiere al estilo intramitocondrial único de esta bacteria.

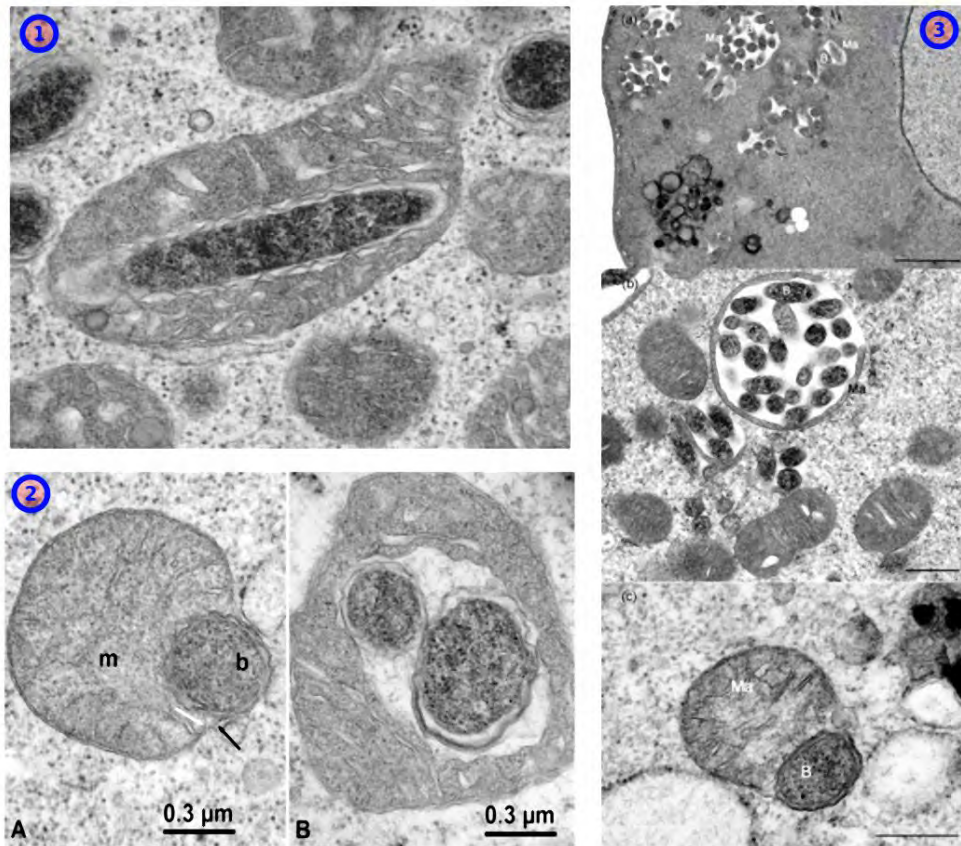


Figura 4. 1) Mitocondria de un oocito de *I. ricinus* “infectada” por “*Ca. M. mitochondrii*” modificada de Pistone et al., (2012). 2) Imagen de microscopia electrónica de transmisión que muestra en (A) inicio de la infección de la bacteria [b] entre las membranas de la mitocondria [m] y (B) dos bacterias dentro de la mitocondria modificada de Sasser et al., (2011). 3) Micrografías electrónicas de células de ovario de *I. ricinus* con “*Ca. M. mitochondrii*” modificada de Sasser et al., (2006).

HIPÓTESIS

El modo de vida endosimbionte tiene como consecuencia la pérdida de genes por evolución reductiva y la transferencia horizontal de genes, si analizamos el genoma de "*Candidatus* Midichloria mitochondrii", habrá pérdida de funciones metabólicas y genes heterólogos.

OBJETIVOS

- -Comparar las capacidades metabólicas de "*Candidatus* Midichloria mitochondrii" con especies relacionadas filogenéticamente, simbioses y de vida libre.
- -Buscar transferencia horizontal en "*Candidatus* Midichloria mitochondrii".

MATERIALES Y MÉTODOS

Genomas utilizados

Tabla 2. Organismos utilizados para los análisis con los códigos de acceso de las bases de datos a las que pertenecen.

Organismo	PGDB BioCyc	RefSeq
<i>Agrobacterium</i> sp. H13-3	GCF_000192635	pAspH13-3 NC_015184.1 Chr linear NC_015508.1 Chr circular NC_015183.1
<i>Anaplasma phagocytophilum</i> HZ	APHA212042	NC_007797.1
<i>Bacillus subtilis subtilis</i> 168	BSUB	NC_000964.3
<i>Bartonella henselae</i> Houston-1	BHEN283166	NC_005956.1
<i>Brucella abortus</i> A13334	BABO1104320	Chr1 NC_016795.1 Chr2 NC_016777.1
' <i>Candidatus Carsonella ruddii</i> ' HT isolate Thao2000	CCAR1202539	NC_018417.1
" <i>Ca. M. mitochondrii</i> " IricVA	CMID696127	NC_015722.1
' <i>Candidatus Nasuia deltocephalinicola</i> ' NAS-ALF	CNAS1343077	NC_021919.1
' <i>Candidatus Pelagibacter ubique</i> ' HTCC1062	CPEL335992	NC_007205.1
<i>Ehrlichia</i> sp. HF	GCF_000632845	NZ_CP007474.1
<i>Escherichia coli</i> K-12 substr. MG1655	ECOLI	NZ_CP012868.1
<i>Orientia tsutsugamushi</i> Boryong	OTSU357244	NC_009488.1
<i>Rickettsia prowazekii</i> Madrid E	RPRO272947	NC_000963.1
<i>Sinorhizobium meliloti</i> 1021	SMEL266834	Chr NC_003047.1 pSymA NC_003037.1 pSymB NC_003078.1
<i>Wolbachia</i> sp. wRi	WSP66084	NC_012416.1
endosymbiont of <i>Acanthamoeba</i> sp. UWC8	GCF_000730245	NZ_CP004403.1
Genoma		GenBank
<i>Ixodes ricinus</i> strain Charles River, whole genome shotgun sequencing project		JXMZ00000000.2
" <i>Ca. M. mitochondrii</i> " IricVA, complete genome		CP002130.1
Estudio		SRA
midgut transcriptome shotgun sequencing from unfed <i>Ixodes ricinus</i> females		SRP051469, SRX821392 35,343,347
<i>Ixodes ricinus</i> strain: Charles River Genome sequencing		SRP051465, SRX821226 998,100,906

Comparación de capacidades metabólicas y genomas

Se compararon las capacidades metabólicas de los organismos listados en la Tabla 2. Para cada organismo se obtuvo el total de las rutas metabólicas del metabolismo central en la base de datos 'BioCyc Database Collection' (<http://biocyc.org/comp-genomics>) (Caspi et al. 2016). El total de cada una de las categorías metabólicas que son representativas del metabolismo central se encuentran en el Apéndice A, Categorías metabólicas (Tabla 4 y Tabla 5).

La base de datos 'BioCyc' es generada a partir de 'Pathway Tools software', este contiene el programa 'PathoLogic' que predice e infiere las rutas metabólicas a partir de la anotación de un genoma, el programa evalúa las enzimas presentes en la anotación del genoma y las contrasta con la base de datos 'MetaCyc'. Esta base de datos contiene la anotación funcional de más de 450 rutas metabólicas de todos los dominios de la vida (Karp et al. 2002; Karp et al. 2016).

Las rutas metabólicas a comparar fueron elegidas en base a la clasificación jerárquica de las rutas metabólicas en la base de datos 'MetaCyc' (<http://metacyc.org/META/class-tree?object=Pathways>): Biosíntesis, Degradación / Utilización / Asimilación y Generación de Metabolitos Precursores y Energía. De estas jerarquías se eligieron, de forma manual, las rutas de los principales grupos de moléculas en el metabolismo: nucleótidos, aminoácidos, carbohidratos, ácidos grasos y lípidos, precursores de energía y cofactores (Caspi et al. 2016). Los resultados se graficaron en un 'heatmap' (Figura 5) con la librería gplots (Warnes et al. 2016) del entorno del lenguaje R (R Core Team 2016).

De la base de datos de secuencias de referencias (RefSeq) (<http://www.ncbi.nlm.nih.gov/refseq>) (Pruitt et al. 2002; Tatusova et al. 2014) se extrajo la información de los organismos referente a: tamaño del genoma, % GC, número de genes y proteínas codificantes (Apéndice B. Resumen de genomas. Tabla 6).

Los datos se graficaron en R, se compararon estas características entre todos los organismos y su forma de vida por medio del paquete 'ggplot2' (Wickham 2009). La forma de vida de cada organismo fue tomada de la literatura disponible (Gosalbes et al. 2010; Martínez-Cano et al. 2014; Moran y Bennett 2014; Cramaro et al. 2015; Di Venere et al. 2015).

Análisis de transferencia horizontal de genes

Se hizo el análisis de transferencia horizontal de genes al genoma de “*Ca. M. mitochondrii*” en el programa DarkHorse-1.5_rev170 (Podell y Gaasterland 2007; Podell et al. 2008) con los siguientes parámetros: corte de 0.3, la exclusión de los términos: clone, cloning, construct, contaminant, cosmid, expression, synthetic, unidentified, unknown, unspecified, untyped, vector.

El resultado fue una tabla que resume las proteínas homólogas que se encuentran en otros organismos con sus respectivos estadísticos: longitud de alineamiento, porcentaje de identidad, ‘E-value’, ‘bitscore’ y el porcentaje de cobertura del alineamiento; la clasificación de coincidencias, las frecuencias de probabilidad y la clasificación por géneros y especies de las posibles especies candidatas de transferencia horizontal (Apéndice C. Resultados ‘DarkHorse’.Tabla 7).

De estos resultados se realizó un histograma con las especies de mayor coincidencia encontrada en el análisis de transferencia y una gráfica con las frecuencias de probabilidad (Figura 6) se usó el paquete 'ggplot2' (Wickham 2009).

El genoma de *I. ricinus* fue parcialmente secuenciado recientemente, junto con un análisis transcriptómico (Cramaro et al. 2015), los ‘contigs’ se encuentran depositados en DDBJ / EMBL / GenBank (código de acceso: JXMZ00000000.2, 204,516 contigs, 514,506,549 bp).

Se descargaron los ‘contigs’ y se buscaron los marcos de lectura (‘orf’) con el programa ‘Exonerate’ (<http://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate>) (Slater y Birney 2005). Se hizo una búsqueda de homología localmente (en un servidor GNU/Linux con 32 núcleos y 120 gb de memoria) de los marcos de

lectura abiertos ('Open Reading Frames'-'ORF') con el programa 'GHOSTX' (Suzuki et al. 2014) contra la base de datos no redundante (nr) de BLAST (Altschul et al. 1990; Madden 2003). Las proteínas de "Ca. M. mitochondrii" que dieron positivo en los 'orfs' de *I. ricinus* fueron identificadas y se hizo una búsqueda de homología en la plataforma web de BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) contra la anotación del genoma y los transcritos de *I. ricinus* (SRX821226, SRX821392).

Análisis filogenético

Los genes potencialmente transferidos de "Ca. M. mitochondrii" a *I. ricinus* se analizaron con el programa 'HMMER' en su plataforma web (<http://www.ebi.ac.uk/Tools/hmmer/>) (Finn et al. 2015) contra la base de datos de 'UniProt' (<http://www.uniprot.org/>) (UniProt Consortium 2015) para buscar homólogos.

Las secuencias peptídicas de los homólogos identificados se descargaron, y se alinearon con el programa 'MUSCLE' (<http://www.ebi.ac.uk/Tools/msa/muscle/>) (Edgar 2004; McWilliam et al. 2013). El análisis filogenético se hizo en el programa 'PhyML' (filogenia por máxima verosimilitud) (Guindon et al. 2010), con 1000 'bootstraps' (con los modelos de sustitución generados en ProtTest3 (Darriba et al. 2011) LG+I+G+F y LG+I+G basados en el criterio de información de Akaike 'AIC'). Los árboles filogenéticos se dibujaron con el programa 'FigTree' (<http://tree.bio.ed.ac.uk/software/figtree>), con los parámetros 'midpoint root', nodos de forma creciente y las ramas transformadas en cladograma (Figura 11 y 12).

RESULTADOS Y DISCUSIÓN

Capacidades metabólicas y genomas

La mayor cantidad de rutas en todos los organismos corresponde a la biosíntesis de cofactores, grupos prostéticos y acarreadores de electrones; la categoría degradación de ácidos grasos y lípidos tiene el menor número de rutas. La Figura 5 muestra la agrupación de los organismos en función de la cantidad de rutas metabólicas. Se puede apreciar un incremento en la cantidad de rutas en función del tamaño del genoma de los organismos, '*Candidatus Nasuia deltocephalinicola*' es el organismo con el tamaño del genoma más pequeño y la menor cantidad de rutas metabólicas en contraste con *Escherichia coli* K12 con el mayor número de rutas metabólicas (Apéndice A. Categorías metabólicas. Tabla 4 y 5. Apéndice B. Resumen de genomas. Tabla 6).

Se visualizan dos grupos principales (Figura 5): el primero de '*Candidatus Nasuia deltocephalinicola*' hasta *Bartonella henselae* y el segundo de '*Candidatus Pelagibacter ubique*' hasta *Escherichia coli* K12. El primer grupo posee pocas rutas metabólicas de degradación y biosíntesis de metabolitos no esenciales y corresponde a los organismos de forma de vida obligada intracelular y facultativa. Probablemente estos organismos han perdido la capacidad metabólica de degradar compuestos no esenciales y solo poseen rutas metabólicas de biosíntesis de compuestos indispensables para sus funciones celulares (Batut et al., 2014). Las probables causas: una estrecha relación con su hospedero en el intercambio metabólico y la reducción del nicho que favorece la reducción del genoma y por ende de las capacidades metabólicas (Dunning Hotopp et al. 2006; Moran et al. 2008; Sassera et al. 2011; Moran y Bennett 2014).

El segundo grupo '*Candidatus Pelagibacter ubique*' hasta *Escherichia coli* K12, representan a organismos que son de vida libre y patógenos, tienen mayor número de rutas en las categorías metabólicas. En general se ve la relación del tamaño del genoma respecto al número de rutas metabólicas.

“*Ca. M. mitochondrii*”, posee varias rutas de biosíntesis principalmente cofactores, grupos prostéticos y acarreadores de electrones, nucleótidos y nucleósidos, aminoácidos, y rutas generadoras de energía, dato que concuerda con lo analizado en Dunning Hotopp et al., 2006; (Sassera et al. 2011).

