

0302/1
11



**UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO**

INSTITUTO DE INVESTIGACIONES BIOMEDICAS

CITOCROMOS P450 BACTERIANOS: EVIDENCIA DE UNA
TRANSFERENCIA HORIZONTAL

T E S I S

QUE PARA OBTENER EL GRADO DE
LICENCIADO EN INVESTIGACION
BIOMEDICA BASICA

P R E S E N T A :

ROBERTO JOSE RISCO FLORES

ASESORES: DRA. JULIETA RUBIO LIGHTBOURN

DR. RAFAEL CAMACHO CARRANZA



MEXICO, D. F.

2003

TESIS CON
FALLA DE ORIGEN



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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.

Agradecimientos

Dedico este trabajo a mis padres, por ellos estoy aquí, porque no perdieron la esperanza, por su amor y su entrega. Esta tesis es mía, pero es dedicada a Uds.

A mis sinodales, gracias por su consejo y por tomarse el tiempo de evaluar este trabajo.

A la Dra. Julieta Rubio, por su paciencia, comprensión, ayuda y consejo a lo largo de los años.

Al Dr. Rafael Camacho por creer en mí; y por sus enseñanzas, algunas de ciencia, y las más importantes, acerca de la vida.

A la técnica académica Clementina Castro por su compañerismo que nunca mengua, su optimismo y su ayuda.

Al todo personal académico y administrativo del IIB que de alguna manera intervino en todo el tiempo que he estado en el instituto, y que sin duda me ayudaron mucho.

Al Lic. Brito, a la Sra. Mary y todas las personas en la biblioteca que me ayudaron cuando hizo falta.

A mi Jéssica por ser una oreja amiga, una cómplice y un verdadero equipo.

A mi Gus, porque él es mi inspiración y mi motor, me recuerda todos los días que nunca me rendiré.

A Eduardo, Manya, Manyita y Sergio por estar siempre cerca con una sonrisa, por el amor que siempre me han ofrecido, en el tiempo y la distancia.

A toda la demás familia, mis abuelos, tíos, primos, sobrinos, etc., que no puedo mencionar por motivo de espacio y tiempo, pero que sin duda han participado en mi vida de muchas maneras.

A todo mis amigos porque se que sus críticas y burlas solo me dieron más coraje para no aventar la toalla y seguir.

A Mario y Norma por los consejos certeros, los regaños necesarios y haberme escuchado siempre que lo necesité.

A Luis y Ana por ser mis amigos y echarme la mano cuando tengo momentos difíciles.

A Paula, porque estuviste cuando tenías que estar, porque nunca dudaste que lo lograría.

A María Luisa, que ha llenado un espacio de amistad que nadie podría llenar.

A los amigos de mi generación, Eileen, Sergio y Emilio (porque uds. si supieron de que lado estaba la canela fina). A Adriana (quien increíblemente sigue siendo mi amiga, después de tanto tiempo); a Oscar P., Oscar R., Michael, Juan Ramón, Charis, y todos los que son mis amigos, amigas, conocidos y querencias a quienes no menciono por el temor de olvidar algún nombre.

A conocidos y desconocidos que me ofrecieron una sonrisa cuando yo tenía una cara larga. Y a todos los que sin querer o queriendo me ayudaron a llegar, les dedico parte de este esfuerzo.

TESIS CON
FALLA DE ORIGEN

ÍNDICE

CITOCROMOS P450 BACTERIANOS: EVIDENCIA DE UNA TRANSFERENCIA HORIZONTAL	3
RESUMEN	3
INTRODUCCIÓN	4
ANTECEDENTES	4
LAS BACTERIAS <i>CAMPYLOBACTER JEJUNI</i> Y <i>HELICOBACTER PYLORI</i>	5
LA TRANSFERENCIA GENÓMICA HORIZONTAL	8
LA TRANSFERENCIA GENÉTICA HORIZONTAL DIRIGE LA FORMACIÓN DE OPERONES	9
ANÁLISIS FILOGENÉTICOS	13
LA APROXIMACIÓN FILOGENÉTICA	16
HIPÓTESIS	19
OBJETIVOS	19
OBJETIVO GENERAL	19
OBJETIVOS PARTICULARES	19
MATERIAL Y MÉTODOS	21
ANÁLISIS TEÓRICOS	21
DEFINICIONES	21
MATERIAL	23
MÉTODOS	27
BÚSQUEDA DE CITOCROMOS P450 EN <i>H. PYLORI</i> POR ANÁLISIS CON BLAST ("WESTERN COMPUTACIONAL")	27
ÁRBOLES FILOGENÉTICOS	28
FRECUENCIA DE SECUENCIAS ESPECÍFICAS DE NUCLEÓTIDOS EN EL CROMOSOMA	31
DISTRIBUCIÓN DE ORFs EN EL GENOMA DE <i>CAMPYLOBACTER JEJUNI</i>	32
FRECUENCIA DE ORTÓLOGOS	32
POSIBLE DONADOR	33
RESULTADOS	34
CITOCROMOS P450 EN <i>H. PYLORI</i> Y EN <i>C. JEJUNI</i>	34
ANÁLISIS FILOGENÉTICO DE <i>H. PYLORI</i> Y <i>C. JEJUNI</i>	35
DISTRIBUCIÓN DE LOS CITOCROMOS P450 BACTERIANOS	38
CITOCROMOS P450 EN <i>MYCOBACTERIUM TUBERCULOSIS</i>	40
ÁRBOLES FILOGENÉTICOS DE CITOCROMOS P450 Y PROTEÍNAS CONSERVADAS	42
FRECUENCIA DE SECUENCIAS ESPECÍFICAS DE NUCLEÓTIDOS EN EL CROMOSOMA	44
DISTRIBUCIÓN DE ORFs EN EL GENOMA DE <i>CAMPYLOBACTER JEJUNI</i>	47

TESIS CON
FALLA DE ORIGEN

FRECUENCIA DE ORTÓLOGOS	50
POSIBLE DONADOR	54
DISCUSIÓN	56
CONCLUSIONES	66
APÉNDICE 1	67
APÉNDICE 2	98
APÉNDICE 3	107
BIBLIOGRAFÍA	114

TESIS CON
FALLA DE ORIGEN

CITOCROMOS P450 BACTERIANOS: EVIDENCIA DE UNA TRANSFERENCIA HORIZONTAL

RESUMEN

Los fenómenos de Transferencia Genética Horizontal (TGH) pueden permitir que proteínas pertenecientes a un grupo bacteriano sean compartidas con otros grupos bacterianos, los cuales anteriormente no las poseían. Las bacterias *C. jejuni* y *H. pylori* pertenecen ambas al grupo delta-epsilon de las proteobacterias y guardan una cercana relación filogenética, más del 50% de sus genes son ortólogos, los cuales se encuentran distribuidos a lo largo del genoma. A pesar de eso *C. jejuni* tiene una región de aprox. 40Kb en la cual sólo es posible encontrar genes ortólogos con *H. pylori*, y que contiene un gene para un citocromo P450, del cual carece *H. pylori*. La presencia de dicha región de aprox. 40Kb podría deberse a un fenómeno de TGH en *C. jejuni* o la ausencia de estos genes en *H. pylori* podría deberse a la pérdida de los mismos. La ocurrencia de fenómenos de TGH suelen ser difíciles de detectar, se pueden obtener conclusiones erradas si los análisis se basan únicamente en búsquedas de secuencias a través de BLAST o debido a las limitaciones intrínsecas de las reconstrucciones filogenéticas. En este trabajo, buscamos construir un cuerpo de evidencia a través de distintos enfoques. Utilizamos un análisis filogenético con citocromos P450 utilizando dos distintos modelos de reconstrucción filogenética, y analizamos la frecuencia de ortólogos, la distribución de genes y la frecuencia acumulada de oligonucleótidos a lo largo de todo el genoma de *C. jejuni*. Cada uno de estos análisis proveyó evidencia de que los ORFs desde Cj1410 hasta Cj1449 de *C. jejuni* son resultado de una TGH la cual ocurrió posteriormente a que *C. jejuni* y *H. pylori* divergieron.

TESIS CON
FALLA DE ORIGEN

INTRODUCCIÓN

ANTECEDENTES

Existe una asociación clara y reconocida de la bacteria *Helicobacter pylori* con el cáncer gástrico. *Helicobacter pylori* es la principal causa de gastritis tipo B, úlcera gástrica y úlcera duodenal, y es factor de riesgo tipo I para cáncer gástrico (13, 42). La infección por *H. pylori* podría estar favoreciendo una reacción carcinogénica a través de diversos mecanismos, entre ellos la acción metabólica de los propios citocromos P450 del *Helicobacter pylori*. Dado que la infección por *H. pylori* es una infección crónica, que puede durar desde la niñez hasta la vejez, y que la mayoría de los carcinógenos y pro-carcinógenos que entran en el cuerpo humano lo hacen por medio del sistema digestivo, se presentaría un escenario idóneo en el que el citocromo P450 de *H. pylori* podría ser importante en la aparición del cáncer gástrico. Previamente a la publicación del genoma de *H. pylori* (48) intentamos buscar evidencia de algún citocromo P450 en el genoma de esta bacteria utilizando técnicas de western blot, sin éxito.

Una vez publicada la secuencia completa del genoma de *H. pylori* pudimos saber con certeza que no tiene ningún citocromo P450, sin embargo, *C. jejuni* que es una bacteria cercana sí tiene un citocromo P450. Es posible conocer si el ancestro común de ambas bacterias poseía un citocromo y *H. pylori* perdió este gene, o si por el contrario el ancestro común carecía de este gene y *C. jejuni* lo adquirió por THG, utilizando herramientas filogenéticas y análisis del entorno genómico.

TESIS CON
FALLA DE ORIGEN

LAS BACTERIAS *Campylobacter jejuni* y *Helicobacter pylori*

C. jejuni y *H. pylori* están cercanamente relacionados filogenéticamente en base a su rRNA 16S (32), comparten muchas propiedades biológicas y fueron originalmente clasificadas dentro del género *Campylobacter*. Bacterias con forma de espiral fueron encontradas por vez primera en la mucosa gástrica desde hace casi cien años por Pel, en el año de 1899; sin embargo estas bacterias fueron aisladas por primera vez en 1982 en cultivos de biopsias de pacientes que sufrían de gastritis y úlcera péptica (22,51). En 1989 Goodwin propuso un nuevo género, *Helicobacter*, y las bacterias conocidas como *Campylobacter pylori* fueron rebautizadas como *Helicobacter pylori* (9).

Campylobacter jejuni es causa importante de la patología diarreica bacteriana en todo el mundo debido a la ingesta de alimentos contaminados, y es el antecedente más frecuente de una forma de parálisis neuromuscular conocida como síndrome de Guillian-Barré. Tanto *Campylobacter jejuni* como *Helicobacter pilory* pertenecen al grupo delta-epsilon de proteobacterias. Ambas son bacterias bacilares, Gram-negativas, espirales o ligeramente curvas y flageladas; son estas características que comparten la razón por la cual se clasificaron como pertenecientes al mismo género cuando *H. pylori* fue inicialmente descrita. A pesar de la cercana relación filogenética de *C. jejuni* y *H. pylori* (32), las similitudes más importantes están confinadas a funciones de mantenimiento (housekeeping); sólo 55.4% de los genes de *C. jejuni* tienen ortólogos en *H. pylori*. En la mayoría de las funciones de sobrevivencia, transmisión y patogénesis, estos organismos tienen poco en común. Lo anterior indica que presiones de selección han generado profundos cambios evolutivos para crear dos organismos patógenos distintos, que se

TESIS CON
FALLA DE ORIGEN

han desarrollado específicamente en distintos nichos a partir de un ancestro común relativamente cercano.

Los genomas de *Campylobacter jejuni* y *Helicobacter pylori* están secuenciados en su totalidad (32). Esto permite realizar análisis y búsquedas por genes conocidos en el genoma. En el genoma de *Campylobacter jejuni* se encontró un gene putativo localizado en la posición 1342550 – 1343911 que codifica para un citocromo P450, y en *H. pylori* no es posible encontrar ningún gene que corresponda con este tipo de citocromo. En el gene puede reconocerse la firma específica de unión al grupo hemo (PS00086) propia de los citocromos P450, y empalma con el dominio PF00067 (Pfam) de citocromos P450.

El genoma de *C. jejuni* se compone de un cromosoma circular de 1,641,481 pares de bases (*C. Jejuni* NCTC11168), con un 30.6% de G+C, predicho a contener 1654 proteínas y 54 especies de RNA estables. En estas 1654 secuencias codificantes predichas (CDS), la longitud promedio de los genes es de 948 pares de bases, y el 94.3% del genoma codifica para proteínas, haciéndolo el genoma bacteriano más denso secuenciado a la fecha (32). En general el 61.1% de las secuencias codificantes son transcritos en la misma dirección que la replicación y el origen de replicación está cerca del inicio del gene *dnaA*. Se identificó un probable sitio de terminación de la replicación en la posición 814319..814352, en el cual se forma una horquilla de 15 pares de bases. Se encontraron dos grandes regiones de bajo contenido de G+C, que involucran las secuencias codificantes Cj1135 – Cj1148 (25.4%) y Cj1421 – Cj1442 (26.5%); estos CDS corresponden a genes de los grupos de biosíntesis para lipooligosacáridos (LOS) y polisacáridos extracelulares (EP) respectivamente. El citocromo P450 corresponde a la secuencia codificante Cj1411, la cual se encuentra cercana a la segunda región con un bajo contenido de G+C.

TESIS CON
FALLA DE ORIGEN

Se puede deducir información funcional (homología con genes de función conocida o perfiles de hidrofobicidad) para 77.8% de los 1654 CDs, mientras que el 13.5% tienen homología con genes de función desconocida en la base de datos y 8.7% no tienen secuencias similares en la base de datos o cualquier otra información funcional (32).

Otra característica interesante del genoma de *C. jejuni* es que carece casi completamente de secuencias repetitivas de DNA; hay nueve secuencias repetidas en el genoma entero. Exceptuando a Cj0742, que es similar a una parte de IS605 tnpB de *H. pylori*, no existe evidencia para ninguna secuencia de elementos de inserción (IS) funcionales, transposones, retrones o profágos en todo el genoma.

Debido a su cercanía filogenética y a la relación que guardan los genomas de *H. pylori* y *C. jejuni*, habría la posibilidad de que *H. pylori* tuviera un gene ortólogo al menos distantemente relacionado con el gene para citocromo P450 de *C. jejuni* o con el dominio Pfam 00086 común a todos los citocromos P450. Dos distintos escenarios podrían explicar la evidencia de que *H. pylori* no tiene ningún gene relacionado con citocromos P450. Los fenómenos de pérdida de genes por procesos estocásticos ocurren de manera normal en los genomas bacterianos, cuando se trata de genes con una baja selección (4), si el ancestro común de ambas bacterias tenía un citocromo P450 la pérdida estocástica del mismo podría explicar la ausencia del citocromo P450 en *H. pylori*. También es posible que el ancestro común de *H. pylori* y *C. jejuni* careciera de un citocromo P450 y ambas bacterias después que divergieron evolutivamente por supuesto carecían de él, sin embargo en algún momento de su historia *C. jejuni* pudo haber adquirido un citocromo P450 de otra bacteria, a través de un fenómeno de transferencia horizontal.

TESIS CON
FALLA DE ORIGEN

LA TRANSFERENCIA GENÓMICA HORIZONTAL

El genoma de los organismos se deriva de una transmisión vertical, la herencia de los cromosomas de una generación a la siguiente. Sin embargo, en raras instancias, los genomas pueden ser también modificados por la transferencia horizontal (también llamada lateral) de genes (TGH); la adquisición de material genético de un organismo distinto, este material genético se convierte más tarde en parte permanente del genoma recipiente (46). La transferencia de información genética intraespecie típicamente se conoce como recombinación entre aislados bacterianos, la TGH involucra a organismos de distintas especies. A pesar de que estos intercambios (TGH) no ocurren con mucha frecuencia en un período de tiempo de generación a generación, una cantidad significativa de estos eventos pueden ocurrir a lo largo de cientos de millones de años. Un ejemplo extremo de este fenómeno es el origen endosimbionte propuesto para las mitocondrias en las células eucariontes y los cloroplastos en las plantas (43, 46).

La transferencia genética horizontal es una fuente importante de variación genética en bacterias, que les permite incurrir en nuevos y variados ambientes (26, 27, 28). Tales transferencias son posibles por una variedad de mecanismos naturales de las bacterias para transferir DNA de una especie a otra. La detección de TGH puede realizarse debido al hecho de que el genoma de cada especie bacteriana tiene una composición de bases única, por esto, la transferencia de una porción de genoma de un organismo a otro puede generalmente ser detectada como una región con una composición diferente del recipiente. Las transferencias muy antiguas podrían no ser detectables de esta manera debido a que la composición de bases y el uso de codones del DNA transferido eventualmente se homogeneizarán con el del organismo

TESIS CON
FALLA DE ORIGEN

receptor. El momento de la transferencia podría ser estimado por el grado al cual la composición del DNA de la TGH se ha confundido con el genoma del receptor. Las comparaciones de genomas bacterianos que han sido secuenciados en su totalidad han revelado que estos genomas son en realidad mosaicos de transferencias ancestrales (20). La proporción del genoma que puede deberse a TGH varía además de acuerdo aproximadamente al tamaño del genoma; en *E. coli* el 12.8% del genoma se debe a TGH (el mayor nivel encontrado), mientras que en *Mycoplasma genitalium*, cuyo genoma es de menos de un cuarto del tamaño del de *E. coli*, parece ser de 0%. *Mycoplasma* perdió muchos de sus genes necesarios para ser un organismo de vida libre y depende de los nutrientes proveídos por la célula huésped; esta clase de organismo no puede contener material genético extra o que no es indispensable. La TGH participa también de manera importante en la patogenicidad de muchas bacterias, este DNA frecuentemente cuenta con secuencias repetidas en los extremos, características de los elementos transponibles.

LA TRANSFERENCIA GENÉTICA HORIZONTAL DIRIGE LA FORMACIÓN DE OPERONES

La TGH junto con la agrupación de genes podría ser un participante en los mecanismos evolutivos, dado que si los genes se encuentran agrupados en el genoma del organismo donador, el organismo receptor podría adquirir toda una vía metabólica de otro organismo por TGH. Los primeros ejemplos de genes agrupados fueron encontrados en bacterias por Demerec y Hartman en 1959, casos tales como *E. coli* y *Salmonella typhimurium*. Entre los procariontes las enzimas requeridas para una vía metabólica particular se codifican por genes que se encuentran frecuentemente físicamente cercanos. Aún más sorprendente es el hallazgo de que genes pertenecientes a una misma vía bioquímica

TESIS CON
FALLA DE ORIGEN

tienden a estar ubicados en el mismo orden de sus reacciones metabólicas. El orden de los 10 genes dentro de los operones de la biosíntesis de histidina en *E. coli* y *S. typhimurium* es casi idéntico al orden de sus reacciones químicas (19). Hay cuatro modelos propuestos para explicar el agrupamiento que puede verse en los procariontes: el modelo Natal, el modelo de Fisher, el de Co-regulación y el del Operón "Egoísta".

El modelo Natal postula que los genes están agrupados porque "nacieron" de esa manera. Horowitz propuso en 1945 que las vías de síntesis pudieron haberse formado de manera escalonada, el gene para una nueva proteína surgía cuando un nutriente se volvía escaso en el medio ambiente, cada nueva enzima permitía el uso de otro compuesto como precursor para sintetizar el nutriente. Dado que estas enzimas actuarían sobre sustratos similares, el acomodo de genes se veía como algo favorable (21). En 1951 Lewis propuso que la duplicación y diferenciación de genes permitiría tener a estos loci unidos con funciones similares. Estas ideas perdieron mérito cuando pudieron conocerse las secuencias de aminoácidos. Prácticamente todos los operones bacterianos están compuestos por genes que no muestran homología o que tienen una similitud muy baja, muchos están relacionados con genes que no forman parte del operón, y que codifican proteínas que catalizan reacciones mecanísticamente similares (21).

El modelo de Fisher postula que los genes se agrupan si alelos específicos trabajan bien juntos, los grupos de genes pueden reflejar coadaptación. Fisher observó en 1930 que cuando alelos específicos de dos genes trabajaban bien juntos, esto aumentaría su unión. Este aumento de unión podía observarse en la selección de ciertos genotipos y la contraselección de los recombinantes. Esta idea fue extendida más tarde sugiriendo que tal selección podría llevar al agrupamiento físico de los genes. La proximidad física reduciría la frecuencia de eventos de

TESIS CON
FALLA DE ORIGEN

recombinación que pudieran romper el loci coadaptado. En bacteriófagos los genes que codifican para proteínas que funcionan unas con otras en grupos (en la cabeza o en la cola) se encuentran agrupados. De acuerdo al modelo de Fisher, estos arreglos permitirían recombinación entre bacteriófagos para generar nuevas combinaciones de grupos pero no romperían los agrupamientos de genes cuyos productos deben trabajar con mucha cercanía. Es difícil vislumbrar cómo el modelo de Fisher pudiera dirigir el agrupamiento de genes para los procesos metabólicos bacterianos. Las diferentes enzimas en una vía metabólica, en el caso de que interaccionaran físicamente, lo hacen en un grado mucho menor que las proteínas estructurales. Es poco probable que los alelos de genes metabólicos estén coadaptados para trabajar con alelos particulares de otros genes en la misma vía metabólica.

El modelo de correulación supone que cuando los genes se encuentran juntos pueden ser inducidos o reprimidos simultáneamente con el control en un solo sitio, llamado operador. Este control ofrecería un mecanismo selectivo por el cual un grupo de genes proveerían de una ventaja selectiva sobre los mismos genes en lugares dispersos; la ventaja regulatoria de un operón se deriva de la cotranscripción de múltiples genes con un solo promotor. Sin embargo la fuerza de este modelo como la entidad causal del agrupamiento de genes disminuyó al encontrarse genes correulados que no se localizan juntos; estos casos mostraron que el agrupamiento no es un requisito para la correulación (21). Otro problema serio de este modelo es que no provee de una serie de pasos intermedios plausibles de la evolución de los grupos de genes. Si el valor selectivo del agrupamiento de genes fuera sólo el operón, el proceso de formación del operón tendría que ocurrir en un solo paso, hay una paradoja en este modelo debido a que un promotor bien regulado responsable por el agrupamiento de genes no puede proveer del

TESIS CON
FALLA DE ORIGEN

beneficio máximo hasta que todos los genes de un vía metabólica estuviesen agrupados. El beneficio de la contrainscripción de producir cantidades equimolares de las proteínas no se emplea, debido a las distintas eficiencias metabólicas de las diferentes enzimas, no es probable que cantidades equimolares producidas por un solo transcrito sean benéficas. En realidad se ha observado que genes de un mismo operón frecuentemente muestran diferentes eficiencias de traducción y diferentes vidas medias de sus RNAs mensajeros, lo que contribuye a niveles no uniformes de las proteínas codificadas en el mismo operón (21).

El modelo del operón "Egoísta" explica la formación de grupos de genes a través de fenómenos de TGH entre los organismos. Las transferencias horizontales entre los procariontes son mediadas por procesos comunes (46) que incluyen bacteriófagos de trasducción, plásmidos conjugativos o la toma de material genético extraño directo del medio (21). Los genomas bacterianos codifican tanto funciones críticas como funciones meramente útiles, estas últimas proveen beneficios esporádicos pero no son necesarias continuamente. Tales funciones no críticas incluyen la degradación de compuestos inusuales como fuentes de carbono. Dentro de este modelo estas funciones se definen como "seleccionadas débilmente" (FSD). Durante períodos de relajación de la selección, los genes de FSD pueden acumular sustituciones de bases rápidamente dado que la selección natural no funcionaría para eliminar a los alelos nulos de la población. Inclusive las células con múltiples defectos en una vía metabólica podrían tener una ventaja selectiva sobre aquellas que tienen defectos en un solo gene de la vía, ya que una sola mutación puede permitir la acumulación de productos metabólicos intermedios tóxicos (27,36). La TGH le permite a genes para FSD, escapar de la extinción, movilizándolo a genes al organismo receptor antes de que desaparezcan del organismo donador por procesos estocásticos.

TESIS CON
FALLA DE ORIGEN

Generalmente se necesita más de un gene de FSD para que el organismo receptor tenga completa la FSD. La transferencia horizontal de un solo gene para una FSD no supondría ninguna ventaja selectiva para el receptor, y los genes adquiridos que no confieren un beneficio selectivo no pueden elevar su frecuencia en la población, lo que llevaría de nuevo a su pérdida por deleción o acumulación de mutaciones. Para que el organismo receptor obtenga una ventaja selectiva por una nueva vía metabólica es necesario que todos los genes para una FSD se transfieran simultáneamente.

El agrupamiento es una propiedad "egoísta" de los genes que lo forman. Si los genes necesarios para una FSD están regados por el cromosoma, la única forma en que pueden propagarse es por transferencia vertical. Sin embargo, si estos mismos genes se encuentran arreglados en un mismo grupo pueden propagarse tanto vertical como horizontalmente, asegurando así escapar de la extinción (21). La proximidad física provee de una fuerte ventaja a los genes mismos en la competencia contra genes dispersos en el cromosoma. La TGH acelera la formación de estos grupos puesto que la eficiencia de transferencia horizontal aumenta proporcionalmente con la proximidad de los genes. Además la TGH puede contribuir a la formación de operones compactos haciendo al DNA espaciador entre dos genes no esencial; cuando una célula recibe material genético, éste será extraño y todo aquel material que no sea esencial para el crecimiento de la célula se perderá (21,27).

ANÁLISIS FILOGENÉTICOS

Para hacer posible la detección de TGHs es necesario poder distinguir apropiadamente una especie bacteriana de otra y cómo se relacionan evolutivamente entre ellas. Las bacterias parecen tener pocas

TESIS CON
FALLA DE ORIGEN

características que puedan utilizarse como base para establecer relaciones filogenéticas que sirvan para su clasificación. Tradicionalmente los bacteriólogos han confiado en las características morfológicas y algunas propiedades metabólicas para poder establecer una clasificación taxonómica y diferenciar distintos aislados bacterianos en especies, pero es difícil saber hasta qué grado esta clasificación refleja realmente las relaciones evolutivas que hay entre los diferentes tipos de bacterias (29); se ha presentado el caso de bacterias que ahora se sabe que pertenecen a distintas especies, y que cuando fueron descritas por primera vez fueron confundidas como una misma, por ejemplo cuando *Helicobacter pylori* fue caracterizado por primera vez (51) se pensó que pertenecía al género *Campylobacter*, debido a su parecido morfológico con las bacterias de este género, como *Campylobacter jejuni*.

A pesar de las dificultades inherentes al intentar elucidar la relación evolutiva entre bacterias, la comparación de secuencias de macromoléculas, principalmente RNA ribosomal, ha logrado disipar muchas de las imprecisiones que se presentaban en estos estudios. Con la ayuda de estas secuencias se ha logrado identificar especies íntimamente relacionadas y el agrupamiento jerárquico de las especies de una manera que refleje cercanamente las historias evolutivas de los organismos.

La confianza que se deposita en la comparación de secuencias de macromoléculas descansa en el modelo que tenemos de cómo opera la evolución: las células adquieren la mayor parte de su material genético de sus ancestros (transferencia vertical), y estos ácidos nucleicos acumulan cambios con el paso del tiempo. El grado de similitud entre secuencias pertenecientes a distintas bacterias refleja qué tan cercanas están en tiempo evolutivo de un ancestro común.

Todas las formas de vida extantes en la tierra comparten un origen común y sus ancestros pueden ser rastreados a uno o algunos organismos.

TESIS CON
FALLA DE ORIGEN

Organismos relacionados cercanamente descienden de ancestros comunes más recientes que aquellos relacionados lejanamente. Los objetivos de los análisis filogenéticos son: reconstruir la genealogía correcta entre los organismos y estimar el tiempo de divergencia entre ellos. Un árbol filogenético es un gráfico compuesto de nodos y ramas, en el cual una sola rama conecta con dos nodos adyacentes. Los nodos representan las unidades taxonómicas, las cuales pueden ser especies, poblaciones, individuos o genes; y las ramas definen las relaciones entre ellas. El patrón de las ramas de un árbol filogenético es su topología. Los nodos internos de un árbol representan las unidades taxonómicas ancestrales; los nodos externos representan las unidades taxonómicas que se están comparando y se refieren como Unidades Taxonómicas Operacionales (OTU)(15). Además de establecer relaciones evolutivas entre las especies, los árboles filogenéticos pueden proveer información acerca de patrones de cambio, distribución y tiempos relativos de los eventos que ocurren en la evolución bacteriana.

Los árboles filogenéticos pueden o no tener raíz. En un árbol con raíz existe un nodo particular, llamado raíz, del cual una dirección única lleva a cualquier otro nodo; la dirección de este camino corresponde al tiempo evolutivo y la raíz es el ancestro común a todos los OTUs en estudio. Un árbol sin raíz sólo representa las relaciones entre los OTUs y no define un camino evolutivo.

La secuencia de los eventos de especiación que lleva a la formación de un grupo de OTUs históricamente única, por lo tanto, sólo uno de todos los posibles árboles que se pueden construir con un cierto número de OTUs representa la historia evolutiva *real*. Un árbol obtenido utilizando un cierto grupo de datos y un método de reconstrucción filogenética se llama árbol inferido. Un árbol inferido puede o no ser idéntico al árbol real.

TESIS CON
FALLA DE ORIGEN

LA APROXIMACIÓN FILOGENÉTICA

La comparación de secuencias, como cambios en la estructura primaria de RNA, DNA, proteínas; puede proveer de la información necesaria para cuantificar las relaciones evolutivas entre genomas bacterianos. La información genotípica tiene varias ventajas en relación a la fenotípica para establecer relaciones filogenéticas (29,30) : 1) los cambios en las secuencias suelen ser discretos e identificables; 2) cada secuencia contiene cientos o miles de caracteres que pueden proveer de información potencialmente útil; 3) las secuencias acumulan cambios consistentemente, aun cuando el fenotipo no cambie; y 4) las manifestaciones fenotípicas son generalmente multifactoriales, lo cual puede llevar a interpretaciones incorrectas haciéndolas inadecuadas para las inferencias filogenéticas.

Se debe tener en cuenta que los análisis moleculares no pueden eliminar por completo los problemas de reconstruir una relación filogenética (38). A pesar de que se supone que la mayoría de la información genética refleja fielmente la historia evolutiva (43), parte de la información molecular puede ser contradictoria debido a la acumulación estocástica de cambios. Las moléculas de DNA llevan sólo cuatro caracteres distintos, bases idénticas no necesariamente representan una similitud ancestral. Dado que una porción pequeña de DNA puede proveer de una vasta información, sólo se usan pequeños segmentos de un genoma, que pueden ser o no representativos de la totalidad del material genético.

Para minimizar los problemas que se presentan, el análisis filogenético se puede resumir en los siguiente pasos (29):

- a) Seleccionar la molécula a analizar. Dado que los genomas contienen elementos que evolucionan a diferentes velocidades, el alcance de la

TESIS CON
FALLA DE ORIGEN

investigación determina la molécula de elección. Los rRNA evolucionan lentamente y pueden ser utilizados para determinar la relación entre especies (taxa) distantes; pero estas moléculas son ineficaces para establecer relaciones entre especies cercanas. Las secuencias codificadoras comprenden dos clases de mutaciones (sinónimas y no-sinónimas) que acumulan cambios a distintas velocidades y que por ello son útiles para establecer relaciones cercanas entre especies. Las sustituciones sinónimas no alteran la composición de aminoácidos de la proteína y se considera que la presión de selección es mínima y se supone que se acumulan las sustituciones a una velocidad cercana al de la tasa de mutación neutral.

- b) Alineamiento de las secuencias. La confiabilidad de un análisis filogenético descansa en la asignación de los caracteres homólogos. Ya que el alineamiento que obtiene el mayor grado de correspondencia no representa necesariamente sitios ortólogos, es necesario revisar la información cuidadosamente, las posiciones nucleotídicas a comparar deben estar relacionadas por herencia.
- c) Inferir relaciones filogenéticas. Una vez que han sido identificados los sitios homólogos, se aplican distintos modelos matemáticos para traducir las diferencias en la secuencias a distintos eventos evolutivos. Estos resultados se representan generalmente en forma de árboles filogenéticos, que muestran el grado de divergencia de las especies en el tiempo, desde una especie primordial reconstruida. Los métodos más frecuentemente usados incluyen máxima parsimonia, que examina características únicas que son compartidas por las especies y frecuentemente usado cuando las diferencias son pocas y generalmente, irreversibles; parsimonia evolutiva, que se basa en el análisis de mutaciones transversas para discriminar entre tres posibles

TESIS CON
FALLA DE ORIGEN

relaciones para cuatro especies y se usa frecuentemente para elucidar relaciones entre especies muy distantes; métodos de distancia fenética (por ejemplo el de "neighbor joining") que se basan en la consolidación de la extensión de la diferencia entre especies; y métodos de máxima verosimilitud, que intentan inferir las relaciones que con mayor probabilidad puedan haber producido los datos, a diferencia de los métodos anteriores que intentan deducir la filogenia a partir de los datos.

- d) Asegurarse de la confiabilidad de la filogenia. Después de inferir las relaciones filogenéticas, es posible revisar hasta un cierto punto su confianza estadística. La utilidad de las filogenias para identificar diferencias biológicas a partir la historia evolutiva molecular radica en la robustez de las relaciones inferidas. El método más usado para revisar la integridad de un árbol filogenético es el de "bootstrap", el cual sistemáticamente infiere nuevas relaciones revolviendo el grupo de datos y comparándolas con la filogenia inferida del grupo de datos original. El análisis de bootstrap proporciona un resultado para cada uno de los nodos de un árbol; los valores de confiabilidad del método de bootstrap dependen del número de iteraciones que se usen para el análisis, cuanto más cercano sea el resultado del bootstrap en un nodo al número total de iteraciones que se usaron para el análisis, mayor es la confiabilidad para ese nodo en particular.

TESIS CON
FALLA DE ORIGEN

HIPÓTESIS

El ancestro común de *H. pylori* y *C. jejuni* no tuvo un citocromo P450, pero *C. jejuni* recibió un fragmento con un citocromo P450 a través de una Transferencia Genética Horizontal.

OBJETIVOS

OBJETIVO GENERAL

Investigar si *C. jejuni* pudo haber adquirido por una TGH una región conteniendo a un citocromo P450.

OBJETIVOS PARTICULARES

Determinar si los genes colindantes al gene para citocromo P450 de *C. jejuni* tienen homólogos en el genoma de *H. pylori*

Realizar análisis filogenéticos para evaluar la posibilidad de una TGH en *C. jejuni* y si es posible que los citocromos P450 bacterianos estén involucrados en fenómenos de TGH.

Analizar la frecuencia y distribución de genes de *C. jejuni* que tengan genes ortólogos en *H. pylori*.

Analizar el patrón de distribución de oligonucleótidos en el genoma de *C. jejuni*, buscando zonas anómalas en el genoma.

Analizar la organización de genes en el genoma de *C. jejuni* buscando grupos de genes en "tandem".

TESIS CON
FALLA DE ORIGEN

Evaluar si es posible definir la extensión de la TGH dado que encontremos evidencia de alguna.

Buscar algún posible donador del gene del citocromo P450 de *C. jejuni*.

TESIS CON
FALLA DE ORIGEN

MATERIAL Y MÉTODOS

ANÁLISIS TEÓRICOS

DEFINICIONES

Los parámetros necesarios para el uso de programas de comparación de secuencias son los siguientes:

MATRICES DE PUNTUACIÓN O DE SUBSTITUCIÓN (SCORING MATRICES)

Las matrices de sustitución utilizan un modelo evolutivo para otorgar puntuaciones a las sustituciones de aminoácidos, las cuales no ocurren con la misma frecuencia. Las matrices BLOSUM (Blocks Amino Acid Substitution Matrices) asignan una puntuación probabilística para cada posición en el alineamiento que está basada en la frecuencia con la que se supone que cada sustitución ocurre dentro de bloques conservados de proteínas relacionadas. También son utilizadas las matrices de sustitución de aminoácidos de Dayhoff o matrices PAM (Percent Accepted Mutation), estas matrices hacen un listado de la probabilidad de cambio de un aminoácido a otro en secuencias proteicas homólogas durante la evolución. Gonnet, Cohen y Brenner desarrollaron un método diferente para medir las diferencias entre aminoácidos haciendo alineamientos de las bases de datos de proteínas existentes hasta ese momento.

- Las matrices BLOSUM y de Gonnet han mostrado empíricamente ser las mejores para detectar similitudes débiles entre proteínas

PENALIDADES DE APERTURAS POR CORTES Y DE LA EXTENSIÓN DE LOS CORTES

TESIS CON
FALLA DE ORIGEN

Un corte (gap) es introducido en un alineamiento para compensar por las inserciones y pérdidas entre una secuencias y otra. Para prevenir que se acumulen muchos cortes en un alineamiento, la introducción de un corte causa que la puntuación del alineamiento se reduzca en una cantidad fija. La extensión del corte para que se puedan considerar nucleótidos o aminoácidos adicionales también se penaliza en la puntuación del alineamiento

- BLAST 2.0 por defecto permite el uso de cortes en los alineamientos. La opción de cortes permite al programa BLAST detectar tanto alineamientos locales como globales y se recomienda para casi cualquier aplicación.