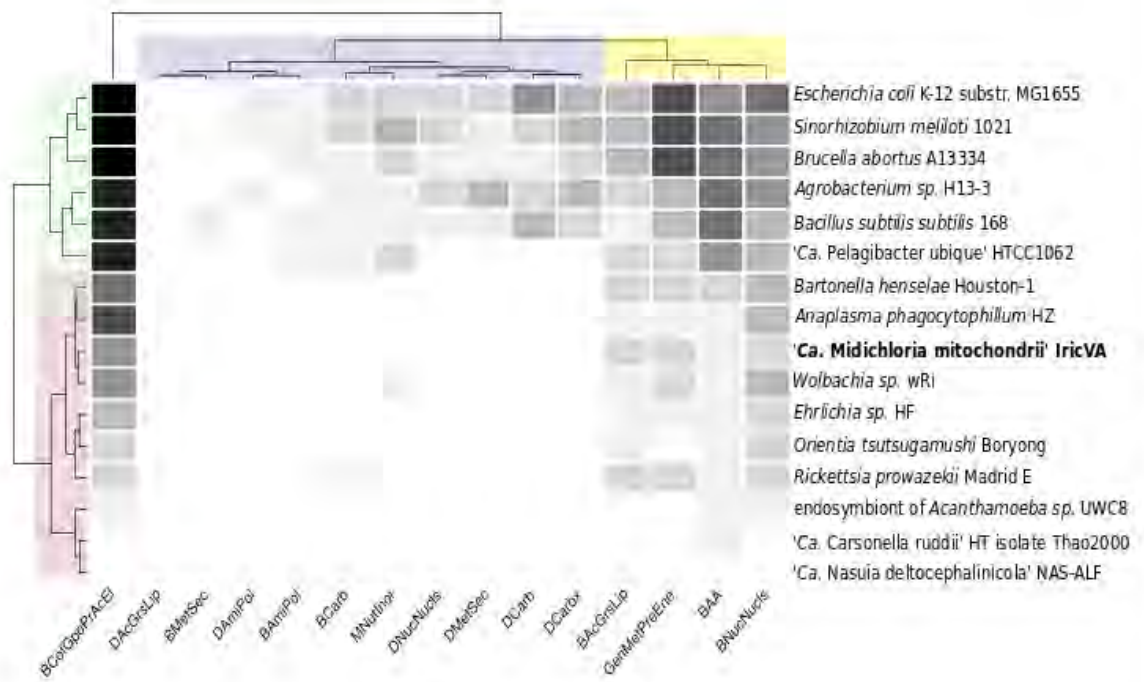


Figura 5. ‘Heatmap’ que muestra el número de rutas en distintas categorías metabólicas afines al metabolismo central de diferentes bacterias, la intensidad del tono, en las celdas del ‘heatmap’, está relacionada con la cantidad de rutas en cada categoría. Los grupos principales en los dendrogramas (izquierda y arriba) están resaltados. BcofGpoPrAcEl = Biosíntesis Cofactores Grupos Prostéticos y Acarreadores de Electrones, DAcGrsLip = Degradación Ácidos Grasos y Lípidos, BMetSec = Biosíntesis Metabolitos Secundarios, DAmiPol = Degradación Aminas y Poliaminas, BAmiPol = Biosíntesis Aminas y Poliaminas, BCarb = Biosíntesis Carbohidratos, MNutInor = Metabolismo Nutrientes Inorgánicos, DNucNucls = Degradación Nucleótidos Nucleósidos, DMetSec = Degradación Metalitos Secundarios, DCarb = Degradación Carbohidratos, DCarbx = Degradación Carboxilatos, BAcsGrsLip = Biosíntesis Ácidos Grasos Lípidos, GenMetPreEne = Generación Metabolitos Precursores Energía, BAA = Biosíntesis Aminoácidos, BNucNucls = Biosíntesis Nucleótidos Nucleósidos.

La Figura 6 muestra el tamaño del genoma de cada organismo como variable principal, contra la cantidad de genes y proteínas codificadas de cada organismo y también por forma de vida. Se visualiza que el tamaño genómico está relacionado directamente con el número de genes y proteínas en todas las bacterias analizadas.

El tamaño del genoma no se correlaciona con su complejidad. C.A. Thomas Jr. denominó como la paradoja C ('C-value paradox') a esta variación en la relación entre genes y tamaño genómico (Eddy 2012). En los genomas bacterianos hay distintos intervalos de tamaño genómico, ~0.12 Mb a ~13 Mb (Ussery et al. 2009). Generalmente los genomas de bacterias no varían en cuanto a la relación genoma/gen, es claro que el tamaño del genoma es directamente proporcional al número de genes, una tendencia que no aplica en genomas eucariontes (Bentley y Parkhill 2004).

En bacterias con genomas grandes tienden a aumentar los genes de regulación y metabolismo secundario, no acumulan DNA no codificante, porque la porción de genoma para estas funciones se mantiene constante. Los genomas reducidos de bacterias tienden a perder genes de biosíntesis de compuestos esenciales (Konstantinidis y Tiedje 2004).

Los genomas bacterianos más pequeños pertenecen a endosimbiontes de organismos eucariontes, estos están bajo un proceso degenerativo debido a una evolución reductiva. En general en bacterias existe una relación entre el tamaño del genoma, la cantidad de genes y proteínas, la cual posiblemente se deba al tipo de ambiente o nicho donde habitan y las presiones ambientales a las que están sometidas (Bentley y Parkhill 2004; Batut et al. 2014). Los organismos que tienen más genes y proteínas son organismos de vida libre y patógenos, estos probablemente habitan un amplio número de nichos, al contrario de los organismos obligados intracelulares y facultativos que se encuentran confinados a sus hospederos (Moran y Bennett 2014).

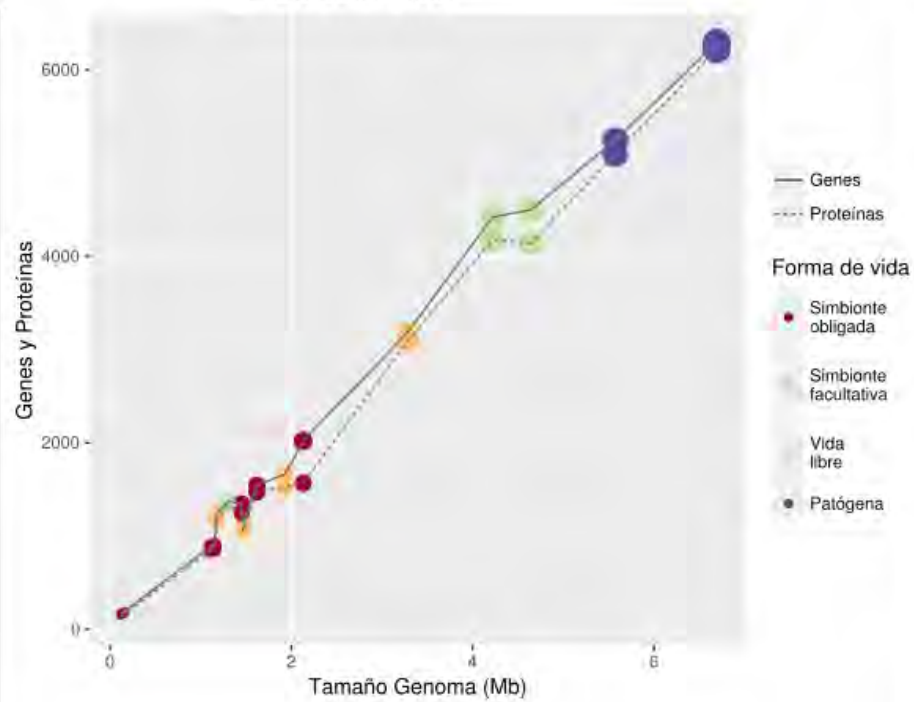
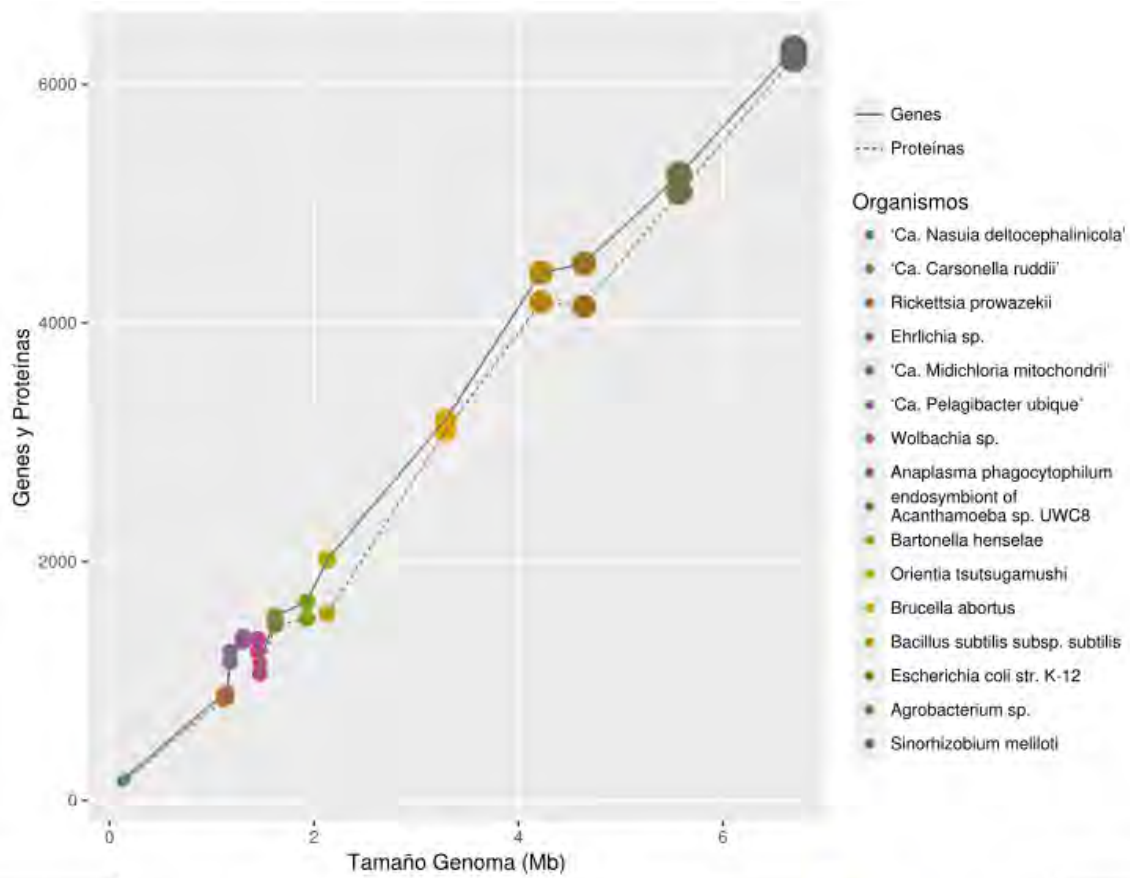


Figura 6. Gráficas comparativas entre la cantidad de genes/proteínas respecto al tamaño del genoma por cada organismo (arriba) y por forma de vida (abajo). Datos para análisis tomados de 'RefSeq' (Pruitt et al. 2002; Tatusova et al. 2014) Apéndice B Tabla 6.

La Figura 7 muestra la relación entre el tamaño del genoma y el contenido de GC, esta relación se ha observado en bacterias obligadas que tienen bajo porcentaje GC (Bentley y Parkhill 2004; McCutcheon et al. 2009). Una explicación a este fenómeno es que los nucleótidos GTP y CTP son más costosos energéticamente que ATP y UTP, un hábitat que limite solo la producción de ATP puede propiciar esta tendencia. Existe en “*Ca. M. mitochondrii*” y en varias Rickettsiales genes de translocasas ATP/ADP que podrían utilizar para importar ATP de *I. ricinus* y sus hospederos; con esto se podría pensar en un metabolismo restringido o de predilección de ATP en “*Ca. M. mitochondrii*” (Sassera et al. 2011; Le et al. 2014). Otra posibilidad es que los genes de reparación de ‘DNA’ en organismos con genomas reducidos están, en varios casos, ausentes y esto tiene un alto porcentaje AT en los genomas de estos organismos (Bentley y Parkhill 2004).

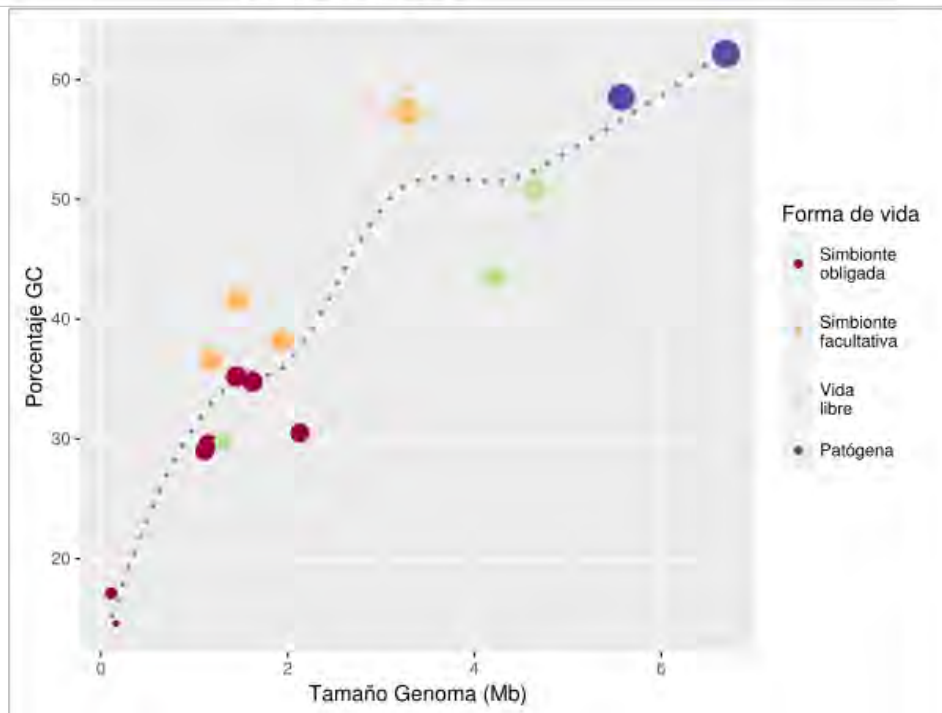
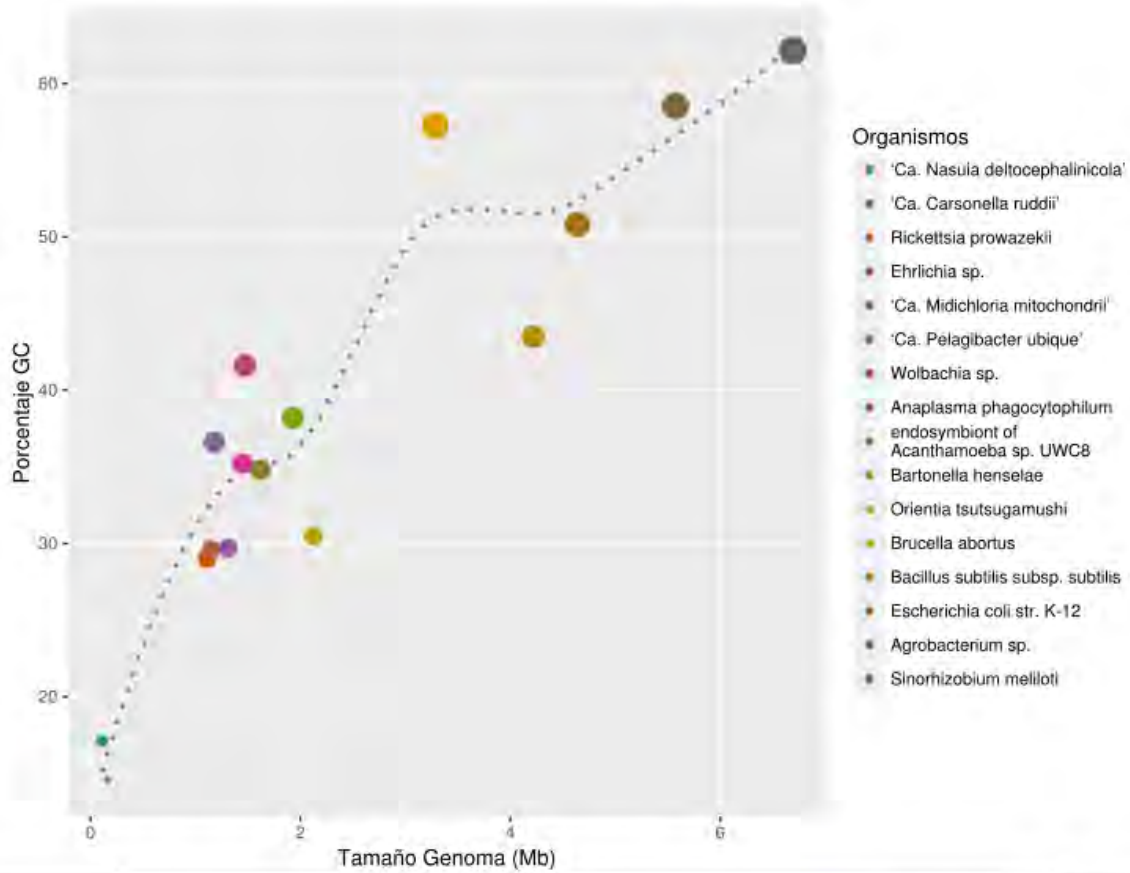


Figura 7. Gráficas comparativas del tamaño de genoma en relación con el porcentaje GC por cada organismo (arriba) y respecto a la forma de vida (abajo). Datos para análisis gráfico tomados de (Pruitt et al. 2002; Tatusova et al. 2014) Apéndice B Tabla 6.

Transferencia horizontal

El análisis de la transferencia horizontal en “*Ca. M. mitochondrii*” seleccionó potenciales proteínas ortólogas de organismos candidatos a transferencia horizontal por medio de la búsqueda de homología de sus secuencias peptídicas. El algoritmo del programa ‘DarkHorse’ usó estas coincidencias (‘hits’) de las secuencias para calcular el índice de probabilidad de linaje (‘lineage probability index’ LPI) para calificar cada proteína en el genoma. El ‘LPI’ es una medida que identifica posibles candidatos de transferencia horizontal en todo un genoma (Podell y Gaasterland 2007).

Fueron removidas 39 auto-coincidencias y 27230 coincidencias de otros linajes en la búsqueda. 1205/1210 proteínas consultadas no fueron auto-coincidencias. Se filtraron de 343074 ‘blast hits’ a 9775 ‘besthits’, ~35% de las coincidencias obtuvieron el mejor valor (score). El tamaño máximo de coincidencias candidatas fue de 480 ~0.14% del total de las base de datos donde se buscaron las secuencias totales del genoma. La proteína más conservada en la búsqueda fue gi:338226103 gb:AEI88987.1 “Dihydrolipoamide acetyltransferase” “*Ca. M. mitochondrii*”.

Solo se detectaron 43 secuencias de aminoácidos atípicas, ~0.012% del total de las secuencias en las bases de datos consultadas (Apéndice C. Resultados ‘DarkHorse’, Tabla 7) en donde el mayor número de posibles secuencias se localizaron en *Microvirga sp.* BSC39. La mayor cantidad de coincidencias se encuentran en ~0.75 ‘LPI’ con 25 coincidencias (Figura 8).

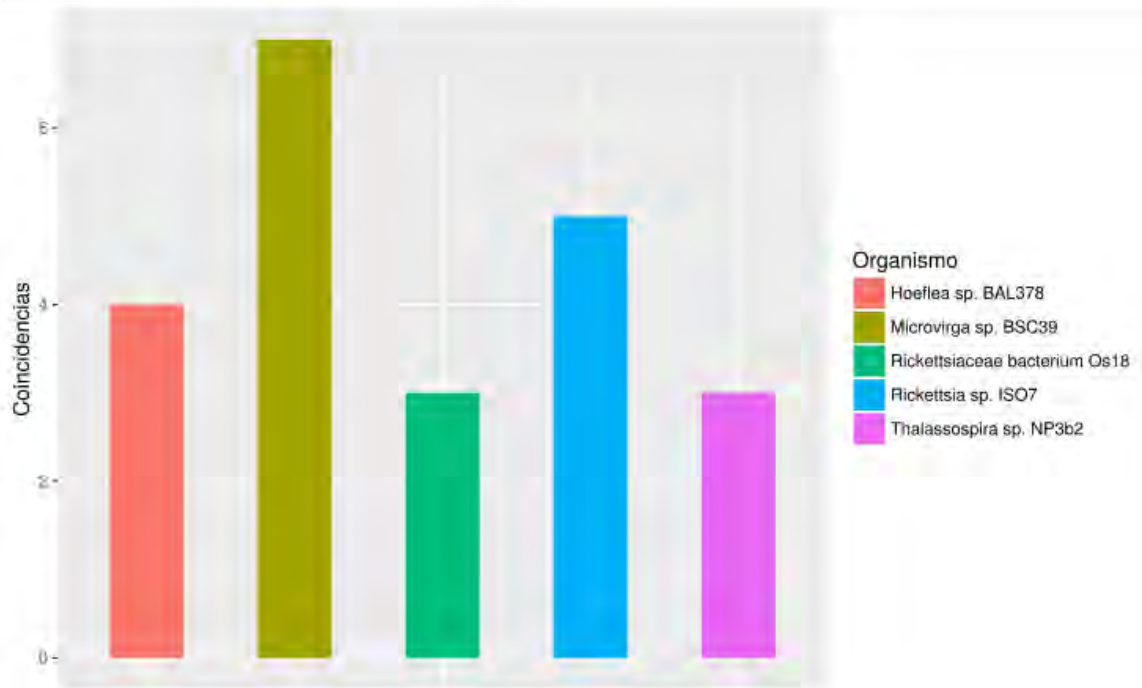
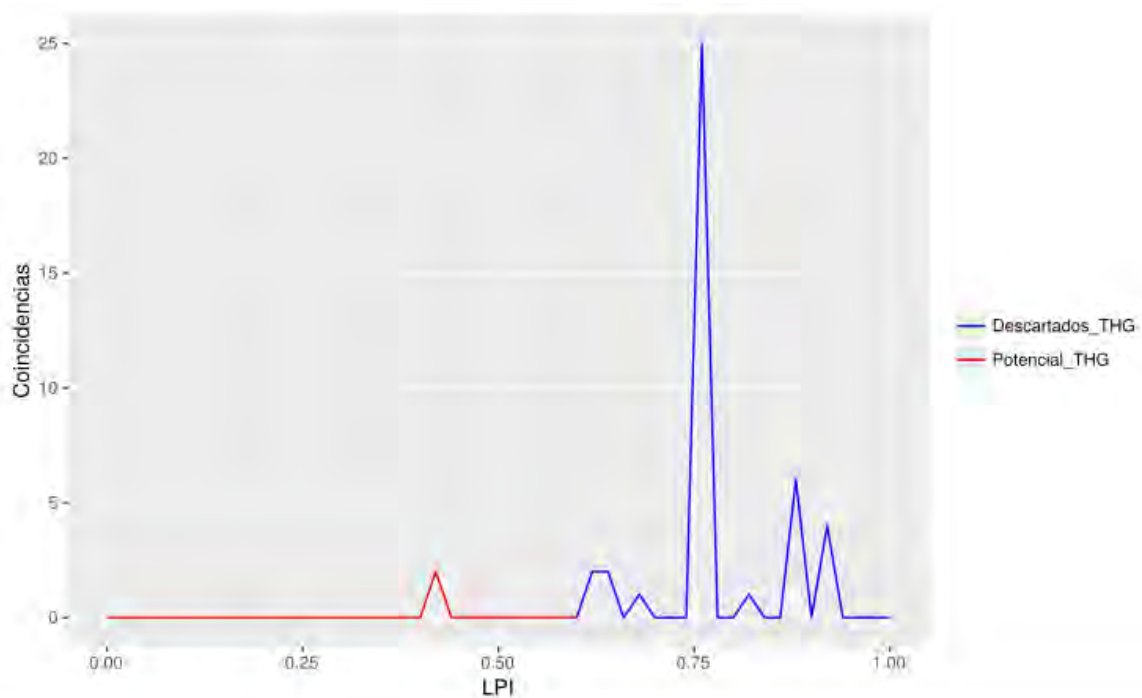


Figura 8. Resultado de la búsqueda de transferencia horizontal en 'DarkHorse'. Arriba, Frecuencias 'LPI' ('lineage probability index') respecto a las coincidencias del análisis de transferencia horizontal en 'DarkHorse', el color rojo indica el número de coincidencias con alta posibilidad de transferencia horizontal, las de color azul se descartan o se analizan individualmente. Abajo, Histograma de las mayores coincidencias encontradas por organismo.

La mayoría de las alphaproteobacterias viven en simpatría lo que favorece la transferencia horizontal. Existe evidencia que demuestra la transferencia entre miembros de la misma familia y otras familias cercanas, por ejemplo, se ha observado que *Rickettsia felis* ha adquirido genes de su propio plásmido y de otras especies de Rickettsiales: *R. prowazekii*, *R. typhi*, *R. canadensis*, *R. bellii*, *Orientia spp.* y *Wolbachia spp.*. Varios genomas de rickettsias (Alphaproteobacteria) han recibido genes de gammaproteobacterias, deltaproteobacterias, bacteroidetes y cianobacterias. Esta transferencia de genes se da al compartir un hospedero en común. También han compartido genes con eucariontes, varias proteínas mitocondriales de *Saccharomyces cerevisiae* parecen haber sido transferidas de *R. prowazekii*, *Rhodospirillum rubrum*, *Rhodopseudomonas palustris*, *Rhodobacter sphaeroides* y *Ochrobactrum anthropi* (Le et al. 2014).

“*Ca. M. mitochondrii*” se localiza en distintos tipos de hospederos por lo que comparte el hábitat con distintos tipos de bacterias (Ahantari et al. 2013). Esto daría la posibilidad de transferencia horizontal de genes con las especies que cohabita.

En el análisis con ‘DarkHorse’, el mayor número de coincidencias tuvieron un ‘LPI’ (‘lineage probability index’) de ~0.75 en promedio, la más baja de 0.409 y la mayor de 0.907, ver figura 8. Entonces, se descartan las coincidencias con altos valores de ‘LPI’ >0.6, porque se considera una mejor ponderación a transferencia horizontal en niveles taxonómicos iniciales de un linaje (reino, phylum, clase) que corresponde a los valores más bajos de ‘LPI’ y menor ponderación al final del linaje (familia, género, especie); los valores intermedios se podrían analizar de forma individual y los valores más bajos <0.6 se pueden considerar como los mejores candidatos de posible transferencia horizontal (Podell y Gaasterland 2007).

‘*Candidatus Amoebophilus asiaticus*’ es la especie con la coincidencia de ‘LPI’ más baja 0.415 (Apéndice C. Resultados ‘DarkHorse’.Tabla 7) y la posibilidad de que se encontrara como ejemplar candidato a transferencia quizá se deba a

que posee un porcentaje de secuencias de proteínas atípicas que comparte con bacterias intracelulares, entre ellas las de género *Rickettsia sp.* que está emparentada con “*Ca. M. mitochondrii*” (Schmitz-Esser et al. 2010).

La segunda especie con mayor coincidencia y bajo ‘LPI’ 0.415 es *Anoxybacillus sp.* KU2-6(11), para los fines de este análisis, esta posibilidad puede descartarse debido a que no comparte el mismo hábitat que “*Ca. M. mitochondrii*”; *Anoxybacillus sp.* KU2-6(11) habita principalmente fuentes geotérmicas. Sin embargo también existe la posibilidad que *Anoxybacillus sp.* KU2-6(11) pudiera incorporar DNA externo debido a que posee proteínas (‘competence protein’) que facilitan la incorporación de DNA externo (Goh et al. 2014).

Con este análisis, no se pudo determinar si existe algún intercambio de genes entre “*Ca. M. mitochondrii*” y alguna otra bacteria, ya que los candidatos posibles son filogenéticamente cercanos, por lo que sugerimos que no existe transferencia horizontal bacteria-bacteria en el genoma de “*Ca. M. mitochondrii*”. Esto puede deberse a la reducción de su genoma, donde posiblemente pierda genes en lugar de ganarlos, un fenómeno común en los endosimbiontes (Khachane et al. 2007).

Recientemente Cramaro y colaboradores (2015), secuenciaron parcialmente el genoma de *I. ricinus*, junto a un análisis transcriptómico y proteómico del intestino medio. Los datos de secuencias generados fueron liberados en la base de datos del GenBank. En el análisis de búsqueda de homología con el programa GHOSTX de los ‘orfs’ determinados en los contigs de *I. ricinus*, se identificaron tres ‘ORF’ que dieron positivo a secuencias de proteínas del genoma de “*Ca. M. mitochondrii*”, las que se muestran en la Tabla 3.

Tabla 3. Posibles proteínas transferidas de “*Ca. M. mitochondrii*” a *I. ricinus*.