LÍMITE DEL VALOR E

El valor E para la puntuación de un alineamiento "S", representa el número de resultados con una puntuación igual a o mejor que "S" que serían esperados por azar (el ruido de fondo) cuando se hace una búsqueda en una base de datos de una longitud determinada.

- En BLAST el valor E es usado en lugar del valor P (probabilidad) para reportar la significancia de un resultado.
- El valor E puede ser incrementado o disminuido para modificar la astringencia del análisis.

PUNTUACIÓN (S SCORE)

La puntuación S de un alineamiento es calculado sumando la puntuación para cada alineamiento letra-con-letra y letra-con-espacio vacío.

TESIS CON
FALLA DE ORIGEN

- La puntuación o Score para cada alineamiento se deriva de la matriz de puntuación o de sustitución, las más populares son las matrices BLOSUM y PAM.
- Cuanto mayor sea la puntuación, es decir el score S , mejor es el alineamiento, pero la significancia de un alineamiento no puede ser deducida de la puntuación solamente.
- Las posiciones en las que una letra se alinea con un espacio se llama gap o "corte" y su puntuación es negativa.
- La puntuación en bits (bit score) toma en cuenta el tipo de sistema de puntuación usado, por ello puede ser más informativo. El bit score se calcula a partir del score S normalizado con las variables estadísticas que definen a un sistema de puntuación dado. Por ello los bit score pueden ser comparados entre distintos alineamientos.

MATERIAL

Para los análisis computacionales se utilizaron las secuencias completas de DNA de las bacterias *Helicobacter pylori* 26695 / J99 (<http://www.tigr.org/tigr-scripts/CMR2/GenomePage3.spl?database=ghp> / <http://scriabin.astrazeneca-boston.com/hpylori/>) y *Campylobacter jejuni* NCTC 11168 (http://www.sanger.ac.uk/Projects/C_jejuni/). Así como secuencias depositadas en la base de datos del NCBI (National Center of Biotechnology Information, <http://www.ncbi.nlm.nih.gov/>) de otras varias bacterias.

Se utilizó una computadora personal tipo IBM/PC ATX Pentium IV@ 1.8 GHz con 256 MB de memoria RAM y los programas que se detallan adelante.

BLAST: El programa BLAST también conocido como "Gapped BLAST" está diseñado para tomar secuencias proteicas o de ácidos nucleicos y

TESIS CON
FALLA DE ORIGEN

compararlas contra secuencias en una base de datos. El algoritmo de BLAST fue escrito con un balance entre velocidad y sensibilidad incrementada para relaciones de secuencias distantes. En lugar de basarse en alineamientos globales (lo cual puede verse en programas de alineamiento múltiple de secuencias como Clustal), BLAST enfatiza regiones de alineamiento local para detectar relaciones entre secuencias que comparten sólo regiones aisladas de similitud (1). Por lo anterior BLAST es más que una herramienta para ver secuencias alineadas o encontrar homologías, este programa puede además localizar regiones de similitud con una visión para comparar estructura y función.

En la página de BLAST se pueden elegir varios programas que realizan tareas específicas. De esos programas fueron utilizados los siguientes:

blastn	Compara una secuencias de nucleótidos contra una base de datos de secuencias de nucleótidos
tblastn	Compara una secuencia proteica contra una base de datos de secuencias de nucleótidos dinámicamente traducida en todos los marcos de lectura

En la página de BLAST del NCBI se encuentran varios recursos que proveen de ayuda en el uso de los programas.

ClustalW/ClustalX 1.81 : ClustalW (47) es un programa de alineamiento global para DNA o proteína. Usa un algoritmo de alineamiento progresivo y un árbol guía basado en similitud de secuencia para alinear las secuencias de DNA o aminoácidos. El alineamiento puede hacerse utilizando una de varias matrices de substitución distintas, el programa utiliza de manera predeterminada las matrices de Gonnet para los alineamientos proteicos. Este algoritmo alinea primero secuencias similares y después las secuencias más distantes. Este programa utiliza el algoritmo

TESIS CON
FALLA DE ORIGEN

de Neighbor Joining (NJ) (38) para producir los árboles filogenéticos a partir del alineamiento de las secuencias.

Clustal X es una interface de gráfica para el programa de alineamiento múltiple con ClustalW. Provee de un ambiente integrado para realizar análisis múltiples de secuencias y alineamientos de perfiles y analizar los resultados.

Se puede obtener la última versión del programa ClustalX vía ftp anónimo en:

- <ftp.igbmcu-strasbg.fr>
- <ftp.embl-heidelberg.de>
- <ftp.ebi.ac.uk>

O en la siguiente dirección WWW:

- <http://www-igbmc.u-strasbg.fr/BioInfo/>

Niplotwin95 : Este programa acompaña al programa ClustalX o ClustalW. Árboles filogenéticos contenidos en archivos con formato Newick pueden ser desplegados en pantalla, re-enraizados, guardados e impresos en archivos PostScript

Niplot está disponible vía ftp anónimo en: <biom3.univ-lyon1.fr> en el directorio /pub/mol_phylogeny/njplot.

Tree Explorer : Programa para visualizar, editar, imprimir y exportar árboles filogenéticos. Este programa fue obtenido de la siguiente dirección en internet: http://evolgen.biol.metro-u.ac.jp/TE/TE_man.html

TESIS CON
FALLA DE ORIGEN

Artemis v. 4 Copyright © 1999, 2000 by Genome Research Limited : Artemis es un visor gratuito de secuencias de DNA y una herramienta para anotaciones. Permite visualizar las características de una secuencia y los resultados de los análisis dentro del contexto de la secuencia, así como los seis marcos de traducción. Artemis está escrito en Java, y está disponible para UNIX, GNU/Linux, Macintosh y MS Windows. Puede leer entradas completas de las bases de datos EMBL y GENBANK, secuencias en formato FASTA o sin formato. Algunas características extra pueden ser vistas si se utilizan los formatos EMBL, GENBANK o GFF. Artemis puede trabajar con secuencias de cualquier tamaño, desde unos cuantos Kb hasta genomas enteros de más de 5Mb. En la siguiente dirección puede descargarse este programa, así como leer su manual y los requisitos de sistema:

- <http://www.sanger.ac.uk/Software/Artemis/>

Phylip: Es un paquete gratuito, diseñado para realizar diversos análisis filogenéticos. Incluye varios programas que están diseñados para tareas discretas, los cuales se integran a través del uso de un formato similar de archivos. Lo utilizamos para los análisis de máxima parsimonia y máxima verosimilitud. Puede obtenerse de la siguiente dirección: <http://evolution.genetics.washington.edu/phylip.html>

BioEdit (11): Es un editor de alineamientos de secuencias biológicas escrito para Windows®. Utiliza una interfaz gráfica y posee muchas características para edición. Puede ser obtenido de la siguiente dirección: <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>

MÉTODOS

BÚSQUEDA DE CITOCROMOS P450 EN *H. PYLORI* POR ANÁLISIS CON BLAST ("WESTERN COMPUTACIONAL")

Se recolectaron diversas secuencias de citocromos P450 eucariontes y procariontes, obtenidas del sitio de NCBI. Estas secuencias se utilizaron para hacer alineamientos de secuencia usando el programa BLAST (1,22), tanto contra la base de procariontes en el NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>) como con el genoma completo de *H. pylori* en la dirección <http://tigrblast.tigr.org/cmrbblast/index.cgi?database=GHP.seq>, en la página de red de The Institute for Genomic Research (TIGR).

Para utilizar el programa BLAST en NCBI se selecciona el algoritmo o programa que se desea utilizar. Para estos análisis se utilizaron blastn, tblastx o tblastp para secuencias proteicas. Esto nos lleva a una página en donde es posible introducir la secuencia que se desea comparar. Se selecciona la base de datos nr y más adelante pueden cambiarse los parámetros del programa en la sección de opciones localizada más abajo en la misma página. Los parámetros de "word size" y "choose filter" se utilizaron con los valores preestablecidos, 11 y "low complexity", respectivamente. En esta región puede seleccionarse limitar la búsqueda para orientar los resultados de acuerdo a lo que se está buscando, en este caso se realizaron las búsquedas seleccionando el grupo de bacterias (Bacteria [ORGN]). Una vez que han sido seleccionadas todas las opciones simplemente debe presionarse el botón "BLAST!" y los resultados se despliegan a través del navegador.

TESIS CON
FALLA DE ORIGEN

Para utilizar el programa BLAST comparando una secuencia escogida contra el genoma completo de *H. pylori* se utilizó el servicio ofrecido por TIGR en la página <http://tigrblast.tigr.org/cmrbblast/index.cgi?database=GHP.seq>. Una vez más se elige primero el algoritmo a utilizar, la base de datos (*Helicobacter pylori* 26695) y el número de hits que se deben presentar. Las opciones de los parámetros se encuentran en la parte de abajo de la página, los cuales se utilizaron con los valores preestablecidos. La secuencia debe suministrarse en formato FASTA. Una vez que se ha introducido la secuencia debe presionarse el botón "submit BLAST job" en la parte más baja de la página, para que se realice el alineamiento.

El valor E para los alineamientos fue reducido a 1 para la mayoría de los análisis, debido a que la secuencia a buscar era larga. Cuando se realizan búsquedas con secuencias cortas (apenas unas decenas de aminoácidos) es necesario aumentar el valor de E para obtener suficientes resultados y no eliminar otros alineamientos que pudieran tener significancia biológica, al contrario, cuando la secuencia a buscar es larga, se puede reducir el valor de E para eliminar resultados que no poseen ningún significado biológico.

ÁRBOLES FILOGENÉTICOS

Para los análisis filogenéticos realizados se utilizaron los algoritmos de neighbor joining (NJ) con el programa Clustal y, de máxima parsimonia (MP) y máxima verosimilitud ("maximum likelihood", ML) del paquete Phylip. Para utilizar cualquiera de estos métodos filogenéticos es indispensable que las secuencias se encuentren alineadas, para lo cual se utilizó en todos los casos el programa Clustal.

Las secuencias de DNA o aminoácidos deben tener un formato especial para ser analizadas con el programa ClustalX. Las secuencias

TESIS CON
FALLA DE ORIGEN

deben ser colocadas todas juntas en un mismo archivo y pueden tener alguno de los siguientes formatos: FASTA (Pearson), NBRF/PIR, EMBL/Swiss Prot, GDE, CLUSTAL, GCG/MSF, GCG9/RSF. El carácter '>' señala el principio del archivo y de cada secuencia, ésta puede llevar un identificador después del carácter '>' y después del primer carácter 'entrar' debe estar propiamente la secuencia. El programa intenta "adivinar" si se trata de una secuencia de DNA o de una secuencia de aminoácidos con los primeros caracteres de la secuencia.

Una vez que las secuencias estén dentro del programa, pueden ajustarse los parámetros para el alineamiento y el formato del archivo de salida (en este caso Clustal o Phylip, según se requirió).

Se solicita al programa que realice un alineamiento completo, tras lo cual hay que esperar un cierto período de tiempo para que el programa realice los cálculos, este tiempo puede variar considerablemente de acuerdo con el poder de cómputo con el que se cuente. Este análisis produce el archivo con las secuencias alineadas, un archivo con el árbol guía y un archivo con el árbol que se genera por el análisis de Neighbor Joining. Si se utiliza la opción preestablecida para el archivo de salida, el archivo de alineamiento tendrá una extensión *.aln, el árbol guía *.dnd.

Una vez que el programa Clustal alinea automáticamente las secuencias, es necesario revisarlas manualmente para asegurarnos de no pasar por alto sitios donde el algoritmo no hizo el alineamiento óptimo de las secuencias. Esto se realizó con ayuda del programa BioEdit, el cual está especialmente diseñado para esta función y además permite al usuario guardar el alineamiento en una gran variedad de formatos.

Para hacer el análisis de Bootstrap del árbol obtenido con NJ, se abre el archivo con el alineamiento deseado después de la edición con BioEdit en Clustal. Para nuestros análisis se utilizó siempre encendida la opción "Correct for Multiple Substitutions", esta opción realiza un ajuste

TESIS CON
FALLA DE ORIGEN

debido a que las distancias observadas suelen subestimarse por el hecho de que a lo largo de la divergencia de las moléculas ocurre más de una sustitución en cada sitio, esto aumenta la eficiencia del algoritmo. Los demás parámetros se utilizaron con sus valores preestablecidos. Para el número de iteraciones de la prueba de Bootstrap la opción por defecto es de 100, sin embargo para este trabajo se realizaron 1000 iteraciones para cada prueba de bootstrap. Una vez más el tiempo para hacer este análisis puede variar significativamente según el poder de cómputo con el que se cuenta y el número de secuencias involucradas en el análisis. Este análisis produce un archivo *.phb (árbol filogenético con valores de bootstrap). Este tipo de archivo puede ser abierto con la aplicación Nplotwin95, con la cual podemos imprimir el árbol o exportarlo a un archivo de árbol (tree file) con extensión *.phr, que podrá ser abierto a su vez con el programa TreeExplorer. Si no se realiza el cambio de formato del archivo *.phb a *.phr con el Nplotwin95, el TreeExplorer no podrá desplegar los valores de bootstrap contenidos en el archivo.

Una vez que el archivo *.phr está listo, en el TreeExplorer se pide la opción importar archivo de árbol en formato Newick Standard. Una vez que el árbol filogenético ha sido cargado puede editarse para brindar una mayor claridad y exportarlo como un archivo gráfico *.emf para su utilización en cualquier otra aplicación que posea el filtro para este tipo de gráfico.

El programa Clustal utiliza el algoritmo de Neighbor Joining de Saitou y Nei de 1987. Este algoritmo se basa en el algoritmo de Fitch y Margoliash (38).

Para los análisis con MP se utilizan los programas <dnapars.exe> y <protpars.exe> del paquete phylip. Ambos cuentan con una serie de opciones descritas en la documentación del paquete phylip. Los archivos a utilizar con estos programas deben estar en el formato del phylip, los

TESIS CON
FALLA DE ORIGEN

archivos tienen una extensión .phy. Para el análisis de secuencias de nucleótidos se utiliza el programa dnapars y para secuencias de proteínas el protpars. Si se trabaja con un computador personal tipo IBM/PC es importante que el archivo de salida se configure de acuerdo a ello. Ambos programas entregan como salida un archivo <outfile>, el cual contiene la información del análisis que se acaba de correr y un árbol representado en código ASCII, y un archivo <treefile> que contiene los datos para construir el árbol en formato phylip/newick standard. Estos árboles no tienen distancias porque el algoritmo de MP no calcula distancias entre los nodos.

Para ML el phylip cuenta con un programa llamado <dnaml>, el cual funciona en general igual que los arriba mencionados, excepto por supuesto que utiliza el algoritmo de ML para inferir un árbol filogenético. Desafortunadamente el paquete phylip no cuenta con un programa para hacer análisis de ML con proteínas, aunque en la documentación recomiendan el uso del programa protml, el cual está incluido en el paquete Molphy, pero este programa no utiliza los formatos de phylip para los archivos de salida, lo que hace difícil trabajar con él y mucho más manejar los archivos con los resultados.

FRECUENCIA DE SECUENCIAS ESPECÍFICAS DE NUCLEÓTIDOS EN EL CROMOSOMA

Para este análisis se utilizó la secuencia completa del genoma de *C. jejuni* (Sanger Center). Se utilizó la paquetería MathLab, con la cual se buscaron a lo largo de toda la secuencia del cromosoma distintos tripletes, tetrapletes o pentapletes específicos. Se diseñó un programa que posee un contador que aumenta en uno cada vez que es localizado el triplete, tetraplete o pentaplete particular en el genoma; después anota a cada evento su localización en el genoma. Esta paquetería colecciona todos

TESIS CON
FALLA DE ORIGEN

los datos y puede mostrarlos en una gráfica, la cual es posible exportar como un archivo gráfico para su utilización posterior en otras aplicaciones. En las gráficas se buscaron tendencias a lo largo de toda la gráfica que se diferenciaran de su entorno, lo cual podía identificarse por cambios en la pendiente de la gráfica.

DISTRIBUCIÓN DE ORFs EN EL GENOMA DE *CAMPYLOBACTER JEJUNI*

Con ayuda del programa Artemis se pudieron visualizar todos las secuencias codificantes (ORFs) del genoma de *C. jejuni* con su orientación a la traducción.

Todos los ORFs contenidos en el genoma de *C. jejuni* se importaron a un archivo de Microsoft Excel ® desde la tabla de codificación de proteínas de la página del NCBI. El archivo se formateó para poder acomodar todos los datos según la orientación a la traducción. Se realizó un análisis en el cual se obtuvo la frecuencia de eventos en donde podía encontrarse un solo ORF seguido por otro en dirección contraria a lo largo de todo el genoma, después la frecuencia de dos ORFs en la misma dirección seguidos de otro en dirección contraria y así sucesivamente.

FRECUENCIA DE ORTÓLOGOS

El 55.4% de los genes de *C. jejuni* tienen ortólogos en *H. pylori* (32). Sanger Center en su página electrónica tiene varios recursos disponibles para trabajar con el genoma de *C. jejuni*. Utilizamos la lista de genes ortólogos de Sanger Center para hacer un análisis de la frecuencia de genes ortólogos en ventanas de diversos tamaños del genoma de *C. jejuni*.

TESIS CON
FALLA DE ORIGEN

Se dividió el genoma entero en fragmentos de 10kb, 20Kb y 40Kb y se hizo un conteo en cada ventana del número de ORFs que tienen ortólogos en *H. pylori*.

POSIBLE DONADOR

Usando el programa TBLASTX se buscaron dobles mejores alineamientos (dual best hit) (18) de cada uno de ORFs desde Cj1410 hasta Cj1449 de *C. jejuni*. El doble mejor alineamiento implica que cuando se hace una búsqueda con la secuencia que da mejor resultado de un alineamiento, la secuencia original del primer alineamiento debe ser a su vez el mejor resultado. Se realiza la primera búsqueda con el ORF de *C. jejuni*, el programa entrega una serie de secuencias, se toma la que tiene el mejor score y con ella se realiza otra búsqueda. Los resultados de ambas búsquedas se comparan, si el mejor resultado de la segunda búsqueda es la primera secuencia de la que partimos, se considera eso como un "best dual hit".

TESIS CON
FALLA DE ORIGEN

RESULTADOS

CITOCROMOS P450 EN *H. PYLORI* Y EN *C. JEJUNI*

Debido a la relación tan estrecha que guarda el *Helicobacter pylori* con la aparición de cáncer gástrico, y a que los citocromos P450 tienen a su vez una relación cercana con el desarrollo de distintos tipos de cánceres en el ser humano, era plausible suponer que *H. pylori* tuviera algún citocromo P450. Además, *Campylobacter jejuni* tiene un gene para un citocromo P450, lo cual apoya la idea de que *H. pylori* pudiera haber tenido un gene para P450 debido a que ambas bacterias comparten un ancestro común. A pesar de lo anterior, *H. pylori* no posee un citocromo P450; esta evidencia nos permitió hacer nuestra pregunta experimental: el ancestro común de ambas bacterias poseía un citocromo P450 o no, luego entonces, si *H. pylori* perdió su citocromo P450 o si *C. jejuni* lo adquirió.

No se ha reportado ningún gene con la firma PS0086 propia de los citocromos P450 en *H. pylori*, para descartar que el gene para el citocromo P450 en *H. pylori* hubiera sufrido tantas mutaciones que no fuera reconocible, buscamos en las secuencias depositadas en el NCBI que pudieran servir como pista para encontrar algún gene o gene putativo de *H. pylori* que tuviera alguna relación, inclusive una muy débil, con un citocromo P450. Tras realizar una búsqueda con el gene del citocromo P450 de *C. jejuni* por medio de TBLASTN, todos los resultados que se obtuvieron con un valor E de no más de $10E-18$ fueron introducidos como elemento para una nueva búsqueda con TBLASTN. Para estas búsquedas se utilizaron las secuencias proteicas para minimizar las diferencias entre las secuencias debidas a los cambios en un solo nucleótido y que sin embargo no codifican para otro aminoácido y no se traducen en un cambio en la proteína. Estas búsquedas no nos llevaron a encontrar ninguna proteína perteneciente a *Helicobacter pylori*.

TESIS CON
FALLA DE ORIGEN

ANÁLISIS FILOGENÉTICO DE *H. pylori* y *C. jejuni*

H. pylori y *C. jejuni* proceden de un ancestro cercano común, por ello que sean cercanas filogenéticamente. Realizamos una reconstrucción filogenética tomando como base los RNAs 16S de varias especies bacterianas, escogidas de entre los distintos grupos bacterianos, de manera de representar con el uso de tres algoritmos de construcción filogenética distintos, la relación de *C. jejuni* con *H. pylori* y probar si nos era posible obtener construcciones cercanas a la aceptada.

Según se propuso en la sección "La Aproximación Filogenética" la primera tarea a resolver para realizar un análisis filogenético es elegir la molécula más adecuada para el problema particular. El RNA ribosomal (RNAr) es la molécula que más se ha utilizado para establecer relaciones filogenéticas debido a varias razones: 1) Es una molécula cuya secuencia es bastante conservada, ya que está sometida a una fuerte presión selectiva (5), debido a su estabilidad los cambios en ella ocurren muy lentamente y permite observar reconstrucciones filogenéticas mucho más confiables o cercanas a la filogenia verdadera (23), que las que son construidas con otras moléculas cuya velocidad de mutación puede ser mayor o que pueden ser susceptibles en mayor grado a TGH; 2) Por su longitud provee suficiente información; 3) Su transferencia debería ser en la mayoría de los casos vertical, dada la dificultad de que un RNAr extraño pueda acoplarse eficientemente a la maquinaria celular del organismo receptor (27).

La segunda tarea para realizar un análisis filogenético es hacer el alineamiento de las secuencias. Para que sea posible obtener un análisis confiable es necesario utilizar varias secuencias además de las dos de interés, ya que esto provee del marco de comparación entre las secuencias. El alineamiento de secuencias (Apéndice 1) se realizó como

TESIS CON
FALLA DE ORIGEN

se describe en Métodos. Para el análisis se utilizaron las siguientes secuencias de RNA ribosomal 16s:

- >Actinobacillus_sp
- >Amycolatopsis_mediterranei
- >Bacillus_halodurans
- >Bacillus_megaterium_IAM13418
- >Bacillus_subtilis
- >Bradyrhizobium_japonicum_PRY65
- >Buchnera_sp
- >Campylobacter_jejuni
- >Caulobacter_crescentus_CB2A
- >Chlamydia_muridarum
- >Chlamydia_pneumoniae_AR39
- >Deinococcus_radiodurans_R1
- >Escherichia_coli_K12
- >Escherichia_coli_O157
- >Helicobacter_pylori
- >Mesorhizobium_loti_LMG6123
- >Moritella_yanosii
- >Mycobacterium_leprae_TN
- >Mycobacterium_smegmatis_ATCC19420
- >Mycobacterium_tuberculosis_H37Rv
- >Mycobacterium_tuberculosis_CDC1551
- >Mycoplasma_pneumoniae
- >Myxococcus_xanthus_MD207
- >Neisseria_meningitidis_MC58
- >Pasteurella_multocida_PM70
- >Pseudomonas_aeruginosa_PA01
- >Rhodococcus_sp_X309
- >Rickettsia_prowazekii_Brient
- >Saccharopolyspora_erythraea
- >Salmonella_typhimurium_LT2
- >Staphylococcus_aureus_N315
- >Streptococcus_pyogenes
- >Streptomyces_avermitilis
- >Streptomyces_clavuligerus_JCM4710
- >Streptomyces_coelicolor_cosmid_D72AA3
- >Streptomyces_griseus_subsp_griseus_IFO13550
- >Streptomyces_lavendulae_IFO13710
- >Streptomyces_venezuelae_JCM4526
- >Synechocystis_PCC6803
- >Thermotoga_maritima
- >Treponema_pallidum
- >Ureaplasma_urealyticum
- >Vibrio_cholerae
- >Xylella_fastidiosa
- >Mus musculus *

* La secuencia de ratón se agrego con el fin de poder agregar la raíz del árbol.

TESIS CON
FALLA DE ORIGEN

El tercer paso, tal como se describió anteriormente, es la aplicación de un algoritmo matemático que nos permita hacer inferencias evolutivas.

Se utilizaron tres algoritmos distintos para este análisis: Neighbor Joining (Clustal), Maximum parsimony and Maximum likelihood (Phylip), y se compararon los árboles resultantes (Apéndice 2).

El método de neighbor joining (NJ) entrega un solo árbol el cual representa un único árbol que resulta de medir las distancias entre las secuencias según el número de diferencias o similitudes que tienen, formando grupos que más tarde se colocan en las ramas del árbol. El método de máxima parsimonia (MP) opera seleccionando aquellos árboles que minimizan en número de pasos evolutivos necesarios para explicar un grupo dado de secuencias, por ello puede dar como resultado varios árboles. El método de máxima verosimilitud (maximum likelihood, ML) evalúa la probabilidad de que un modelo del proceso evolutivo (árbol) diera lugar a los datos observados (secuencias); se asume que la historia inferida con mayor probabilidad de dar origen a los datos reales es preferible a una que tiene una menor probabilidad de alcanzar el estado actual de los datos. En el árbol filogenético de ML fue necesario quitar el OTU de ratón, debido a que su distancia de los demás OTUs era tal que no los demás nodos quedaban demasiado cerca impidiendo el poder verlos.

Como último paso para tener un análisis filogenético confiable, hay que revisar su confianza estadística. Para ello se utilizó el método de bootstrap como se describe en Métodos, los valores resultantes del bootstrap están representados en cada nodo del árbol construido con el método de NJ. El método de bootstrap sólo se aplicó con el algoritmo de NJ debido a que el programa misma ejecuta el análisis; con el paquete phylip no nos fue posible realizar el análisis de bootstrap.

Las árboles inferidos con los tres distintos métodos son bastante consistentes entre sí, con sólo pequeñas diferencias. Los tres modelos

TESIS CON
FALLA DE ORIGEN

fueron consistentes con los datos utilizados para reconstruir la taxonomía de las bacterias utilizadas para este análisis.

El método de MP sólo encontró dos árboles como posibles modelos para explicar los datos, la única diferencia entre estos dos árboles es el lugar donde se sitúa el nodo de *Staphylococcus*, en un caso comparte la rama con *Streptococcus* y *B. megaterium* y en el otro comparte la rama con *B. halodurans* y *B. subtilis*; de acuerdo a la taxonomía publicada en NCBI, el modelo más preciso sería el segundo caso. Con el método de ML encontramos un único árbol con una máxima probabilidad. El árbol que obtuvimos usando NJ tiene valores altos de bootstrap (1000 iteraciones) para la mayoría de los nodos, lo cual nos brinda confianza estadística. Con este análisis fuimos capaces de reproducir la taxonomía aceptada para estas especies y familiarizarnos con el uso de estas metodologías.

DISTRIBUCIÓN DE LOS CITOCROMOS P450 BACTERIANOS

Los citocromos se encuentran distribuidos en distintas formas de vida, incluyendo animales, plantas y microorganismos; sin embargo, aunque existe un gran número de citocromos en los otros reinos, los citocromos que más se han estudiado, que mejor se conocen y de los que más se han descrito, son los pertenecientes al reino animal.

Debido a que los citocromos P450 se pueden encontrar en todos los reinos, esta evidencia podría llevar a pensar que se trata de una proteína ancestral común a todos los seres vivos. Sin embargo, no se han encontrado genes parálogos para citocromo en todos los genomas bacterianos secuenciados hasta el momento. Se pueden observar familias bacterianas en las cuales es posible encontrar citocromos P450 en todas

TESIS CON
FALLA DE ORIGEN

las especies analizadas (Tabla 3), aun cuando sus genomas no estén totalmente secuenciados; sin embargo hay familias donde han sido secuenciados los genomas de varios miembros y en las cuales no se encuentra ninguna especie que tenga algún gene relacionado con los citocromos P450. Lo que resulta aún más interesante, es que hay grupos bacterianos en donde hay tanto especies que tienen citocromos P450 como especies que carecen de ellos. En bacterias cuyo genoma no ha sido secuenciado por completo, es posible que la región que contiene el citocromo P450 no haya sido secuenciada a la fecha, y que por ello no se ha identificado.

	Genoma Completo	Presencia de P450
Proteobacteria (purple bacteria)		
alpha subdivision		
Caulobacter crescentus	+	
Agrobacterium tumefaciens plasmado Ti		+
Rhizobium sp. NGR234 Plasmido pNGR234a, section 19 of 46, section 20 of 46		+
Bradyrhizobium japonicum		+
Mesorhizobium loti	+	+
Rickettsia prowazekii	+	
beta subdivision		
Hydrogenophaga intermedia		
Neisseria meningitidis Z249 I	+	
Neisseria meningitidis MC58	+	
delta subdivision		
Sorangium cellulosum (Polyangium cellulosum)		+
Myxococcus xanthus		+
epsilon subdivision		
Campylobacter jejuni	+	+
Helicobacter pylori 26695	+	
Helicobacter pylori J99	+	
gamma subdivision		
Escherichia coli K12	+	
Escherichia coli O157:H7	+	
Escherichia coli O157:H7 EDL933	+	
Buchnera	+	
Pasteurella multocida	+	
Haemophilus influenzae Rd	+	
Pseudomonas aeruginosa	+	+

TESIS CON
FALLA DE ORIGEN

<i>Pseudomonas incognita</i>		+
<i>Pseudomonas putida</i>		+
<i>Pseudomonas sp.</i>		+
<i>Vibrio cholerae</i>	+	
<i>Xylella fastidiosa</i>	+	+
<i>Borrelia burgdorferi</i>	+	
<i>Treponema pallidum</i>	+	
Thermotogales		
<i>Thermotoga maritima</i>	+	
<i>Deinococcus radiodurans</i>	+	+

Tabla 3. Resumen de las bacterias cuyo genoma ha sido secuenciado en su totalidad y/o que ha sido encontrado algún citocromo P450.

La distribución dispersa de los citocromos P450 en las distintas familias de Eubacterias no es consistente con una premisa que establezca a los citocromos P450 como una proteína que evolucionó desde un ancestro común. Sin embargo la distribución observada de los citocromos P450 podría explicarse si los genomas bacterianos estuvieran experimentando constantemente tanto fenómenos de transferencia genómica horizontal, como pérdida de genes para los citocromos P450.

CITOCROMOS P450 EN *MYCOBACTERIUM TUBERCULOSIS*

Como ya mencioné anteriormente, los citocromos P450 están bastante distribuidos entre los procariontes. El caso de *Mycobacterium tuberculosis*, en cuyo genoma es posible encontrar 21 citocromos P450, es interesante porque podría hablarnos de una gran movilidad de los citocromos P450 entre distintos grupos bacterianos. Construimos un árbol con los citocromos P450 de *M. tuberculosis* para compararlo después con otro árbol en el cual inferimos las relaciones filogenéticas de estos citocromos con otros pertenecientes a otras bacterias (los alineamientos se encuentran en el apéndice 1, los árboles en el apéndice 2). El resultado nos sugiere que los citocromos P450 pudieran estar involucrados en muchos eventos de THG entre los procariontes, y que debido a las características metabólicas propias de estos citocromos se estabilizan en las poblaciones

TESIS CON
FALLA DE ORIGEN

bacterianas, haciendo posible su detección; tal podría ser el caso del citocromo P450 de *C. jejuni*.

En la comparación de ambos arboles (árbol 6 y 7, Apéndice 2) podemos observar que los grupos que arbitrariamente formamos con los citocromos P450 de *M. tuberculosis*, se disgregaron en el árbol que contiene los citocromos P450 de otras especies. Esto sugiere que los algunos de los citocromos que vemos en *M. tuberculosis* pueden haber sido adquiridos por TGH, ya sea muy recientemente o proveen a la bacteria de alguna ventaja por lo que se han perpetuado en la población. Si *M. tuberculosis* pudo adquirir tantos citocromos P450 por TGH, podría ser también que *C. jejuni* haya adquirido el suyo a través de un evento de TGH.

TESIS CON
FALLA DE ORIGEN

ÁRBOLES FILOGENÉTICOS DE CITOCROMOS P450 Y PROTEÍNAS CONSERVADAS

El genoma de una bacteria acumula mutaciones a lo largo del tiempo, con lo que va adquiriendo paulatinamente ciertas características particular a esa especie. La tasa de mutación y el/los lugares donde se acumulan dentro de una proteína nos permiten inferir las relaciones filogenéticas. Cuando se comparan las relaciones filogenéticas inferidas en árboles hechos con proteínas muy conservadas, éstas se asemejan mucho a aquellas obtenidas utilizando RNAr 16S. Al parecer los citocromos P450 están involucrados en eventos de TGH que les permiten perpetuarse en distintas especies. Un árbol hecho con citocromos P450 podría asemejar las relaciones filogenéticas vistas con proteínas conservadas, de lo contrario sería una evidencia indirecta de TGHs (35).

Se realizaron alineamientos automáticos utilizando el Clustal y después fueron revisados y editados más finamente con el programa BioEdit (los alineamientos pueden revisarse en el apéndice 1). Tomando como base el árbol con citocromos P450 se construyó un árbol con un citocromo P450 de cada especie, tomando al azar sólo un citocromo P450 de aquellas especies que tienen varios. Se hicieron también árboles filogenéticos utilizando las proteínas dnaA, polA, recA y rpoA de las mismas especies. En estos análisis se utilizaron algoritmos de NJ y MP como se describe en métodos; en todos los casos el algoritmo de MP encontró sólo un árbol con máxima parsimonia.

Los árboles obtenidos con MP muestran bifurcaciones sucesivas en todos los nodos de los árboles, esto podría deberse a las limitaciones del algoritmo, es posible que con esta longitud y número de secuencias, no haya sido del todo consistente. Sin embargo, la mayoría de las relaciones pueden verse representadas tanto en los árboles con MP como con NJ. En el árbol construido con los citocromos P450 las relaciones entre las especies

TESIS CON
FALLA DE ORIGEN

utilizadas son totalmente diferentes a las que vemos con proteínas conservadas, brindándonos una evidencia de que estos citocromos han sufrido eventos de TGH; el citocromo de *C. jejuni* se relaciona cercanamente con especies las cuales son lejanas vistas con RNAr 16S.

TESIS CON
FALLA DE ORIGEN

FRECUENCIA DE SECUENCIAS ESPECÍFICAS DE NUCLEÓTIDOS EN EL CROMOSOMA

La secuencia de las bases a lo largo de todo el genoma está determinado principalmente por la codificación de proteínas. Esto es particularmente cierto para el genoma de *C. jejuni*, que es uno de los genomas secuenciados a la fecha que tienen menos secuencias no codificantes (32). Sin embargo, hay otras variables que pueden afectar esta secuencia, como son el sesgo de mutación (mutation bias), uso preferente de codones, tensiones estructurales y ameliorización (20,29,46). El efecto de estos fenómenos es consistente a lo largo del genoma entero en procariontes (20,28,46), y estas características se perpetúan verticalmente en las poblaciones bacterianas. Su resultado sobre el DNA puede ser medido a través de la frecuencia de aparición de un dinucleótido, trinucleótido, etc., en el cromosoma completo (di-, tri-, tetra-, etc., nucleotide fingerprints); el contenido de G+C; el índice de uso de codones (CAI, Codon Adaptation Index) (41); y la variación de bases en la primera, segunda y tercera posición (ameliorización) (20,29,44). El análisis de cualquiera de estas variables puede constituir una herramienta útil cuando se investiga la posible historia evolutiva de una región de un cromosoma, revelando posibles transferencias horizontales, en las cuales las características arriba mencionadas son diferentes de las del genoma completo (23, 28).

Se realizó un análisis del genoma completo de *C. jejuni* buscando distintas distribuciones de dinucleótidos que pudieran revelar un posible origen externo de la región que nos interesa. Este análisis se hizo según se describe en Métodos. El dinucleótido de citosinas mostró un patrón de distribución ligeramente distinto de los demás dinucleótidos probados, en la región de interés, por lo que se decidió buscar secuencias más largas de

C con el fin de hacer más clara la diferencia de esa región. Conforme extendemos la longitud del oligonucleótido de citosinas se hace más clara la diferencia en el patrón de distribución en la región con el citocromo P450 con respecto al resto del genoma. El resultado óptimo de este análisis se obtuvo utilizando una secuencia de 5 citosinas como se muestra en la figura 9. Con el pentaplete de citosina se observa claramente un cambio en la pendiente de la gráfica en la región que corresponde del pentaplete 155 hasta el pentaplete 200 en el genoma. Hasta antes del pentaplete número 154, la pendiente no tiene cambios importantes, los cambios en la pendiente sólo perduran en pequeñas extensiones del genoma. Entre el pentaplete 155 y el 200 hay un cambio dramático en la pendiente, formándose una meseta, ésta representa una acumulación rápida de dicho pentaplete en esa región, la cual contiene casi la cuarta parte de

TESIS CON
FALLA DE ORIGEN

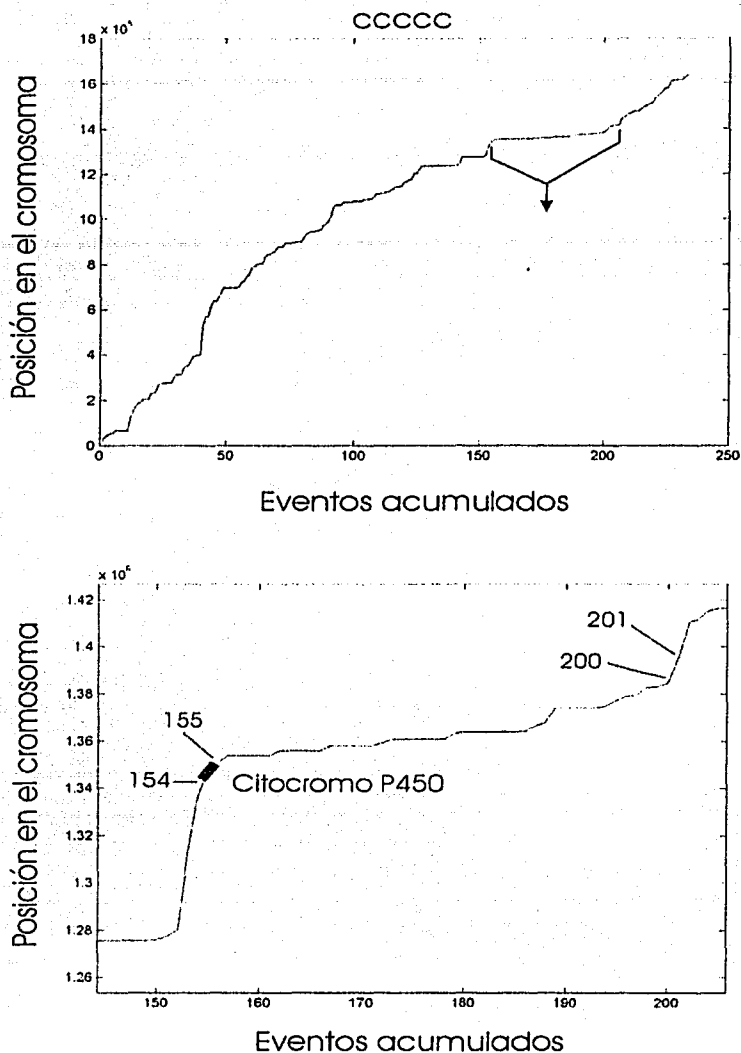


Figura 9. Frecuencia del pentaplete de citosina a lo largo del genoma de *C. jejuni*. En A se muestra la gráfica de todo el cromosoma con la zona propuesta como TGH señalada. En B se muestra un acercamiento de la zona señalada en A, la región propuesta como TGH queda localizada entre los eventos 154 y 201, abarcando toda la meseta.

todos los pentapletes encontrados en el genoma en una región de apenas 40 Kpb, que representa el 2.44% del cromosoma completo. La región que

forma la meseta corresponde al fragmento que proponemos como TGH el cual va desde la posición 1342197 hasta 1388235, quedando localizada entre los pentapletes 154 y 201. El citocromo P450 va desde la posición 1,342,550 hasta la posición 1,343,911, ubicado entre los pentapletes 154 y 155.

La frecuencia de acumulación de distintos dinucleótidos, triplete, tetrapletes y pentapletes se conserva a lo largo de todo el genoma de *C. jejuni*, con excepción del fragmento propuesto como TGH en el caso particular del pentaplete de citosinas. Como se dijo al principio de este capítulo, los genomas tienden a conservar las mismas características estructurales en todo el cromosoma el cual transfieren verticalmente a su descendencia. El proceso de ameliorización, tal como describen Ochman y Lawrence (29), a lo largo de millones de años ejerce una presión en el DNA para que aquellos fragmentos extraños, que ahora forman parte del genoma, adquieran las características del DNA circundante y finalmente sean tan parecidos al DNA del huésped que es imposible saber si ese DNA fue adquirido por transferencia vertical o por un evento de TGH. La diferencia tan marcada en la distribución del pentaplete de C en toda una región del genoma de *C. jejuni*, es un fuerte indicador de que toda esta región pudo haber sido adquirida por un evento de transferencia horizontal (13), el cual debe haber ocurrido después de que *C. jejuni* y *H. pylori* divergieron. Este evento de TGH debe ser relativamente reciente en tiempos evolutivos, dado que el fragmento no ha tenido tiempo suficiente para sufrir un evento de ameliorización y aún conserva características extrañas al genoma de *C. jejuni*.

DISTRIBUCIÓN DE ORFs EN EL GENOMA DE *Campylobacter jejuni*

El genoma de *C. jejuni* es el más compacto de los secuenciados a la fecha, como ya se mencionó el 94.3% del genoma codifica para proteínas.

TESIS CON
FALLA DE ORIGEN

De todas estas CDS el 61.1% se transcriben en la dirección de la replicación.

Los genes que se localizan en una o en otra cadena de DNA están salteados a lo largo del genoma, en algunas ocasiones se encuentran en grupos de varios de ellos seguidos. Se realizó un análisis para observar la frecuencia y la cantidad de estas secuencias codificantes (ORFs, putativos y de función identificada) que se encuentran en la misma cadena y que se localizan juntos. Si fuera posible observar una distribución de grupos de ORFs a lo largo del genoma, ello nos permitiría separar regiones que salieran de esta distribución, como regiones que podrían ser extrañas a este genoma y buscar en ellas características distintas al genoma completo (35). El evento de encontrar una región que salga de la distribución de ORFs del genoma entero por si mismo puede ser una evidencia de que esa región proviene de un genoma extraño (3).

En una base de datos que contiene todos los ORFs que se han encontrado en el genoma de *C. jejuni*, se localizaron y cuantificaron todos los ORFs que se transcriben en la cadena *Leading* y en la cadena *lagging*, y de todos ellos los que se encuentran localizados en sucesión (contiguos en la misma cadena) formando estructuras que recuerdan a los operones. También se localizaron los sitios más probables para la posición de *ori* y *terminus*. El sitio *ori* se encuentra cercano al gene *dnaA*. El lugar más probable para el sitio *terminus* se encuentra cerca de la mitad del genoma en la secuencia 814319..814352

TESIS CON
FALLA DE ORIGEN

Frecuencia de ORFs contiguos en el mismo sentido

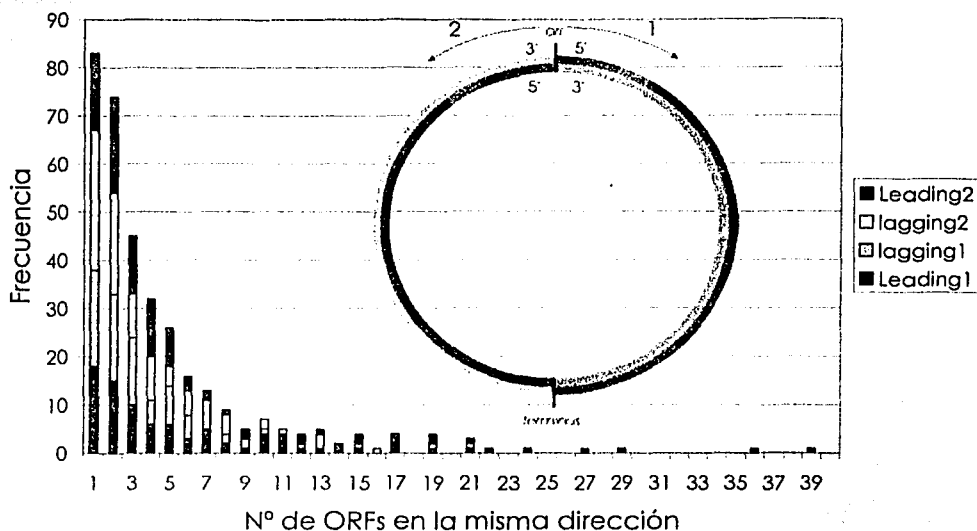


Figura 9. Frecuencia con la que es posible encontrar en el genoma de *C. jejuni* un solo ORF seguido por otro con su orientación contraria, dos ORFs con la misma orientación y seguidos por otro en dirección contraria, y así sucesivamente. En el recuadro se muestra una representación del cromosoma de *C. jejuni*, están señaladas las dos regiones con más ORFs contiguos en la misma orientación.

La base de datos con estos análisis se encuentra en el apéndice localizado en la página electrónica del laboratorio de la Dra. Julieta Rubio, en el sitio del Instituto de Investigaciones Biomédicas.

Al hacer la cuantificación de todos los genes localizados en tandem con la misma orientación con respecto a su transcripción, se obtiene una frecuencia de grupos con el mismo número de genes alineados, esto se muestra en la figura 9. Aquí están representadas tanto la cadena Leading como la cadena lagging del genoma, los números 1 y 2 se refieren a la primera mitad del cromosoma desde ori hasta terminus y a la segunda mitad del cromosoma desde terminus hasta ori, ambas en dirección 5' a 3'.

Se observa que la mayoría de los ORFs se encuentran muy dispersos, formando grupos de apenas algunos de ellos alineados o inclusive solos. Cuanto más aumenta el número de ORFs orientados, hay menos casos a lo largo del genoma, describiendo una gráfica negativa binomial. Llama la atención que hay realmente pocos grupos de ORFs orientados en la misma dirección en números mayores de 21. En el extremo de la gráfica se pueden ver dos grupos que se alejan bastante del común, el grupo más alejado, de 39 genes seguidos abarca las posiciones 1342134..1388184 (46050 pb), va del ORF Cj1410 al Cj 1449, el citocromo P450 es el ORF Cj1411. La otra región con un número alto de genes juntos es la región que va desde la base 96074 hasta la base 125478 (29404), involucrando desde Cj0087 hasta Cj0122. Es interesante que estas dos regiones que contienen aproximadamente el mismo número de ORFs, parecen corresponderse si obtuviésemos la imagen especular de la mitad del cromosoma, se encuentran a distancias similares del punto *ori*.

La región que proponemos como TGH está delimitada por los ORFs Cj1410 hasta el Cj1449 de acuerdo a este criterio, y corresponde con el punto en la figura 9 de 39 ORFs ordenados en la misma orientación, que como se señaló con anterioridad, no es un evento frecuente en el genoma de *C. jejuni*.

FRECUENCIA DE ORTÓLOGOS

C. jejuni y *H. pylori* son bacterias cercanas filogenéticamente, comparten un 55.4% de sus ORFs, esto quiere decir que de manera general podríamos encontrar un ortólogo por cada dos ORFs en el genoma de *C. jejuni*. Por supuesto existen regiones donde los ORFs ortólogos se encuentran más cercanos los unos de los otros y que contienen por ello un mayor número de ORFs ortólogos en secciones cortas del genoma, pero

TESIS CON
FALLA DE ORIGEN

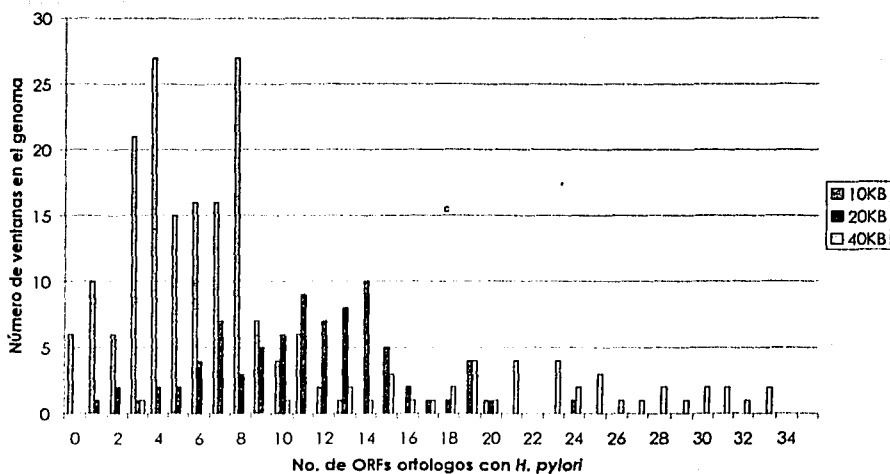
este sesgo puede disminuirse considerando ventanas grandes para contar el número de ORFs.

Se analizó la frecuencia o densidad de ORFs de *C. jejuni* que poseen ORFs ortólogos en el genoma de *H. pylori*. Para determinar si un ORF es ortólogo a otro se utilizan parámetros de alineación de secuencias, así como análisis filogenéticos. Para el presente trabajo se utilizaron las definiciones de ortología para *C. jejuni* reportadas por el Sanger Center (<http://www.sanger.ac.uk/>).

Se realizó la cuenta de ORFs ortólogos con *H. pylori* en el genoma de *C. jejuni*, en ventanas de 10, 20 y 40 Kb y en ventanas de 10, 20 y 40 ORFs. El genoma se dividió por completo en ventanas de los tamaños especificados, es decir, el desplazamiento de cada ventana con respecto de la anterior fue exactamente del tamaño de la ventana misma (10, 20 o 40 Kb). Es conveniente recordar que el fragmento que se propone como TGH comprende 40 ORFs y se extiende sobre aproximadamente 40 kbp. Se graficaron la cantidad de ORFs que una ventana contenía contra el número de ventanas que contenían ese mismo número de ORFs, como se muestra en la figura 10. Al graficar los datos obtenidos con las ventanas de 10 y 20 kbp, se observaron distribuciones homogéneas, similares a una distribución normal, para el número de ORFs ortólogos contenidos en las ventanas de esos tamaños; lo mismo ocurrió cuando se tomaron ventanas de 10 y 20 ORFs (figura 10). Cuando se utilizaron ventanas tanto de 40 kbp como de 40 ORFs, puede notarse claramente una distribución que se aleja de la distribución observada para el resto del genoma. La región propuesta como TGH contiene solamente 3 ORFs para los cuales se ha reportado un ortólogo en *H. pylori*, y queda afuera de campana de distribución para el resto del genoma (figura 10, 11).

TESIS CON
FALLA DE ORIGEN

No. de ORFs que tienen ortología con *H. pylori* a lo largo del genoma de *C. jejuni*



No. de ORFs de *C. jejuni* con ortología en *H. pylori*, en ventanas de 10, 20 y 40 ORFs de longitud

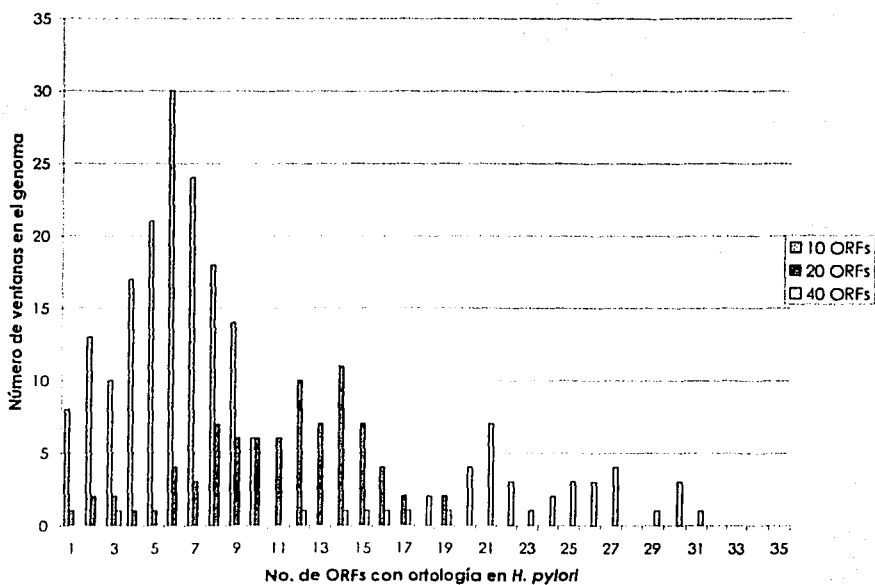


Figura 10. A) Número de ventanas de 10, 20 y 40 kbp en *C. jejuni*, y su correspondiente número de ORFs ortólogos con *H. pylori*. B) Número de ventanas de 10, 20 y 40 ORFs consecutivos en *C. jejuni*, y su correspondiente número de ORFs ortólogos con *H. pylori*.

Por sí sola esta es una fuerte evidencia, para apoyar la hipótesis de que el fragmento contenido entre Cj1410 y Cj1449 es una transferencia horizontal. De estos experimentos se deduce que el genoma de *C. jejuni*, como el de todas las bacterias, tiene una organización similar a lo largo de todo el genoma. Esta organización es debida en parte a la presión selectiva sobre genes y regiones del genoma que conservan características similares. El hecho de que hayamos encontrado esta región anómala en el genoma de *C. jejuni* podría deberse a que la transferencia horizontal sea relativamente reciente, de manera que las presiones selectivas de la maquinaria celular de esta bacteria no ha tenido el tiempo para modificar una región tan grande de DNA, o que los genes contenidos dentro de esta región confieren a *C. jejuni* tal ventaja adaptativa que esto ha perpetuado a la región misma dentro del genoma.

No. de ORFs que tienen ortología con *H. pylori* a lo largo del genoma de *C. jejuni*, en ventanas de 40 Kbp

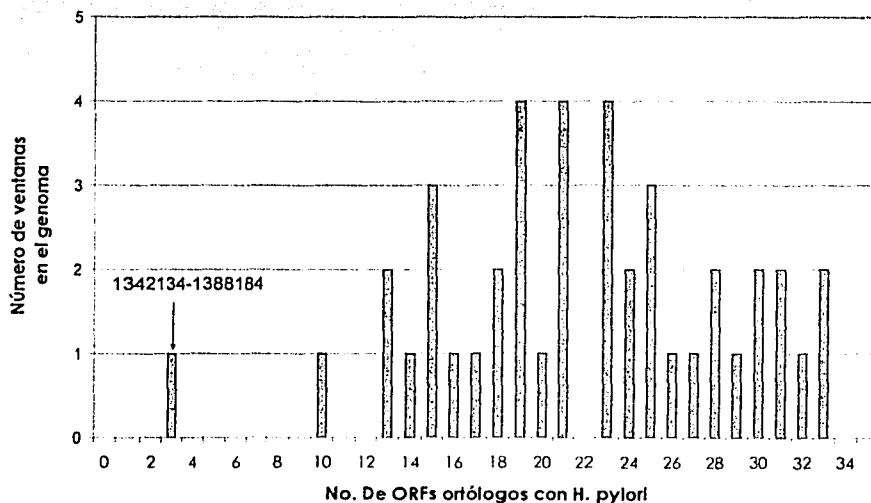


Figura 11. Número de ventanas de 40 kbp en *C. jejuni*, y su correspondiente número de ORFs ortólogos con *H. pylori*. Está señalado el fragmento propuesto como TGH, el cual contiene solamente 3 ORFs ortólogos.

Los genes que flanquean al fragmento propuesto como TGH en *C. jejuni*, 10 genes antes del gene Cj1410 y 10 genes arriba del gene Cj1439 fueron comparados con genes de *H. pylori*, para ubicar su posición en éste genoma. Todos ellos se encuentran distribuidos en todo el genoma de *H. pylori*, no se encuentran agrupados. Si estos genes estuvieran cercanos en el genoma de *H. pylori* esto nos podría ayudar a identificar un sitio donde podría haberse perdido todo el fragmento propuesto como TGH en el genoma de *H. pylori*. Debido a que no pudimos encontrar a los genes ortólogos de aquellos genes flanqueantes del fragmento de TGH agrupados en *H. pylori* no tenemos evidencia para soportar la idea de que *H. pylori* hubiera perdido los genes ortólogos de los genes que forman parte del fragmento de TGH.

POSIBLE DONADOR

Una vez realizados los análisis para descubrir si la presencia de un citocromo P450 en *C. jejuni* se debía a una transferencia horizontal, nuestra siguiente tarea fue intentar encontrar un posible donador para este fragmento. Tras un análisis de "Dual Best Hit" (Como se describe en Análisis Teóricos – Métodos) no logramos obtener resultados conclusivos. Para la mayoría de los ORFs no es posible obtener un "dual best hit", en la siguiente tabla presentamos cuales de los ORFs pertenecientes al fragmento obtuvieron un "dual best hit" y con qué especie:

ORF	Especie
Cj1414	<i>Pasteurella multocida</i>
Cj1415	<i>Pseudomonas fluorescens</i>
Cj1416	<i>Pseudomonas fluorescens</i>
Cj1417	<i>Pseudomonas fluorescens</i>
Cj1418	<i>Pseudomonas fluorescens</i>
Cj1419	<i>Pseudomonas fluorescens</i>
Cj1420	<i>Pseudomonas fluorescens</i>
Cj1423	<i>Aneurinibacillus thermoaerophilus</i>

TESIS CON
FALLA DE ORIGEN

Cj1424	<i>Aneurinibacillus thermoaerophilus</i>
Cj1427	<i>Synechocystis</i> sp.
Cj1432	<i>Salmonella typhi</i> strain CT18
Cj1433	<i>Desulfitobacterium hafniense</i>
Cj1436	<i>Microbulbifer degradans</i>
Cj1445	<i>Escherichia coli</i>

En la región que va desde Cj1414 hasta Cj1420 obtuvimos un fuerte indicativo de una posible ortología con *Pseudomonas fluorescens*, en ninguna otra parte de la región de TGH obtuvimos un resultado tan claro. No fue posible identificar cuáles son los productos de las secuencias de *Pseudomonas fluorescens* debido posiblemente a que los productos no estén caracterizados o simplemente a la manera en que están indexados algunas entidades dentro de la página del NCBI.

En los análisis de homología con los demás ORFs no obtuvimos ningún "dual best hit". En el apéndice 3 presentamos los primeras 10 bacterias con las cuales obtuvimos un "hit" contra cada uno de los ORFs de la región que nos interesa; sin embargo, en muchos casos los resultados tienen niveles de E muy altos o puntuaciones S muy bajas. Es común observar valores muy bajos de E que tienen puntuaciones S bajas (menores a 100), por esa razón es importante tomar en consideración la puntuación S de un resultado además del valor E. Por supuesto donde obtenemos valores E muy altos o S muy bajos es imposible inferir algún significado biológico.

Pudimos observar que los resultados o "hits" más comunes varían de ORF en ORF y si éste está localizado hacia el principio, en medio o hacia el final de la región de interés. Los resultados de estos análisis pueden variar en poco tiempo debido a que la base de datos del NCBI se actualiza constantemente, y continuamente se reportan nuevas secuencias.

TESIS CON
FALLA DE ORIGEN

DISCUSIÓN

Las dificultades para clasificar a las bacterias basándose únicamente en diferencias morfológicas han llevado al desarrollo de herramientas filogenéticas diseñadas para evidenciar las diferencias entre los genomas bacterianos (29). Estas herramientas se basan en la idea de que los genomas son relativamente estables en su estructura y secuencia, que sus cambios lentos, acumulados en el tiempo nos permiten distinguir una especie de otra. Los más recientes trabajos de secuenciación de genomas completos han aportado evidencia de que las transferencias horizontales pueden ser fenómenos más o menos frecuentes entre los procariontes, lo que pudiera llevar a pensar que las diferencias entre las especies se han ido perdiendo a lo largo del tiempo borrando la línea divisoria entre sí (3). Sin embargo la estabilidad en general de los genomas bacterianos es lo que nos permite detectar las TGHs (27).

Existe un número creciente de trabajos en los que se proponen eventos de TGH (3, 9, 19, 22, 27), pero estos fenómenos son difíciles de probar lo cual puede llevar a errores; a pesar de esto, se ha reportado una evidencia experimental de una TGH (2). Se han publicado trabajos en los cuales se estudian TGH basándose solamente en análisis de homología (18, 37, 44), este método ha llevado a problemas en las interpretaciones de los resultados, debido a que la ausencia de homologías en macromoléculas pertenecientes a especies relativamente cercanas, aún cuando sea posible encontrar homologías entre especies distantes, no es una evidencia directa de TGH; como se ha observado, semejantes resultados pueden ocurrir cuando hay pérdida de genes o de grupos de genes en familias enteras, o debido a la esporádica representación de los genomas completos en las bases de datos, llevando a interpretaciones erróneas (20, 27, 37, 44). La aproximación filogenética ha probado ser bastante útil en la resolución de filogenias, sin embargo, por sí sola puede llevar a errores

debido a su sensibilidad relativa a la cantidad de especies involucradas en el análisis, se requiere un gran poder computacional para realizar dichos análisis. Dependen por completo de los algoritmos matemáticos en los cuales están basados, dichos algoritmos tienen limitaciones inherentes asociadas a la reconstrucción de los posibles caminos evolutivos que reflejen de mejor manera los datos disponibles.

Se utilizaron herramientas filogenéticas y diversas estrategias de análisis de contexto genómico (20) con el objetivo de demostrar que la región del cromosoma de *C. jejuni* que va desde el gen putativo Cj1410 hasta el Cj1449, es producto de una TGH; y que la presencia de un citocromo P450 en *C. jejuni* probablemente se deba a este fenómeno y no a su herencia vertical desde un ancestro común con *H. pylori*.

Los análisis filogenéticos se realizaron utilizando un método de múltiple alineamiento progresivo de secuencias con Clustal, los algoritmos utilizados fueron el neighbor joining, máxima parsimonia y máxima verosimilitud (likelihood) (38,49). Se conoce que las bacterias *C. jejuni* y *H. pylori* están cercanas filogenéticamente y *Campylobacter* y *Helicobacter* son dos de los géneros más representativos del grupo delta/epsilon subdivisión epsilon. Para las clasificaciones taxonómicas la macromolécula más utilizada ha sido el RNA ribosomal 16s (27,49), por tratarse de una molécula altamente conservada. Realizamos un análisis filogenético utilizando secuencias de DNA para RNAr 16s pertenecientes a bacterias de varios grupos de eubacterias. Este análisis también nos sirvió para conocer la solidez teórica de los métodos que utilizamos, dadas las limitaciones inherentes de los métodos filogenéticos, era necesario corroborar nuestros resultados con la taxonomía convencional. El árbol resultante (apéndice) se asemeja a la taxonomía publicada en el NCBI, lo cual nos habla de la consistencia de nuestra metodología (7). Además de esto, resulta claro que *C. jejuni* y *H. pylori* son cercanas filogenéticamente entre sí, aún a

TESIS CON
FALLA DE ORIGEN

pesar de que es imposible en este momento establecer distancias de divergencia absolutas en nuestro árbol debido a las dificultades metodológicas y teóricas para establecer relojes moleculares (44). Con el método de NJ, obtuvimos un porcentaje de divergencia entre los RNAr 16s de *C. jejuni* y *H. pylori* de aproximadamente 0.14; el porcentaje de divergencia entre *S. typhimurium* y *E. coli* que obtuvimos fue de aprox. 0.02, para las cuales se ha calculado un tiempo de divergencia evolutiva de 100 millones de años (5), lo cual nos daría un estimado del tiempo de divergencia para *C. jejuni* y *H. pylori* de unos 700 millones de años, este tiempo es demasiado largo y no es consistente con el árbol de la vida (5), lo que demuestra la dificultad para establecer relojes moleculares.

El caso de *M. tuberculosis* H37Rv es peculiar ya que contiene el mayor número de citocromos P450 de las bacterias cuyos genomas han sido secuenciados a la fecha. Se desconoce hasta el momento cuál pudiera ser la implicación biológica de que *Mycobacterium* posea un tan alto número de citocromos P450 y se carecía de evidencia acerca del origen de estas proteínas en su genoma. Hicimos un análisis filogenético de citocromos P450 bacterianos incluyendo a los citocromos P450 de *M. tuberculosis*, esto nos permitió corroborar indirectamente que los citocromos P450 participan ampliamente en eventos de TGH, los cuales se perpetúan en las poblaciones bacterianas, probablemente debido a las características metabólicas de estos citocromos. Los citocromos P450 de *M. tuberculosis* se disgregan en un árbol en presencia de otros citocromos P450 debido a su origen distinto (21,22).

Los métodos filogenéticos tienen limitaciones inherentes a las suposiciones teóricas en base a los cuales se diseñan, por ello un análisis filogenético basado en una sola macromolécula podría no reflejar la verdadera historia filogenética de las especies involucradas en el análisis (49). Se hicieron árboles filogenéticos con varias proteínas que se

TESIS CON
FALLA DE ORIGEN

consideran conservadas con el objetivo de contar con una reconstrucción filogenética más completa de las especies consideradas para el análisis de citocromos P450; los alineamientos se encuentran en el apéndice 1 y los árboles en el apéndice 2.

La parsimonia ignora la información de la longitud de las ramas, mientras que ML (maxima verosimilitud) considera que los cambios tienen mayor probabilidad de ocurrir en las ramas más largas. Esta diferencia explica la mayor consistencia de ML en casos donde MP es inconsistente (14). Los métodos de distancias (como NJ) pueden tener ser muy útiles en estos casos también, aunque son menos eficientes que ML (14). Probablemente el método más confiable y consistente sea ML, sin embargo no contamos con un programa que nos permita utilizar este método de manera accesible.

Las topologías de los árboles son muy consistentes entre sí, entre los obtenidos con NJ y entre los obtenidos con MP. Los árboles que obtuvimos con MP se bifurcan constantemente mostrando una topología poco común, esto podría deberse a que por el tipo de secuencias, la longitud de las mismas o su número, hayamos caído dentro de una zona de inconsistencia para este algoritmo. En general las relaciones filogenéticas se conservan hasta cierto grado, las especies que aparecen cercanas con NJ, lo están también con MP. Los árboles de citocromos P450 (con NJ o MP) muestran distintas relaciones entre las especies consideradas, lo que supone que estas proteínas están involucradas en TGHs y da soporte a la idea de una TGH en *C. jejuni*.

Los árboles de citocromos P450 y de proteínas conservadas se construyeron utilizando las secuencias de aminoácidos ya que de esta manera es posible reducir las diferencias provocadas por las mutaciones puntuales, que en muchas ocasiones son mutaciones silenciosas. Esto permite concentrar el análisis sólo en cambios de aminoácidos los cuales

confieren estructura y función a la proteína. Las variaciones en el árbol pueden proveer información concerniente a la influencia de los aminoácidos en la función de la proteína y de las mutaciones asociadas con la conservación de la función o con cambios en la función (22).

Los análisis de contexto genómico que se hicieron posteriormente se basaron en el estudio de la estructura peculiar del fragmento que proponemos como TGH comparada con el resto del cromosoma de *C. jejuni*.

Se realizó una búsqueda de oligonucleótidos en el genoma de *C. jejuni* con la esperanza de hallar algún patrón distintivo en la distribución de alguno de ellos (13). La frecuencia acumulada de todos los tripletes excepto CCC es homogénea a lo largo de todo el cromosoma, con el triplete de citosinas pudimos observar una acumulación más rápida del mismo en la zona de la gráfica correspondiente al fragmento producto de TGH (Datos no mostrados), lo cual se refleja en un ligera declinación de la pendiente. Decidimos continuar aumentando nucleótidos de citosina a nuestra "sonda" con el fin de incrementar el efecto que observamos con el triplete CCC, con cuatro nucleótidos de citosina pudo verse un efecto más marcado y finalmente con el pentaplete de citosina se obtuvo el mejor resultado, cuando se aumentó a seis nucleótidos el patrón se rompió por completo. La meseta que observamos en la gráfica de la figura 9 indica una rápida acumulación de este pentaplete en esa región del cromosoma, y esta acumulación corresponde con la TGH propuesta, en esta región se acumula una cuarta parte de todos los pentapletes de este tipo en el genoma, a pesar de ser una región de apenas 40 Kb. Este resultado nos brinda una fuerte evidencia en apoyo de nuestra hipótesis, dado que esta organización es totalmente anómala con respecto al resto del genoma, lo cual es interesante además porque esta región contiene

TESIS CON
FALLA DE ORIGEN

un fragmento de 28351 bases de longitud con un contenido bajo de G+C (26.5%) (46). Desconocemos cual pueda ser el significado biológico de esta alta frecuencia de pentapletes de citosina, pero es una característica particular sólo de esta región lo cual indica un probable origen externo. Esta frecuencia de pentapletes de citosina tan característica del fragmento proveniente de una posible TGH, y tan distinta al resto del genoma de *C. jejuni*, podría servir como un marcador para buscar regiones con frecuencias acumuladas igualmente altas en otros genomas (39). En un futuro la búsqueda de acumulaciones de frecuencia de este pentaplete en los posibles candidatos donadores de la TGH, podría aportar evidencia más profunda sobre el posible donador, aunado a los resultados de los análisis de alineamientos de secuencias que se realizaron.

Analizamos la orientación de las secuencias codificantes con respecto a la dirección de su transcripción (tanto de función conocida como genes putativos sin función identificada) debido al hallazgo de que todos los ORFs que se localizan en el fragmento producto de TGH tienen la misma orientación de la transcripción, sin ningún ORF intercalado con su orientación contraria, lo cual parece no corresponder a un proceso estocástico (8). Tras de realizar este análisis pudimos observar que es común encontrar en el resto del genoma de uno a dos ORFs con la misma orientación seguidos de un ORF en la dirección contraria; es menos frecuente encontrar casos conforme aumenta el número de ORFs contiguos con la misma orientación, más allá de 21 ORFs contiguos con la misma orientación es raro encontrar grupos más grandes en el cromosoma, la región que es producto de una TGH sale por completo de la curva como el grupo más grande de ORFs con la misma dirección contiguos. Encontramos un segundo grupo con un alto número de ORFs contiguos orientados en la misma dirección, el cual va desde el ORF Cj0087 hasta el

Cj0122; el significado biológico de este segundo grupo con un alto número de ORFs contiguos orientados en la misma dirección, que se encuentra localizado a una distancia relativamente similar del punto ori, es desconocido, los ORFs localizados en este grupo no guardan ninguna relación con aquellos pertenecientes a la región propuesta como TGH.

Se analizó también la frecuencia de ORFs en el cromosoma de *C. jejuni* que tienen un ortólogo en *H. pylori*. Dado que un 55.4% del genoma de *C. jejuni* tiene ortología con *H. pylori*, podríamos esperar que de cada dos ORFs, uno tuviera ortología con *H. pylori* en una distribución normal. Nuestro análisis mostró una distribución cercana a una distribución de Poisson independientemente del tamaño de ventana que se utilizó, algunos datos que no ajustan perfectamente con la distribución de Poisson podrían deberse meramente a un sesgo debido al corrimiento de la ventana, y seguramente desaparecerían moviendo un poco el marco de las ventanas. Una vez más la región propuesta como TGH se sale por completo de la distribución observada para el resto del genoma. La región de TGH contiene solamente tres ORFs con ortología en *H. pylori* según los datos de ortología reportados por el Sanger Center (figura 12), este número es inusualmente bajo considerando el gran número de ORFs ortólogos con *H. pylori* en *C. jejuni*, además a ambos lados de esta región encontramos ORFs con ortología en *H. pylori*. Estos datos son consistentes con la hipótesis de que la región completa proviene de un fenómeno de transferencia horizontal; la ortología con *H. pylori* que poseen los tres ORFs dentro de la región podría deberse a un evento casual como podría ser la ausencia en las bases de datos de un ORF en otra especie con mayor similitud (27), y no al origen hereditario desde un ancestro común con *H. pylori*.

TESES CON
FALLA DE ORIGEN

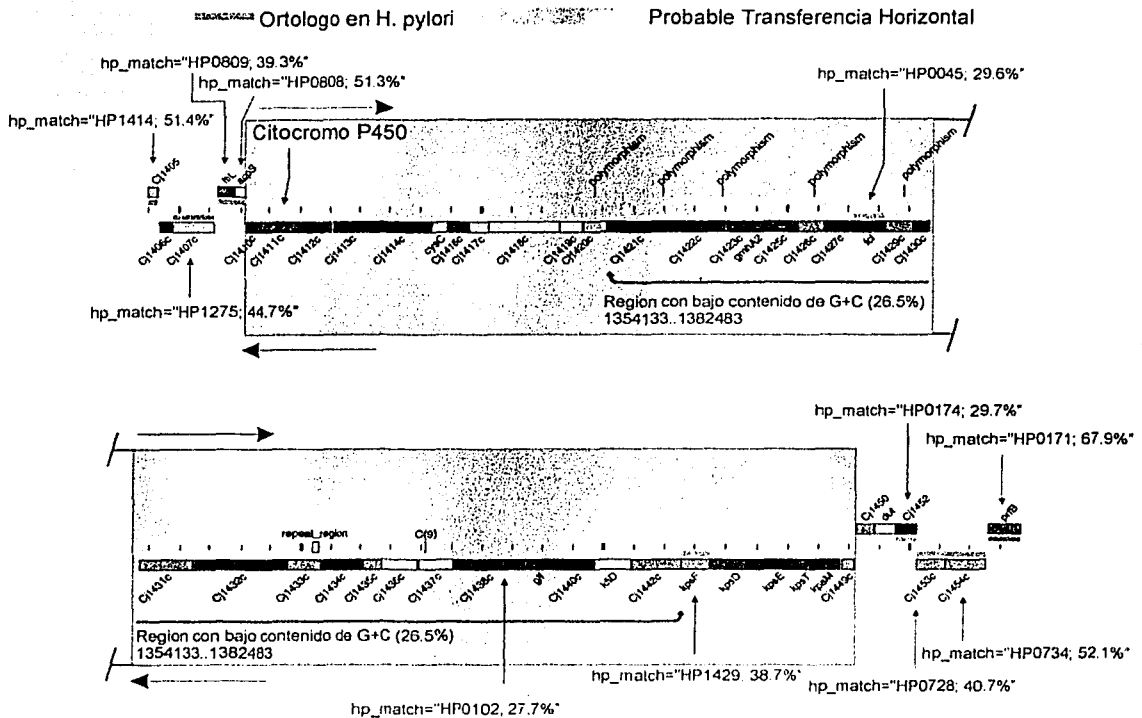


Figura 12. Esquema de la región producto de una TGH. Los bloques representan ORFs, se muestran los ORFs transferidos horizontalmente dentro del recuadro azul. Los rectángulos rosas de algunos ORFs indican ortología con *H. pylori*, y para estos casos está señalado el ORF ortólogo en *H. pylori* y su porcentaje de similitud. También está indicada la posición del citocromo P450 y de la región de bajo contenido de G+C dentro de la TGH.