‘Contigs’ <i>Ixodes ricinus</i>	gb	‘ORF’	Genes identificados “ <i>Candidatus</i> <i>Midichloria mitochondrii</i> ”
136930	JXMZ02161538	Número 2 al leer el marco 2 en la cadena inversa.	mmn:midi_00391 hypothetical protein (N)
8928	JXMZ02193468	Número 4 al leer el marco 3 en la cadena directa.	mmn:midi_00973 hypothetical protein; K09949 uncharacterized protein (N)
109948	JXMZ02149861	Número 5 al leer el marco 2 en la cadena directa.	mmn:midi_00900 hypothetical protein (N): Flagellar_assmbl_FliX_class2

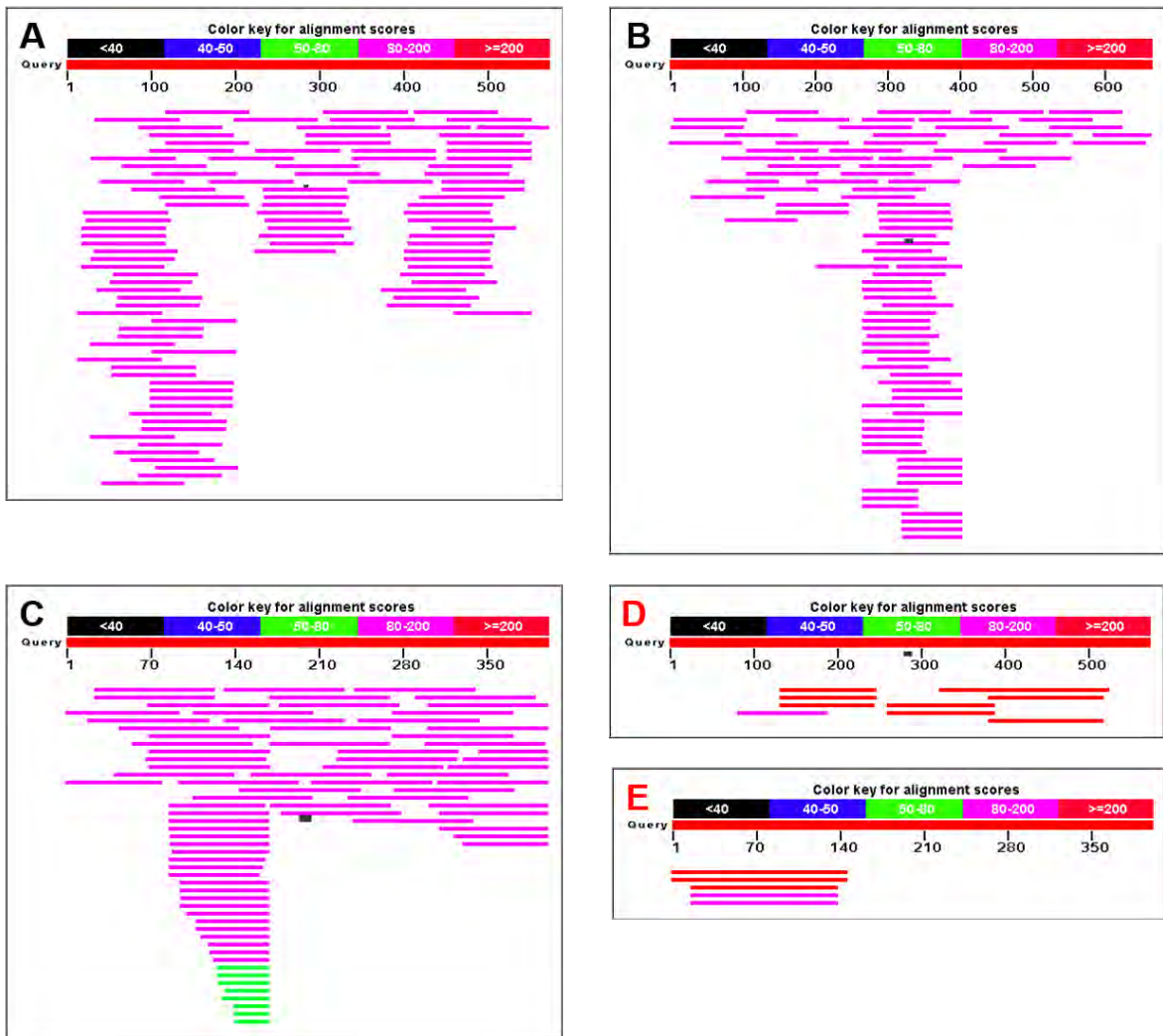


Figura 9. Resumen gráfico de las alineaciones de la búsqueda en la plataforma web de BLAST, de las secuencias de nucleótidos de los genes mmn:midi_00391 (A, D), mmn:midi_00900 (B, E) y mmn:midi_00973 (C) de “*Candidatus* *Midichloria mitochondrii*” contra la base de datos de las secuencias genómicas (SRX821226) <ABC> y las secuencias del transcriptoma (SRX821392) del intestino medio <DE> de *I. ricinus*.

Se realizó una búsqueda de las secuencias de nucleótidos de “Ca. M. mitochondrii” localizadas en los "orf" de *I. ricinus*, en la base de datos del genoma parcial y el transcriptoma del intestino medio de *I. ricinus* en la plataforma web de BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Figura 9). La búsqueda se realizó con el programa BLASTN 2.5.0+ (Zhang et al. 2000).

La Figura 9 muestra el resumen gráfico de las secuencias alineadas con el genoma y el transcriptoma del intestino medio de *I. ricinus*. Las descripciones de las secuencias que presentan alineaciones significativas están en el Apéndice D. BLAST. (Tablas 8 a 12).

La alineación de las secuencias en el genoma de *I. ricinus* presentan los siguientes valores, (Apéndice D. BLAST. Tablas 8 a 10):

- mmn:midi_00391 e-value de 4.25E-43 a 1.2E-33 y ~100% de identidad.
- mmn:midi_00900 e-value de 2.9E-43 a 0.003 y ~100% de identidad.
- mmn:midi_00973 e-value de 4.6E-43 a 1.42E-23 y ~100% de identidad.

La alineación de las secuencias en la anotación del transcriptoma del intestino medio presentan los siguientes valores (Apéndice D. BLAST. Tablas 11 y 12):

- mmn:midi_00391 e-value de 4.19E-101 a 1.63E-45 y ~100% de identidad.
- mmn:midi_00900 e-value de 6.56E-68 a 2.43E-47 y ~100% de identidad.

En la búsqueda de la secuencia nucleotídica mmn:midi_00973 contra la base de datos de la anotación del transcriptoma del intestino medio no se encontró similitud significativa en la anotación.

La transferencia de material genético entre bacterias es común y es un mecanismo que juega un papel importante en la dinámica de sus genomas. En eucariontes es poco común, sin embargo existe varios estudios de transferencia

entre bacterias y eucariontes, generalmente en estudios de aplicaciones biotecnológicas (Marcet-Houben y Gabaldón 2010; Coelho et al. 2013).

Aparte de estos estudios también existe evidencia de transferencia de material genético en Metazoarios y Algas (Boschetti et al. 2012; Schönknecht et al. 2013), varios insectos que comparten el holobioma con bacterias intracelulares poseen fragmentos del material genético de sus endosimbiontes (Dunning Hotopp et al. 2007), pero los mecanismos por los que se transfiere el material genético es desconocido (Suzuki et al. 2015).

Con este análisis proponemos una posible transferencia de genes entre “*Ca. M. mitochondrii*” e *I. ricinus*.

Análisis filogenético

Después de identificar a los posibles genes transferidos en la tabla 3, se buscaron los homólogos para establecer la función que realizan, de acuerdo al nivel de similitud y para asegurar que se trata de genes homólogos. En la Figura 10, la búsqueda en ‘HMMER’ con un corte de búsqueda de 0.003, indica que solo en las proteínas que codifican los genes *mmn:midi_00900* y *mmn:midi_00973* existen homólogos para realizar una reconstrucción filogenética de ambos genes y corroborar que son de origen bacteriano y no eucarionte (heterólogo en *I. ricinus*).

El gen *mmn:midi_00391* no tiene opciones de coincidencias de homólogos y en la búsqueda solo está identificada la misma secuencia de la proteína en “*Ca. M. mitochondrii*” con un ‘E-value’ 1.6E-119. En los dos genes *mmn:midi_00900* y *mmn:midi_00973* si existen homólogos con ‘E-value’ 2.6E-74 y 1.5E-135 respectivamente. Para la secuencia *mmn:midi_00900* solo se encontraron 6 homólogos, por lo que para hacer la reconstrucción filogenética se tomaron los miembros que forman la familia de la proteína FliX (PF10768) de la base de datos ‘Pfam’ (Finn et al. 2016). En cuanto a la secuencia *mmn:midi_00973* se utilizaron los primeros 100 resultados de los 1476 en total, de los homólogos encontrados.

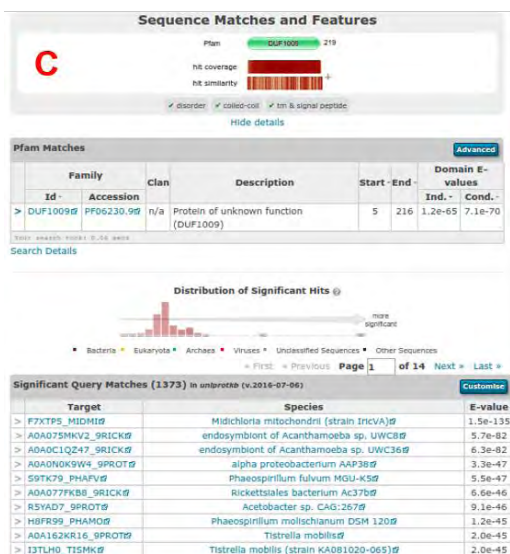
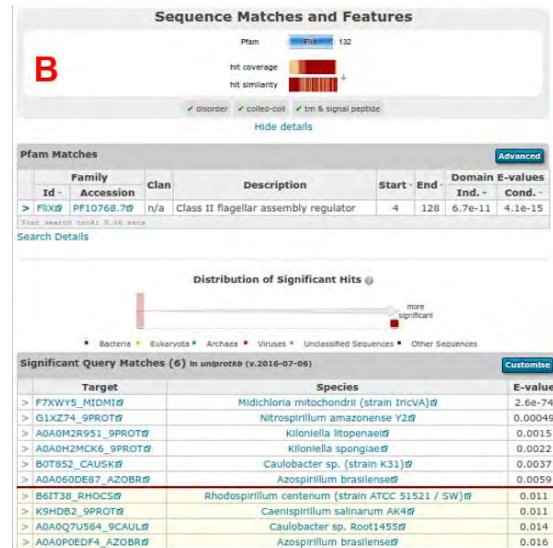
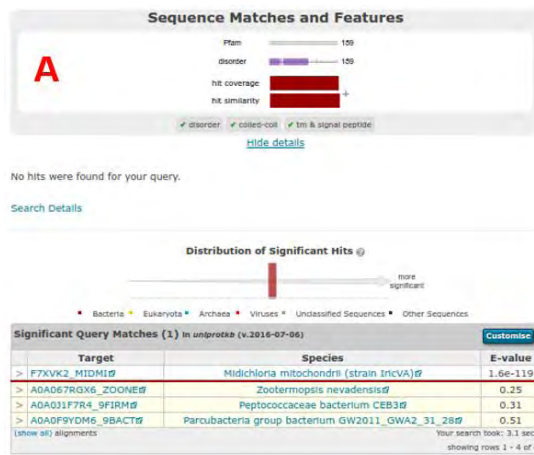


Figura 10. Resultados de la búsqueda de homólogos contra de la base de datos 'UniProtKB' en la plataforma web de HMMER (<http://hmmer.org/>) de las secuencias de aminoácidos identificadas en los 'contigs' de *I. ricinus*. A) mmn:midl_00391; B) mmn:midl_00900 y C) mmn:midl_00973. Muestra parte de los resultados de las secuencias, las características, una gráfica de barras y un listado de los resultados más significativos de acuerdo al 'e-value'.

El gen mmn:midl_00900 tiene dominios de la proteína FliX, la cual posiblemente es un componente transitorio del flagelo y se requiere para el proceso de montaje del mismo (Mohr et al. 1998). Cabe destacar que "Ca. M. mitochondrii" posee varios genes con funciones de ensamblado flagelar no encontrados en otros ejemplares de las Rickettsiales, pero no existe evidencia de la presencia de un aparato flagelar en "Ca. M. mitochondrii" (Cramaro et al. 2015).

Sobre el gen mmn:mid_00973 existen un gran número de homólogos con 'E-values' que van de los 2.0E-45 a 5.7E-82, un par de endosimbiontes de *Acanthamoeba* son los que tienen los valores más significativos. El gen codifica una proteína de funciones desconocidas (DUF1009).

Después de encontrar homólogos de las proteínas codificadas en "Ca. M. mitochondrii" se alinearon y se construyeron los árboles filogenéticos (Figura 11 y 12). La posición de "Ca. M. mitochondrii" se muestra resaltada en ambos árboles.

En el caso del árbol filogenético del gen mmn:mid_00900 FliX "Ca. M. mitochondrii" se encuentra agrupada en las ramas más profundas junto a *Desulfovibrio piezophilus*, que es una bacteria que posee elementos flagelares perteneciente a la clase Deltaproteobacteria. Esta proteína se encuentra también en "Ca. M. mitochondrii" a diferencia de otros miembros de las Rickettsiales que es el grupo más cercano filogenéticamente a "Ca. M. mitochondrii" (Ferla et al. 2013).

La presencia de este gen flagelar podría deberse a que el genoma está en transición de pérdida de genes debido al estilo de vida intramitocondrial, ya que este tipo de genes se pierden en endosimbiontes obligados (Toft y Fares 2008).

El resto de las ramas en orden creciente, presenta miembros de la clase Alphaproteobacteria que se caracterizan por ser bacterias móviles de ambientes acuáticos con presencia de aparato flagelar.

En el árbol filogenético del gen mmn:mid_00973 (Figura 12), "Ca. M. mitochondrii" está agrupada con otros miembros de las Rickettsiales, el gen codifica una proteína conservada, pero de funciones desconocidas (DUF1009).



Figura 11. Árbol filogenético de los homólogos de la proteína codificada por el gen midi_00900 FliX de “*Candidatus Midichloria mitochondrii*”. Se muestran los valores de ‘bootstrap’ en los nodos y se colapsaron ramas con un corte de ≤ 50 .

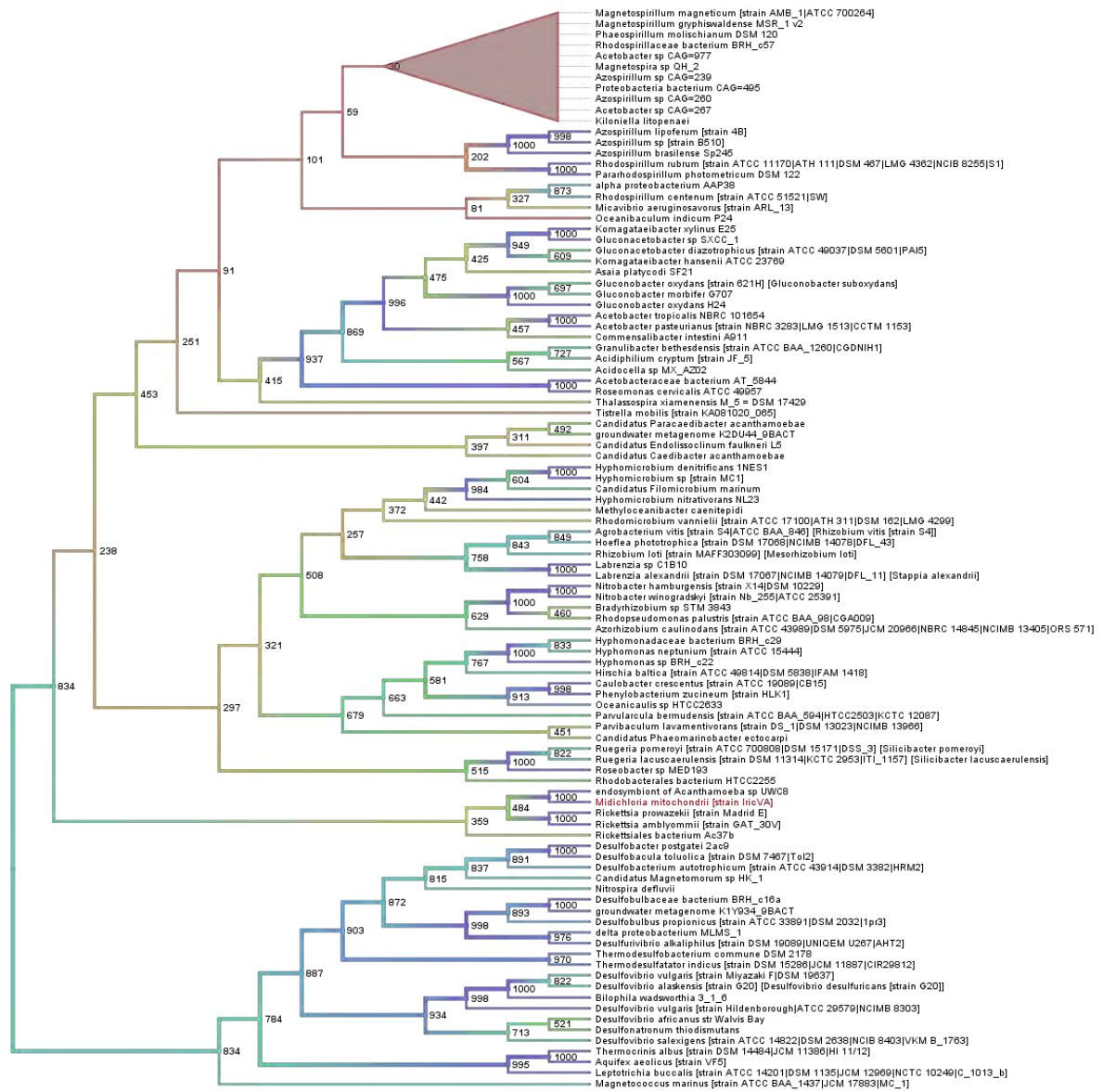


Figura 12. Árbol filogenético de los homólogos de la proteína codificada por el gen *midi_00973* de “*Candidatus Midichloria mitochondrii*”. Se muestran los valores de ‘bootstrap’ en los nodos y se colapsaron ramas con un corte de ≤ 50 .