La búsqueda de un posible donador para el fragmento que contiene al citocromo P450 de *C. jejuni* no nos brindó ninguna evidencia clara su origen. Obtuvimos un fuerte indicador en la región desde el ORF 1414 hasta el 1420 de *Pseudomonas fluorescens*, sin embargo, a lo largo de todo el fragmento no obtuvimos ningún otro resultado que apuntara fuertemente a esta bacteria perteneciente al grupo gama. A todo lo largo del fragmento se obtuvieron resultados muy variados, pero las bacterias que en la mayoría de los casos estuvieron presentes con buenos scores fueron: *Mesorhizobium loti* y *Agrobacterium tumefaciens*, *Escherichia coli*, *Mycobacterium tuberculosis*, *Bacillus subtilis*, *Bacillus halodurans*, *Neisseria*

meningitidis, *Pasteurella multocida* y *Pseudomonas fluorescens*. La bacteria *Pseudomonas fluorescens* parece ser un candidato para ser la donadora, pero de los resultados no puede concluirse que esta bacteria sea la donadora de este fragmento del cromosoma de *C. jejuni*. La base de datos del NCBI se encuentra en constante crecimiento, además el genoma de *Pseudomonas fluorescens* no está secuenciado por completo, por ello es posible que más adelante se puedan obtener resultados más claros. Otra evidencia interesante es que el plásmido Ti de *A. tumefaciens* contiene a los genes *virH1* y *virH2* que codifican un par de citocromos (mono-oxigenasas) P450 (53), y que ambos están flanqueados por elementos IS (secuencias de inserción), sin embargo no nos fue posible localizar elementos IS cercanos al fragmento propuesto como una TGH en *C. jejuni*.

Las Transferencias Genéticas Horizontales podrían estar participando activamente en los procesos evolutivos de muchas especies (27,46), tales transferencias pueden dotar a una bacteria de cualidades fenotípicas que podrían resultar en la rápida adaptación de dichas especies a nuevos ambientes (26,27). De acuerdo a la teoría del "Operón Egoísta" (21) cuando los genes están agrupados cercanamente esto representa una ventaja para que dichos genes puedan transferirse y conservarse en el genoma recipiente. Los operones introducidos recientemente pueden ser detectados por sus diferencias con el cromosoma del organismo recipiente. Si los genes nuevos tienen una estructura de DNA significativamente diferente, pueden ser identificados por su composición de bases (% G+C), sesgo en el uso de codones (50), huellas de polinucleótidos y diferencias estructurales con el DNA circundante (21). Si la transferencia es reciente, el DNA extraño no habrá tenido tiempo para sufrir ameliorización, esto significa ajustar su composición a la del DNA huésped (23, 29). Cuando se trata de una transferencia muy antigua es

posible que no pueda ser detectada o que sea muy difícil lograrlo, debido a la homogenización de la secuencia transferida con el DNA del recipiente. Esta es una de las razones por las cuales es conveniente implementar varias técnicas de detección de TGH cuando se analiza un posible evento de esta naturaleza (43), de manera que pueda minimizarse el riesgo de error de una falsa detección, debida a una evolución convergente y no a un evento de TGH.

Nuestros resultados muestran que la región que analizamos cuenta con características que corresponden con la estructura de un operón candidato para una TGH: a) los genes están agrupados muy cercanamente (21); b) debido a las propiedades catalíticas de los citocromos P450, éste puede proveer una ventaja adaptativa a *C. jejuni* permitiéndole utilizar nuevas fuentes de carbono o brindándole protección en medios hostiles; c) varios de los ORFs ubicados en la región que hemos propuesto como una TGH parecen tener funciones involucradas en el metabolismo de diversos azúcares, los cuales en conjunto pudieran inclusive constituir una vía metabólica completa (27).

Además de las estrategias más usadas en la búsqueda de transferencias horizontales, como son homología y relaciones filogenéticas, buscamos en el contexto genómico de *C. jejuni* otras evidencias que pudieran ayudarnos a responder nuestra pregunta experimental: la presencia de un citocromo P450 en *C. jejuni* se debe a un fenómeno de transferencia genómica horizontal, o si es parte de su herencia genómica y *H. pylori* perdió el gene del citocromo P450 después de que ambas divergieron de su ancestro común.

CONCLUSIONES

En el presente trabajo logramos llegar a las siguientes conclusiones:

1. La región del cromosoma de *C. jejuni* que involucra desde el ORF Cj1410 hasta el Cj1449 es resultado de un fenómeno de transferencia genómica horizontal.
2. Este fenómeno de transferencia horizontal debió presentarse en algún momento de la evolución de la bacteria *C. jejuni* posterior a que ella y *H. pylori* divergieron de su ancestro común.
3. La presencia de un gene que codifica para un citocromo P450 en el genoma de *C. jejuni* es resultado de un fenómeno de transferencia horizontal.
4. El análisis del contexto genómico de un fragmento de un cromosoma bacteriano puede brindarnos mayor evidencia en la búsqueda de transferencias horizontales, aunado a los análisis de homología y filogenia.

TESIS CON
FALLA DE ORIGEN

	10	20	30	40	50	60
Actinobaci				GAACGCTGGCGG	GC TAA	ACATGCA
Amycolatop		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Bhaloduran		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Emegateriu				GAACGCTGGCGG	GC TAA	ACATGCA
Bsubtilis		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Bjaponicum					GC TAA	ACATGCA
Buchnerasp				GAACGCTGGCGG	GC TAA	ACATGCA
C_jejuni		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Caulobacte		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Chlamydia		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Cpneumonia					GC TAA	ACATGCA
Deinococcu		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Ecoli_K12		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Ecoli_O157		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
H_pylori		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
M_loti					GC TAA	ACATGCA
Moritellay		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
M_leprae		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
M_smegmati					GC TAA	ACATGCA
M_tuber_H3		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
M_tuber_CD		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Mycoplasma		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Myxococcus		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Neisseria		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Pasteurell		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Pseudomona		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Rhodococcu					GC TAA	ACATGCA
Rickettsia		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Styphimuri		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Staphyloco		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Streptococ					GC TAA	ACATGCA
Savermitil		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Sclavulige					GC TAA	ACATGCA
Scoelicolo		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Sgriseus_s					GC TAA	ACATGCA
Slavendula		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Svenezuela					GC TAA	ACATGCA
Synechocys		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Thermotoga		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Treponema					GC TAA	ACATGCA
Ureaplasma		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Vibrio_cho		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Xylella_fa		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA
Mus_muscul		ATC	GCTCAG	GAACGCTGGCGG	GC TAA	ACATGCA

	70	80	90	100	110	120
Actinobaci	AGTCCG	CG			AG GCGG	ACGGGTG
Amycolatop	AGTCCG	CG			AG GCGG	ACGGGTG
Bhaloduran	AGTCCG	CG			AG GCGG	ACGGGTG
Emegateriu	AGTCCG	CG			AG GCGG	ACGGGTG
Bsubtilis	AGTCCG	CG			AG GCGG	ACGGGTG
Bjaponicum	AGTCCG	CG			AG GCGG	ACGGGTG
Buchnerasp	AGTCCG	CG			AG GCGG	ACGGGTG
C_jejuni	AGTCCG	CG			AG GCGG	ACGGGTG
Caulobacte	AGTCCG	CG			AG GCGG	ACGGGTG
Chlamydia	AGTCCG	CG			AG GCGG	ACGGGTG
Cpneumonia					AG GCGG	ACGGGTG
Deinococcu	AGTCCG	CG			AG GCGG	ACGGGTG
Ecoli_K12	AGTCCG	CG			AG GCGG	ACGGGTG
Ecoli_O157	AGTCCG	CG			AG GCGG	ACGGGTG
H_pylori	AGTCCG	CG			AG GCGG	ACGGGTG
M_loti	AGTCCG	CG			AG GCGG	ACGGGTG

TEXAS A&M
 FALLA DE ORIGEN

Alineamiento de rRNA 16S

Moritellay	AGTCGA	CG							AG	GGCG	ACGGGTG
M_leprae	AGTCGA	CG							AG	GGCG	ACGGGTG
M_smegmati	AGTCGA	CG							AG	GGCG	ACGGGTG
M_tuber_H3	AGTCGA	CG							AG	GGCG	ACGGGTG
M_tuber_CD	AGTCGA	CG							AG	GGCG	ACGGGTG
Mycoplasma	AGTCGA	CG							AG	GGCG	ACGGGTG
Myxococcus	AGTCGA	CG							AG	GGCG	ACGGGTG
Neisseria	AGTCG	CG							AG	GGCG	ACGGGTG
Pasteurell	AGTCGA	CG							AG	GGCG	ACGGGTG
Pseudomona	AGTCGA	CG							AG	GGCG	ACGGGTG
Rhodococcu	AGTCGA	CG							AG	GGCG	ACGGGTG
Rickettsia	AGTCGA	CG							AG	GGCG	ACGGGTG
Styphimuri	AGTCGA	CG							AG	GGCG	ACGGGTG
Staphyloco	AGTCGA	CG							AG	GGCG	ACGGGTG
Streptococ											ACGGGTG
Savermitil	AGTCGA	CG							AG	GGCG	ACGGGTG
Sclavulige	AGTCGA	CG							AG	GGCG	ACGGGTG
Scoelicolo	AGTCGA	CG							AG	GGCG	ACGGGTG
Sgriseus_s	AGTCGA	CG							AG	GGCG	ACGGGTG
Slavendula	AGTCGA	CG							AG	GGCG	ACGGGTG
Svenezuela	AGTCGA	CG							AG	GGCG	ACGGGTG
Synechocys	AGTCGA	CG							AG	GGCG	ACGGGTG
Thermotoga	AGTCGA	CG							AG	GGCG	ACGGGTG
Treponama		CG							AG	GGCG	ACGGGTG
Ureaplasma	AGTCGA	CG							AG	GGCG	ACGGGTG
Vibrio_cho	AGTCGA	CG							AG	GGCG	ACGGGTG
Xylella_fa	AGTCG	CG							AG	GGCG	ACGGGTG
Mus_muscul	AGTCGA	CG									CG

	130	140	150	160	170	180	
Actinobaci	AGTAA	TGGG	AA	CTG	GGGATAAC	GGAAAC	CTAA
Amycolacop	AGTAA	TGGG	AA	CTGCC	GGGATAA	GGAAAC	CTAA
Bhaloduran	AGTAA	TGGG	AA	CTGCC	GGGATAAC	GAAA	CTAA
Bmegateriu	AGTAA	TGGG	AA	CTGCC	GGGATAAC	GGAAAC	CTAA
Bsubtilis	AGTAA	TGGG	AA	CTGCC	GGGATAAC	GGAAAC	CTAA
Bjaponicum	AGTAA	TGGG	AA	CTCC	GGATAAC	GGAAAC	CTAA
Buchnerasp	AGTAA	TGGG	AA	CTCC	GGGATAAC	GAAA	CTAA
C_jejuni	AGTAA	TGGG	AA	CTGCC	GGATAAC	GGAAAC	CTAA
Caulobacte	AGTAA	TGGG	AA	TGCC	GGATAAC	GGAAAC	CTAA
Chlamydia	AGTAA	TGGG	AA	TGCC	GGATAAC	GGAAAC	CTAA
Cpneumonia	AGTAA	TGGG	AA	TGCC	GGGATAAC	GGAAAC	CTAA
Deinococcu	AGTAA	TGGG	AA	CTCC	GGATAAC	GAAA	CTAA
Ecoli_K12	AGTAA	TGGG	AA	CTGCC	GGGATAAC	GGAAAC	CTAA
Ecoli_O157	AGTAA	TGGG	AA	CTGCC	GGGATAAC	GGAAAC	CTAA
H_pylori	AGTAA	TGGG	AA	TGCC	GGGATAAC	GGAAAC	CTAA
M_loti	AGTAA	TGGG	AA	CTCC	GGATAAC	GGAAAC	CTAA
Moritellay	AGTAA	TGGG	AA	CTGCC	GGGATAAC	GGAAAC	CTAA
M_leprae	AGTAA	TGGG	AA	CTGCC	GGGATAAC	GGAAAC	CTAA
M_smegmati	AGTAA	TGGG	AA	CTGCC	GGGATAAC	GGAAAC	CTAA
M_tuber_H3	AGTAA	TGGG	AA	CTGCC	GGGATAAC	GGAAAC	CTAA
M_tuber_CD	AGTAA	TGGG	AA	CTGCC	GGGATAAC	GGAAAC	CTAA
Mycoplasma	AGTAA	TGGG	AA	CTCC	GGGATAAC	GAAA	CTAA
Myxococcus	AGTAA	TGGG	AA	CTGCC	GGGATAAC	GAAA	CTAA
Neisseria	AGTAA	TGGG	AA	CTCC	GGGATAAC	GAAA	CTAA
Pasteurell	AGTAA	TGGG	AA	CTGCC	GGGATAAC	GGAAAC	CTAA
Pseudomona	AGTAA	TGGG	AA	CTGCC	GGGATAAC	GGAAAC	CTAA
Rhodococcu	AGTAA	TGGG	AA	CTGCC	GGGATAAC	GGAAAC	CTAA
Rickettsia	AGTAA	TGGG	AA	CTCC	GGATAAC	GAAA	CTAA
Styphimuri	AGTAA	TGGG	AA	CTGCC	GGGATAAC	GGAAAC	CTAA
Staphyloco	AGTAA	TGGG	AA	CTCC	GGGATAAC	GGAAAC	CTAA
Streptococ	AGTAA	TGGG	AA	CTCC	GGGATAAC	GGAAAC	CTAA
Savermitil	AGTAA	TGGG	AA	CTGCC	GGGATAAC	GGAAAC	CTAA
Sclavulige	AGTAA	TGGG	AA	CTGCC	GGGATAAC	GGAAAC	CTAA
Scoelicolo	AGTAA	TGGG	AA	CTGCC	GGGATAAC	GGAAAC	CTAA

TESIS CON
FALLA DE ORIGEN

Alineamiento de rRNA 16S

Sgriseus_s	AGTAA	GGGG	CTGCC	GGGA	GGAAAC	CTAA
Slavendula	AGTAA	GGGG	CTGCC	GGGA	GGAAAC	CTAA
Svenezuela	AGTAA	GGGG	CTGCC	GGGA	GGAAAC	CTAA
Synechocys	AGTAA	GGGG	CTGCC	GGGA	GGAAAC	CTAA
Thermotoga	AGTAA	GGGG	CTGCC	GGGA	GGAAAC	CTAA
Treponema	AGTAA	GGGG	CTGCC	GGGA	GGAAAC	CTAA
Ureaplasma	AGTAA	GGGG	CTGCC	GGGA	GGAAAC	CTAA
Vibrio_cho	AGTAA	GGGG	CTGCC	GGGA	GGAAAC	CTAA
Xylella_fa	AGTAA	GGGG	CTGCC	GGGA	GGAAAC	CTAA
Mus_muscul	LA	EG	CC	GAT	AA	IAA

190 200 210 220 230 240

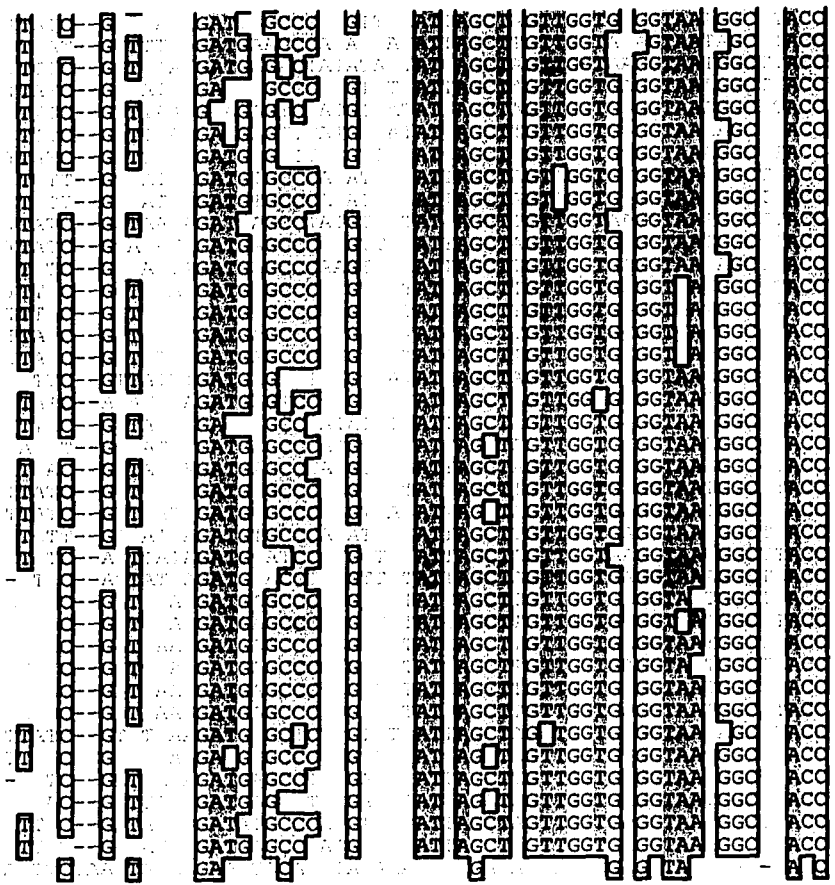
Actinobaci	TACCG	ATA				
Amycolatop	TACCG	ATA				
Bhaloduran	TACCG	ATA				
Bmegateriu	TACCG	ATA				
Bsubtilis	TACCG	ATA				
Bjaponicum	TACCG	ATA				
Buchnerasp	TACCG	ATA				
C_jejuni	TAC	ATA				
Caulobacte	TACCG	ATA				
Chlamydia	TACCG	ATA				
Cpneumonia	TACCG	ATA				
Deinococcu	TAC	ATA				
Ecoli_K12	TACCG	ATA				
Ecoli_O157	TACCG	ATA				
H_pylori	TACC	ATA				
M_loti	TACCG	ATA				
Moritellay	TACCG	ATA				
M_leprae	TACCG	ATA				
M_smeigmati	TACCG	ATA				
M_tuber_H3	TACCG	ATA				
M_tuber_CD	TACCG	ATA				
Mycoplasma	TACCG	ATA				
Myxococcus	TACCG	ATA				
Neisseria	TACCG	ATA				
Pasteurell	TACCG	ATA				
Pseudomona	TACCG	ATA				
Rhodococcu	TACCG	ATA				
Rickettsia	TACCG	ATA				
Styphimuri	TACCG	ATA				
Staphyloco	TACCG	ATA				
Streptococ	TACCG	ATA				
Savermitil	TACCG	ATA				
Sclavulige	TACCG	ATA				
Scoalicolo	TACCG	ATA				
Sgriseus_s	TACCG	ATA				
Slavendula	TACCG	ATA				
Svenezuela	TACCG	ATA				
Synechocys	TACC	ATA				
Thermotoga	TACC	ATA				
Treponema	TACCG	ATA				
Ureaplasma	TACCG	ATA				
Vibrio_cho	TACCG	ATA				
Xylella_fa	TACCG	ATA				
Mus_muscul	TACCG	ATA				

250 260 270 280 290 300

Actinobaci	AT	AG	GTGGTG	GGTAA	GGC	ACC
Amycolatop	AT	AGCT	GTGGTG	GGTAA	GGC	ACC
Bhaloduran	AT	AGCT	GTGGTG	GGTAA	GGC	ACC
Bmegateriu	AT	AGCT	GTGGTG	GGTAA	GGC	ACC
Bsubtilis	AT	AGCT	GTGGTG	GGTAA	GGC	ACC

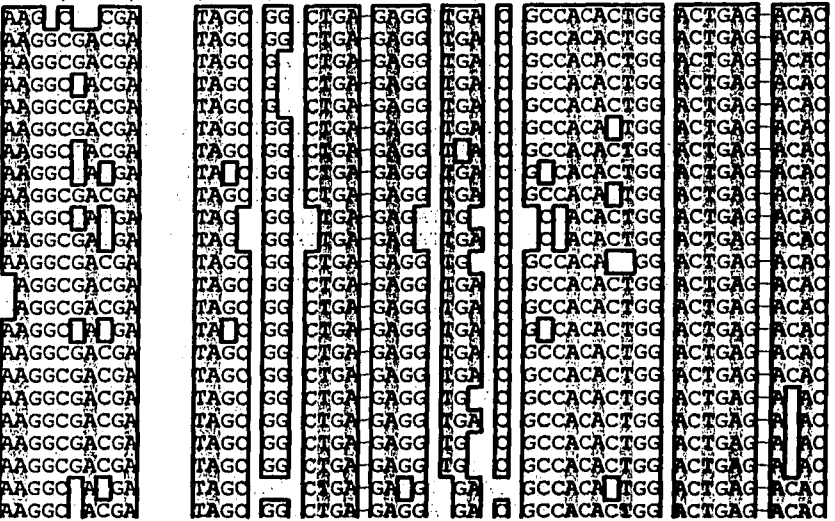
TESIS CON
FALLA DE ORIGEN

Bjaponicum
 Buchnerasp
 C_jejuni
 Caulobacte
 Chlamydia
 Cpnueumonia
 Deinococcu
 Ecoli_K12
 Ecoli_O157
 H_pylori
 M_loti
 Moritellay
 M_leprae
 M_smegmati
 M_tuber_H3
 M_tuber_CD
 Mycoplasma
 Myxococcus
 Neisseria
 Pasteurell
 Pseudomona
 Rhodococcu
 Rickettsia
 Styphimuri
 Staphyloco
 Streptococ
 Savermitil
 Sclavulige
 Scoelicolo
 Sgriseus_s
 Slavendula
 Svezuela
 Synachocys
 Therमतoga
 Treponema
 Ureaplasma
 Vibrio_cho
 Xylella_fa
 Mus_muscul



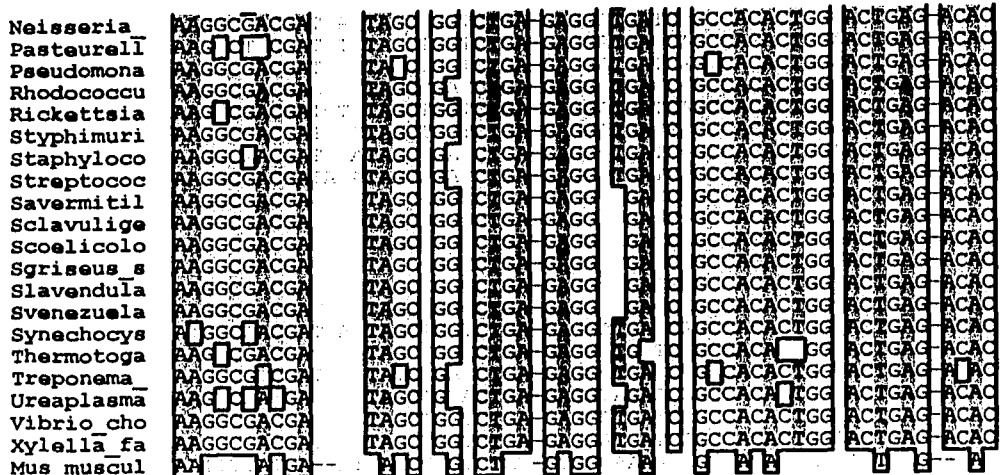
310 320 330 340 350 360

Actinobaci
 Amycolatop
 Bhaloduran
 Bmegateriu
 Bsubtilis
 Bjaponicum
 Buchnerasp
 C_jejuni
 Caulobacte
 Chlamydia
 Cpnueumonia
 Deinococcu
 Ecoli_K12
 Ecoli_O157
 H_pylori
 M_loti
 Moritellay
 M_leprae
 M_smegmati
 M_tuber_H3
 M_tuber_CD
 Mycoplasma
 Myxococcus



TESIS CON
 FALLA DE ORIGEN

Alineamiento de rRNA 16S



370 380 390 400 410 420



TESS CON
FALLA DE ORIGEN

Alineamiento de rRNA 16S

Vibrio cho GG CCA GAC CCA CGGG GGC GCG GGG -- CAC A GGGCG AAGCC
 Xylella_fa GG CCA GACT CCT ACGGG GGC AGC GGG -- CCA A GGGCG AAGCC
 Mus_muscul G G CCA G CCG A GGC GCG GGG -- C A A G GGG C AAG

	430	440	450	460	470	480
Actinobaci	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Amycolatop	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Bhaloduran	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Bmegateriu	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Bsubtilis	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Bjaponicum	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Buchnerasp	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
C_jejuni	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Caulobacte	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Chlamydia	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Cpnemonia	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Deinococcu	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Ecoli_K12	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Ecoli_O157	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
H_pylori	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
M_loti	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Moritellay	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
M_leprae	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
M_smeagmati	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
M_tuber_H3	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
M_tuber_CD	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Mycoplasma	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Myxococcus	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Neisseria	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Pasteurell	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Pseudomona	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Rhodococcu	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Rickettsia	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Styphimuri	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Staphyloco	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Streptococ	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Savermitil	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Sclavulige	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Scoelicolo	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Sgriseus_s	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Slavendula	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Svenezuela	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Synechocycs	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Thermotoga	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Treponema	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Ureaplasma	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Vibrio cho	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Xylella_fa	GATGCA GC	GCCGCGTG	GA	GG	GG	GG
Mus_muscul	GATGCA GC	GCCGCGTG	GA	GG	GG	GG

	490	500	510	520	530	540
Actinobaci	GA GAA			AGACG TAC	GAA AAGC	
Amycolatop	GA GAA			AGACGGTACC	GA AAGC	
Bhaloduran	GA GAA			AGACGGTACC	GA AAGC	
Bmegateriu	GA GAA			AGACGGTACC	GA AAGC	
Bsubtilis	GA GAA			AGACGGTACC	GA AAGC	
Bjaponicum	GA GAA			AGACGGTACC	GA AAGC	
Buchnerasp	GA GAA			AGACGGTACC	GA AAGC	
C_jejuni	GA GAA			AGACGGTACC	GA AAGC	
Caulobacte	GA GAA			AGACGGTACC	GA AAGC	
Chlamydia	GA GAA			AGACGGTACC	GA AAGC	
Cpnemonia	GA GAA			AGACGGTACC	GA AAGC	
Deinococcu	GA GAA			AGACGGTACC	GA AAGC	

TESIS CON
FALLA DE ORIGEN

Alineamiento de rRNA 16S

Ecoli_K12	GA	GAA			GA	CG	TACC	GA	AGC
Ecoli_O157	GA	GAA			GA	CG	TACC	GA	AGC
H_pylori	GA	GA			GA	CGGT		GA	AGC
M_loti	GA	GA			GA	CGGT		GA	AGC
Moritellay	GA	GAA			GA	CG	TACC	GA	AGC
M_leprae	GA	GAA			GA	CGGT		GA	AGC
M_smegmati	GA	GAA			GA	CGGT		GA	AGC
M_tuber_H3	GA	GAA			GA	CGGT		GA	AGC
M_tuber_CD	GA	GAA			GA	CGGT		GA	AGC
Mycoplasma	GA	GAA			GA	CG	TACC	GA	AGC
Myxococcus	GA	GAA			GA	CGGT	CC	GA	AGC
Neisseria	GA	GAA			GA	CGGT	CC	GA	AGC
Pasteurell	GA	GAA			GA	CG	TACC	GA	AGC
Pseudomona	GA	GAA			GA	CGGT	CC	GA	AGC
Rhodococcu	GA	GAA			GA	CGGT	CC	GA	AGC
Rickettsia	GA	GA			GA	CG	TACC	GA	AGC
Styphimuri	GA	GAA			GA	CGGT	CC	GA	AGC
Staphyloco	GA	GAA			GA	CGGT	CC	GA	AGC
Streptococ	GA	GAA			GA	CGGT	CC	GA	AGC
Savermiitil	GA	GAA			GA	CGGT	CC	GA	AGC
Sclavulige	GA	GAA			GA	CGGT	CC	GA	AGC
Scoelicolo	GA	GAA			GA	CGGT	CC	GA	AGC
Sgriseus_s	GA	GAA			GA	CGGT	CC	GA	AGC
Slavendula	GA	GAA			GA	CGGT	CC	GA	AGC
Svenezuela	GA	GAA			GA	CGGT	CC	GA	AGC
Synechocys	GA	GAA			GA	CGGT	CC	GA	AGC
Thermotoga	GA	GAA			GA	CGGT	CC	GA	AGC
Treponema	GA	GAA			GA	CG	TACC	GA	AGC
Ureaplasma	GA	GAA			GA	CG	TACC	GA	AGC
Vibrio_cho	GA	GAA			GA	CGGT	CC	GA	AGC
Xylella_fa	GA	GAA			GA	CGGT	CC	GA	AGC
Mus_muscul	GA	GAA			GA	CGGT	CC	GA	AGC

	550	560	570	580	590	600		
Actinobaci	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Amycolatop	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Bhaloduran	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Bmegateriu	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Bsubtilis	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Bjaponicum	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Buchnerasp	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
C_jejuni	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Caulobacte	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Chlamydia	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Cpneumonia	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Deinococcu	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Ecoli_K12	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Ecoli_O157	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
H_pylori	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
M_loti	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Moritellay	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
M_leprae	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
M_smegmati	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
M_tuber_H3	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
M_tuber_CD	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Mycoplasma	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Myxococcus	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Neisseria	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Pasteurell	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Pseudomona	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Rhodococcu	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Rickettsia	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Styphimuri	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT
Staphyloco	CGGCTAACT	CGG	GCCAGCAGCCGCGG	CG	AGGG	GC	AGCGT	CGGAAT

TESIS CON
FALLA DE ORIGEN

Alineamiento de rRNA 16S

Streptococ	CGGC	TAA	CG	GCC	GC	GCCGCGG	CG	GGG	GC	GCG	CGG
Savermitil	CGGC	TAA	CG	GCC	GC	GCCGCGG	CG	GGG	GC	GCG	CGG
Sclavulige	CGGC	TAA	CG	GCC	GC	GCCGCGG	CG	GGG	GC	GCG	CGG
Scoelicolo	CGGC	TAA	CG	GCC	GC	GCCGCGG	CG	GGG	GC	GCG	CGG
Sgriseus_s	CGGC	TAA	CG	GCC	GC	GCCGCGG	CG	GGG	GC	GCG	CGG
Slavendula	CGGC	TAA	CG	GCC	GC	GCCGCGG	CG	GGG	GC	GCG	CGG
Svenezuela	CGGC	TAA	CG	GCC	GC	GCCGCGG	CG	GGG	GC	GCG	CGG
Synechocys	CGGC	TAA	CG	GCC	GC	GCCGCGG	CG	GGG	GC	GCG	CGG
Thermotoga	CGGC	TAA	CG	GCC	GC	GCCGCGG	CG	GGG	GC	GCG	CGG
Treponema_	CGGC	TAA	CG	GCC	GC	GCCGCGG	CG	GGG	GC	GCG	CGG
Ureaplasma	CGGC	TAA	CG	GCC	GC	GCCGCGG	CG	GGG	GC	GCG	CGG
Vibrio_cho	CGGC	TAA	CG	GCC	GC	GCCGCGG	CG	GGG	GC	GCG	CGG
Xylella_fa	CGGC	TAA	CG	GCC	GC	GCCGCGG	CG	GGG	GC	GCG	CGG
Mus_muscul	TAA	TAA	CG	GCC	GC	GCCGCGG	CG	GGG	GC	GCG	CGG

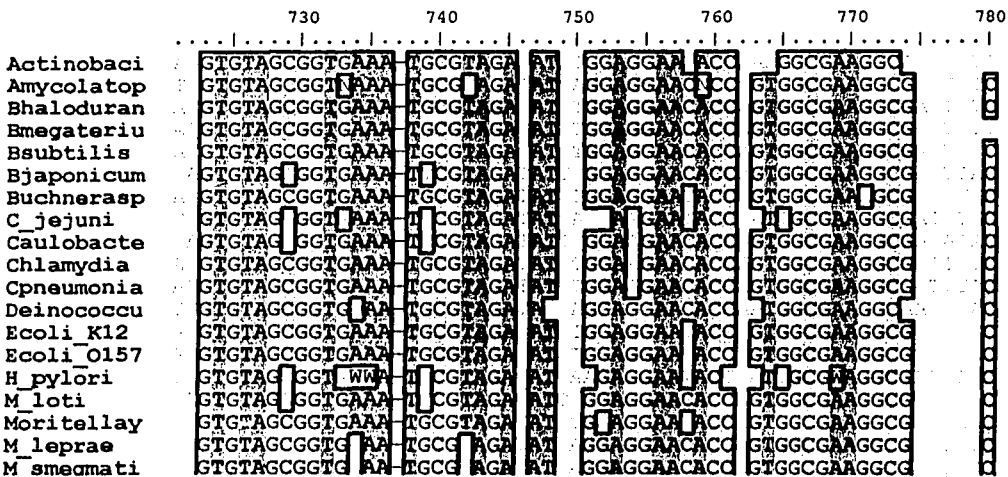
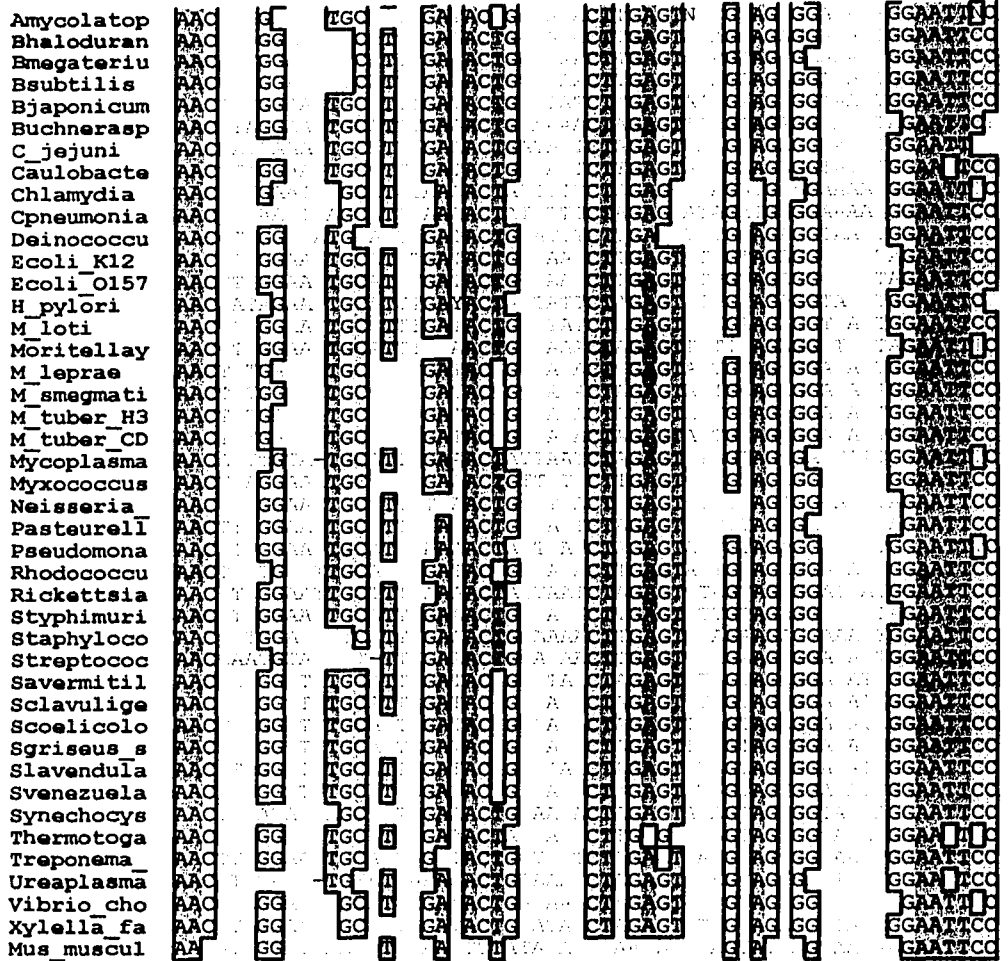
610 620 630 640 650 660

Actinobaci	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Amycolatop	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Bhaloduran	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Bnegateriu	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Bsubtilis	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Bjaponicum	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Buchnerasp	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
C_jejuni	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Caulobacte	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Chlamydia	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Cpneumonia	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Deinococcu	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Ecoli_K12	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Ecoli_O157	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
H_pylori	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
M_loti	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Moritellay	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
M_leprae	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
M_smegmati	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
M_tuber_H3	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
M_tuber_CD	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Mycoplasma	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Myxococcus	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Neisseria	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Pasteurell	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Pseudomona	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Rhodococcu	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Rickettsia	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Styphimuri	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Staphyloco	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Streptococ	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Savermitil	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Sclavulige	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Scoelicolo	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Sgriseus_s	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Slavendula	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Svenezuela	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Synechocys	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Thermotoga	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Treponema_	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Ureaplasma	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Vibrio_cho	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Xylella_fa	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC
Mus_muscul	TGGGCGTAAAG	GC	CG	TGGGCGTAAAG	AG	TGGGCGTAAAG	CCG	GGC

670 680 690 700 710 720

Actinobaci AAC GG TGC T A ACTG CA GAGT AG G GAATCC

TESIS CON
 FALLA DE ORIGEN



TESIS CON FALLA DE ORIGEN

Alineamiento de rRNA 16S

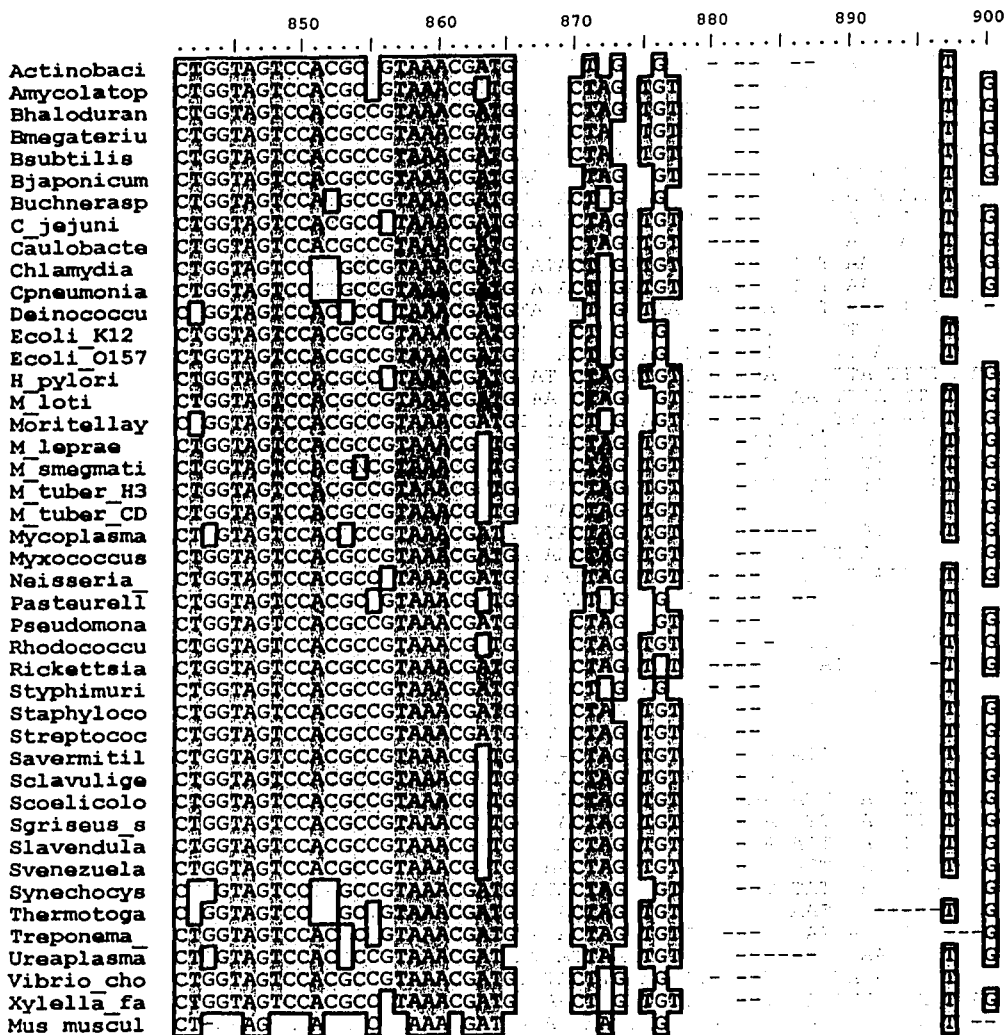
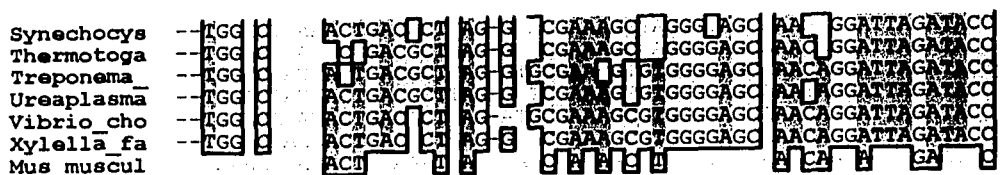
M_tuber_H3	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
M_tuber_CD	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Mycoplasma	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Myxococcus	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Neisseria	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Pasteurell	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Pseudomona	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Rhodococcu	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Rickettsia	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Styphimuri	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Staphyloco	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Streptococ	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Savermitil	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Sclavulige	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Scoelicolo	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Sgriseus s	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Slavendula	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Svenezuela	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Synechocys	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Thermotoga	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Treponema	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Ureaplasma	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Vibrio_cho	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Xylella_fa	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G
Mus_muscul	G T G T A G C G G G G G A A T G C C G G G	G G G G C C C C G G G C G G C C G

790 800 810 820 830 840

Actinobaci	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Amycolatop	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Bhaloduran	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Bmegateriu	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Bsubtilis	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Bjaponicum	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Buchnerasp	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
C_jejuni	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Caulobacte	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Chlamydia	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Cpneumonia	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Deinococcu	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Ecoli_K12	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Ecoli_O157	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
H_pylori	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
M_loti	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Moritellay	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
M_leprae	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
M_smegmati	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
M_tuber_H3	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
M_tuber_CD	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Mycoplasma	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Myxococcus	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Neisseria	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Pasteurell	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Pseudomona	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Rhodococcu	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Rickettsia	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Styphimuri	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Staphyloco	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Streptococ	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Savermitil	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Sclavulige	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Scoelicolo	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Sgriseus s	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Slavendula	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC
Svenezuela	ACTGACGCT	AGG	GCGAAGCGTGGGGAGC	AACAGGATTAGATACC

TESIS CON
 FALLA DE ORIGEN

Alineamiento de rRNA 16S



TESIS CON
FALLA DE ORIGEN

Alineamiento de rRNA 16S

Caulobacte	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Chlamydia	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGAGTAC	CGCAAG
Cpneumonia	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGAGTAC	CGCAAG
Deinococcu	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGAGTACGG	CGCAAG
Ecoli_K12	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Ecoli_O157	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
H_pylori	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
M_loti	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Moritellay	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
M_leprae	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
M_smegmati	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
M_tuber_H3	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
M_tuber_CD	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Mycoplasma	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGAGTAC	CGCAAG
Myxococcus	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGAGTACGG	CGCAAG
Neisseria	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Pasteurell	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Pseudomona	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Rhodococcu	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Rickettsia	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Styphimuri	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Staphyloco	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Streptococ	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Savermitil	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Sclavulige	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Scoelicolo	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Sgriseus_s	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Slavendula	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Svenezuela	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Synechocys	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Therमतoga	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Treponama	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Ureaplasma	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGAGTAC	CGCAAG
Vibrio_cho	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Xylella_fa	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG
Mus_muscul	CG	AGCTAACGG	ATAAG	CCGCC	---	GGGGAGTACGG	CGCAAG

970 980 990 1000 1010 1020

Actinobaci	ACTCAAA	GAATTGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
Amycolatop	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	NNNNNNNNNNNNNN	GC
Bhaloduran	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
Bmegateriu	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
Bsubtilis	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
Bjaponicum	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
Buchnerasp	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
C_jejuni	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
Caulobacte	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
Chlamydia	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
Cpneumonia	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
Deinococcu	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
Ecoli_K12	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
Ecoli_O157	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
H_pylori	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
M_loti	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
Moritellay	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
M_leprae	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
M_smegmati	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
M_tuber_H3	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
M_tuber_CD	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
Mycoplasma	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
Myxococcus	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
Neisseria	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
Pasteurell	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC
Pseudomona	ACTCAAA	GGAAATGACGGGGGCCCGCAC	AGCGGTGGAGCATGTGG	TTAATTCGA	GC

TESIS CON FALLA DE ORIGEN

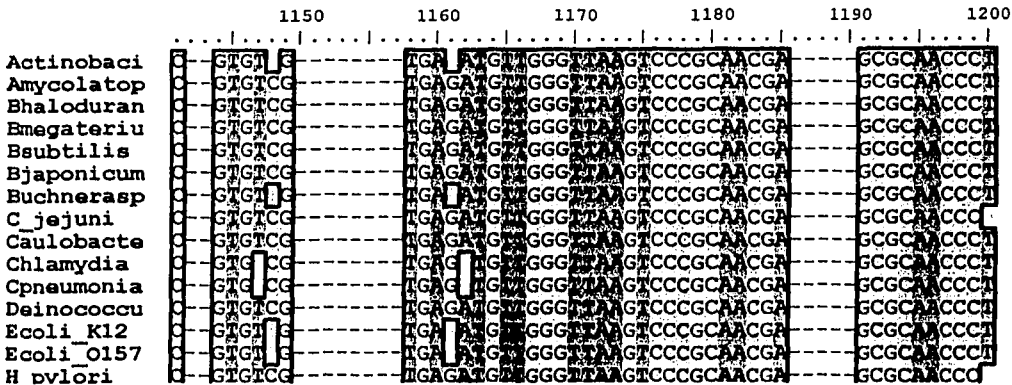
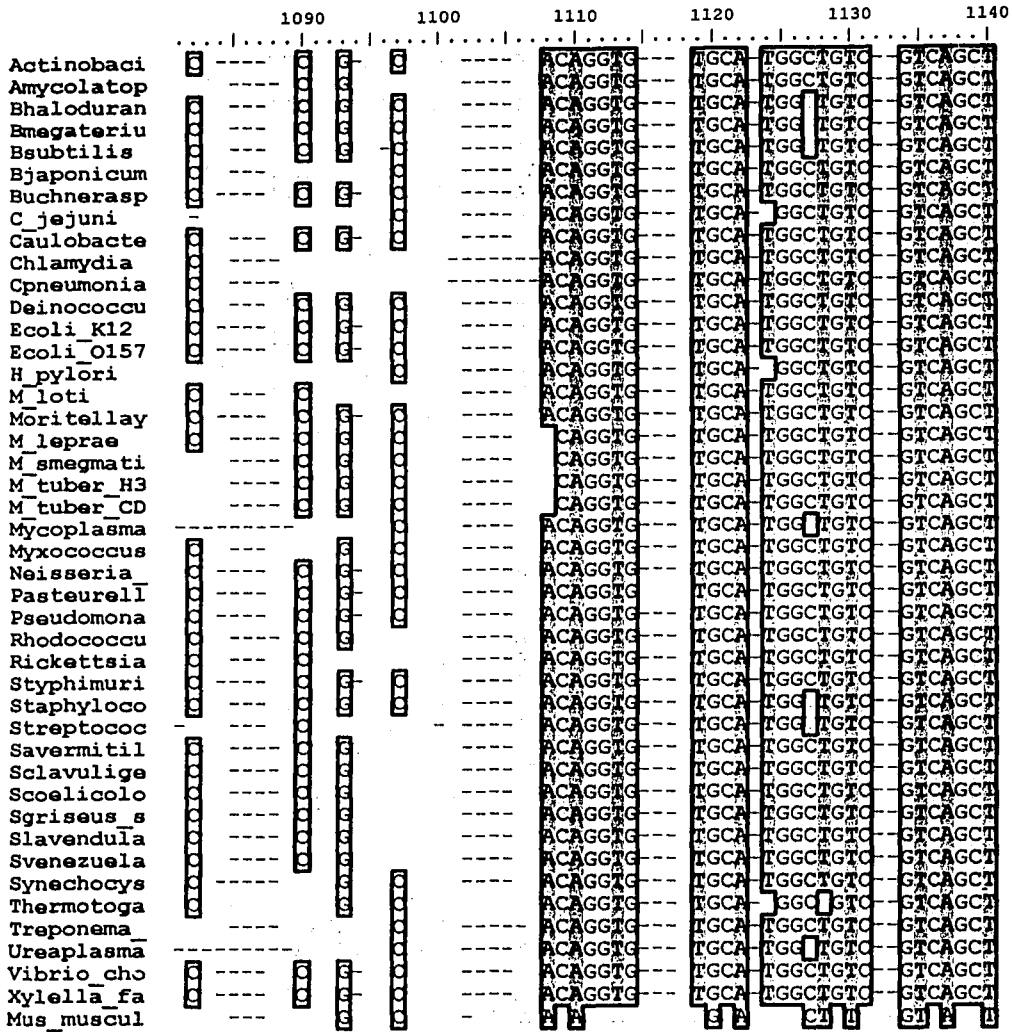
Alineamiento de rRNA 16S

Rhodococcu	ACTCAA	GGAAATGACGGGGGCCCGCAC	AGCGG	GGAGCATGTGG	TTAATTCGA	GC
Rickettsia	ACTCAA	GGAAATGACGGGGGCCCGCAC	AGCGG	GGAGCATGTGG	TTAATTCGA	GC
Styphimuri	ACTCAA	GGAAATGACGGGGGCCCGCAC	AGCGG	GGAGCATGTGG	TTAATTCGA	GC
Staphyloco	ACTCAA	GGAAATGACGGGGGCCCGCAC	AGCGG	GGAGCATGTGG	TTAATTCGA	GC
Streptococ	ACTCAA	GGAAATGACGGGGGCCCGCAC	AGCGG	GGAGCATGTGG	TTAATTCGA	GC
Savermitil	ACTCAA	GGAAATGACGGGGGCCCGCAC	AGCGG	GGAGCATGTGG	TTAATTCGA	GC
Sclavulige	ACTCAA	GGAAATGACGGGGGCCCGCAC	AGCGG	GGAGCATGTGG	TTAATTCGA	GC
Scoelicolo	ACTCAA	GGAAATGACGGGGGCCCGCAC	AGCGG	GGAGCATGTGG	TTAATTCGA	GC
Sgriseus_s	ACTCAA	GGAAATGACGGGGGCCCGCAC	AGCGG	GGAGCATGTGG	TTAATTCGA	GC
Slavendula	ACTNNP	GGAAATGACGGGGGCCCGCAC	AGCGG	GGAGCATGTGG	TTAATTCGA	GC
Svenezuela	ACTCAA	GGAAATGACGGGGGCCCGCAC	AGCGG	GGAGCATGTGG	TTAATTCGA	GC
Synechocys	ACTCAA	GGAAATGACGGGGGCCCGCAC	AGCGG	GGAGCATGTGG	TTAATTCGA	GC
Thermotoga	ACTCAA	GGAAATGACGGGGGCCCGCAC	AGCGG	GGAGCATGTGG	TTAATTCGA	GC
Treponema	ACTCAA	GGAAATGACGGGGGCCCGCAC	AGCGG	GGAGCATGTGG	TTAATTCGA	GC
Ureaplasma	ACTCAA	GGAAATGACGGGGGCCCGCAC	AGCGG	GGAGCATGTGG	TTAATTCGA	GC
Vibrio_cho	ACTCAA	GGAAATGACGGGGGCCCGCAC	AGCGG	GGAGCATGTGG	TTAATTCGA	GC
Xylella_fa	ACTCAA	GGAAATGACGGGGGCCCGCAC	AGCGG	GGAGCATGTGG	TTAATTCGA	GC
Mus_muscul	UAA	GG	U	C	G	C

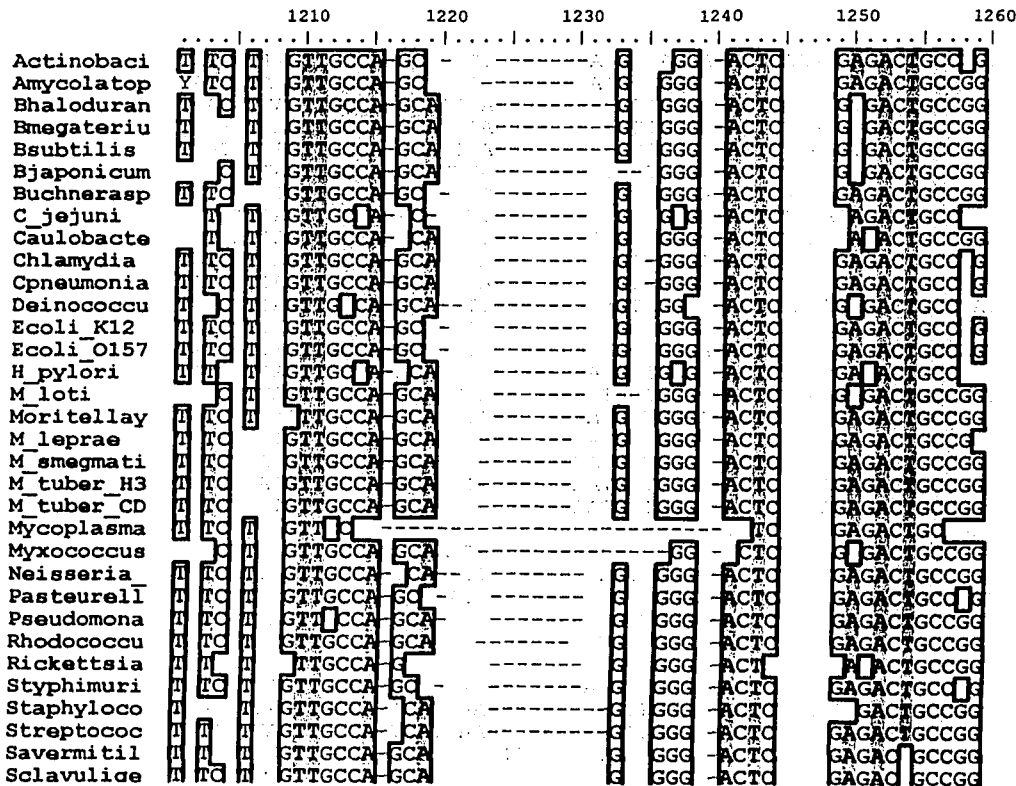
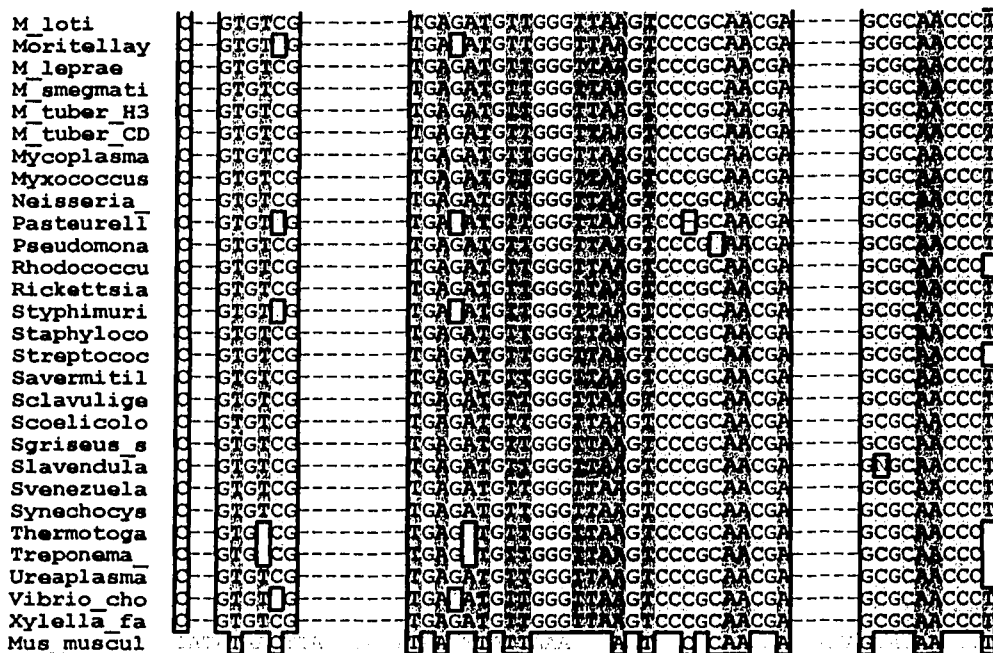
1030 1040 1050 1060 1070 1080

Actinobaci	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Amycolatop	NNNCGGAAGAACCTTACC	CTMGACAT				AGAGAT
Bhaloduran	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Bmegateriu	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Bsubtilis	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Bjaponicum	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Buchnerasp	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
C_jejuni	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Caulobacte	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Chlamydia	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Cpneumonia	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Deinococcu	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Ecoli_K12	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Ecoli_O157	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
H_pylori	ACCGAAGAACCTTACC	CTMGACAT				AGAGAT
M_loti	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Moritellay	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
M_leprae	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
M_smeigmati	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
M_tuber_H3	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
M_tuber_CD	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Mycoplasma	ACCGAAGAACCTTACC	CTMGACAT				AGAGAT
Myxococcus	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Neisseria	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Pasteurell	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Pseudomona	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Rhodococcu	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Rickettsia	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Styphimuri	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Staphyloco	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Streptococ	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Savermitil	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Sclavulige	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Scoelicolo	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Sgriseus_s	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Slavendula	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Svenezuela	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Synechocys	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Thermotoga	AGGAGAACCTTACC	CTMGACAT				AGAGAT
Treponema	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Ureaplasma	ACCGAAGAACCTTACC	CTMGACAT				AGAGAT
Vibrio_cho	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Xylella_fa	AACGCGAAGAACCTTACC	CTMGACAT				AGAGAT
Mus_muscul	AGGAGAACCTTACC	CTMGACAT				AGAGAT

TESIS CON FALLA DE ORIGEN

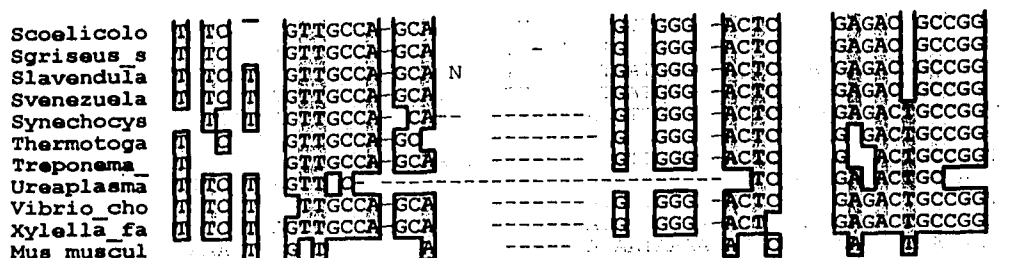


TESIS CON
FALLA DE ORIGEN

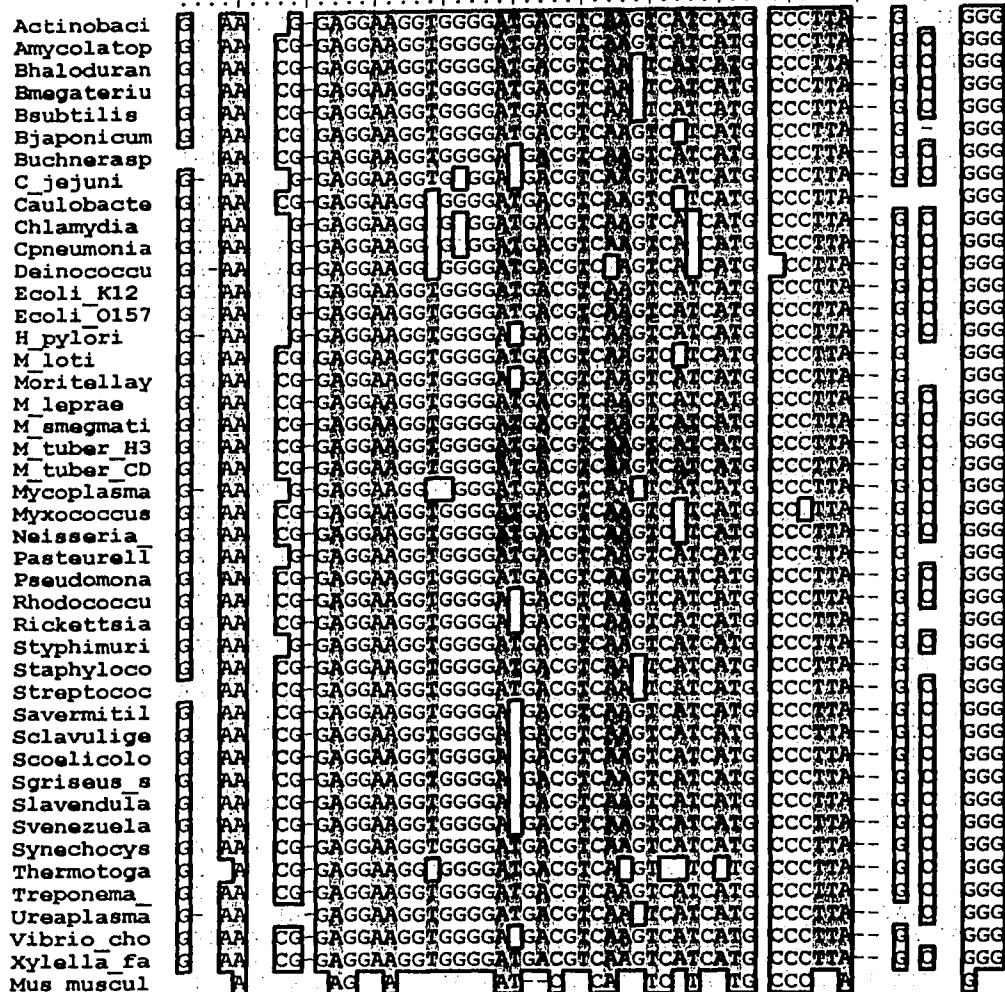


TESIS CON
FALLA DE ORIGEN

Alineamiento de rRNA 16S



1270 1280 1290 1300 1310 1320

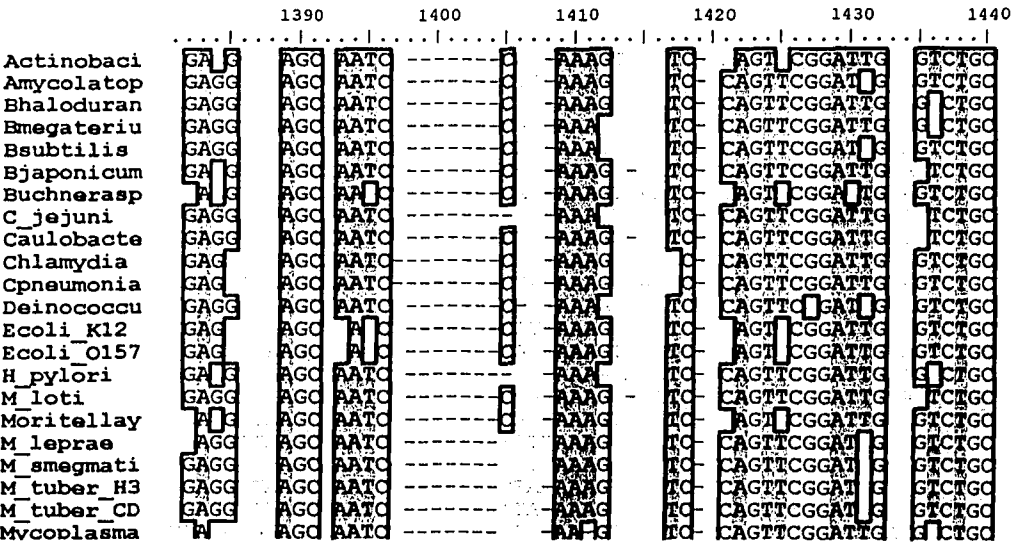
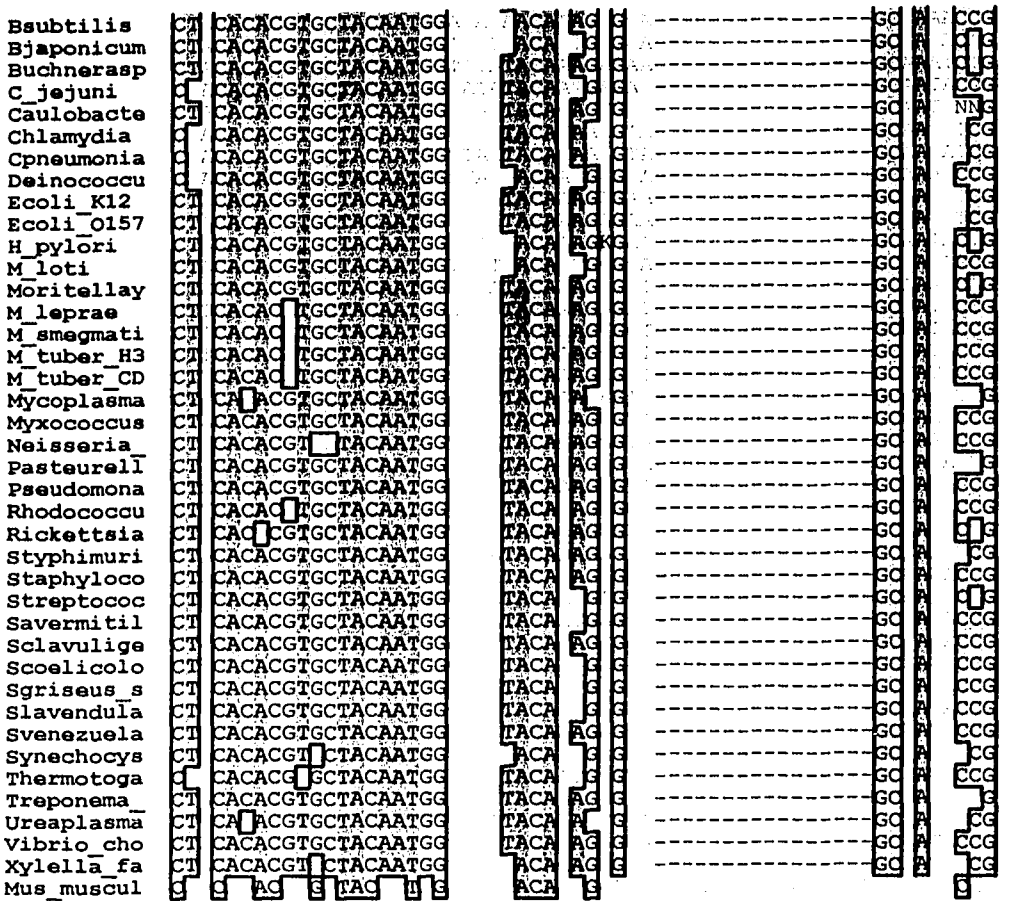


1330 1340 1350 1360 1370 1380



TESIS CON
 FALLA DE ORIGEN

Alineamiento de rRNA 16S



TESIS CON
FALLA DE ORIGEN

Alineamiento de rRNA 16S

Myxococcus	GAGG	AGC	AATC			TC	CAGTTC	GATTG	GTCTGC
Neisseria	GAGG	AGC	AATC			TC	AGT	CGGATG	TCCTGC
Pasteurell	GAG	AGC	AATC			TC	AGT	CGGATG	GTCTGC
Pseudomona	GAGG	AGC	AATC			TC	AGT	CGGATG	GTCTGC
Rhodococcu	GAGG	AGC	AATC			TC	CAGTTCGGAT	G	GTCTGC
Rickettsia	GA	AGC	AATC			TC	CAGTTCGGATG		TCCTGC
Styphimuri	GAG	AGC	AATC			TC	AGT	CGGATG	GTCTGC
Staphyloco	GAGG	AGC	AATC			TC	CAGTTCGGATG		GTCTGC
Streptococ	GA	AGC	AATC			TC	CAGTTCGGATG		GTCTGC
Savermartil	AGG	AGC	AATC			TC	CAGTTCGGATG		GTCTGC
Sclavulige	EAGG	AGC	AATC			TC	CAGTTCGGATG		GTCTGC
Scoelicolo	AGG	AGC	AATC			TC	CAGTTCGGATG		GTCTGC
Sgriseus_s	GAGG	AGC	AATC			TC	CAGTTCGGATG		GTCTGC
Slavandula	GAGG	AGC	AATC			TC	CAGTTCGGATG		GTCTGC
Svenezuela	GAGG	AGC	AATC			TC	CAGTTCGGATG		GTCTGC
Synechocys	GAG	AGC	AATC			TC	CAGTTC	GATTG	GTCTGC
Thermotoga	GAGG	AGC	AATC			TC	AGT	CGGATG	GTCTGC
Treponema	GA	AGC	AATC			TC	CAGTTCGGAT	G	GTCTGC
Ureaplasma	AG	AGC	AATC			TC	AGT	CGGATG	GTCTGC
Vibrio_cho	GAGG	AGC	AATC			TC	CAGTTCGGATG		GTCTGC
Xylrella_fa	GAGG	AGC	AATC			TC	CAGTTCGGATG		GTCTGC
Mus_muscul	AGG	AGC	AATC			TC	CAGTTCGGATG		GTCTGC

1450 1460 1470 1480 1490 1500

Actinobaci	AACTCGAC	CATGAAGT	GGATCGCTAG	GAATCGG	ATCAG	G	CGGGTGAAT
Amycolatop	AACTCGAC	CTGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Bhaloduran	AACTCGC	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Emegateriu	AACTCGC	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Bsubtilis	AACTCGAC	CTGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Bjaponicum	AACTCGA	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Buchnerasp	AACTCGAC	CA	GAAGT	GGATCGCTAG	GAATCGG	GC	CGGGTGAAT
C_jejuni	AACTCGA	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Caulobacte	AACTCGA	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Chlamydia	AACTCGAC	CATGAAGT	GGATCGCTAG	GAATCGG	ATCAG	GC	CGGGTGAAT
Cpneumonia	AACTCGAC	CATGAAGT	GGATCGCTAG	GAATCGG	ATCAG	GC	CGGGTGAAT
Deinococcu	AACTCGAC	CTGAAGT	GGATCGCTAG	GAATCGG	ATCAGCA	GC	CGGGTGAAT
Ecoli_K12	AACTCGAC	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Ecoli_O157	AACTCGAC	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
H_pylori	AACTCGC	CATGAAGT	GGATCGCTAG	GAATCGG	ATCAGCA	GC	CGGGTGAAT
M_loti	AACTCGA	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Moritellay	AACTCGAC	CATGAAGT	GGATCGCTAG	GAATCGG	ATCAGCA	GC	CGGGTGAAT
M_leprae	AACTCGAC	CTGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
M_smagmati	AACTCGAC	CTGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
M_tuber_H3	AACTCGAC	CTGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
M_tuber_CD	AACTCGAC	CTGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Mycoplasma	AACTCGC	CATGAAGT	GGATCGCTAG	GAATCGG	ATCAGCA	GC	CGGGTGAAT
Myxococcus	AACTCGAC	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Neisseria	AACTCGA	CATGAAGT	GGATCGCTAG	GAATCGG	ATCAGCA	GC	CGGGTGAAT
Pasteurell	AACTCGAC	CATGAAGT	GGATCGCTAG	GAATCGG	ATCAGCA	GC	CGGGTGAAT
Pseudomona	AACTCGAC	CTGAAGT	GGATCGCTAG	GAATCGG	ATCAGCA	GC	CGGGTGAAT
Rhodococcu	AACTCGAC	CTGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Rickettsia	AACTCGA	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Styphimuri	AACTCGAC	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Staphyloco	AACTCGAC	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Streptococ	AACTCGC	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Savermartil	AACTCGAC	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Sclavulige	AACTCGAC	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Scoelicolo	AACTCGAC	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Sgriseus_s	AACTCGAC	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Slavandula	AACTCGAC	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Svenezuela	AACTCGAC	CATGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Synechocys	AACTCGC	CATGAAGT	GGATCGCTAG	GAATCGG	ATCAGCA	GC	CGGGTGAAT
Thermotoga	AACTCGC	CTGAAGT	GGATCGCTAG	GAATCGG	SATCAGCA	GC	CGGGTGAAT
Treponema	AACTCGAC	CATGAAGT	GGATCGCTAG	GAATCGG	ATCAGCA	GC	CGGGTGAAT

TESIS CON
FALLA DE ORIGEN

Alineamiento de rRNA 16S

Ureaplasma AACTCGAC CATGAAGT GGAATCGCTAGTAATCGG ATCAG GCGGTGAAT
 Vibrio_cho AACTCGAC CATGAAGT GGAATCGCTAGTAATCGG ATCAG GCGGTGAAT
 Xylella_fa AACTCGAC CATGAAGT GGAATCGCTAGTAATCGG ATCAG GCGGTGAAT
 Mus_muscul AACTCGAC CATGAAGT GGAATCGCTAGTAATCGG ATCAG GCGGTGAAT