CONCLUSIONES

Después de revisar las características metabólicas y genómicas de “*Ca. M. mitochondrii*” comparándolas con diferentes organismos (de estilos de vida facultativa y vida libre), concluimos que el tipo de vida endosimbionte intracelular se ve reflejado en la pérdida de funciones metabólicas y que “*Ca. M. mitochondrii*” posee características genómicas que son propias en este tipo de organismos.

A pesar de no haber encontrado evidencia de transferencia horizontal de genes bacteria-bacteria, las secuencias homologas de “*Ca. M. mitochondrii*” identificadas en el DNA genómico de *I. ricinus*, indican una posible transferencia horizontal bacteria-eucarionte, ya que el análisis filogenético de los genes identificados de “*Ca. M. mitochondrii*” en el genoma de *I. ricinus* corroboraron que los genes son de origen bacteriano.

REFERENCIAS

- Ahantarig A, Trinachartvanit W, Baimai V, Grubhoffer L. 2013. Hard ticks and their bacterial endosymbionts (or would be pathogens). *Folia Microbiol. (Praha)* 58:419–428.
- Altschul S, Gish W, Miller W, Myers E, Lipman D. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.
- Batut B, Knibbe C, Marais G, Daubin V. 2014. Reductive genome evolution at both ends of the bacterial population size spectrum. *Nat. Rev. Microbiol.* 12:841–850.
- Bentley SD, Parkhill J. 2004. Comparative Genomic Structure of Prokaryotes. *Annu. Rev. Genet.* 38:771–791.
- Boschetti C, Carr A, Crisp A, Eyres I, Wang-Koh Y, Lubzens E, Barraclough TG, Micklem G, Tunnacliffe A. 2012. Biochemical Diversification through Foreign Gene Expression in Bdelloid Rotifers. Zhang J, editor. *PLoS Genet.* 8:e1003035.
- Carrapiço F. 2015. Can We Understand Evolution Without Symbiogenesis? En: *Reticulate Evolution, Interdisciplinary Evolution Research*. Vol. 3. p. 81–105.
- Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, et al. 2016. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 44:D471–D480.
- Cavalier-Smith T. 2003. Microbial Muddles. *Bioscience* 53:1008–1010,1013.
- Cerqueda-García D, Falcón LI. 2016. La construcción del nicho y el concepto de holobionte, hacia la reestructuración de un paradigma. *Rev. Mex. Biodivers.* 87:239–241.
- Coelho MA, Gonçalves C, Sampaio JP, Gonçalves P. 2013. Extensive Intra-Kingdom Horizontal Gene Transfer Converging on a Fungal Fructose Transporter Gene. Heitman J, editor. *PLoS Genet.* 9:e1003587.
- Cramaro WJ, Revets D, Hunewald OE, Sinner R, Reye AL, Muller CP. 2015. Integration of *Ixodes ricinus* genome sequencing with transcriptome and proteome annotation of the naïve midgut. *BMC Genomics* 16:871.
- Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinforma. Oxf. Engl.* 27:1164–5.
- Delmotte F, Rispé C, Schaber J, Silva FJ, Moya A. 2006. Tempo and mode of early gene loss in endosymbiotic bacteria from insects. *BMC Evol. Biol.* 6:56.

- Dunning Hotopp JC. 2011. Horizontal gene transfer between bacteria and animals. *Trends Genet.* 27:157–163.
- Dunning Hotopp JC, Clark ME, Oliveira DCSG, Foster JM, Fischer P, Muñoz Torres MC, Giebel JD, Kumar N, Ishmael N, Wang S, et al. 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 317:1753–6.
- Dunning Hotopp JC, Lin M, Madupu R, Crabtree J, Angiuoli S V., Eisen J, Seshadri R, Ren Q, Wu M, Utterback TR, et al. 2006. Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genet.* 2:208–223.
- Eddy SR. 2012. The C-value paradox, junk DNA and ENCODE. *Curr. Biol.* 22:R898–R899.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–7.
- Ferla MP, Thrash JC, Giovannoni SJ, Patrick WM. 2013. New rRNA gene-based phylogenies of the Alphaproteobacteria provide perspective on major groups, mitochondrial ancestry and phylogenetic instability. Badger JH, editor. *PloS One* 8:e83383.
- Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR. 2015. HMMER web server: 2015 update. *Nucleic Acids Res.* 43:W30–W38.
- Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44:D279–D285. [consultado 2016 nov 10]. <http://nar.oxfordjournals.org/content/44/D1/D279>
- Gil R, Latorre A, Moya A. 2004. Bacterial endosymbionts of insects: Insights from comparative genomics. *Environ. Microbiol.* 6:1109–1122.
- Gofton A, Oskam C, Lo N, Beninati T, Wei H, McCarl V, Murray D, Papparini A, Greay T, Holmes A, et al. 2015. Inhibition of the endosymbiont “*Candidatus Midichloria mitochondrii*” during 16S rRNA gene profiling reveals potential pathogens in Ixodes ticks from Australia. *Parasit. Vectors* 8:345.
- Goh KM, Gan HM, Chan KG, Chan GF, Shahar S, Chong CS, Kahar UM, Chai KP. 2014. Analysis of *Anoxybacillus* genomes from the aspects of lifestyle adaptations, prophage diversity, and carbohydrate metabolism. Zhang R, editor. *PLoS ONE* 9:e90549.
- Gosalbes MJ, Latorre A, Lamelas A, Moya A. 2010. Genomics of intracellular symbionts in insects. *Int. J. Med. Microbiol. IJMM* 300:271–8.

- Guerrero R, Margulis L, Berlanga M. 2013. Symbiogenesis: The holobiont as a unit of evolution. *Int. Microbiol.* 16:133–143.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* 59:307–321.
- Juhas M. 2015. Type IV secretion systems and genomic islands-mediated horizontal gene transfer in *Pseudomonas* and *Haemophilus*. *Microbiol. Res.* 170:10–17.
- Karp PD, Latendresse M, Paley SM, Krummenacker M, Ong QD, Billington R, Kothari A, Weaver D, Lee TT, Subhraveti P, et al. 2016. Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology. *Brief. Bioinform.* 17:877–890.
- Karp PD, Paley S, Romero P. 2002. The Pathway Tools software. *Bioinformatics* 18:S225–S232.
- Khachane AN, Timmis KN, Martins Dos Santos VAP. 2007. Dynamics of reductive genome evolution in mitochondria and obligate intracellular microbes. *Mol. Biol. Evol.* 24:449–456.
- Konstantinidis KT, Tiedje JM. 2004. Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci. U. S. A.* 101:3160–3165.
- Le PT, Pontarotti P, Raoult D. 2014. Alphaproteobacteria species as a source and target of lateral sequence transfers. *Trends Microbiol.* 22:147–156.
- Madden T. 2003. The BLAST Sequence Analysis Tool. En: McEntyre J, Ostell J, editores. *The NCBI Handbook*[internet]. Bethesda (MD): National Center for Biotechnology Information (US). p. 1–15.
- Marcet-Houben M, Gabaldón T. 2010. Acquisition of prokaryotic genes by fungal genomes. *Trends Genet.* 26:5–8.
- Margulis L. 1990. Words as Battle Cries: Symbiogenesis and the New Field of Endocytobiology. *BioScience* 40:673.
- Margulis L, Sagan D. 2002. *Acquiring Genomes: A Theory of the Origins of Species*. Basic Books.
- Martínez-Cano DJ, Reyes-Prieto M, Martínez-Romero E, Partida-Martínez LP, Latorre A, Moya A, Delaye L, Martínez-Cano DJ, Reyes-Prieto M, Martínez-Romero E, et al. 2014. Evolution of small prokaryotic genomes. *Front. Microbiol.* 5:742.

- McCutcheon JP, McDonald BR, Moran NA. 2009. Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. Matic I, editor. *PLoS Genet.* 5:e1000565.
- McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R. 2013. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* 41:W597–W600.
- Mohr CD, MacKichan JK, Shapiro L. 1998. A membrane-associated protein, fliX, is required for an early step in *Caulobacter* flagellar assembly. *J. Bacteriol.* 180:2175–2185.
- Moran NA. 2002. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* 108:583–6.
- Moran NA, Bennett GM. 2014. The Tiniest Tiny Genomes. *Annu. Rev. Microbiol.* 68:195–215.
- Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. *Ann Rev Genet* 42:165–190.
- Moya A, Peretó J, Gil R, Latorre A. 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat. Rev. Genet.* 9:218–29.
- Podell S, Gaasterland T. 2007. DarkHorse: a method for genome-wide prediction of horizontal gene transfer. *Genome Biol.* 8:R16.
- Podell S, Gaasterland T, Allen EE. 2008. A database of phylogenetically atypical genes in archaeal and bacterial genomes, identified using the DarkHorse algorithm. *BMC Bioinformatics* 9:419.
- Pruitt K, Brown G, Tatusova T, Maglott D. 2002. The Reference Sequence (RefSeq) Database. En: McEntyre J, Ostell J, editores. *The NCBI Handbook*. Bethesda (MD): National Center for Biotechnology Information (US). p. 1–24.
- R Core Team. 2016. R: A language and environment for statistical computing. R. Foundation for Statistical Computing.
- Ravenhall M, Škunca N, Lassalle F, Dessimoz C. 2015. Inferring Horizontal Gene Transfer. Wodak S, editor. *PLOS Comput. Biol.* 11:e1004095.
- Rosenberg E, Sharon G, Zilber-Rosenberg I. 2009. The hologenome theory of evolution contains Lamarckian aspects within a Darwinian framework. *Environ. Microbiol.* 11:2959–2962.
- Sassera D, Beninati T, Bandi C, Bouman EAP, Sacchi L, Fabbi M, Lo N. 2006. “Candidatus *Midichloria mitochondrii*”, an endosymbiont of the tick *Ixodes ricinus* with a unique intramitochondrial lifestyle. *Int. J. Syst. Evol. Microbiol.* 56:2535–40.

- Sassera D, Lo N, Epis S, D'Auria G, Montagna M, Comandatore F, Horner D, Peretó J, Luciano AM, Franciosi F, et al. 2011. Phylogenomic evidence for the presence of a flagellum and cbb(3) oxidase in the free-living mitochondrial ancestor. *Mol. Biol. Evol.* 28:3285–96.
- Schmitz-Esser S, Tischler P, Arnold R, Montanaro J, Wagner M, Rattei T, Horn M. 2010. The genome of the amoeba symbiont “Candidatus Amoebophilus asiaticus” reveals common mechanisms for host cell interaction among amoeba-associated bacteria. *J. Bacteriol.* 192:1045–1057.
- Schonknecht G, Chen W-H, Ternes CM, Barbier GG, Shrestha RP, Stanke M, Brautigam A, Baker BJ, Banfield JF, Garavito RM, et al. 2013. Gene Transfer from Bacteria and Archaea Facilitated Evolution of an Extremophilic Eukaryote. *Science* 339:1207–1210.
- Simpson I. 2009. Heuristic methods, clustering and gene feature finding. *Bioinformatics 2 - Lecture 7*.
- Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31.
- Soucy SM, Huang J, Gogarten JP. 2015. Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.* 16:472–482.
- Stewart FJ. 2013. Where the genes flow. *Nat. Geosci.* 6:688–690.
- Suzuki K, Moriguchi K, Yamamoto S. 2015. Horizontal DNA transfer from bacteria to eukaryotes and a lesson from experimental transfers. *Res. Microbiol.* 166:753–763.
- Suzuki S, Kakuta M, Ishida T, Akiyama Y. 2014. GHOSTX: An improved sequence homology search algorithm using a query suffix array and a database suffix array. Moustafa A, editor. *PLoS ONE* 9:e103833.
- Tatusova T, Ciufu S, Fedorov B, O'Neill K, Tolstoy I. 2014. RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res.* 42:D553-9.
- Toft C, Fares MA. 2008. The evolution of the flagellar assembly pathway in endosymbiotic bacterial genomes. *Mol. Biol. Evol.* 25:2069–2076.
- UniProt Consortium TU. 2015. UniProt: a hub for protein information. *Nucleic Acids Res.* 43:D204-12.
- Ussery DW, Wassenaar TM, Borini S. 2009. *Computing for Comparative Microbial Genomics*. 1a ed. Ussery DW, Wassenaar TM, Borini S, editores. London: Springer London (Computational Biology).

Di Venere M, Fumagalli M, Cafiso A, De Marco L, Epis S, Plantard O, Bardoni A, Salvini R, Viglio S, Bazzocchi C, et al. 2015. *Ixodes ricinus* and Its Endosymbiont *Mitochondria mitochondrii*: A Comparative Proteomic Analysis of Salivary Glands and Ovaries. Munderloh UG, editor. PLOS ONE 10:e0138842.

Warnes GR, Bolker B, Bonebakker L, Gentleman R, Wolfgang H, Liaw A, Lumley T, Maechler M, Magnusson A, Moeller S, et al. 2016. *gplots*: Various R. Programming Tools for Plotting Data.

Wickham H. 2009. *ggplot2: Elegant Graphics for Data Analysis*.

Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A Greedy Algorithm for Aligning DNA Sequences. *J. Comput. Biol.* 7:203–214.

APÉNDICES

Apéndice A Categorías metabólicas

Tabla 4. El total de rutas de Biosíntesis y Generación de Metabolitos Precusores y Energía recuperadas de la base de datos de 'BioCyc'.