1510 1520 1530 1540 1550 1560

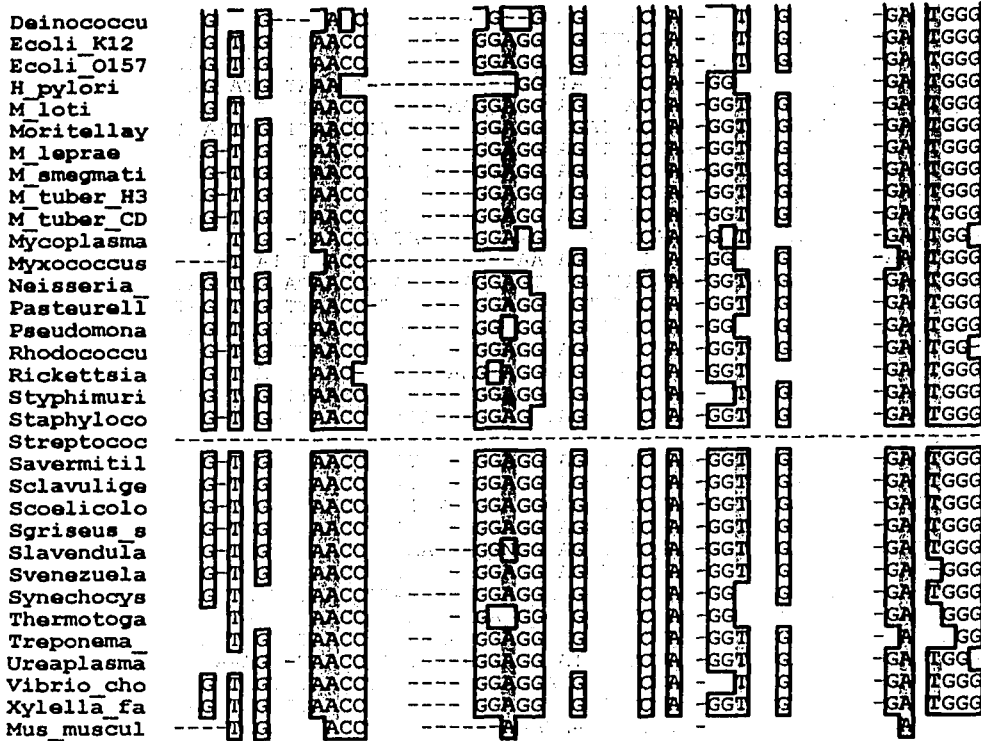
Actinobaci ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Amycolatop ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Bhaloduran ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Emegateriu ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Bsubtilis ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Bjaponicum ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Buchnerasp ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 C_jejuni ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Caulobacte ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Chlamydia ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Cpnemonia ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Deinococcu ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Ecoli_K12 ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Ecoli_O157 ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 H_pylori ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 M_loti ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Moritellay ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 M_leprae ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 M_smeigmati ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 M_tuber_H3 ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 M_tuber_CD ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Mycoplasma ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Myxococcus ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Neisseria ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Pasteurell ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Pseudomona ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Rhodococcu ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Rickettsia ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Styphimuri ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Staphyloco ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Streptococ ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Savermitil ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Sclavulige ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Scoelicolo ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Sgriseus_s ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Slavendula ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Svenzuela ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Synchocys ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Thermotoga ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Treponema ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Ureaplasma ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Vibrio_cho ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Xylella_fa ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG
 Mus_muscul ACGTTCGGGCGCTTGTACACACCGCCCGTCAO CATG AGT GGT ACC GAAG

1570 1580 1590 1600 1610 1620

Actinobaci AACC GAG GGG
 Amycolatop AACC GAG GGG
 Bhaloduran AACC GAG GGG
 Emegateriu AACC GAG GGG
 Bsubtilis AACC GAG GGG
 Bjaponicum AACC GAG GGG
 Buchnerasp AACC GAG GGG
 C_jejuni AACC GAG GGG
 Caulobacte AACC GAG GGG
 Chlamydia AACC GAG GGG
 Cpnemonia AACC GAG GGG

TESIS CON FALLA DE ORIGEN

Alineamiento de rRNA 16S



1630 1640 1650 1660 1670

Actinobaci	GAAGTCGTAACAAGGTA	CC	TA	E	A
Amycolatop	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Bhaloduran	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Bmegateriu	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Bsubtilis	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Bjaponicum	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Buchnerasp	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
C_jejuni	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Caulobacte	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Chlamydia	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Cpneumonia	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Deinococcu	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Ecoli_K12	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Ecoli_O157	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
H_pylori	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
M_loti	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Moritellay	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
M_leprae	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
M_smeagmati	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
M_tuber_H3	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
M_tuber_CD	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Mycoplasma	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Myxococcus	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Neisseria	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Pasteurell	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Pseudomona	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Rhodococcu	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Rickettsia	GAAGTCGTAACAAGGTA	CC	TA	GGAA	
Styphimuri	GAAGTCGTAACAAGGTA	CC	TA	GGAA	

TESIS CON
 FALLA DE ORIGEN

10 20 30 40 50 60

Mus_muscul M SMYGL AFVSATELLAVTVECLGFVWVRA TRTWV KGLKT G WGL FIG-HMLTV
 B_halodura -----M-TNTM TG KGN VLG-NTIEF
 Agrobacter -----MTEIGERT STDTTGAQ VSKLATRALALS
 Campylobac -----MSEC FF K YKNKASTLLTFLLK
 Bacillus_s -----MKQASAI Q KTYG LKNL HLE
 M_tubercul -----M
 Pseudomona -----M DRKLRGEEELISLHALYDGL
 Rhizobium -----M EQLLPTLFMRVVD
 Xylella_fa -----MKLTDLSNPAFLE
 A_tumefaci -----MIANSSTDVSVADQKFLNVAKSNQIDDAVFISR
 Pseudomona -----MTTETIQSNANLA LSHVPEHLVDFD

70 80 90 100 110 120

Mus_muscul GKN HLG PRLSQOYGDVLQIRIG-STVVVLSGLNTIKQALVRQDDDFKGRFDLYSFTL
 B_halodura GKD LQF TRCSQOYEIVRLRFEERDTELLNDPKHIQYVFMNKGGEFSKGYQDDPIMG
 Agrobacter IRN LKA EIFSE AVFTRLGG--VMRVHLAD VLIHEALVKNAALLGKGEDVRRALG
 Campylobac RRSWLDQYERSYKMQTGYVKM N--FDLVINDTKEVKRMVDEVREF KSAFLHELLS
 Bacillus_s KEQLSQD VRIADELG IFRDFE -GVSSVFVSGHNLVAEVCDEKRFDKNLGKGLQKVRE
 M_tubercul RY LGEALALALYRWRG LINAGVG--GHGYTYLLGAEANRFVFNADAFSWSQTFESLVI
 Pseudomona QVDGA R AHRAAEH VVVVTRYR---DARKVLNH GVRRDARQAELYAKRTG-S RAG
 Rhizobium HIE S T LALRANG IHNVRF RGHEG VVTGYDEAKAVLSDAAFR AGM AAF TDC
 Xylella_fa N Y LYETLRAQA FVSGI NALMTGRYSLVDSLHNRMNGKKYMESMRVRYGDSADM
 A_tumefaci LDSEGH TFAEWR KR FLRREDGIFLVLRADHIFLLGTD RTRQIETELMLNRGVKAGA
 Pseudomona DMYN SN SAGVQEA AVLQESNV DLVWTRCNGGHIA TRGQLIREAYEDYRHFSSC

130 140 150 160 170 180

Mus_muscul ITNGKSMTEFN DSG VWAARR DQNLKSEFSLASD TSASSCYEEHVSKEANVYLVSKL
 B_halodura LVFGNGLTAE--GSFLRRLRRLYASH KRIA-DYADTVGYCERHLNTNDNDTRDI
 Agrobacter ALGQGLTAD--GDHKKGRQSVAAAE RHEKLL-ELL VMIETARRTKQRWRSSTADI
 Campylobac LLGESIFPTN--GEVKKGRRLR R EEMTRLN----KFNEL EAVADMDREFSKY
 Bacillus_s FG-GDGLFSTWTHE NQKAKHLL LRSQAKK--GYHEMLD ATQLIQKSRNLN NE
 M_tubercul VDG TALVSD--GADHRRR R VVA GLRHHVQ----RYVAV VANIDTVIDESQ GQ
 Pseudomona IGEGLSHMLNLD DHTRLR LVRGAFTRQEE----RER ERITEALIDAMAGRE
 Rhizobium VILGS GILVSHGEHEHARLRITVA AEDRRYK----ILAQY EATAAQLETLAAQ
 Xylella_fa LFAAFRHFITIN ANTHIRLVGAGAFGRESE----SR LAIDEAHQLIDNEFOK
 A_tumefaci VFDFID NLF NBE KKRK R K R FRV E----R ER E KTECLD LDKVD
 Pseudomona FI REAEAYDFI MD E L R FR L AN V M VDKKENR ELAC LIE LR QG

190 200 210 220 230 240

Mus_muscul QKVMAEVGEFD YKLVV VANVICAICFGRYDHDDELL IVNL NEFGEVTVGSGY A
 B_halodura NDEMQLTMAIAFKTFLEDLDLHKGDTEA RSLDTVMPAFN--EQMNVFRHVLHLIGL
 Agrobacter DPEENQ----R VEDTIVEVNEI QYGDIDARVE TIDYK DQ TFALAMLGA E
 Campylobac NNAVIEVDEAN FFAADVIFRIMS KLDEEKGKKILNAFV FLEL VHTANRRMFRF K
 Bacillus_s EIDVADDNRLELDITGLCGFNFRNFYRDSQH FITMLRALKEAMNQSKRLGLQDKM
 M_tubercul RLDIYGE LRAVRRSTAESLFGRLAVHEDFLGEQLL LLDLTRR QVIRLQORVNS G
 Pseudomona QADLMADFAI LIAVIFELGI EAEREHARQF ERBAELIE EEAQALADAQ-----
 Rhizobium Q ADLRRRL F L ANVI ALMGVLYEDNAFFAGL DEVMRHEHG REARL A----
 Xylella_fa VVDVAEFAF F M IICKMDVDIGDAVTLGIAV KIAKVF D MEADELVHAS----
 A_tumefaci DFNFIEMYANQL ALDIAVIGL SEDT FFTRLVYKY RCL R RDEEFEEIEA----
 Pseudomona LCNFEEDYAE F IRIFMLIAGL EEDI HLKYLIDNR DFN PAEA-----

250 260 270 280 290 300

Mus_muscul DFI VIRYL N L DAFKQ NDKEY KKKIKERYR FEK R D D D EHC DRKL
 B_halodura KLV VRELR--EAVE DKKY DEERKK E-----DR DEY-----Y
 Agrobacter EL E GRRK R--AAVDYRA A A RRK E-----DR DMV EAK
 Campylobac LAYVLDCKR--AKAGDVR V D K RYDMADN--AEFE DAE EALVV----
 Bacillus_s NVK K L F KD-----EEN I DR RERKANEDE-----NIK D YAK
 M_tubercul RRAMA-----ARKR DDI DA L DNR R ER-----DDI D D R
 Pseudomona -----VDYR EAKRR E-----A D R K A A

TESIS CON FALLA DE ORIGEN

Alineamiento de citocromos P450

Rhizobium -----WEEI RAYI RGV RDKRQDE-----GNIITDIAAVD-----
 Xylella_fa -----TAYEET AQYFTKTEI RRTHE-----GTDITMEIRAE-----
 A_tumefaci -----AIEIDDYVRSVITADSGRRMR-----DDELSRYIKAVR-----
 Pseudomona -----KEALDYDIITIEQRROKE-----GIDATSIIVANGQVN-----

310 320 330 340 350 360

Mus_muscul DENANVQSDDKVITITLDLDFGAGEDVITATSESEIYVVTNRVQRKIQEELDTVIGRD
 B_halodura DEDDGSYMDRQI RDEIITLLEIAGFETTANTLSAFYLLSQEPPHVEEKLYQEVSVQLG-N
 Agrobacter D ETGRMMSDEEIIDNLLTEI IAGFETTALGLATFHLLSQNPFETERKAVEEIEAVTG-G
 Campylobac DADTNKRSEFEEILDQVAMLEIAGFETTASSLITLILYLSLYEKEQEKAAYEITQVLQ-G
 Bacillus_s DVTGETLDENIRYCIITELIAGETTSGILSEFAYICLLTHEEKLKKAQEEADRVLTD
 M_tubercul CSEGETTLDNEIRDSIVSLI IAGYETTSGALATAYALLTVGWTWESAASEVARVLG-G
 Pseudomona ---DESGCLSEAEIVSMAHLMMSGEETTMMNIGNALVILLVNFEE-----
 Rhizobium ---RGEAESEBAIGLAAGMLVAGFESTVACIEFEELAMLRFEEQ-----
 Xylella_fa ---EDGEKLDHDEIVSNVIMLLIAGYETTSMNIGNALLALHRFEE-----
 A_tumefaci ---EAGTLESIEEIMGLMLIILAGSDTTRTAMVMVTALALCNFA-----
 Pseudomona ---GR TISDEAKRMCGILLVGLDITVNVNELSEVEEFAKSEE-----

370 380 390 400 410 420

Mus_muscul RQRLSDRQLYEAETLETERHSSF--VFTIHSTRDITNGFYIKGCCVFNQW
 B_halodura R ATLEDMKLSYAEHVIKETLRVQ T--VWLISRR-AEKDVTLEDYHLSAGSEIISQW
 Agrobacter E VAAEHIANLAYRQVFSAMRLY --AVITRT-ALCDFRGEHDIAGVIVYIY
 Campylobac GVIEISHLRQFKYLTNIEKESLRLY --VGFFARE-AKGDTCVRDKIKKSGCVIAE
 Bacillus_s DT EYKIQQLKYTRMVLNEILRLY T--A AFSLYAKEDTVLGGEYISKQDVTLLI
 M_tubercul RV AADDLSALTYLNGVVEELRLYS --GVISARR-VLRDLLEDGHRFRACRLEIFSA
 Pseudomona ---QLALLRAQELLNAMEEIVRHS --VRASMLRFETVEDVEEDGVIAAGEYIISNL
 Rhizobium ---QRERLVGNLVDKAVEEELRHY GAGDGDIMRYRFDVYTAGVIAEKVIL
 Xylella_fa ---QLALLKSDLSIQAVEECLRYDGS---VOFTMRAAMDIEVEGEIIRGTVVEMLG
 A_tumefaci ---LSSLRINQSYAAAVVEEELREE ---VGFRLALKDIDIDGYVIEKSIILAVM
 Pseudomona ---HRQELIERERAAACEELERREEL---VADRILTDYEERGVIAKKEIDVILQW

430 440 450 460 470 480

Mus_muscul QVN DREIAG-DNEFRERELTSGTLDKRLEEKVYI EGLGKRKCTGETIGR:EVFEET
 B_halodura GMRNRYEN-DLTFIPERDND---NKSKYVVF EGG RVCHGERFELMBAIETI
 Agrobacter AVRHITAIQD-EERFEEERFE EK--VKARHRYAYM EGAG RVCHGNAFVMBEAVAI
 Campylobac LIRHEEET-NRGFNRRESEY-----KKDAKLEEGGERICHGEMEMBAIHI
 Bacillus_s KIRDNATVDAEDFRPERED ---LWFK EGNLRCHGAGVEMBAIHI
 M_tubercul VRLIEIT-E EFRERD NAADYRK AHEE EAGGRCHGAGVEMBAIHI
 Pseudomona TANDAERE-----DEPRLDTR-----NEDDLEKGEVHYCAGASIRIERRI
 Rhizobium ANFDRIE-----DEIFDGR-----DAKLAIEYGHYCHGAIIRIELKVE
 Xylella_fa ANRDAGS-----DEPGLDTR-----KGRIDPONGGIBHCHGARIIELECA
 A_tumefaci GLRDEK:Y-----DEIFDVR-----LNRHAEFGNGHRCHEGTERIELEEG
 Pseudomona LDELDERENA-----CEAHVDEFR-----CKVAVIRG BHCCHGIBHIREIIV

490 500 510 520 530

Mus_muscul AITGQIEFKVGEKVDQTYGHLKARCEHFVLMRISILIA-----
 B_halodura AITGREFRMEVDELTKVEIIRIKGVEKLRKR-----
 Agrobacter AVITKHELENRTMAAE LMRVIRERELKIRNKAV-----
 Campylobac ANIKYKLEEESEV DVGRLVRAANFHKFKRK-----
 Bacillus_s GIVKHEELINEHGYELKKEALKK DDFKIVK RKAAINV RKEADIKAE
 M_tubercul ARSRRMMLA RNRRAANFAIR QVEERKA A-----
 Pseudomona QRMRFDEIAYHAEVLLLELRALIVRGGCAANTASANIERA
 Rhizobium GIEFRFAIRIAYA EELKLRKEITGGEEFVL-----
 Xylella_fa GIEEERLARIADALN RNRNRGVNIIIDIAKN-----
 A_tumefaci RIRRA NAAVAD RAGGGRRAADDEKLFEL-----
 Pseudomona KEIIRIDELIAGVIRKIGIIVGVALVDAKAV-----

TESIS CON FALLA DE ORIGEN

10 20 30 40 50 60

Pseudomona -----MSVELHQQCVDIQRDIT S-----QCENR
 Pasteurell -----MK DLSSLWQECILQLQDQDITSL-----TDFSR
 Xylella_fa -----MESWSRCLERPEIEE -----EDVFR
 Bacillus_s -----MENILDLWNQALACTEKIKSK-----SEEFK
 Bacillus_h -----MENIHDLWERAIAKSMKKVSK-----SEEFK
 Mycobacter -----MTDD: GSGFTTVWNAVWSEINGD KVDDGT QORAL
 Nostoc_sp -----MMEM: IDNLWSQVLERLQIEISR-----TEFEI
 Campylobac -----MN: SQILENFKKEISE-----NEYENYL
 Agrobacter MQLNLMGAVADGDRA: KACDAACSEAAGDKSAMHHDALFERESARLKAQVG EYVAG
 Sinorhizob MRMNLATAFGG-----FQAGSNQSQAAAGEKHMHRHDALFERV SARTKAQVG DVEAG
 Mesorhizob -----MQSGIERELTGDLEFF: GTLIGANEMAVSSDAEQKFDRLKIQIKARLGAEVVSS

70 80 90 100 110 120

Pseudomona R: LCVEAEGD--ELRYVAENRFVLLVNEKVI SRLEI D BERGEGQL ALSLLIGSKR
 Pasteurell R: LQADFSVQN--TILYASNVFKQKQDESVAQLTKVAQELSGNAELVQVQKVG----
 Xylella_fa R: LQADQRGD--SVVEYAFN FIEIIVEERYLGRTRREIISYFSGIREVVLAIISR--- K
 Bacillus_s KSTKAHSLQG-DTILITAFNEFARDLESRYLHITADTIYELTG--EESIKFEVI--- Q
 Bacillus_h KQTKANSIED-STIITAFNEFARDLEKHYDELISEITDDLTG--VRYIKFEVI--- T
 Mycobacter NIVQ LTIIVE-GFALLSVESSVQNEIERHLRA ITDALSRRLLGHQICLVRIA --- A
 Nostoc_sp KTANAERLEN-NQVITITFN ERNQLQKYVISTIANVQSILGH VEYITVAK---GE
 Campylobac SNLKFNEKQSKDILVENAENELMAKETQTKYEKKAHFYEVQSGNKAIINIAQSS---AK
 Agrobacter ARKLLHTVSKS-VREIVVETTELKSTNNRYDILTSIVQSED DVLKVEILVRSASR V
 Sinorhizob GRKLLHSVSKS-VREIVVETTELKSTNNRYDILTSIVQSED: EILKVEILVRTATRGH
 Mesorhizob GRMKVAEASKG-IYRSVETAFLRSITNGHYDILISELKKQEDADLLKTEIVVRTATRQG

130 140 150 160 170 180

Pseudomona VA A VQ VSAA VVV REEL VTTA SVSSAAAM AGAA AVRTERNVQVEGALKH
 Pasteurell VK E K AQ SAL THHNKEENK QTVIR-----
 Xylella_fa TTEL V VDTTGRLSSTV FN-----
 Bacillus_s NQDVEDEM K QVKKAVKEDTS--DF QN-----
 Bacillus_h SOLDE FVEQELKK MKQ AQNGEM NN-----
 Mycobacter TDEADDTV SEN ATTS DTFDNDIEDDAAAARGDNQHSW SYFTER HNTDSATAG
 Nostoc_sp EFEEIGGGGEMEL TTNIINET NQNRQ N-----
 Campylobac QSNKSTKIDIAHKAQST-----
 Agrobacter R AQTEERAQ VQEVGAA RNKRFI SQWATA A-----AQ MAAQATLROGGSG LF
 Sinorhizob R TA EEVAAMAAEAAYV R---RFAA TVA-----IAANAVVAAA AR VEA LF
 Mesorhizob RSHAE ELA ARKTRRITLAAAGVY RVER-----V R GAVE EFRNVL

190 200 210 220 230 240

Pseudomona TSYLNRFIEEENEVQKSNQIAFAAAKQADN--LKSGVNPFLFYG GEGKTHILAVG
 Pasteurell --SYLN KIEEENEVQKSNQIAFAAAKQADN--GEKSNPFLFYG GEGKTHILAVG
 Xylella_fa --GNLDIEYEDNEYEQKSNQIAFAAAKQADN--GDRTPNPLFYG GEGKTHILAVG
 Bacillus_s --MLN KYEEDIEVIGKSNRBAFAAALVVEA --AKAVNPFLFYG GEGKTHILAVG
 Bacillus_h --MNDKYIEDIEVIGKSNRBAFAAALVVEA --AKAVNPFLFYG GEGKTHILAVG
 Mycobacter VTNLNRYIEDIEVIGKSNRBAFAAALVVEA --ARAVNPFLFYG GEGKTHILAVG
 Nostoc_sp --TELNAKVEEIEVIGANNRFAAALVVEA --GREENPFLFYG GEGKTHILAVG
 Campylobac --ILN SEIEEIEVIGSNKAYGACKVIAKDKLGGKIVNPIFYG GEGKTHILAVG
 Agrobacter GNLDIEEIEEIEVIGSNRVAIAAKIIEACAGAVRENDEIEV GEGKTHILAVG
 Sinorhizob GNLDIEEIEEIEVIGSNRVAIAAKIIEACAGAVRENDEIEV GEGKTHILAVG
 Mesorhizob GNLDIEEIEEIEVIGSNRVAIAAKIIEACAVRENDEIEV GEGKTHILAVG

250 260 270 280 290 300

Pseudomona NELLKKN NAKVYVLEEREVADMKVQVAINAINEKREYR QDNLDDDIQEFARKER
 Pasteurell NELLRNINARVYVLEEREVADMKVQVAINAINEKREYR QDNLDDDIQEFARKER
 Xylella_fa NVMRQVN YKQVYVLEEREFAMIRADDKMDIEKREYR QDNLDDDIQEFARKER
 Bacillus_s BYVTDHN NAKVYVLEEREFNEFINRDNKAVDERNKYR QDNLDDDIQEFARKER
 Bacillus_h BYVTDHN NAKVYVLEEREFNEFINRDNKAVDERNKYR QDNLDDDIQEFARKER
 Mycobacter NYAORLF GNRVYVLEEREFNEFINRDNRKYRERKREYR QDNLDDDIQEFARKER
 Nostoc sp BYR EIC NAKVYVLEEREFNDLITARDNDMDIEEERARNDI QDNLDDDIQEFARKER

TESIS CON FALLA DE ORIGEN

Alineamiento de dnaA

Campylobac NNAETEMG--KKVYVAGEEINDFTSNENKNSLDKRETEKVRNCDVLELIDDVCELEKIDKI
 Agrobacter NAAIDSRNRVYVYVEYEMRFAIATRDNDALTIKQDTRIDLLIIDDVCELEIGKM-I
 Sinorhizob LAALQSARA RVYVYVEYEMRFAIATRDNDALSIKESIRIDLLIIDDVCELEIGKS-I
 Mesorhizob AESLKONKSRVYVYVEYEMRFAIATRDNNALTIKQDIRIDLLIIDDVCELEIGKS-I

310 320 330 340 350 360

Pseudomona QEEFFHFNALIEECSQVVIHSDRYEKETESSEERLKERFGGLTVAVEFELETRVAIL
 Pasteurell QEEFFHFNALIEDTGRQIILLSDRYEKETEKLEERLKERFGGLTTAIVEFELETRVAIL
 Xylella_fa QEEFFHFNALIEDGKQIILLSDRYEREVNGLERLKSRLLAGLSVAIDFDETRIAIV
 Bacillus_s QEEFFHFNALIEEESKQIVISDRERKEITLEDRLRERFEVGLITDITFDETRIAII
 Bacillus_h QEEFFHFNALIEDNKQIVISDRERKEITLEDRLRERFEVGLITDITFDETRIAII
 Mycobacter QEEFFHFNALLNANKQIVISDRERKQATLEDRLRTRFEVGLITDVOFELETRIAII
 Nostoc_sp_ QEEFFHFNALIEEACKQIVISDRERKQATLEDRLRTRFEVGLITDVOFELETRIAII
 Campylobac QEEFFHFNALIEEACKQIVISDRERKQATLEDRLRTRFEVGLITDVOFELETRIAII
 Agrobacter QEEFFHFNALIEEACKQIVISDRERKQATLEDRLRTRFEVGLITDVOFELETRIAII
 Sinorhizob QEEFFHFNALIEEACKQIVISDRERKQATLEDRLRTRFEVGLITDVOFELETRIAII
 Mesorhizob QEEFFHFNALIEEACKQIVISDRERKQATLEDRLRTRFEVGLITDVOFELETRIAII

370 380 390 400 410 420

Pseudomona MKKAEQAK-----EPEEQAFFETADFRSNVRELEGALRVIAISHFMGRVLELELIFE
 Pasteurell IKKAEQKN-----EPEEQAFFETADFRSNVRELEGALRVIAISHFMGRVLELELIFE
 Xylella_fa LAKARERG-----ATPEEQAFFETADFRSNVRELEGALRVIAISHFMGRVLELELIFE
 Bacillus_s RKKAKAEG-----EPEEQAFFETADFRSNVRELEGALRVIAISHFMGRVLELELIFE
 Bacillus_h RKKAKAEN-----EPEEQAFFETADFRSNVRELEGALRVIAISHFMGRVLELELIFE
 Mycobacter RKKAKMER-----EPEEQAFFETADFRSNVRELEGALRVIAISHFMGRVLELELIFE
 Nostoc_sp_ RKKAEYEK-----EPEEQAFFETADFRSNVRELEGALRVIAISHFMGRVLELELIFE
 Campylobac RKKAEYND-----EPEEQAFFETADFRSNVRELEGALRVIAISHFMGRVLELELIFE
 Agrobacter NRRMGSARQDDPEEPEDEILLTFVAKFASERELEGAFNOLMFRFSE-NLSVDRVDE
 Sinorhizob KRRLEAARQDDAEEPEPEDEILLTFVAKFASERELEGAFNOLMFRFSE-NLSVDRVDE
 Mesorhizob KIRLAVARIDDAEEPEPEDEILLTFVAKFASERELEGAFNOLMFRFSE-NLSVDRVDE

430 440 450 460 470 480

Pseudomona LKDKLALD-KVYVENIGKRVVAYYKIKDIAEKRRRERVAPERGOAMVSEKEKLVNH
 Pasteurell TEKDKLALD-KVYVENIGKRVVAYYKIKDIAEKRRRERVAPERGOAMVSEKEKLVNR
 Xylella_fa LKDKLALD-KVYVENIGKRVVAYYKIKDIAEKRRRERVAPERGOAMVSEKEKLVNH
 Bacillus_s AKDKLALD-KVYVENIGKRVVAYYKIKDIAEKRRRERVAPERGOAMVSEKEKLVNH
 Bacillus_h AKDKLALD-KVYVENIGKRVVAYYKIKDIAEKRRRERVAPERGOAMVSEKEKLVNH
 Mycobacter VAKDKLALD-KVYVENIGKRVVAYYKIKDIAEKRRRERVAPERGOAMVSEKEKLVNH
 Nostoc_sp_ VAKDKLALD-KVYVENIGKRVVAYYKIKDIAEKRRRERVAPERGOAMVSEKEKLVNH
 Campylobac VAKDKLALD-KVYVENIGKRVVAYYKIKDIAEKRRRERVAPERGOAMVSEKEKLVNH
 Agrobacter LKDKLALD-KVYVENIGKRVVAYYKIKDIAEKRRRERVAPERGOAMVSEKEKLVNH
 Sinorhizob LKDKLALD-KVYVENIGKRVVAYYKIKDIAEKRRRERVAPERGOAMVSEKEKLVNH
 Mesorhizob LKDKLALD-KVYVENIGKRVVAYYKIKDIAEKRRRERVAPERGOAMVSEKEKLVNH

490 500 510 520 530

Pseudomona LPEEYGNFEGGRDQVYVYHARKKAEKREEDADYREDYKNEGLRQVIT
 Pasteurell LPEEYGNFEGGRDQVYVYHARKKAEKREEDADYREDYKNEGLRQVIT
 Xylella_fa LPEEYGNFEGGRDQVYVYHARKKAEKREEDADYREDYKNEGLRQVIT
 Bacillus_s LPEEYGNFEGGRDQVYVYHARKKAEKREEDADYREDYKNEGLRQVIT
 Bacillus_h LPEEYGNFEGGRDQVYVYHARKKAEKREEDADYREDYKNEGLRQVIT
 Mycobacter LPEEYGNFEGGRDQVYVYHARKKAEKREEDADYREDYKNEGLRQVIT
 Nostoc_sp_ LPEEYGNFEGGRDQVYVYHARKKAEKREEDADYREDYKNEGLRQVIT
 Campylobac LPEEYGNFEGGRDQVYVYHARKKAEKREEDADYREDYKNEGLRQVIT
 Agrobacter LPEEYGNFEGGRDQVYVYHARKKAEKREEDADYREDYKNEGLRQVIT
 Sinorhizob LPEEYGNFEGGRDQVYVYHARKKAEKREEDADYREDYKNEGLRQVIT
 Mesorhizob LPEEYGNFEGGRDQVYVYHARKKAEKREEDADYREDYKNEGLRQVIT

TESIS CON FALLA DE ORIGEN

Alineamiento de pola

```

      10          20          30          40          50          60
-----
Bacillus_h  -----MKKVLLVDGNEIAYRAFAEALFLLN-----NDK--GVYVAVYGEFAMMLKVV
Bacillus_s  -----MTRKKVLLVDGNEIAYRAFAEALFLLS-----NDK--GVYVAVYGEFAMMLKVV
Mycobacter  TTASASEDRAKTEMLLDGNSLAERAEVALEAEN---FKTRCGLTINAVYGFAMMLINI
Nostoc_sp_  ENSTVTTTTTRTEIIVLDGSELAERSYAEFAKGRDGGRLRTKIGIISVCGEGLKSLLEV
Pseudomona  -----MSQAIVLVDGSEIYVRAFAHALF-----LFTT-----STCKVCAVKGVLNMLLSL
Pasteurell  -----MAKIAENIVLVDGSEIYVRAFAHAE-----LTN-----SLNEVCAVYGVNMLKSL
Xylella_fa  -----MSKIVLVDGSEIYVRAFAHALF-----LSN-----AAGEVCALEGVNMLRRTT
Sinorhizob  -----MKNGDPEIIVLDGSEIFERAFHAE-----LNRK-----SDGLVNAVAVGFCNMLNKL
Agrobacter  SGALSMLMKKGDPEIIVLDGSEIFERAFHAE-----LNRK-----SDGLVNAVAVGFCNMLNKL
Campylobac  -----MKIVLVDTEGFFERIVYALKGFKNS-----CGDAEEMISGEFANEIYSL
  
```

```

      70          80          90          100         110         120
-----
Bacillus_h  CE-----DEKETHMVAFDAGKTERHDTYDAKSGRQKTEELSECEIIRELLDAF
Bacillus_s  CE-----DEKETHMVAFDAGKTERHGTEKEYKEGRQKTEELSECEMFIRELLDAY
Mycobacter  LR-----DEAPTHIAAFDVSRCDFERLQRYEYKANRSTSEDFACDIDITKEVLGAL
Nostoc_sp_  MA-----TQCEDEMTAFDGLLTERHEADDTYKADRGTEDDEVDBKNEHELLEGF
Pseudomona  RK-----QYDEBAVMFDAGKTERDELEAEYKANRSMEDDLRVQVEIHAVSRAL
Pasteurell  TS-----QVESHIAVMFDAGKTERDEIFEQYKSHRMEEELRSCILLNHISRAL
Xylella_fa  LK-----ERDVAEVIDAGKTERDALDSEYKSHRMEDDLRVQVEMCCIVGAL
Sinorhizob  LDARDTENVGVTEHDAVIDEDYSSKTERNGIYDQYKANRTAEDELIQFGIRHATRAF
Agrobacter  LKDARNTDVGVTEHDAVIDEDYSSKTERNELYDIYKANRSAEEDLVQFGITREATRAF
Campylobac  KN-----EYKSEVVFALDSKGMERSEIDNYKANRTAEELLEQLICIKWIEKM
  
```

```

      130         140         150         160         170         180
-----
Bacillus_h  NARVEVENEAEADDEIETLAKKQENFEKTEGDKDLQERDHIQVALRKKIQCID
Bacillus_s  QFRYEIEEAEADDEIETLAKKQEKDQFEKVEEGDKDLQLATDKWTVAIRKKQTDVE
Mycobacter  QITVAEIEEAEADDEIETLAKKQENEGYRIVVTEGDRDAIQLVEDQVLYRCKVSELY
Nostoc_sp_  NQIYVAEVEAEADDEIETLAKKQAEAGYKQMLNGDRDLEQLVD EKETVTLKRSTNSIT
Pseudomona  QIITCEVGEAEADDEIETLAKKQAAADRVISAGDKDMAQLVQGHVILVNAVITGRR
Pasteurell  QIITLAVGEAEADDEIETLAKKQFRACKKQMLTSGDKDMAQLVDENVILNINNTNLI
Xylella_fa  QIDITCEIDGAEADDEIETLAKKQAVADGIAVITSGDKDEAQLVRCQELVNTVGRR
Sinorhizob  NIQIEKEGEAEADDEIETLAKKQEEAGADVIMSEDKDLQLVTKVMYDEKDKQ
Agrobacter  NEQIEKEGEAEADDEIETLAKKQAEAGADVIMSEDKDLQLVIANVMYDAKDKQ
Campylobac  GFIITCEKGEAEADDEIETLAKKQCENRDIETLITLTKDQLQLIKDKGKTSIYBISKN
  
```

```

      190         200         210         220         230         240
-----
Bacillus_h  VYDEKAEERKCYLFEKYLIDKQDGDSDNIEPGVPGIGKTAIKLAKBYG
Bacillus_s  FYTEAEERKCYLFEKYLIDKQDGDSDNIEPGVPGVGEKTAIKLAKBYG
Mycobacter  RFEAEAEERKCYLFEKYLIDKQDGDSDNIEPGVPGVGEKTAIKLAKBYG
Nostoc_sp_  EVTEAEERKCYLFEKYLIDKQDGDSDNIEPGVPGVGEKTAIKLAKBYG
Pseudomona  LDVDEAEERKCYLFEKYLIDKQDGDSDNIEPGVPGVGEKTAIKLAKBYG
Pasteurell  LDREAEAEERKCYLFEKYLIDKQDGDSDNIEPGVPGVGEKTAIKLAKBYG
Xylella_fa  MDAAAEERKCYLFEKYLIDKQDGDSDNIEPGVPGVGEKTAIKLAKBYG
Sinorhizob  ITVDEAEERKCYLFEKYLIDKQDGDSDNIEPGVPGVGEKTAIKLAKBYG
Agrobacter  IGVDEAEERKCYLFEKYLIDKQDGDSDNIEPGVPGVGEKTAIKLAKBYG
Campylobac  DYDEACAEERKCYLFEKYLIDKQDGDSDNIEPGVPGVGEKTAIKLAKBYG
  
```

```

      250         260         270         280         290         300
-----
Bacillus_h  EKKELEIDEIKKIKESIEEFDKIVKKAIAIYEAQLDADIFEGEYDKE
Bacillus_s  EKKELEIDEIKKIKESIEEFDKIVKKAIAIYEAQLDADIFEGEYDKE
Mycobacter  ARVYVNDKDAIRKIKKQIRANLAVVNRREITDVRDIEEEDLRIGORDR
Nostoc_sp_  LDVYKANEKKAIAIKKIEKKAIAIAKAVDVEVNEEDCKVGEDEI
Pseudomona  LDKVETIRAKIKKIKKIEENREAFVYLYKIKDVEEDVEKLYGEVRE
Pasteurell  LOKVETIRAKIKKIKKIEAKDIEKIKKIKKIDVDDV EALFEGVANKDA
Xylella_fa  LDVYKANEKKAIAIKKIEENREAFVYLYKIKDVEEDVEKLYGEVRE
Sinorhizob  LDVYKANEKKAIAIKKIEENREAFVYLYKIKDVEEDVEKLYGEVRE
Agrobacter  LDVYKANEKKAIAIKKIEENREAFVYLYKIKDVEEDVEKLYGEVRE
Campylobac  LDVYKANEKKAIAIKKIEENREAFVYLYKIKDVEEDVEKLYGEVRE
  