	Aminas y Poliaminas	Aminoácidos	Carbohidratos	Cofactores, Grupos prostéticos y acarreadores de electrones	Ácidos grasos y Lípidos	Nucleósidos y Nucleótidos	Metabolitos secundarios	Generación de Metabolitos precusores y Energía
' <i>Candidatus Carsonella ruddii</i> ' HT isolate Thao2000	0	9	0	1	0	5	0	3
" <i>Candidatus Midichloria mitochondrii</i> " IricVA	1	11	5	28	14	14	2	12
' <i>Candidatus Nasuia deltocephalinicola</i> ' NAS-ALF	0	5	0	4	0	0	0	1
' <i>Candidatus Pelagibacter ubique</i> ' HTCC1062	6	26	9	40	15	22	2	14
<i>Agrobacterium</i> sp. H13-3	5	33	10	42	17	25	3	20
<i>Anaplasma phagocytophilum</i> HZ	0	7	1	36	10	19	1	9
<i>Bacillus subtilis subtilis</i> 168	7	29	6	41	11	19	9	20
<i>Bartonella henselae</i> Houston-1	2	16	3	33	13	19	1	13
<i>Brucella abortus</i> A13334	8	33	10	51	21	27	3	39
<i>Ehrlichia</i> sp. HF	0	9	1	21	6	16	0	6
endosymbiont of <i>Acanthamoeba</i> sp. UWC8	0	6	1	8	5	10	0	5
<i>Escherichia coli</i> K-12 substr. MG1655 (EcoCyc)	8	28	15	47	19	29	3	38
<i>Orientia tsutsugamushi</i> Boryong	0	7	2	15	9	13	0	4
<i>Rickettsia prowazekii</i> Madrid E	0	7	6	17	13	14	0	14
<i>Sinorhizobium meliloti</i> 1021	9	31	16	46	20	28	2	35
<i>Wolbachia</i> sp. wRi	1	9	2	26	10	18	1	17

Tabla 5. El total de rutas de degradación recuperadas de las categorías de la base de datos de 'BioCyc'.

	Aminas y Poliaminas	Carbohidratos	Carboxilatos	Ácidos grasos y Lípidos	Metabolismo de Nutrientes Inorgánicos	Nucleósidos y Nucleótidos	Metabolitos secundarios
' <i>Candidatus Carsonella ruddii</i> ' HT isolate Thao2000	0	0	0	0	1	0	0
' <i>Candidatus Midichloria mitochondrii</i> ' IricVA	0	1	1	0	2	1	3
' <i>Candidatus Nasuia deltocephalinicola</i> ' NAS-ALF	0	0	0	0	0	0	0
' <i>Candidatus Pelagibacter ubique</i> ' HTCC1062	5	4	5	1	13	2	2
<i>Agrobacterium</i> sp. H13-3	10	15	18	2	11	12	18
<i>Anaplasma phagocytophilum</i> HZ	0	0	1	0	5	0	0
<i>Bacillus subtilis subtilis</i> 168	5	20	14	3	6	10	11
<i>Bartonella henselae</i> Houston-1	0	2	2	3	2	0	1
<i>Brucella abortus</i> A13334	5	8	13	4	16	7	8
<i>Ehrlichia</i> sp. HF	0	0	0	0	3	0	0
endosymbiont of <i>Acanthamoeba</i> sp. UWC8	0	0	0	0	2	0	0
<i>Escherichia coli</i> K-12 substr. MG1655 (EcoCyc)	10	23	22	4	15	13	16
<i>Orientia tsutsugamushi</i> Boryong	0	0	0	0	1	1	1
<i>Rickettsia prowazekii</i> Madrid E	0	0	3	0	0	1	0
<i>Sinorhizobium melloti</i> 1021	11	17	18	4	20	12	11
<i>Wolbachia</i> sp. wRi	0	0	2	0	6	0	1

Apéndice B Resumen de genomas

Tabla 6. El resumen de las características de cada uno de los genomas utilizados.

	Tamaño (Mb)	GC%	Proteínas	Genes	rRNA	tRNA	Otros RNA	Pseudogenes	Forma de vida
' <i>Candidatus Carsonella ruddii</i> ' HT isolate Thao2000	0.16	14.6	166	192	2	23		1	Simbionte obligada
' <i>Candidatus Midichloria mitochondrii</i> ' IricVA	1.18	36.6	1161	1246	3	35	1	46	Simbionte obligada
' <i>Candidatus Nasuia deltocephalinicola</i> ' NAS-ALF	0.112091	17.1	137	169	3	29			Simbionte obligada
' <i>Candidatus Pelagibacter ubique</i> ' HTCC1062	1.31	29.7	1333	1370	3	32	1	1	Simbionte obligada
<i>Agrobacterium</i> sp. H13-3	5.57377	58.5333	5097	5244	15	57	1	74	Simbionte facultativa
<i>Anaplasma phagocytophilum</i> HZ	1.47	41.6	1058	1141	3	37	1	42	Vida libre
<i>Bacillus subtilis subtilis</i> 168	4.22	43.5	4175	4421	30	86	62	68	Simbionte obligada
<i>Bartonella henselae</i> Houston-1	1.93	38.2	1525	1665	7	43	3	87	Simbionte facultativa
<i>Brucella abortus</i> A13334	3.28608	57.2354	3095	3188	9	55	1	27	Simbionte obligada
<i>Ehrlichia</i> sp. HF	1.15	29.6	855	903	3	36	1	8	Simbionte facultativa
endosymbiont of <i>Acanthamoeba</i> sp. UWC8	1.62	34.8	1470	1544	3	36	1	34	Simbionte obligada
<i>Escherichia coli</i> K-12 substr. MG1655 (EcoCyc)	4.64	50.8	4140	4498	22	87	67	184	Simbionte facultativa
<i>Orientia tsutsugamushi</i> Boryong	2.13	30.5	1565	2019	3	34	1	416	Vida libre
<i>Rickettsia prowazekii</i> Madrid E	1.11	29	843	889	3	33	1		Vida libre
<i>Sinorhizobium meliloti</i> 1021	6.69169	62.1591	6218	6292	9	54	7	4	Patógena
<i>Wolbachia</i> sp. wRi	1.45	35.2	1248	1350	3	34	2	63	Patógena

Apéndice C Resultados 'DarkHorse'

Tabla 7. Resumen de resultados de la búsqueda de transferencia horizontal 'DarkHorse'

query_annotation	besthit_annotation	norm_LPI	pct_id	q_len	align_len	pct_cov	evaluate	bitscore
GTP binding protein typA	GTP-binding protein TypA [<i>Candidatus Amoebophilus asiaticus</i>][#gij189502125]ref YP	0.415	71.93	606	602	99	0	846
succinyl-CoA synthetase, alpha subunit	succinyl-CoA synthetase subunit alpha [<i>Anoxybacillus</i> sp. KU2-6(11)]	0.415	67.36	259	242	93	3E-104	317
chaperonin GroS	molecular chaperone GroES [<i>Smithella</i> sp. F21]	0.62	71.28	97	94	96	2E-40	139
hypothetical protein midi_00061	pyrroline-5-carboxylate dehydrogenase [<i>Vibrio</i> sp. B183]	0.62	62.5	49	48	97	6E-13	72
50S ribosomal protein L36	50S ribosomal protein L36 [<i>Achromobacter</i> sp. RTa]	0.621	68.29	41	41	100	4E-11	61
translation elongation factor Tu	elongation factor Tu [<i>Massilia</i> sp. BSC265][#gij672865253]gb KF108113.1 elongation	0.621	76.52	392	396	101	0	582
FeS cluster assembly scaffold IscU	scaffolding protein [<i>Aquabacterium</i> sp. NJ1]	0.666	72.31	131	130	99	3E-63	201
ribosomal protein L2	50S ribosomal protein L2 [<i>Ahrensia</i> sp. 13_GOM-1096m]	0.757	68.36	275	275	100	4E-115	344
ribosomal protein S12	30S ribosomal protein S12 [<i>Paracoccus</i> sp. 10990]	0.757	85.33	76	75	98	3E-38	134
translation initiation factor IF-1	translation initiation factor IF-1 [<i>Novosphingobium</i> sp. AP12][#gij398032609]gb EJ	0.757	76.47	79	68	86	2E-31	115
ribosomal protein L34P	50S ribosomal protein L34 [<i>Sphingopyxis</i> sp. LC81]	0.757	83.72	46	43	93	7E-15	71
NADH dehydrogenase subunit I	NADH dehydrogenase [<i>Sphingopyxis</i> sp. MVB1]	0.757	71.6	162	162	100	2E-77	239
thioredoxin reductase	thioredoxin reductase TrxB [<i>Acidiphilium</i> sp. JA12-A1]	0.757	67.65	308	306	99	1E-152	442
cysteine desulfurase	cysteine desulfurase [<i>Dongia</i> sp. URHE0060]	0.757	67.76	396	397	100	0	588
NADH dehydrogenase I subunit F	NADH dehydrogenase [<i>Inquilinus limosus</i> MP06]	0.757	69.93	434	419	96	0	630
ribosomal protein L20	50S ribosomal subunit protein L20 [<i>Magnetospira</i> sp. QH-2]	0.757	68.7	120	115	95	8E-52	171
NADH dehydrogenase subunit B	NADH dehydrogenase [<i>Thalassospira</i> sp. NP3b2]	0.758	75.15	177	165	93	4E-80	247
Cbb3-type cytochrome oxidase, subunit I	cytochrome oxidase [<i>Thalassospira</i> sp. NP3b2]	0.758	68.45	343	336	97	1E-167	488
pyruvate dehydrogenase subunit beta	pyruvate dehydrogenase [<i>Thalassospira</i> sp. NP3b2]	0.758	72.5	327	320	97	9E-160	467
ATP/ADP translocase class 1	ATPase AAA [<i>Candidatus Liberibacter solanacearum</i>]	0.758	73.08	497	494	99	0	715
ribosomal protein L27	50S ribosomal protein L27 [<i>Rhizobium</i> sp. YS-1r]	0.758	69.51	85	82	96	3E-34	123
ribosomal protein L14	50S ribosomal protein L14 [<i>Hoeflea</i> sp. BAL378]	0.758	64.75	122	122	100	1E-35	129
NADH:ubiquinone oxidoreductase subunit D	NADH dehydrogenase [<i>Hoeflea</i> sp. BAL378]	0.758	70.92	391	392	100	0	595
ribosomal protein S19	30S ribosomal protein S19 [<i>Hoeflea</i> sp. BAL378]	0.758	69.66	93	89	95	2E-38	134
recombinase A	recombinase RecA [<i>Hoeflea</i> sp. BAL378]	0.758	69.05	352	349	99	6E-177	508
GTP-binding protein LepA	GTP-binding protein LepA [<i>Microvirga</i> sp. BSC39]	0.758	70.83	557	545	97	0	785
ATP synthase subunit beta	ATP synthase F0F1 subunit beta [<i>Microvirga</i> sp. BSC39]	0.758	77.83	470	469	99	0	685

query_annotation	besthit_annotation	norm_LPI	pct_id	q_len	align_len	pct_cov	evaluate	bitscore
molecular chaperone DnaK	molecular chaperone DnaK [Microvirga sp. BSC39]	0.758	67.61	631	636	100	0	839
dihydrolipoamide acetyltransferase	dihydrolipoamide succinyltransferase [Microvirga sp. BSC39]	0.758	68.89	225	225	100	2E-113	342
succinate dehydrogenase flavoprotein subunit	fumarate reductase [Microvirga sp. BSC39]	0.758	70.13	597	596	99	0	861
FOF1-type ATP synthase, alpha subunit	ATP FOF1 synthase subunit alpha [Microvirga sp. BSC39]	0.758	67.21	431	427	99	0	587
ribosomal protein S10	50S ribosomal protein L16 [Microvirga sp. BSC39]	0.758	69.34	138	137	99	6E-63	200
enoyl-(acyl-carrier-protein) reductase	enoyl-(acyl-carrier-protein) reductase (NADH) [Rhodobacteraceae bacterium HIMB11]	0.812	69.53	263	256	97	2E-129	379
NADH dehydrogenase I subunit H	NADH:ubiquinone oxidoreductase subunit H [Candidatus Rickettsia gravesii]	0.865	73.79	333	309	92	1E-139	411
type IV secretion system protein VirD4	Conjugal transfer protein traG [Rickettsia sp. ISO7]	0.865	81.01	461	437	94	0	691
lipoyl synthase	Lipoyl synthase [Rickettsia sp. ISO7]	0.865	75	245	236	96	2E-132	389
type IV secretion system protein VirB11	Type IV secretion system protein virB11 [Rickettsia sp. ISO7]	0.865	71.34	327	321	98	7E-142	416
DNA-directed RNA polymerase, beta' subunit	DNA-directed RNA polymerase subunit beta [Rickettsia sp. ISO7]	0.865	72.62	730	683	93	0	1007
elongation factor G	Vegetative protein 19 [Rickettsia sp. ISO7]	0.865	72.85	680	685	100	0	1001
ferredoxin	2FE-2S ferredoxin [Candidatus Paracaedibacter symbiosus]	0.919	68.81	110	109	99	6E-48	160
ribonucleotide-diphosphate reductase subunit beta	ribonucleotide-diphosphate reductase subunit beta [Rickettsiaceae bacterium Os18]	0.919	76.18	341	319	93	1E-169	487
cytochrome b	cytochrome B [Rickettsiaceae bacterium Os18]	0.919	68.37	406	392	96	0	522
Biotin synthase	biotin synthase [Rickettsiaceae bacterium Os18]	0.919	68.4	312	307	98	6E-156	451

Apéndice D BLAST

Tabla 8. Resultado BLAST de la búsqueda de la secuencia mmn:midi_00391 contra la anotación del genoma de *I. ricinus*.

queryid	subjectids	% ident.	align.length	Mismatches	gap opens	q. start	q. end	s. start	s. end	evaluate	bit score
mmn:midi_00391	gnl SRA SRR1739958.433183888.2	100	100	0	0	304	403	1	100	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.350883237.1	100	100	0	0	118	217	100	1	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.337319625.1	100	100	0	0	312	411	1	100	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.313809994.2	100	100	0	0	199	298	1	100	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.313809994.1	100	100	0	0	273	372	100	1	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.277709928.1	100	100	0	0	35	134	100	1	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.263136713.1	100	100	0	0	86	185	1	100	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.245478269.2	100	100	0	0	99	198	1	100	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.245478269.1	100	100	0	0	284	383	100	1	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.236091004.1	100	100	0	0	118	217	100	1	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.219821101.2	100	100	0	0	284	383	100	1	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.219821101.1	100	100	0	0	99	198	1	100	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.214086983.2	100	100	0	0	338	437	100	1	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.214086983.1	100	100	0	0	170	269	1	100	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.203744739.2	100	100	0	0	67	166	1	100	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.203744739.1	100	100	0	0	248	347	100	1	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.177367149.2	100	100	0	0	102	201	1	100	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.177367149.1	100	100	0	0	271	370	100	1	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.176499895.1	100	100	0	0	379	478	1	100	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.169018307.1	100	100	0	0	31	130	100	1	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.103389314.1	100	100	0	0	41	140	100	1	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.72870204.2	100	100	0	0	411	510	1	100	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.45198802.2	100	100	0	0	78	177	100	1	4.25E-43	185
mmn:midi_00391	gnl SRA SRR1739958.469547195.1	99	100	1	0	450	549	1	100	1.98E-41	180
mmn:midi_00391	gnl SRA SRR1739958.331863468.1	99	100	1	0	170	269	1	100	1.98E-41	180
mmn:midi_00391	gnl SRA SRR1739958.327620632.2	99	100	1	0	441	540	1	100	1.98E-41	180
mmn:midi_00391	gnl SRA SRR1739958.263136713.2	99	100	1	0	225	324	100	1	1.98E-41	180
mmn:midi_00391	gnl SRA SRR1739958.204676329.1	99	100	1	0	450	549	1	100	1.98E-41	180
mmn:midi_00391	gnl SRA SRR1739958.169158588.2	99	100	1	0	112	211	100	1	1.98E-41	180
mmn:midi_00391	gnl SRA SRR1739958.61283092.2	99	100	1	0	118	217	100	1	1.98E-41	180
mmn:midi_00391	gnl SRA SRR1739958.104880286.1	98	100	2	0	450	549	1	100	9.2E-40	174

queryid	subject ids	% ident.	alig.length	Mismatches	gap opens	q. start	q. end	s. start	s. end	evaluate	bit score
mmn:midi_00391	gnl SRA SRR1739958.490879729.2	97	100	3	0	233	332	100	1	4.28E-38	169
mmn:midi_00391	gnl SRA SRR1739958.464257901.1	97	100	3	0	450	549	1	100	4.28E-38	169
mmn:midi_00391	gnl SRA SRR1739958.441333544.2	97	100	3	0	235	334	1	100	4.28E-38	169
mmn:midi_00391	gnl SRA SRR1739958.381642594.1	97	100	3	0	428	527	1	100	4.28E-38	169
mmn:midi_00391	gnl SRA SRR1739958.280770790.1	97	100	3	0	22	121	1	100	4.28E-38	169
mmn:midi_00391	gnl SRA SRR1739958.268127054.2	97	100	3	0	424	523	1	100	4.28E-38	169
mmn:midi_00391	gnl SRA SRR1739958.260672112.1	97	100	3	0	25	124	1	100	4.28E-38	169
mmn:midi_00391	gnl SRA SRR1739958.256955059.2	97.959	98	1	1	444	541	2	98	4.28E-38	169
mmn:midi_00391	gnl SRA SRR1739958.232658193.2	97	100	3	0	19	118	100	1	4.28E-38	169
mmn:midi_00391	gnl SRA SRR1739958.172541812.2	97	100	3	0	19	118	100	1	4.28E-38	169
mmn:midi_00391	gnl SRA SRR1739958.91762959.1	97	100	3	0	232	331	100	1	4.28E-38	169
mmn:midi_00391	gnl SRA SRR1739958.60033234.2	97	100	3	0	227	326	1	100	4.28E-38	169
mmn:midi_00391	gnl SRA SRR1739958.52049173.1	97	100	3	0	19	118	100	1	4.28E-38	169
mmn:midi_00391	gnl SRA SRR1739958.331863468.2	96.97	99	3	0	339	437	99	1	1.54E-37	167
mmn:midi_00391	gnl SRA SRR1739958.199344342.1	96.97	99	3	0	34	132	1	99	1.54E-37	167
mmn:midi_00391	gnl SRA SRR1739958.193215876.1	96.97	99	3	0	31	129	2	100	1.54E-37	167
mmn:midi_00391	gnl SRA SRR1739958.68357690.1	96.97	99	3	0	236	334	100	2	1.54E-37	167
mmn:midi_00391	gnl SRA SRR1739958.494328340.1	96.939	98	3	0	19	116	98	1	5.54E-37	165
mmn:midi_00391	gnl SRA SRR1739958.207877946.2	96.939	98	3	0	444	541	1	98	5.54E-37	165
mmn:midi_00391	gnl SRA SRR1739958.478880136.1	96.04	101	2	2	334	433	100	1	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.469186623.2	96	100	4	0	57	156	1	100	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.462323963.1	96.907	97	3	0	53	149	3	99	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.426673532.2	96.04	101	3	1	418	518	1	100	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.418111338.2	96	100	4	0	239	338	100	1	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.391352592.2	96	100	4	0	37	136	100	1	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.364233074.2	96	100	4	0	62	161	1	100	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.361825062.1	96	100	4	0	60	159	1	100	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.341170953.1	96.04	101	3	1	404	504	1	100	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.300035894.2	96.04	101	3	1	400	500	1	100	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.291213240.1	96.04	101	3	1	404	504	1	100	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.285893564.1	96	100	4	0	432	531	1	100	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.268320598.1	96	100	4	0	15	114	100	1	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.267513449.1	96	100	4	0	102	201	1	100	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.224365869.2	96	100	4	0	229	328	100	1	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.202424400.1	96.04	101	3	1	406	506	100	1	1.99E-36	163

queryid	subject ids	% ident.	alig.length	Mismatches	gap opens	q. start	q. end	s. start	s. end	evaluate	bit score
mmn:midi_00391	gnl SRA SRR1739958.167554601.1	96	100	4	0	64	163	1	100	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.163147026.2	96.04	101	3	1	404	504	100	1	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.91762959.2	96	100	4	0	63	162	1	100	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.62163258.1	96	100	4	0	30	129	1	100	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.61671155.1	96.04	101	3	1	400	500	1	100	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.58288330.1	96	100	4	0	102	201	1	100	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.49285835.1	96.04	101	3	1	400	500	1	100	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.48689906.1	96	100	4	0	15	114	100	1	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.43687282.1	96	100	4	0	55	154	1	100	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.39239860.1	96.04	101	3	1	404	504	1	100	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.16642095.2	96	100	4	0	55	154	1	100	1.99E-36	163
mmn:midi_00391	gnl SRA SRR1739958.401580600.1	95.96	99	4	0	100	198	2	100	7.16E-36	161
mmn:midi_00391	gnl SRA SRR1739958.485569187.1	95.96	99	3	1	396	494	3	100	2.58E-35	159
mmn:midi_00391	gnl SRA SRR1739958.326537580.1	95.918	98	4	0	100	197	98	1	2.58E-35	159
mmn:midi_00391	gnl SRA SRR1739958.91949174.1	95.918	98	4	0	100	197	98	1	2.58E-35	159
mmn:midi_00391	gnl SRA SRR1739958.14632486.2	95.918	98	4	0	100	197	98	1	2.58E-35	159
mmn:midi_00391	gnl SRA SRR1739958.490879729.1	95.876	97	4	0	76	172	1	97	9.26E-35	158
mmn:midi_00391	gnl SRA SRR1739958.435169555.1	95.05	101	4	1	409	509	100	1	9.26E-35	158
mmn:midi_00391	gnl SRA SRR1739958.414644643.2	95	100	5	0	91	190	100	1	9.26E-35	158
mmn:midi_00391	gnl SRA SRR1739958.337319625.2	100	85	0	0	486	570	100	16	9.26E-35	158
mmn:midi_00391	gnl SRA SRR1739958.280770790.2	95	100	5	0	241	340	100	1	9.26E-35	158
mmn:midi_00391	gnl SRA SRR1739958.260170721.1	95.05	101	3	2	373	472	1	100	9.26E-35	158
mmn:midi_00391	gnl SRA SRR1739958.241563539.2	95.05	101	3	2	388	487	100	1	9.26E-35	158
mmn:midi_00391	gnl SRA SRR1739958.224365869.1	95	100	5	0	90	189	1	100	9.26E-35	158
mmn:midi_00391	gnl SRA SRR1739958.193773725.1	95.05	101	3	2	379	478	1	100	9.26E-35	158
mmn:midi_00391	gnl SRA SRR1739958.163936404.2	95	100	5	0	30	129	100	1	9.26E-35	158
mmn:midi_00391	gnl SRA SRR1739958.155446269.1	95	100	5	0	86	185	100	1	9.26E-35	158
mmn:midi_00391	gnl SRA SRR1739958.148201950.2	95	100	5	0	58	157	100	1	9.26E-35	158
mmn:midi_00391	gnl SRA SRR1739958.137497790.2	95	100	5	0	77	176	100	1	9.26E-35	158
mmn:midi_00391	gnl SRA SRR1739958.74755865.1	95.876	97	4	0	107	203	100	4	9.26E-35	158
mmn:midi_00391	gnl SRA SRR1739958.68357690.2	95	100	5	0	86	185	1	100	9.26E-35	158
mmn:midi_00391	gnl SRA SRR1739958.373091313.2	96.774	93	3	0	458	550	1	93	3.33E-34	156
mmn:midi_00391	gnl SRA SRR1739958.4925561.1	95.833	96	4	0	223	318	5	100	3.33E-34	156
mmn:midi_00391	gnl SRA SRR1739958.406026310.2	94.898	98	5	0	43	140	100	3	1.2E-33	154

Tabla 9. Resultado BLAST de la búsqueda de la secuencia mmn:midl_00900 contra la anotación del genoma de *I. ricinus*.

query id	subject ids	% ident	Alig. length	mismatches	gap opens	q. start	q. end	s. start	s. end	evalue	bit score
mmn:midl_00900	gnl SRA SRR1739958.496603476.1	100	100	0	0	240	339	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.339175127.1	100	100	0	0	290	389	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.336440750.2	100	100	0	0	25	124	100	1	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.336330733.2	100	100	0	0	169	268	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.325946597.1	100	100	0	0	300	399	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.323015893.2	100	100	0	0	132	231	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.323015893.1	100	100	0	0	271	370	100	1	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.322410035.2	100	100	0	0	69	168	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.322410035.1	100	100	0	0	243	342	100	1	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.318568789.2	100	100	0	0	170	269	100	1	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.293049612.1	100	100	0	0	106	205	100	1	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.275963490.2	100	100	0	0	132	231	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.275963490.1	100	100	0	0	271	370	100	1	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.269947559.2	100	100	0	0	169	268	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.266023907.1	100	100	0	0	25	124	100	1	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.265413333.2	100	100	0	0	177	276	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.265413333.1	100	100	0	0	226	325	100	1	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.239009604.1	100	100	0	0	70	169	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.190276890.2	100	100	0	0	300	399	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.177153175.2	100	100	0	0	225	324	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.174385526.1	100	100	0	0	70	169	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.166414806.1	100	100	0	0	213	312	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.163830213.1	100	100	0	0	267	366	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.157377845.2	100	100	0	0	204	303	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.157076387.1	100	100	0	0	67	166	100	1	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.138666162.2	100	100	0	0	154	253	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.115823775.2	100	100	0	0	145	244	100	1	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.115094469.1	100	100	0	0	234	333	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.114425845.2	100	100	0	0	298	397	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.106042620.2	100	100	0	0	170	269	100	1	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.101522792.2	100	100	0	0	45	144	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.101522792.1	100	100	0	0	179	278	100	1	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.99349255.1	100	100	0	0	56	155	1	100	2.9E-43	185

query_id	subject_ids	% ident	Alig. length	mismatches	gap opens	q. start	q. end	s. start	s. end	evaluate	bit score
mmn:midl_00900	gnl SRA SRR1739958.71945632.2	100	100	0	0	70	169	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.28058668.2	100	100	0	0	238	337	100	1	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.28058668.1	100	100	0	0	94	193	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.4706401.2	100	100	0	0	272	371	1	100	2.9E-43	185
mmn:midl_00900	gnl SRA SRR1739958.208090528.1	100	99	0	0	106	204	100	2	1.04E-42	183
mmn:midl_00900	gnl SRA SRR1739958.177153175.1	100	99	0	0	301	399	100	2	1.04E-42	183
mmn:midl_00900	gnl SRA SRR1739958.276524027.2	99.01	101	0	1	19	119	100	1	1.35E-41	180
mmn:midl_00900	gnl SRA SRR1739958.214837818.2	98	100	2	0	41	140	100	1	6.27E-40	174
mmn:midl_00900	gnl SRA SRR1739958.157377845.1	100	92	0	0	308	399	100	9	8.11E-39	171
mmn:midl_00900	gnl SRA SRR1739958.208598648.1	100	90	0	0	310	399	100	11	1.05E-37	167
mmn:midl_00900	gnl SRA SRR1739958.138666162.1	100	90	0	0	310	399	100	11	1.05E-37	167
mmn:midl_00900	gnl SRA SRR1739958.106042620.1	98.936	94	0	1	1	94	8	100	1.05E-37	167
mmn:midl_00900	gnl SRA SRR1739958.277963883.2	100	83	0	0	317	399	90	8	8.17E-34	154
mmn:midl_00900	gnl SRA SRR1739958.422076046.1	98.795	83	1	0	86	168	99	17	3.8E-32	148
mmn:midl_00900	gnl SRA SRR1739958.134576986.1	98.795	83	1	0	86	168	1	83	3.8E-32	148
mmn:midl_00900	gnl SRA SRR1739958.98075043.1	98.795	83	1	0	86	168	1	83	3.8E-32	148
mmn:midl_00900	gnl SRA SRR1739958.240252230.1	98.78	82	1	0	87	168	100	19	1.37E-31	147
mmn:midl_00900	gnl SRA SRR1739958.224455581.1	98.78	82	1	0	87	168	100	19	1.37E-31	147
mmn:midl_00900	gnl SRA SRR1739958.222803799.2	100	78	0	0	322	399	1	78	4.92E-31	145
mmn:midl_00900	gnl SRA SRR1739958.287268598.1	98.75	80	1	0	86	165	80	1	1.77E-30	143
mmn:midl_00900	gnl SRA SRR1739958.265944874.1	98.75	80	1	0	89	168	1	80	1.77E-30	143
mmn:midl_00900	gnl SRA SRR1739958.57893882.1	98.75	80	1	0	86	165	80	1	1.77E-30	143
mmn:midl_00900	gnl SRA SRR1739958.115823775.1	98.75	80	0	1	1	80	22	100	6.36E-30	141
mmn:midl_00900	gnl SRA SRR1739958.290491340.2	98.718	78	1	0	86	163	23	100	2.29E-29	139
mmn:midl_00900	gnl SRA SRR1739958.99097346.1	98.667	75	1	0	86	160	75	1	1.06E-27	134
mmn:midl_00900	gnl SRA SRR1739958.336330733.1	100	71	0	0	329	399	100	30	3.83E-27	132
mmn:midl_00900	gnl SRA SRR1739958.335553876.1	98.649	74	1	0	95	168	1	74	3.83E-27	132
mmn:midl_00900	gnl SRA SRR1739958.269947559.1	100	71	0	0	329	399	100	30	3.83E-27	132
mmn:midl_00900	gnl SRA SRR1739958.222192239.1	98.649	74	1	0	95	168	1	74	3.83E-27	132
mmn:midl_00900	gnl SRA SRR1739958.253304119.2	98.63	73	1	0	96	168	1	73	1.38E-26	130
mmn:midl_00900	gnl SRA SRR1739958.71221803.2	97.297	74	2	0	95	168	1	74	1.78E-25	126
mmn:midl_00900	gnl SRA SRR1739958.492239972.2	98.529	68	1	0	101	168	100	33	8.28E-24	121
mmn:midl_00900	gnl SRA SRR1739958.196045258.1	100	58	0	0	342	399	2	59	6.45E-20	108
mmn:midl_00900	gnl SRA SRR1739958.297142713.1	98.333	60	1	0	109	168	100	41	2.32E-19	106

query id	subject ids	% ident.	Alig. length	mismatches	gap opens	q. start	q. end	s. start	s. end	evaluate	bit score
mmn:midi_00900	gnl SRA SRR1739958.180584517.2	98.333	60	1	0	109	168	100	41	2.32E-19	106
mmn:midi_00900	gnl SRA SRR1739958.460144426.2	98.214	56	1	0	113	168	100	45	3.88E-17	99
mmn:midi_00900	gnl SRA SRR1739958.26043679.1	98	50	1	0	119	168	100	51	8.4E-14	87.9
mmn:midi_00900	gnl SRA SRR1739958.218372323.1	97.959	49	1	0	120	168	1	49	3.02E-13	86.1
mmn:midi_00900	gnl SRA SRR1739958.90346207.1	97.826	46	1	0	123	168	1	46	1.41E-11	80.5
mmn:midi_00900	gnl SRA SRR1739958.335530347.2	97.674	43	1	0	126	168	1	43	6.54E-10	75
mmn:midi_00900	gnl SRA SRR1739958.198139142.1	97.674	43	1	0	126	168	1	43	6.54E-10	75
mmn:midi_00900	gnl SRA SRR1739958.46716766.2	97.619	42	1	0	127	168	1	42	2.35E-09	73.1
mmn:midi_00900	gnl SRA SRR1739958.305723258.1	100	36	0	0	133	168	100	65	0.000000109	67.6
mmn:midi_00900	gnl SRA SRR1739958.140575024.2	97.436	39	1	0	130	168	3	41	0.000000109	67.6
mmn:midi_00900	gnl SRA SRR1739958.423690476.2	100	29	0	0	140	168	100	72	0.000852	54.7
mmn:midi_00900	gnl SRA SRR1739958.146801260.1	100	28	0	0	141	168	1	28	0.003	52.8
mmn:midi_00900	gnl SRA SRR1739958.53680461.2	100	28	0	0	141	168	1	28	0.003	52.8