```

TESIS CON
 FALLA DE ORIGEN

Alineamiento de pola

```

          310          320          330          340          350          360
Bacillus_h  KLEKLEDFNSI[K]KV DADGEGIE-----KEE-LQSLNYE-----TLREVKEAH-
Bacillus_s  AIFKDLSENTI[ER]IGEDSAAE-----QQSLEDINVK-----TVTDVTSDI-
Mycobacter  RLFDLLEERVLRDR[IF]DTLAAAG-----G EVDEGFVDRGGALA GTVQRWLAEH-
Nostoc_sp_   IIE[K]IEFTRE[SI]QINDLQLRFGGTVET. KTETSAIETT[F]FEDDDIAFYTAETQAIQ-
Pseudomona  ALYRETEFKNI[DD]LREAKEAGENG-----AETIQAEVD-----
Pasteurell  EYFARYEE[K]R[NE]VNGGESSVTNGSEQAVKIN--YQAT[ST]NERENTVSVQIDRSQ--
Xylella_fa  GFY[RY]CF[TA]KELDGGSVAVTVAEQE[GR]GRNAGFHVAE[DA]SVAMEINE[AL]SA[FG]E--
Sinorhizob  AFI[K]AMEET[TL]TRRVAAATD[TD]AEAE[PA]HV[VE]WGAQAHG[FD]LDVGEAGG[PT]S[Q]SSS
Agrobacter  AFI[K]AMEET[TL]TRRVAAATD[SD]ASAI[DA]ANV[VQ]WGADAHG[DL]DA[AA]AMAGDVS[AE]GN
Campylobac  EII[E]HYELNTI[K]KLRQ[N]DNKDKNLG

```

```

          370          380          390          400          410          420
Bacillus_h  -----LADKAAI[V]VEVLDENYHQ-----
Bacillus_s  -----IV[S]SAR[V]VECTGDNYHE-----
Mycobacter  -----AGDGRRAGLTV[ST]HL[H]-----
Nostoc_sp_   -----QNOAS[IQ]RINTEAK-----
Pseudomona  -----YDV[VL]DQAGEDA[V]KLEEA-----
Pasteurell  -----YQOLLELSE[KR]FDKLNQA-----
Xylella_fa  -----YETI[F]TVECL[OD]AVRLRVA-----
Sinorhizob  AT[VS]FLSSGQDADTTGAT[TGL]AEARAAYFGKA[FH]SGYRT[IR]DIDT[ER]ADAREA
Agrobacter  NR[AS]AVA[AK]RMIGEGST[AD]LTARQALFAATKIDT[TAY]TAT[RD]IAT[DL]AAAREA
Campylobac  -----FKATI[V]QDENK[FE]I[NT]LDKE

```

```

          430          440          450          460          470          480
Bacillus_h  G---DILGLAIAHNNGCYYT[TDE]I[RS]-----
Bacillus_s  E---ILGFSIVNETC[AYE]I[KDI]AV[SS]-----
Mycobacter  GGDATAMAVAAAD[EG]M[DT]A[LT]DDD-----
Nostoc_sp_   LIELVQILQQTIN[ET]VADTET[SA]LE RDAALVIGIGCCNGTETDANAYI LAHTKG--
Pseudomona  ELIAFDTE[TS]IDAQCA[V]GVSE[AV]KEG-----EAAYV LAHSYMG-----V
Pasteurell  KCIAIDTE[TS]IDYMV[LV]GVSE[AL]ENG-----EAAYL LRHDYLG-----A
Xylella_fa  GQFALDTE[TS]IDM[AV]LIG[SE]FAE[VG]-----CAAYL FGHDY G-----A
Sinorhizob  GLVGFDTQATS[DAM]R[LV]SE[LA]VADYANDSGSRIRAAYV LAHK[GV]S[DL]LGGG V
Agrobacter  GVVAFDTE[TS]IDM[GE]IV[SE]SLA[AD]NGKDSGTDIRAAYV LTHK[GV]S[GD]L[FD]SLA
Campylobac  SIIAFDTE[TS]IDT[KP]K[IV]SE[FC]M[NEN]-----EAFYV LTHNYLG-----

```

```

          490          500          510          520          530          540
Bacillus_h  -----EALIQ[AK]DERK[K]LEDA[KRA]VVA[EG]KN-----
Bacillus_s  -----EVEKEN[V]ENDEQK[VFD]H[KRA]VVA[ER]G-----
Mycobacter  -----AALAA[LD]AK KALHEA[KAA]VHD[AG]R[G]-----
Nostoc_sp_   -ENLNQDIV[AA]R[ILE]SADY[KFF]NA[KER]LVFLV[G]-----
Pseudomona  EQLDRAV[RA]K[ILE]DKKIVGCA[KY]DIN[VA]NAST-----
Pasteurell  EQVDFFVA[K]DK[V]HEN[EIK]K[NI]K[OL]I[EAR]G-----
Xylella_fa  VQLDRG[A]I[AI]I[LE]DAAVR[KV]G[KY]D[HV]R[RY]G-----
Sinorhizob  DRQV GRETRAK[EL]LED SVIKVGNL[KY]BYLV[KR]G-----TAMR[FED]V[LV]SYV
Agrobacter  EGQA FAEA[ER]K[ILE]DAV[KV]ACNL[KY]OYLI[KR]G-----V[VR]FDD[ML]SYV
Campylobac  ---VGEQI[AK]KATE[V]FNHEVIC[NI]K[DF]K[ION]NED[UN]K[AD]V[LA]W[K]

```

```

          550          560          570          580          590          600
Bacillus_h  D[AV]E[EV]DV[K]HRK[G]HE[DA]VY[SG]K[AK]RRI[DE]K[SE]G[AR]DAI[FA]K
Bacillus_s  D[GN]D[DV]VA[K]DY[GL]I[DE]VY[SG]K[AK]RAV[ED]V[LE]L[ER]K[AL]AI[Q]R
Mycobacter  R[GR]E[FE]D[ER]R[RE]LRAET[Q]Q[Q]L[L]LDDDDTDA[ET]T[IL]R[AV]D[FA]
Nostoc_sp_   NDI[EL]H[IR]V[GL]E[N]D[AV]---K[K]IAD[NA]VAVAN[CCL]VYAT[EV]
Pseudomona  D[NT]R[ND]E[AK]K[GL]E[ED]IACK[AK]L[FD]I[TE]SAG[K]E[AD]OV[TR]E
Pasteurell  D[IT]G[R]V[DE]LAK[YL]R[Q]E[ED]IAGK[KA]Q[L]E[NG]E[SA]E[ED]D[IT]M[K]D
Xylella_fa  E[NG]HAR[VD]E[ER]L[VE]E[AV]D[V]K[AK]I[FA]V[LE]E[AT]R[EL]D[IT]UR[Q]
Sinorhizob  D[AG]NG[VD]E[ER]L[GT]E[AV]K[DY]R[GR]S[LE]F[ED]D[K]ATA[V]E[ED]D[IA]L[RI]
Agrobacter  EAGK[E]D[ER]L[KE]H[KY]R[GG]K[IT]E[FD]E[ND]KATA[V]E[ED]OV[TR]E
Campylobac  N[L]R[V]N[DD]B[ER]LENYE[SE]V[KK]---SE[FA]V[ER]KACKY[ED]AYI[TR]EY

```

TESIS CON
 FALLA DE ORIGEN

Alineamiento de pola

```

        610          620          630          640          650          660
    Bacillus_h  EHEEKKENE-QLGFENELEFVSVILGRMETTGVKQDVECTEQGSEDIYARLEIEIET
    Bacillus_s  EKIVGCEENNID-QLLFEELVYELALILGEMESTGVKQDVEDRLKRMSEELGAKIKEYEEK
    Mycobacter  DADDAEFLARID-STALGGELELVDRVIAKMESAGIYDL.MYTELQSOEGGQIDRDAEFA
    Nostoc_sp_   AKRREELAKT.NLQKLLREVECPLESVLAIVEYTGVSINSAYLHELSCOLEIDLAKIQQC
    Pseudomona  QALWCKEAAI.SLAVLTDIEVPLV.VLAKIERNGALVDANLIDGIQSRELKPKMVALEEQ
    Pasteurell  QVLEWKAQAQ.EELVKYQIMELPLASVLSRIERHGVLHSDSALFSQSOIGVRLTALBQQ
    Xylella_fa  CVLAKIKLAE.SLFRVYREIELELVAVIERTANGVITDAEELYKQSOADLSRMVIAQOK
    Sinorhizob  HVLKIPLAAGK-LTRVYERLEKELISVLAGMEERGIIVDRQIILSRISGELAQGAAALEDE
    Agrobacter  MAIK.RVMSER-LTRVYERLEKELV.VLAKMEERGIIVDRQIILSRISGELACKAASFEE
    Campylobac  LYFLKNE--ET.LLELAKNCEPFDIKIIMMEENGIKLDTNALEILMKKFENELKNSSE
  
```

```

        670          680          690          700          710          720
    Bacillus_h  IHELAGVSNINSFKQLGAILFEKIKLLE--VIKKTKTGYS TSA DVLKLAGQ--HEIIDH
    Bacillus_s  IHEIAGEFNINSFKQLGVILFEKIKLLE--VVKKTKTGYS TSA DVLKLAGQ--HDIVDY
    Mycobacter  AYVIGKQINLGSFKQLGVILFEKIKLLE--KTKRKTGYS TSA DVLKLAGQ--FLQY
    Nostoc_sp_   ATELAGEFNLSFKQLGVILFEKIKLLE--VTKRKTGYS TSA DVLKLAGQ--TFVNA
    Pseudomona  AYDLAGQFNLSFKQLGAILYDKIKLLE--VLSKKTGYS TSA DVLKLAGQ--QDFELKV
    Pasteurell  AYELAGQFNLSFKQLGAILYDKIKLLE--VLKKTGYS TSA DVLKLAGQ--LAYERHALKI
    Xylella_fa  AMDLAGEFNLSFKQLGVILFEKIKLLE--VVVKTGYS TSA DVLKLAGQ--EIAVDQHALRV
    Sinorhizob  IYVLAGQFNLSFKQLGAILYDKIKLLE--GGSKKTGYS TSA DVLKLAGQ--DLAAGHDLRK
    Agrobacter  IYELAGEFNLSFKQLGAILYDKIKLLE--GGSKKTGYS TSA DVLKLAGQ--DLAEGAELRK
    Campylobac  IYVLCEDFNLSFKQLGAILYDKIKLLE--SGKKTGYS TSA DVLKLAGQ--NILLDK--HVIK
  
```

```

        730          740          750          760          770          780
    Bacillus_h  ILYRRLKIKLSTYDALLLKLVKKTGYSRVHTRENOALITGRLSSDENLONIEIRLEE
    Bacillus_s  ILYRRLKIKLSTYDALLLKLVKKTGYSRVHTRENOALITGRLSSDENLONIEIRLEE
    Mycobacter  LLAVRDLVRLKVTYDGLDVAADG--RIHTIFNQTIAATGRLSSDENLONIEIRLDA
    Nostoc_sp_   ILYRRLKIKLSTYDALLLKLVKKTGYSRVHTRENOALITGRLSSDENLONIEIRLTA
    Pseudomona  ILYRRLKIKLSTYDALLLKLVKKTGYSRVHTRENOALITGRLSSDENLONIEIRLTA
    Pasteurell  ILYRRLKIKLSTYDALLLKLVKKTGYSRVHTRENOALITGRLSSDENLONIEIRLNE
    Xylella_fa  ILYRRLKIKLSTYDALLLKLVKKTGYSRVHTRENOALITGRLSSDENLONIEIRLTD
    Sinorhizob  ILYRRLKIKLSTYDALLLKLVKKTGYSRVHTRENOALITGRLSSDENLONIEIRLGE
    Agrobacter  ILYRRLKIKLSTYDALLLKLVKKTGYSRVHTRENOALITGRLSSDENLONIEIRLTA
    Campylobac  ILYRRLKIKLSTYDALLLKLVKKTGYSRVHTRENOALITGRLSSDENLONIEIRLHAGQY
  
```

```

        790          800          810          820          830          840
    Bacillus_h  GRKIRKAFVNEKDFELIADYSQIELRVLAHLESEDEKILNAEAFREDNDVHTKAFADVEFV
    Bacillus_s  GRKIRKAFVNEKDFELIADYSQIELRVLAHLESEDEKILNAEAFINDMDIHTKAFADVEFV
    Mycobacter  GRRIRKAFVNGDGYAEIADYSQIELRVLAHLESEDEKILNAEANTGEDLSSEVSRAGFV
    Nostoc_sp_   GRRIRKAFVNEKDFELIADYSQIELRVLAHLESEDEKILNAEAYQNEIDIHTKAFADVEFV
    Pseudomona  GRRIRKAFVNGDGYAEIADYSQIELRVLAHLESEDEKILNAEAFRHDIQVRAFAEVEFV
    Pasteurell  GRRIRKAFVNGDGYAEIADYSQIELRVLAHLESEDEKILNAEAFRHDIQVRAFAEVEFV
    Xylella_fa  GRRIRKAFVNGDGYAEIADYSQIELRVLAHLESEDEKILNAEAFRHDIQVRAFAEVEFV
    Sinorhizob  GRKIRKAFVNGDGYAEIADYSQIELRVLAHLESEDEKILNAEAFADGVDIHTKAFADVEFV
    Agrobacter  GRKIRKAFVNGDGYAEIADYSQIELRVLAHLESEDEKILNAEAFENGDIHTKAFADVEFV
    Campylobac  AKDYKCEFAKDFEFLIADYSQIELRVLAHLESEDEKILNAEAFANDEDIHTKAFADVEFV
  
```

```

        850          860          870          880          890          900
    Bacillus_h  AKDEEENRRKQVNFQIVYGLDYGLGKILNARKEANIGIERYEA V KVKVEYED
    Bacillus_s  AKDEEENRRKQVNFQIVYGLDYGLGKILNARKEANIGIERYEA V KVKVEYED
    Mycobacter  IDEVQCEARRRKAQVNFQIVYGLDYGLGKILNARKEANIGIERYEA V KVKVEYED
    Nostoc_sp_   --EDVNEARRRKAQVNFQIVYGLDYGLGKILNARKEANIGIERYEA V KVKVEYED
    Pseudomona  IEDVQCEARRRKAQVNFQIVYGLDYGLGKILNARKEANIGIERYEA V KVKVEYED
    Pasteurell  IEDVQCEARRRKAQVNFQIVYGLDYGLGKILNARKEANIGIERYEA V KVKVEYED
    Xylella_fa  IDEVQCEARRRKAQVNFQIVYGLDYGLGKILNARKEANIGIERYEA V KVKVEYED
    Sinorhizob  IEDVQCEARRRKAQVNFQIVYGLDYGLGKILNARKEANIGIERYEA V KVKVEYED
    Agrobacter  IEDVQCEARRRKAQVNFQIVYGLDYGLGKILNARKEANIGIERYEA V KVKVEYED
    Campylobac  IEDVQCEARRRKAQVNFQIVYGLDYGLGKILNARKEANIGIERYEA V KVKVEYED
  
```

TESIS CON FALLA DE ORIGEN

Alineamiento de polA

```

          910          920          930          940          950          960
Bacillus_h VIDEAREGCVYVILKRRRVLFDLIGRNENCRSFAERIAMNIEIQGTAADIIRKAMVMA
Bacillus_s SVQEAQKQCYVITLHRRRVLELTSRNENIRSFAPERIAMNIEIQGSAADIIRKAMIDMA
Mycobacter VVERARKIGYTSVIGRRRVLELDSENRQREAAERAAALNAIQGSAADIIRKAMITVD
Nostoc_sp VKKCAISLIGYVETLIGRRRVFDFTNNSLRKIKAGLIRSAANAIQGSADIIIRKAMVTH
Pseudomona TRACAAEAGFVETLIGRRRLVLEIISKNEAMRKAERIAINAIQOGTAADIMKRAMVAVD
Pasteurell IRETAKAQGYVITLFDLRRRLVLEIDICGSSNAIRKKAERVAINAIQOGTAADITKRAMITLD
Xylella_fa TRQCARDFGYVETVEGRRRLVLSNGIASENQTCRAGAEERAAINAIQOGTAADIIRKRAMVAVD
Sinorhizob TKAFARENGYVETLIGRRRAHVEIDIRSN SMRAFNERASINAIQOGSAADIIRRAMVKE
Agrobacter TKAFARENGYVETLIGRRRAHVEIIRSN SIKAFNERAAINAIQOGSAADIIRRAMIIE
Campylobac VKNEAKQNGFIVLISGRKRVEFDENAKMQLA-MYEREISINSLQGSAADVIRKAMLEIN

```

```

          970          980          990          1000          1010          1020
Bacillus_h SRMDEED--LQRRLLQVHDELIFEVKEEVAQMKEIVTSVMEQAVE---LQVPLKVDV
Bacillus_s AKIKKEKQ--LKRLLQVHDELIFEAKKEIEILEKIVLVMEQALA---LQVPLKVDV
Mycobacter KALNEAQ--LARRMLQVHDELIFEIA CERERVEALVRDKMSGAY---LQVPLQVSV
Nostoc_sp EVLKK---YQRRLLQVHDELVFEVQEAELQCIKSVMEQAVK---LQVPLQVVDV
Pseudomona NMLQESG--LDRVILQVHDELVFEVREDIVEQVCECIRLMSGAAAT---LQVPLVVEA
Pasteurell REIAGQ---DIKMIMQVHDELVFEVRTDKIEHFRTIIEKTTMEQAAQ---LQVPLIVDV
Xylella_fa GVLGVGHRERFAMILQVHDELVFEADEGEVTLVHEVSARMAAAQ---LQVPLVVDV
Sinorhizob AIEAAK--LSARRMLQVHDELIFEVEDGEIERTI VVISVMEQAMALDQVPLKQVDA
Agrobacter AIEAAK--LSARRMLQVHDELIFEVEAEIEKTLVVISVMEQAM AISQVPLKQVDA
Campylobac KELNEDK---KLLIQVHDELIFEVKDDLCENFVKTRDIMEQIVK---LQVPLKQVTS

```

1030

```

.....|.....|
Bacillus_h ACCDIFVDAK-
Bacillus_s ASGHEAVDAK-
Mycobacter GYGRERVDAAAH
Nostoc_sp RACENAVETK-
Pseudomona CVGQNDVAHI-
Pasteurell GVCKNVAEHI-
Xylella_fa GVQCHVAEHI-
Sinorhizob RAAFNVAEHI-
Agrobacter RAADNVAEHI-
Campylobac SIAKNVAIDLK-

```

TESIS CON
FALLA DE ORIGEN

Alineamiento de recA

	10	20	30	40	50	60
Bacillus_s	-----MSDRKALDMLKQTEK	FGKGSIMKLG	EKTDRITSTV	EGSLALDI		
Bacillus_h	-----MSDRKALDMLKQTEK	FGKGSIMKLG	EDAEQRVSTI	EGALALDI		
Mycobacter	-----MTQKDRKALDMLKQTEK	YKGSVMRLG	DEARQISTVI	TGSIALDV		
Nostoc_sp	-----MAINTDTSCKKALFMDLQIERS	FGKGAIMRLG	DATRMVETIS	TGALILDL		
Sinorhizob	-----NDKSKALDAALQIERS	FGKGSIMKLG	AKDSVVEIETV	TGSLGLDI		
Agrobacter	MAQNSLRIVEDKSVKSKALDAALQIERS	FGKGSIMKLG	SNENVVEIETV	TGSLGLDI		
Mesorhizob	MAQNSLRIVEDKAVDKSKALDAALQIERS	FGKGSIMRLG	ANEQVVEIETV	TGSLGLDI		
Pasteurell	-----MATKEEKNKALAAALQIETK	FGKGSIMKLG	DTQALDVEAV	TGSLGLDV		
Pseudomona	-----MDENKRALAAALQIETK	FGKGVMRMG	DHERQAIAT	TGSLGLDI		
Xylella_fa	-----MDENKRALAAALQIETK	FGKGSVMRMG	DRVIEAVEAT	TGSLGLDI		
Campylobac	-----MDDNKRKALDAALKSLDKI	FGKGTILRLG	DKEVECLDSI	TGSLGLDI		

	70	80	90	100	110	120
Bacillus_s	ALGIGGVFRGRIIEIVYGPSSGKTTVALHA	IAEVQCR	TS	AFDAEHALDE	YVAKKLGV	
Bacillus_h	ALGVGGYFRGRVIEIVYGPSSGKTTVALHA	IAEVCRNGGQA	AFDAEHALDE	YVAKKLGV		
Mycobacter	ALGIGGFRGRVIEIVYGPSSGKTTVALHAVA	NPQAGGVVAF	DAEHALDE	YVAKKLGV		
Nostoc_sp	ALGIGGLFKRVIIEIVYGPSSGKTTVALHA	IAEVQKEGGIAAF	DAEHALDE	YVAKKLGV		
Sinorhizob	ALGIGGLEKGRVIEIVYGPSSGKTTVALDQI	IAEAPQKGGICCF	DAEHALDE	YVAKKLGV		
Agrobacter	ALGIGGLEKGRVIEIVYGPSSGKTTVALDT	IAEAPQKGGICAF	DAEHALDE	YVAKKLGV		
Mesorhizob	ALGVGGYFRGRIIEIVYGPSSGKTTVALHT	VAEAPQKGGICAF	DAEHALDE	YVAKKLGV		
Pasteurell	ALGIGGLENGRIVEIVYGPSSGKTTLISV	IAEAPQKGGTCAF	DAEHALDE	YVAKKLGV		
Pseudomona	ALGIGGLEKGRVIEIVYGPSSGKTTLISV	IAEAPQKGGTCAF	DAEHALDE	YVAKKLGV		
Xylella_fa	ALGIGGLEKGRVIEIVYGPSSGKTTLISDA	IAEAPQKGGTAF	DAEHALDE	YVAKKLGV		
Campylobac	ALGIGGLEKGRVIEIVYGPSSGKTTLISHI	IAEAPQKGGVCAF	DAEHALDE	YVAKKLGV		

	130	140	150	160	170	180
Bacillus_s	NIEELITLTOPDTEGCALEIADLVRSGAVD	IVVIDSVAAL	VERAEIEGDMGD	IVGLOAR		
Bacillus_h	NIDEELITLTOPDTEGCALEIADLVRSGA	IDVIDSVAAL	VERAEIEGEMGD	IVGLOAR		
Mycobacter	OIIDLITLTOPDTEGCALEIADLIRSGALD	IVVIDSVAAL	VERAEIEGEMGD	IVGLOAR		
Nostoc_sp	DIENLITLTOPDTEGCALEIADLVRSGA	IVVIDSVAAL	VERAEIEGDMGD	IVGLOAR		
Sinorhizob	DIENLITLTOPDTEGCALEIADLVRSGA	IDVIDSVAAL	VERAEIEGEMGD	IVGLOAR		
Agrobacter	DIENLITLTOPDTEGCALEIADLVRSGA	VDVIDSVAAL	VERAEIEGEMGD	IVGLOAR		
Mesorhizob	DIENLITLTOPDTEGCALEIADLVRSGA	IDVIDSVAAL	VERAEIEGEMGD	IVGLOAR		
Pasteurell	NVNEELITLTOPDTEGCALEIADLVRSGA	VDVIDSVAAL	VERAEIEGEMGD	IVGLOAR		
Pseudomona	NVDEELITLTOPDTEGCALEIADLVRSGA	VDVIDSVAAL	VERAEIEGEMGD	IVGLOAR		
Xylella_fa	NVDEELITLTOPDTEGCALEIADLVRSGE	IDVIDSVAAL	VERAEIEGEMGD	IVGLOAR		
Campylobac	NIDEELITLTOPDTEGCALEIADLVRSGA	VDVIDSVAAL	VERAEIEGDMGD	IVGLOAR		

	190	200	210	220	230	240
Bacillus_s	LVGQARKEKLNKKNKVFIFNQIRKIGVMFG	NPENLVGGFATKIFYEVRLEVRR				
Bacillus_h	LVGQARKEKLNKKNKVFIFNQIRKIGVMFG	NPENLVGGFATKIFYEVRLEVRR				
Mycobacter	LVGQARKEKLNKKNKVFIFNQIRKIGVMFG	NPENLVGGFATKIFYEVRLEVRR				
Nostoc_sp	LVGQARKEKLNKKNKVFIFNQIRKIGVMFG	NPENLVGGFATKIFYEVRLEVRR				
Sinorhizob	LVGQARKEKLNKKNKVFIFNQIRKIGVMFG	NPENLVGGFATKIFYEVRLEVRR				
Agrobacter	LVGQARKEKLNKKNKVFIFNQIRKIGVMFG	NPENLVGGFATKIFYEVRLEVRR				
Mesorhizob	LVGQARKEKLNKKNKVFIFNQIRKIGVMFG	NPENLVGGFATKIFYEVRLEVRR				
Pasteurell	LVGQARKEKLNKKNKVFIFNQIRKIGVMFG	NPENLVGGFATKIFYEVRLEVRR				
Pseudomona	LVGQARKEKLNKKNKVFIFNQIRKIGVMFG	NPENLVGGFATKIFYEVRLEVRR				
Xylella_fa	LVGQARKEKLNKKNKVFIFNQIRKIGVMFG	NPENLVGGFATKIFYEVRLEVRR				
Campylobac	LVGQARKEKLNKKNKVFIFNQIRKIGVMFG	NPENLVGGFATKIFYEVRLEVRR				

	250	260	270	280	290	300
Bacillus_s	AEVKKFNDVYGNKIKKWKVKNKVAPEKVE	EDYVYEGG	IKG	GEV	DEG	ELDIYK
Bacillus_h	AEVKKFNDVYGNKIKKWKVKNKVAPEKVE	EDYVYEGG	IKG	GEV	DEG	ELDIYK
Mycobacter	VEVKKFNDVYGNKIKKWKVKNKVAPEKVE	EDYVYEGG	IKG	GEV	DEG	ELDIYK
Nostoc_sp	IEVKKFNDVYGNKIKKWKVKNKVAPEKVE	EDYVYEGG	IKG	GEV	DEG	ELDIYK
Sinorhizob	IEVKKFNDVYGNKIKKWKVKNKVAPEKVE	EDYVYEGG	IKG	GEV	DEG	ELDIYK
Agrobacter	IEVKKFNDVYGNKIKKWKVKNKVAPEKVE	EDYVYEGG	IKG	GEV	DEG	ELDIYK
Mesorhizob	IEVKKFNDVYGNKIKKWKVKNKVAPEKVE	EDYVYEGG	IKG	GEV	DEG	ELDIYK

TESIS CON FALLA DE ORIGEN

```

Pasteurell  TGATKEGEVIGNEIRVKVVKNKVAPPERQVDFIYGGGIEKIGELIETGVKHKIVDKA
Pseudomona TGAVKEGEVVGEEIRVKVVKNKVSPPERQAEFIIYGGKIVRIGEIIDEGVLCIVKES
Xylella_fa  IGATKKEDLIGNIKIKVVKNKLAPPERKQVVTIILYGGEGISREGELIEMGVBAKIVKKA
Campylobac  VATIKKNEELIGNRIRKVKVVKNKVAPPERQAEFDVMEGEGLSREGELIDVGVKLDIVDKS
    
```

310 320 330 340 350 360

```

Bacillus_s  GSYSYSEERIQQGRENAKFLKENKDIMLITQETREHYGLDNNGVVQQQAETQEELE
Bacillus_h  GAVYSENDERLGGGRENAKFLKENEDAAEEIESFIREHYGLNGEIEVEA SEEEFEDL
Mycobacter  GAVFYEGEOLGQKKNARFLVENADVADEIEKKIKEKLGIG-AVVTDD: SNDGVLFA:
Nostoc_sp   GAVSYVNGDNIHQGRDNAIKYLEEKEEESQIKQQVREKLDKGAVVSANSVAKNDEDDE
Sinorhizob  GAVFSYNSORLGGGRENAKFLRENEQLLREIEPALRQNAGLIADRFLENG: ESDGDEA
Agrobacter  GAVFSYNSORLGGGRENAKFLRDNEQIANEIELALRQNAGLIADRFLQNGG: DAGEGDD
Mesorhizob  GAVFSYNSORLGGGRENAKFLRDNEQIANEIELALRQNAGLIAEKFLENGSGEGGDDGF
Pasteurell  GAVYVNGEKIGQKANA MKLEEHSEALADETKLRNELLAN: EKVLAADIAEKNESS
Pseudomona  GAVYSYGGKIGQKANA AKYLEDNEEIGSVLEKTIIRDQLLAKSG: VKADAEVADAEAD
Xylella_fa  GAVYSEGERIGQKDNAREYLRNENHIAAKDADLREKFE: TELSREEGDEDTLIEDTM-
Campylobac  GAVFSYKDKKIGQGRENKFLKENEEIADETKATQNSMGIEGMISGSEDEDEGEE----
    
```

```

Bacillus_s  ...
Bacillus_h  LDL
Mycobacter  VDF
Nostoc_sp   EDV
Sinorhizob  ADM
Agrobacter  GSD
Mesorhizob  EDE
Pasteurell  GLE
Pseudomona  ---
Xylella_fa  ---
Campylobac  ---
    
```

TESIS CON FALLA DE ORIGEN

Alineamiento de rpoA

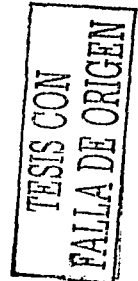
	10	20	30	40	50	60
Bacillus_s	-----M E E K E K I E V E R I D A K A K I K V	EPLE GYG ILG AL F I L L S S I G A V T A				
Bacillus_h	-----M E E K E K I E E E E E E D A K Y K K V	EPLE GYG ILG AL F I L L S S I G A V T A				
Mycobacter	-----M E E K E K I E E E V I D N R V I I I I	EPLE GFG ILG AL F I L L S S I G A V T A				
Agrobacter	K L K N E E E E E K N K E F F I R I K A T I I	EPLE GFG ILG AL F V L L S S L R G A V T A				
Sinorhizob	K L K N E E E E E K N K V E F A F I T R K A T I I	EPLE GFG ILG AL F V L L S S L R G A V T A				
Mesorhizob	K L K N E E E E E K N K I E F K K K I Y T I I	EPLE GFG ILG N A L F V L L S S L R G A V T A				
Pseudomona	M L V N E E F E E F R I D V A V V I T R A K I I	EPLE GFG ILG AL F I L L S S M G C A V E				
Pasteurell	M L V N E E F E E F I R L V D I E S I T H A K I I	EPLE GFG ILG AL F I L L S S M G C A V T E				
Xylella_fa	M L V N E E F E E F I R K E R I E R I E N R A K I I	EPH I G Y I L G N A L F V L L S S I G A V T E				
Nostoc_sp	-----M A L E E E C E V E N T E E R N Y K K I I	EPLE G G I V G A L F V L L S L R G A V T A				
Campylobac	-M R N I E L A Y E E E F I E N I D V A K I I	E P E I G Y G I L L S S I G A V T A				

	70	80	90	100	110	120
Bacillus_s	I Q I D G V L H E F S S I G V E D V I I L H I	L A E I Y E E K I L I A V T A G G G V T A				
Bacillus_h	V Q I D G V L H E F S S I G V E D V I I V L L	L A E I Y I E D K I L I I D G G G V T A				
Mycobacter	I R I D G V L H E F T V G V E D V E I I L L	L V E N T E V M I R K G G G V T A				
Agrobacter	V Q I D G V L H E F S S I G V E D V I I V L I	I A I M D I C K R M V R K G G G V T A				
Sinorhizob	V G I D G V L H E F S S I G V E D V I I V L I	I A I M D I C A K R M V R K G G G V T A				
Mesorhizob	V G I D G V L H E F S S I G V E D V I I V L I	I A I M E I T K R M V R K G G G V T A				
Pseudomona	V E I D G V L H E Y S I I G V C E D V E I D L	I A I E R D E M E E K G G G V T A				
Pasteurell	V E I D G V L H E Y S K I G V C E D V E V L L	L A V V N K L E V N K G G G V T A				
Xylella_fa	V E I D G V L H E Y T V I G L C E D V E V L I	V A I I L E E I R K G G G V T A				
Nostoc_sp	V R I E V H E F A T V G V E D V E I I M Y V	I L D A E T P E K G G G I T A				
Campylobac	L I E D V I E E F I S M G M E D V I I L I D	L R E K I N K I E E E K K E Y K D I N N				

	130	140	150	160	170	180
Bacillus_s	D V E I N E E E E E E N E F R R R A	P G G Y P A D M R D E I I I Y S				
Bacillus_h	D V D V N E E E E E E N E K R R G R A	P G G Y P A E N K D D I G I I Y S				
Mycobacter	V V I I F E E E E E E N E C K E R R V E	P G G Y P A E N R E E E I G I I V E Y				
Agrobacter	D V E I N E E E E E E N E E E R I M E E	P N G G Y P A E R N R E E E E I G I I V				
Sinorhizob	D V E I N E E E E E E N E E E R I M E E	P N G G Y P A E R N R E E E E I G I I V				
Mesorhizob	D V E I N E E E E E E N E E E R I M E E	P N G G Y P A E R N R E E E E I G I I V				
Pseudomona	D V E I N E E E E E E N E D N E M K K K	P G G Y P A D R E E E R I G L L L E A				
Pasteurell	D V E I N E E E E E E N E D N E M R P V	P G G Y P A E R R E E E R I G L L V A				
Xylella_fa	D V E I N E E E E E E N E C K I I N R K	P G G Y P A R R K E E E R I G L L L E A				
Nostoc_sp	E E E E E E E E E E E E E E E E E E	E E E E E E E E E E E E E E E E E				
Campylobac	E E E E E E E E E E E E E E E E E E	E E E E E E E E E E E E E E E E E				

	190	200	210	220	230	240
Bacillus_s	P V R R V E N P A P A D D D D H H R A	D D D D D G G G G E E E V E A A A				
Bacillus_h	P V R R V E N P A P A D D D D H H R A	D D D D D G G G G E E E V E A A A				
Mycobacter	P V R R V E N P A P A D D D D H H R A	D D D D D G G G G E E E V E A A A				
Agrobacter	P V R R V E N P A P A D D D D H H R A	D D D D D G G G G E E E V E A A A				
Sinorhizob	P V R R V E N P A P A D D D D H H R A	D D D D D G G G G E E E V E A A A				
Mesorhizob	P V R R V E N P A P A D D D D H H R A	D D D D D G G G G E E E V E A A A				
Pseudomona	P V R R V E N P A P A D D D D H H R A	D D D D D G G G G E E E V E A A A				
Pasteurell	P V R R V E N P A P A D D D D H H R A	D D D D D G G G G E E E V E A A A				
Xylella_fa	P V R R V E N P A P A D D D D H H R A	D D D D D G G G G E E E V E A A A				
Nostoc_sp	P V R R V E N P A P A D D D D H H R A	D D D D D G G G G E E E V E A A A				
Campylobac	P V R R V E N P A P A D D D D H H R A	D D D D D G G G G E E E V E A A A				

	250	260	270	280	290	300
Bacillus_s	K E E K E E E E E E E E E E E E E	E E E E E E E E E E E E E E E E E				
Bacillus_h	K E E K E E E E E E E E E E E E E	E E E E E E E E E E E E E E E E E				
Mycobacter	K E E K E E E E E E E E E E E E E	E E E E E E E E E E E E E E E E E				
Agrobacter	K E E K E E E E E E E E E E E E E	E E E E E E E E E E E E E E E E E				
Sinorhizob	K E E K E E E E E E E E E E E E E	E E E E E E E E E E E E E E E E E				
Mesorhizob	K E E K E E E E E E E E E E E E E	E E E E E E E E E E E E E E E E E				
Pseudomona	K E E K E E E E E E E E E E E E E	E E E E E E E E E E E E E E E E E				



Alineamiento de rpoA

```

Pasteurell  Q- ENKFF--EK EED  Y R MDDLEEVV  NCIR  E  IGDIV  KBE  EELK  N
Xylella_fa  R- ENK  --A  H R  EDDLEEVV  NCFK  E  IGDLI  KE  EELK  N
Nostoc_sp_  -----E N  C  D  A  E  EEL  LEV  A  NCF  RA  V  V  LLL  D  E  DLE  K  N
Campylobac  -WIKN  A  NE  E  M  F  E  N  C  L  E  S  C  L  E  K  V  I  G  E  L  A  E  E  E  R  L  K  N