Tabla 10. Resultado BLAST de la búsqueda de la secuencia mmn:midi_00973 contra la anotación del genoma de *I. ricinus*.

query id	subject ids	% ident.	Alig length	mismatches	gap opens	q. start	q. end	s. start	s. end	evaluate	bit score
mmn:midi_00973	gnl SRA SRR1739958.338254352.2	100	100	0	0	286	385	1	100	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.310822446.2	100	100	0	0	106	205	1	100	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.310822446.1	100	100	0	0	148	247	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.282759031.2	100	100	0	0	8	107	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.279444913.1	100	100	0	0	4	103	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.275406833.1	100	100	0	0	77	176	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.269993616.1	100	100	0	0	233	332	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.269319706.2	100	100	0	0	148	247	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.269319706.1	100	100	0	0	106	205	1	100	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.257357640.2	100	100	0	0	521	620	1	100	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.234812967.2	100	100	0	0	343	442	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.234812967.1	100	100	0	0	180	279	1	100	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.228069489.2	100	100	0	0	136	235	1	100	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.228069489.1	100	100	0	0	280	379	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.215518061.2	100	100	0	0	73	172	1	100	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.215518061.1	100	100	0	0	268	367	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.200295909.2	100	100	0	0	220	319	1	100	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.193926875.2	100	100	0	0	236	335	1	100	4.96E-43	185

query id	subject ids	% ident	Align length	mismatches	gap opens	q. start	q. end	s. start	s. end	evalue	bit score
mmn:midi_00973	gnl SRA SRR1739958.158491057.2	100	100	0	0	480	579	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.154259998.2	100	100	0	0	188	287	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.154259998.1	100	100	0	0	1	100	1	100	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.146762769.1	100	100	0	0	365	464	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.120662067.2	100	100	0	0	521	620	1	100	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.89510458.2	100	100	0	0	413	512	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.89510458.1	100	100	0	0	261	360	1	100	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.84863386.2	100	100	0	0	252	351	1	100	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.84863386.1	100	100	0	0	363	462	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.82461457.2	100	100	0	0	237	336	1	100	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.73717276.2	100	100	0	0	148	247	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.73717276.1	100	100	0	0	106	205	1	100	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.71696350.2	100	100	0	0	286	385	1	100	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.71696350.1	100	100	0	0	453	552	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.51559652.2	100	100	0	0	286	385	1	100	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.51559652.1	100	100	0	0	431	530	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.29576773.1	100	100	0	0	52	151	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.11530348.2	100	100	0	0	553	652	1	100	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.9604030.2	100	100	0	0	148	247	100	1	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.9604030.1	100	100	0	0	106	205	1	100	4.96E-43	185
mmn:midi_00973	gnl SRA SRR1739958.338254352.1	100	99	0	0	453	551	100	2	1.78E-42	183
mmn:midi_00973	gnl SRA SRR1739958.193926875.1	100	99	0	0	404	502	100	2	1.78E-42	183
mmn:midi_00973	gnl SRA SRR1739958.447455210.2	99	100	1	0	289	388	100	1	2.31E-41	180
mmn:midi_00973	gnl SRA SRR1739958.438717494.2	99	100	1	0	289	388	100	1	2.31E-41	180
mmn:midi_00973	gnl SRA SRR1739958.418140774.1	99	100	1	0	289	388	1	100	2.31E-41	180
mmn:midi_00973	gnl SRA SRR1739958.355243034.1	99	100	1	0	266	365	100	1	2.31E-41	180
mmn:midi_00973	gnl SRA SRR1739958.353635770.1	99	100	1	0	285	384	1	100	2.31E-41	180
mmn:midi_00973	gnl SRA SRR1739958.325106811.2	99	100	1	0	32	131	100	1	2.31E-41	180
mmn:midi_00973	gnl SRA SRR1739958.187190656.2	100	97	0	0	265	361	99	3	2.31E-41	180
mmn:midi_00973	gnl SRA SRR1739958.158491057.1	99	100	1	0	281	380	1	100	2.31E-41	180
mmn:midi_00973	gnl SRA SRR1739958.146762769.2	99	100	1	0	202	301	1	100	2.31E-41	180
mmn:midi_00973	gnl SRA SRR1739958.122004873.1	99	100	1	0	280	379	1	100	2.31E-41	180
mmn:midi_00973	gnl SRA SRR1739958.72902763.1	100	97	0	0	265	361	99	3	2.31E-41	180
mmn:midi_00973	gnl SRA SRR1739958.49454803.1	100	97	0	0	265	361	99	3	2.31E-41	180
mmn:midi_00973	gnl SRA SRR1739958.4347970.1	99	100	1	0	267	366	100	1	2.31E-41	180

query id	subject ids	% ident	Align length	mismatches	gap opens	q. start	q. end	s. start	s. end	evalue	bit score
mmn:midi_00973	gnl SRA SRR1739958.315862035.2	98.98	98	1	0	293	390	100	3	2.99E-40	176
mmn:midi_00973	gnl SRA SRR1739958.83727071.1	98.98	98	1	0	269	366	1	98	2.99E-40	176
mmn:midi_00973	gnl SRA SRR1739958.336270257.2	100	94	0	0	265	358	94	1	1.07E-39	174
mmn:midi_00973	gnl SRA SRR1739958.319088493.2	98	100	2	0	77	176	100	1	1.07E-39	174
mmn:midi_00973	gnl SRA SRR1739958.285827843.1	100	94	0	0	265	358	94	1	1.07E-39	174
mmn:midi_00973	gnl SRA SRR1739958.154491313.1	98	100	2	0	271	370	1	100	1.07E-39	174
mmn:midi_00973	gnl SRA SRR1739958.350914559.1	100	92	0	0	265	356	92	1	1.39E-38	171
mmn:midi_00973	gnl SRA SRR1739958.84923283.2	100	92	0	0	265	356	92	1	1.39E-38	171
mmn:midi_00973	gnl SRA SRR1739958.457936496.2	97	100	3	0	286	385	1	100	5E-38	169
mmn:midi_00973	gnl SRA SRR1739958.232900222.1	97.938	97	2	0	301	397	1	97	5E-38	169
mmn:midi_00973	gnl SRA SRR1739958.589337.1	100	91	0	0	265	355	10	100	5E-38	169
mmn:midi_00973	gnl SRA SRR1739958.290123677.1	96.97	99	3	0	303	401	100	2	1.8E-37	167
mmn:midi_00973	gnl SRA SRR1739958.326544767.2	96	100	4	0	287	386	100	1	2.32E-36	163
mmn:midi_00973	gnl SRA SRR1739958.299467271.2	96.875	96	3	0	306	401	100	5	8.36E-36	161
mmn:midi_00973	gnl SRA SRR1739958.151484234.2	96.875	96	3	0	306	401	100	5	8.36E-36	161
mmn:midi_00973	gnl SRA SRR1739958.248758838.2	100	86	0	0	265	350	86	1	3.01E-35	159
mmn:midi_00973	gnl SRA SRR1739958.231433115.2	96.842	95	3	0	307	401	1	95	3.01E-35	159
mmn:midi_00973	gnl SRA SRR1739958.168020516.2	100	86	0	0	265	350	86	1	3.01E-35	159
mmn:midi_00973	gnl SRA SRR1739958.21317508.1	100	86	0	0	265	350	86	1	3.01E-35	159
mmn:midi_00973	gnl SRA SRR1739958.62823511.2	100	83	0	0	265	347	18	100	1.4E-33	154
mmn:midi_00973	gnl SRA SRR1739958.432202720.1	100	82	0	0	265	346	82	1	5.03E-33	152
mmn:midi_00973	gnl SRA SRR1739958.217358191.2	97.727	88	2	0	265	352	13	100	5.03E-33	152
mmn:midi_00973	gnl SRA SRR1739958.89007022.2	96.667	90	3	0	312	401	100	11	1.81E-32	150
mmn:midi_00973	gnl SRA SRR1739958.459641862.1	96.629	89	3	0	313	401	1	89	6.51E-32	148
mmn:midi_00973	gnl SRA SRR1739958.256558119.1	100	80	0	0	581	660	1	80	6.51E-32	148
mmn:midi_00973	gnl SRA SRR1739958.402695082.2	96.591	88	3	0	314	401	100	13	2.34E-31	147
mmn:midi_00973	gnl SRA SRR1739958.358123609.1	96.591	88	3	0	314	401	1	88	2.34E-31	147
mmn:midi_00973	gnl SRA SRR1739958.192758034.2	95.556	90	4	0	312	401	100	11	2.34E-31	147
mmn:midi_00973	gnl SRA SRR1739958.480073825.1	100	77	0	0	265	341	24	100	3.03E-30	143
mmn:midi_00973	gnl SRA SRR1739958.111238000.1	100	77	0	0	265	341	24	100	3.03E-30	143
mmn:midi_00973	gnl SRA SRR1739958.161269218.1	98.701	77	1	0	265	341	24	100	1.41E-28	137
mmn:midi_00973	gnl SRA SRR1739958.155348461.1	96.386	83	3	0	319	401	100	18	1.41E-28	137
mmn:midi_00973	gnl SRA SRR1739958.55385831.2	96.386	83	3	0	319	401	100	18	1.41E-28	137
mmn:midi_00973	gnl SRA SRR1739958.13593542.2	96.386	83	3	0	319	401	100	18	1.41E-28	137
mmn:midi_00973	gnl SRA SRR1739958.341430134.1	96.341	82	3	0	320	401	1	82	5.07E-28	135

query id	subject ids	% ident	Alig length	mismatches	gap opens	q. start	q. end	s. start	s. end	evalue	bit score
mmn:midl_00973	gnl SRA SRR1739958.66114296.1	96.341	82	3	0	320	401	1	82	5.07E-28	135
mmn:midl_00973	gnl SRA SRR1739958.317796351.1	100	71	0	0	265	335	71	1	6.56E-27	132
mmn:midl_00973	gnl SRA SRR1739958.308656438.2	100	71	0	0	265	335	71	1	6.56E-27	132
mmn:midl_00973	gnl SRA SRR1739958.314697477.2	100	68	0	0	265	332	68	1	3.05E-25	126
mmn:midl_00973	gnl SRA SRR1739958.291819511.1	100	68	0	0	265	332	68	1	3.05E-25	126
mmn:midl_00973	gnl SRA SRR1739958.249782648.1	100	68	0	0	265	332	68	1	3.05E-25	126
mmn:midl_00973	gnl SRA SRR1739958.237643501.2	100	68	0	0	265	332	68	1	3.05E-25	126
mmn:midl_00973	gnl SRA SRR1739958.189648176.2	100	68	0	0	265	332	33	100	3.05E-25	126
mmn:midl_00973	gnl SRA SRR1739958.294044125.1	94.872	78	4	0	324	401	1	78	1.1E-24	124
mmn:midl_00973	gnl SRA SRR1739958.83106313.2	100	67	0	0	265	331	34	100	1.1E-24	124
mmn:midl_00973	gnl SRA SRR1739958.159628462.2	100	65	0	0	265	329	65	1	1.42E-23	121

Tabla 11. Resultado BLAST de la búsqueda de la secuencia mmn:midi_00391 en la anotación del transcriptoma del intestino medio de *I. ricinus*.

query id	subject ids	% ident.	Alig length	mismatches	gap opens	q. start	q. end	s. start	s. end	evalue	bit score
mmn:midi_00391	gnl SRA SRR1768390.2444696.1	100	202	0	0	321	522	1	202	4.19E-101	374
mmn:midi_00391	gnl SRA SRR1768387.1294562.1	99.27	137	1	0	379	515	8	144	2.65E-63	248
mmn:midi_00391	gnl SRA SRR1768392.2204204.1	100	127	0	0	260	386	1	127	2.07E-59	235
mmn:midi_00391	gnl SRA SRR1768390.2802311.1	100	127	0	0	260	386	1	127	2.07E-59	235
mmn:midi_00391	gnl SRA SRR1768387.264332.1	95.652	138	1	5	379	515	8	141	7.48E-54	217
mmn:midi_00391	gnl SRA SRR1768391.4799221.1	100	115	0	0	132	246	1	115	9.68E-53	213
mmn:midi_00391	gnl SRA SRR1768389.3819431.1	100	115	0	0	132	246	11	125	9.68E-53	213
mmn:midi_00391	gnl SRA SRR1768392.4839625.1	100	113	0	0	132	244	1	113	1.25E-51	209
mmn:midi_00391	gnl SRA SRR1768389.2707625.1	98.165	109	0	2	82	188	85	193	1.63E-45	189

Tabla 12. Resultado BLAST de la búsqueda de la secuencia mmn:midi_00900 en la anotación del transcriptoma del intestino medio de *I. ricinus*.

query id	subject ids	% ident.	Alig length	mismatches	gap opens	q. start	q. end	s. start	s. end	evalue	bit score
mmn:midi_00900	gnl SRA SRR1768400.192780.1	99.315	146	0	1	1	146	9	153	6.56E-68	263
mmn:midi_00900	gnl SRA SRR1768387.2741368.1	98.63	146	0	2	1	146	19	162	3.05E-66	257
mmn:midi_00900	gnl SRA SRR1768392.4119227.1	97.581	124	0	3	17	138	1	123	8.67E-52	209
mmn:midi_00900	gnl SRA SRR1768392.1732750.1	95.935	123	0	4	17	138	1	119	2.43E-47	195
mmn:midi_00900	gnl SRA SRR1768391.752964.1	95.935	123	0	4	17	138	1	119	2.43E-47	195