```

310

320

330

340

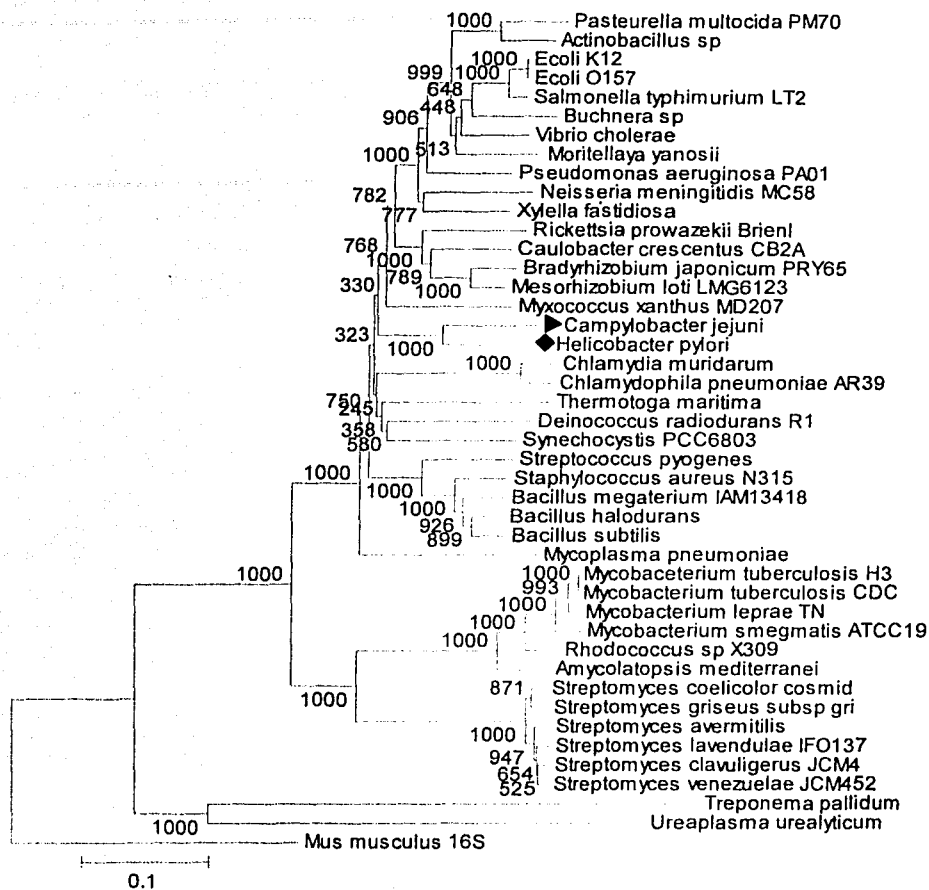
```

.....
Bacillus_s  G  P  S  E  V  K  E  E  E  L  S  E  R  K  D  D  -----
Bacillus_h  G  P  S  E  V  K  E  E  E  L  S  E  R  K  E  E  -----
Mycobacter  G  G  L  E  V  I  K  L  E  E  R  G  L  E  K  D  L  F  D  E  V  A  N  Y  D  A  T  T  A  K
Agrobacter  G  G  R  L  N  E  I  E  L  E  E  M  G  L  E  M  E  V  A  E  E  N  E  D  L  A  K  R  Y  E  D  Y
Sinorhizob  G  G  R  L  N  E  I  E  L  E  E  M  G  L  E  M  E  V  D  E  E  N  E  D  L  A  K  R  Y  E  D  Y
Mesorhizob  G  G  R  L  N  E  I  E  L  E  E  M  G  L  E  M  E  V  D  E  E  N  E  D  L  A  K  R  Y  E  D  Y
Pseudomona  G  G  G  L  E  I  L  L  E  E  R  G  L  E  M  R  L  D  N  A  A  E  K  K  D  D  K  A  T  A  ---
Pasteurell  G  G  G  L  E  I  E  L  E  E  R  G  L  E  M  R  L  E  N  A  A  E  T  A  E  D  -----
Xylella_fa  G  G  G  L  E  I  E  L  E  E  R  G  L  E  M  V  K  L  E  N  A  C  V  N  Y  E  M  I  G  -----
Nostoc_sp_  G  G  G  L  E  V  E  V  E  L  R  E  F  I  C  I  E  M  E  R  L  K  N  -----
Campylobac  G  G  G  L  E  I  E  L  E  E  R  G  L  E  M  K  L  D  N  K  E  L  K  N  K  I  A  E  L  K  A  Q  N

```

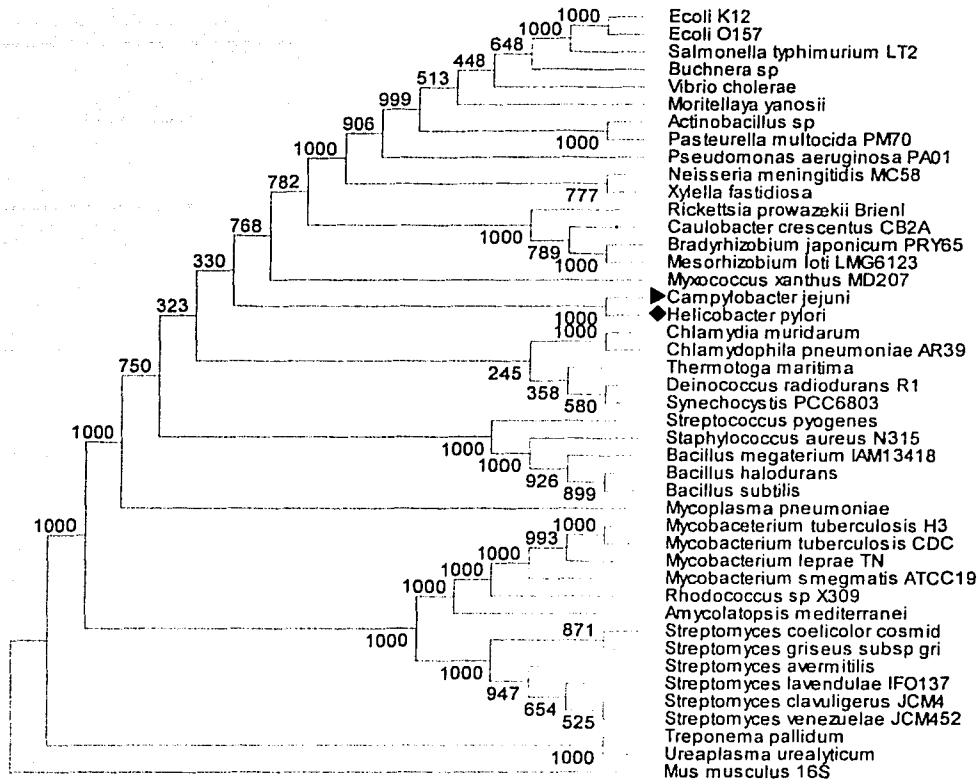
TESIS CON
FALLA DE ORIGEN

APÉNDICE 2



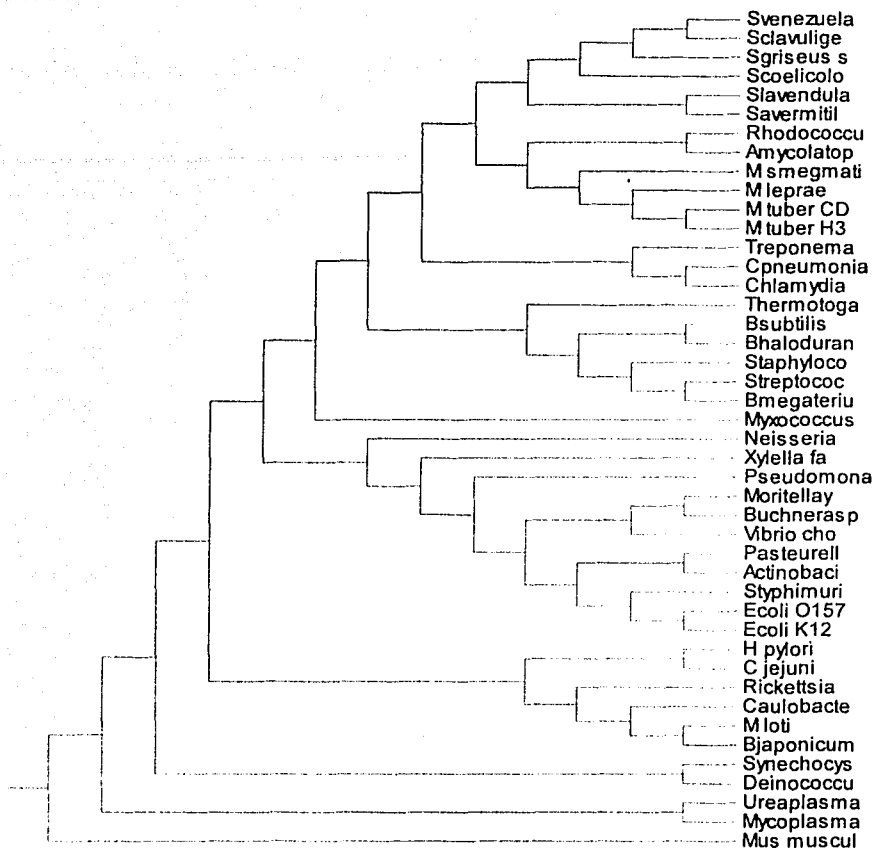
Árbol 1) Utilizando RNAr 16s con el método de neighbor joining. Los valores numéricos se refieren a los valores de bootstrap con 1000 iteraciones, los valores son directamente proporcionales al grado de confianza para ese nodo. El alineamiento, el árbol y el bootstrap fueron realizados con ClustalX 1.81. La escala representa el porcentaje de divergencia entre los OTUs.

TESIS CON
FALLA DE ORIGEN

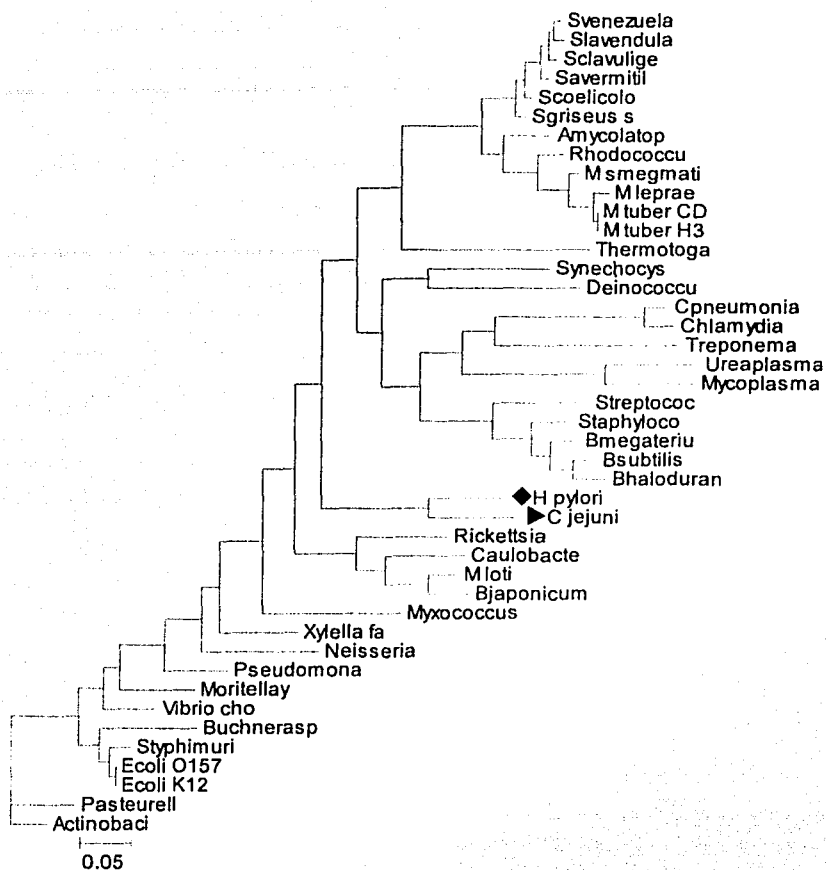


Árbol 2) Topología del árbol de rRNA 16s con el método de neighbor joining. Los valores numéricos se refieren a los valores de bootstrap con 1000 iteraciones, los valores son directamente proporcionales al grado de confianza para ese nodo. El alineamiento, el árbol y el bootstrap fueron realizados con ClustalX 1.81.

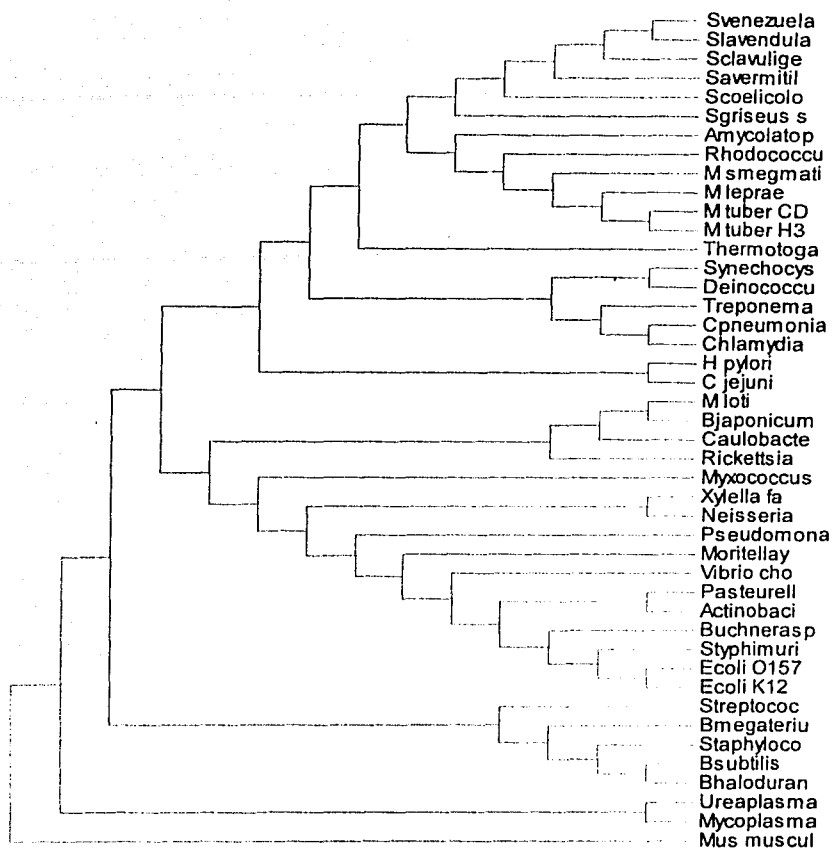
TESIS CON
FALLA DE ORIGEN



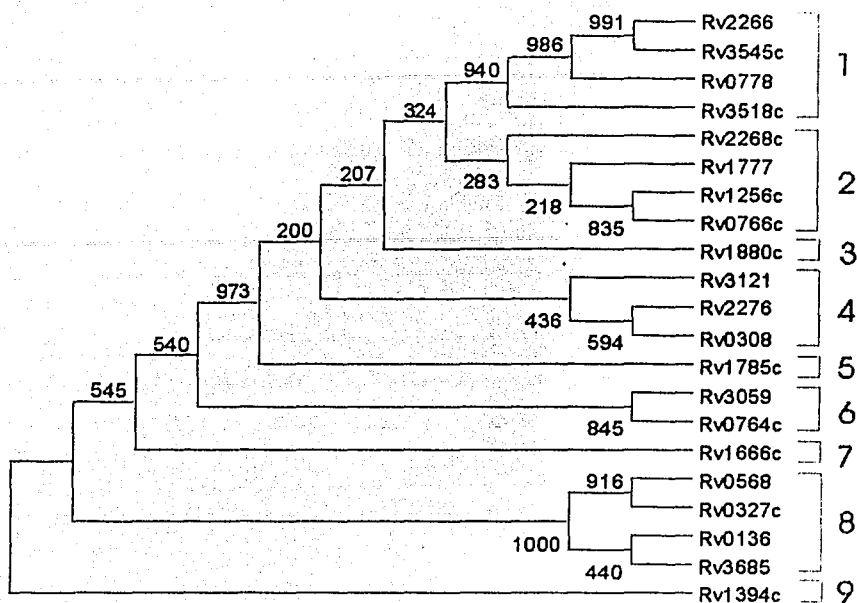
Árbol 3) Utilizando RNAr 16s con el método de máxima parsimonia. El alineamiento se realizó con ClustalX 1.81, el algoritmo se aplicó con el programa dnapsars del paquete Phylip.



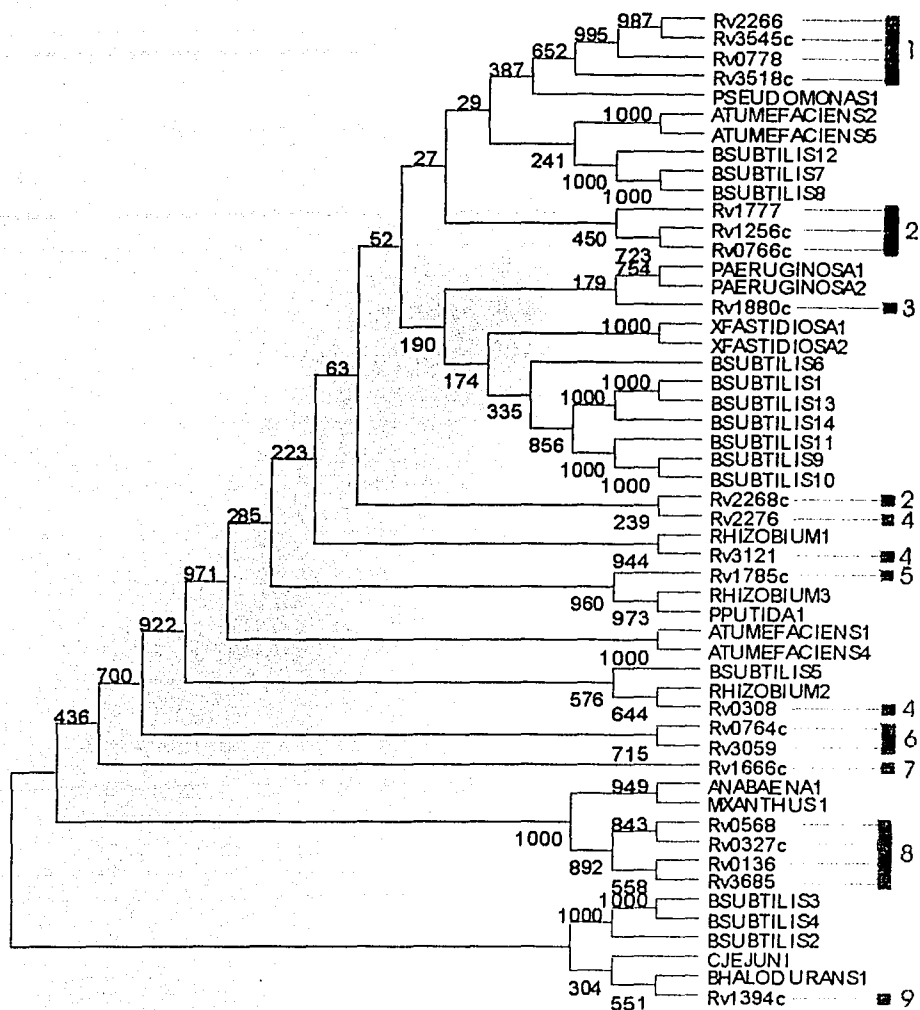
Árbol 4) Utilizando RNAr 16s con el método de máxima verosimilitud. El alineamiento se realizó con ClustalX 1.81, el algoritmo se aplicó con el programa DnaML del paquete Phylip. La escala representa el porcentaje de divergencia entre los OTUs



Árbol 5) Topología del árbol de RNAr 16s con el método de máxima verosimilitud. El alineamiento se realizó con ClustalX 1.81, el algoritmo se aplicó con el programa DnaML del paquete Phylip.

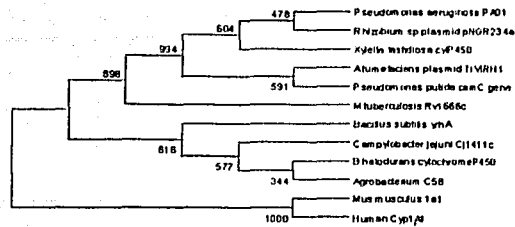


Árbol 6) Citocromos P450 contenidos en el genoma de *Mycobacterium tuberculosis*. Los números en los nodos representan los valores de Bootstrap (confianza estadística) para cada uno. El árbol fue realizado como se describe en Métodos para el método de NJ. La escala representa el porcentaje de divergencia entre los elementos. Este árbol no tiene raíz.

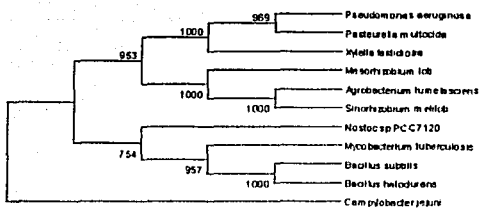


Árbol 7. Topología del árbol filogenético construido a partir de citocromos P450 de *C. jejuni*, *R. melliloti*, *B. subtilis*, *B. halodurans*, *M. xanthus*, *P. putida*, *X. fastidiosa*, *A. tumefaciens*, *P. aeruginosa*, *Anabaena* sp., y los genes de *M. tuberculosis* identificados con el nombre del ORF, que corresponden a los nombres usados en el árbol 6. El árbol y el análisis de bootstrap se realizó como se describe en Métodos para NJ. Este árbol no tiene raíz.

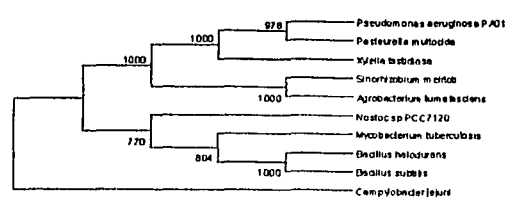
Citocromos P450



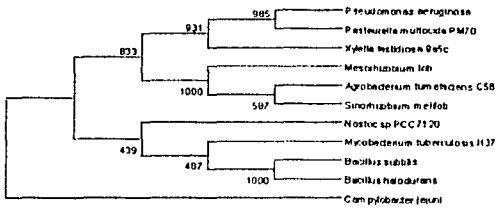
dnaA



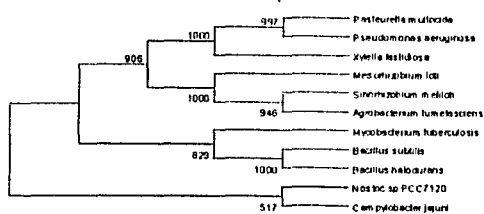
polA



recA

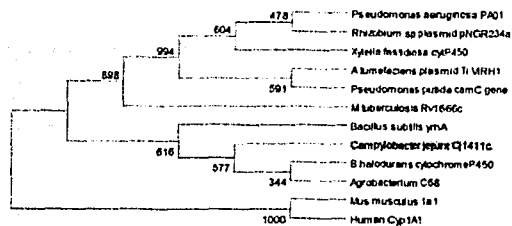


rpoA

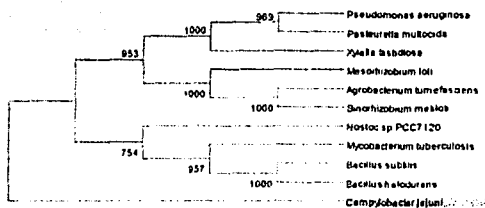


Árboles 8, 9, 10, 11 y 12) Topología de los árboles contruidos a partir de genes conservados (*dnaA*, *polA*, *recA* y *rpoA*) y del árbol construido con un citocromo P450 de dichas especies. Los valores numéricos se refieren a los valores de bootstrap con 1000 iteraciones, los valores son directamente proporcionales al grado de confianza para ese nodo. El alineamiento, el árbol y el bootstrap fueron realizados con ClustalX 1.81, inferidos con el algoritmo de NJ. Estos árboles no tienen raíz.

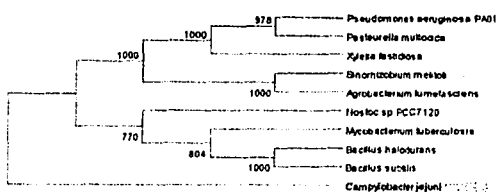
Citocromos P450



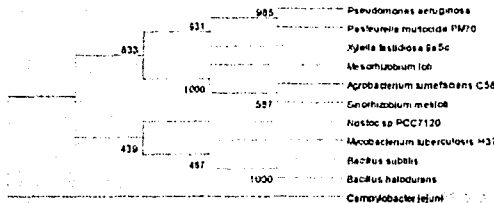
dnaA



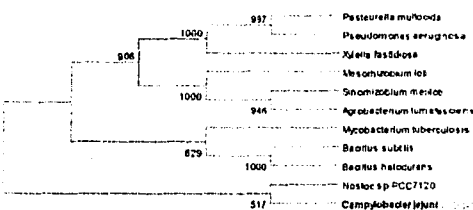
polA



recA



rpoA



Árboles 13, 14, 15, 16 y 17) Topología de los árboles contruidos a partir de genes conservados (dnaA, polA, recA y rpoA) y del árbol construido con un citocromo P450 de dichas especies. El alineamiento fue realizado con ClustalX 1.81, los árboles fueron inferidos con el algoritmo de MP como se describe en métodos. Estos árboles no tienen raíz.

APÉNDICE 3

CJI40

Número de acceso	Organismo	S	Valor E
gi 23130735 ref INZ_AAAX01000001.1	<i>Prochlorococcus marinus</i>	39	0.0821
gi 22987037 ref INZ_AAAC01000276.1	<i>Burkholderia fungorum</i>	31	0.302
gi 23137660 ref INZ_AABE01000066.1	<i>Cytophaga hutchinsonii</i>	35	1.01
gi 3721550 dbj AB010958.1	<i>Clostridium thermocellum</i>	35	1.01
gi 23008679 ref INZ_AAAP01002286.1	<i>Magnetospirillum magnetotacticum</i>	34	2.01
gi 20516987 gb AE013149.1	<i>Thermoanaerobacter tenqcongensis</i>	34	2.71
gi 23027361 ref INZ_AAAT01000005.1	<i>Microbulbifer degradans</i>	33	5.11
gi 4980788 gb AE001711.1 AE001711	<i>Thermotoga maritima</i>	33	5.11
gi 22775678 dbj AP004593.1	<i>Oceanobacillus ihoyensis</i>	32	7.01
gi 10567794 gb AF297457.1 AF297457	<i>Pseudomonas fluorescens</i>	29	7.12

CJI41

gi 23128923 ref INZ_AABC01000189.1	<i>Nostoc punctiforme</i>	128	2e-443
gi 15156656 gb AE008080.1 AE008080	<i>Agrobacterium tumefaciens</i>	111	2e-352
gi 10173176 dbj AP001509.1	<i>Bacillus halodurans</i>	106	6e-352
gi 22960729 ref INZ_AAAF01000001.1	<i>Rhodospseudomonas palustres</i>	73	9e-323
gi 24575076 gb AY048670.1	<i>Streptomyces globisporus</i>	89	1e-303
gi 15072928 emb AL591782.1 SME591782	<i>Sinorhizobium meliloti</i>	65	8e-284
gi 14024756 dbj AP003005.2	<i>Mesorhizobium loti</i>	66	8e-263
gi 13424631 gb AE005963.1	<i>Caulobacter crescentus</i>	84	1e-242
gi 23009801 ref INZ_AAAP01002719.1	<i>Magnetospirillum magnetotacticum</i>	74	2e-232
gi 1934604 gb U93874.1 BSU93874	<i>Bacillus subtilis</i>	72	6e-233

CJI42

gi 23464338 gb AE014590.1	<i>Brucilla suis</i>	42	8e-082
gi 23009495 ref INZ_AAAP01002606.1	<i>Magnetospirillum magnetotacticum</i>	40	4e-073
gi 17984479 gb AE009671.1 AE009671	<i>Brucilla melitensis</i>	39	7e-072
gi 15979410 emb AJ414148.1	<i>Yersinia pestis CO92</i>	32	0.0632
gi 23020176 ref INZ_AAJO1000003.1	<i>Clostridium thermocellum</i>	36	2.01
gi 10863206 gb AF260873.1 AF260873	<i>Enterococcus faecalis</i>	28	4.02
gi 21666291 gb AF389515.1	<i>Streptococcus intermedius</i>	25	4.83
gi 22533669 gb AE014221.1	<i>Streptococcus agalactiae</i>	34	7.01
gi 23129393 ref INZ_AABC01000195.1	<i>Nostoc punctiforme</i>	34	9.61

CJI43

gi 4887089 gb AF127177.1 AF127177	<i>Escherichia coli</i>	136	6e-717
gi 7378778 emb AL162752.2 HMA122491	<i>Neisseria meningitidis</i>	85	7e-657
gi 22979608 ref INZ_AAJO1000356.1	<i>Ralstonia metallidurans</i>	125	2e-625
gi 6424984 gb AF170495.1 AF170495	<i>Pasteurella haemolytica</i>	75	1e-586
gi 7682447 gb AF169324.1 AF169324	<i>Pasteurella multocida</i>	93	3e-566
gi 20502702 gb AF375657.1	<i>Aeromonas hydrophila</i>	122	8e-483
gi 23132699 ref INZ_AABD01000001.1	<i>Synechococcus sp.</i>	139	5e-452
gi 12750966 gb AF228583.1 AF228583	<i>Burkholderia pseudomallei</i>	94	2e-383
gi 13446679 gb AF285636.1	<i>Burkholderia mallei</i>	93	2e-373
gi 1235585 emb X64131.1 RMP1X23	<i>Sinorhizobium meliloti</i>	85	6e-232
gi 23011016 ref INZ_AAAP01003137.1	<i>Magnetospirillum magnetotacticum</i>	80	1e-193

CJI44

gi 7413419 gb AE002367.2 AE002367	<i>Neisseria meningitidis</i>	171	e-1256
gi 7682447 gb AF169324.1 AF169324	<i>Pasteurella multocida</i>	191	e-1146
gi 397404 emb X74567.1 ECKPCLST	<i>Escherichia coli</i>	166	e-1097
gi 6424984 gb AF170495.1 AF170495	<i>Pasteurella haemolytica</i>	173	e-1016
gi 22979688 ref INZ_AAJO1000356.1	<i>Ralstonia metallidurans</i>	120	4e-987
gi 20502702 gb AF375657.1	<i>Aeromonas hydrophila KpsM</i>	135	4e-846
gi 15140405 emb AL603644.1 RME603644	<i>Rhizobium meliloti</i>	146	2e-744
gi 12750966 gb AF228583.1 AF228583	<i>Burkholderia pseudomallei</i>	106	2e-625
gi 9931334 gb AF212041.1 AF212041	<i>Zymomonas mobilis</i>	190	3e-553
gi 23008444 ref INZ_AAAP01002190.1	<i>Magnetospirillum magnetotacticum</i>	143	2e-362

CJI45

gi 23058186 ref INZ_AABA01000065.1	<i>Pseudomonas fluorescens</i>	93	3e-436
gi 19714856 gb AE010629.1	<i>Fusobacterium nucleatum</i>	90	1e-161
gi 45264 emb X55652.1 PAAACC3	<i>Pseudomonas aeruginosa</i>	66	1e-152

TESIS CON
FALLA DE ORIGEN

gi 23132699 ref NZ_AABD01000001.1	Synechococcus sp.	60	1e-154
gi 216997791 gb AF4C2777.2	Streptococcus pneumoniae	82	4e-141
gi 23130735 ref NZ_AAAX01000001.1	Prochlorococcus marinus	56	1e-133
gi 6460468 gb AE001862.1 AE001862	Deinococcus radiodurans	58	3e-132
gi 15022925 gb AE007523.1 AE007523	Clostridium acetobutylicum	62	7e-132
gi 14026998 dbj AP003012.2	Mesorhizobium loti	54	8e-133
gi 23000130 ref NZ_AAAN01000184.1	Magnetococcus sp.	51	9e-132

Cj1416

gi 23058186 ref NZ_AABA01000065.1	Pseudomonas fluorescens	78	8e-385
gi 19714856 gb AE010629.1	Fusobacterium nucleatum	90	5e-171
gi 15458769 gb AE008487.1 AE008487	Streptococcus pneumoniae	82	2e-141
gi 18146725 dbj AP003187.2	Clostridium perfringens	73	6e-121
gi 2983301 gb AE000704.1 AE000704	Aquifex aeolicus	70	8e-111
gi 23103862 ref NZ_AAAD01000082.1	Azotobacter vinelandii	69	1e-101
gi 13422462 gb AE005793.1	Caulobacter crescentum CB15	69	1e-101
gi 23020895 ref NZ_AA AJ01000016.1	Clostridium thermocellum	69	1e-101
gi 15026105 gb AE007802.1 AE007802	Clostridium acetobutylicum	68	2e-101
gi 3322366 gb AE001195.1 AE001195	Treponema pallidum	68	2e-101
gi 15072928 emb AL591782.1 SME591782	Sinorhizobium meliloti	63	6e-091

Cj1417

gi 23058186 ref NZ_AABA01000065.1	Pseudomonas fluorescens	49	8e-123
gi 12725095 gb AE006435.1 AE006435	Lactococcus lactis	49	2e-112
gi 15024728 gb AE007685.1 AE007685	Clostridium acetobutylicum	45	4e-082
gi 23024872 ref NZ_AAAD01000048.1	Leuconostoc mesenteroides	47	1e-072
gi 19713988 gb AE010562.1	Fusobacterium nucleatum	40	2e-073
gi 23003135 ref NZ_AAAB01000011.1	Lactobacillus gasseri	41	4e-062
gi 3868717 emb AJ235271.1 RPXX02	Rickettsia prowazekii	44	2e-042
gi 15619630 gb AE008617.1 AE008617	Rickettsia conorii	40	3e-042
gi 20517707 gb AE013209.1	Thermoanaerobacter tenycongensis	36	0.0032
gi 24419007 emb AL939109.1 SCO939109	Streptomyces coelicolor	42	0.0141

Cj1418

gi 23058186 ref NZ_AABA01000065.1	Pseudomonas fluorescens	330	e-1708
gi 23472805 ref NZ_AABH01000015.1	Pseudomonas syringae	47	0.0021
gi 23104119 ref NZ_AAAD01000084.1	Azotobacter vinelandii	47	0.0021
gi 22958146 ref NZ_AA AE01000120.1	Rhodobacter sphaeroides	47	0.0021
gi 9949712 gb AE004777.1	Pseudomonas aeruginosa	46	0.0041
gi 23335573 ref NZ_AABF01000005.1	Bifidobacterium longum	46	0.0061
gi 14023709 dbj AP003002.2	Mesorhizobium loti	46	0.0061
gi 17429824 emb AL646072.1	Ralstonia solanacearum	45	0.0081
gi 2984342 gb AE000775.1 AE000775	Aquifex aeolicus	44	0.0201
gi 24347032 gb AE015577.1	Shewanella oneidensis	44	0.0201

Cj1419

gi 23058186 ref NZ_AABA01000065.1	Pseudomonas fluorescens	117	4e-423
gi 9946421 gb AE004492.1	Pseudomonas aeruginosa	43	7e-082
gi 21108003 gb AE011813.1	Xanthomonas axonopodis	50	5e-062
gi 21112874 gb AE012280.1	Xanthomonas campestris	48	1e-052
gi 22978195 ref NZ_AA AI01000311.1	Ralstonia metallidurans	52	2e-051
gi 6458805 gb AE001958.1 AE001958	Deinococcus radiodurans	51	3e-051
gi 14523147 gb AE007203.1 AE007203	Sinorhizobium meliloti	42	1e-042
gi 23113498 ref NZ_AABBU01000293.1	Desulfitobacterium hafniense	40	2e-042
gi 3261517 emb AL021841.1 MTV016	Mycobacterium tuberculosis	43	3e-042
gi 24413781 emb AL939116.1 SCO939116	Streptomyces coelicolor	40	4e-042

Cj1420

gi 23058186 ref NZ_AABA01000065.1	Pseudomonas fluorescens	87	7e-161
gi 3261517 emb AL021841.1 MTV016	Mycobacterium tuberculosis	59	3e-103
gi 21112874 gb AE012280.1	Xanthomonas campestris	62	9e-103
gi 24413781 emb AL939116.1 SCO939116	Streptomyces coelicolor	53	2e-093
gi 6458805 gb AE001958.1 AE001958	Deinococcus radiodurans	52	3e-082
gi 21108003 gb AE011813.1	Xanthomonas axonopodis	58	5e-082
gi 22979608 ref NZ_AA AI01000356.1	Ralstonia metallidurans	45	2e-062
gi 9946421 gb AE004492.1	Pseudomonas aeruginosa	47	2e-062
gi 23128557 ref NZ_AABC01000184.1	Nostoc punctiforme	55	2e-061
gi 23493310 dbj AP005219.1	Corynebacterium efficiens	47	4e-062

TESIS CON
FALLA DE ORIGEN

qi 14523147 qb AE007203.1 AE007203	Sinorhizobium meliloti	40	2e-052
------------------------------------	------------------------	----	--------

Cj1421

qi 5757614 qb U34622.1 CJU34622	Campylobacter jejuni	100	7e-344
qi 21326777 dbj AB079602.1	Escherichia coli	35	0.412
qi 18143657 dbj AP003185.1	Clostridium perfringens	39	0.511
qi 2688731 qb AE001178.1 AE001178	Borrelia burgdorferi	38	0.701
qi 13274354 qb AF302465.1 AF302465	Pasteurella multocida	37	1.81
qi 24197103 qb AE011482.1	Leptospira interrogans	35	6.41
qi 4981810 qb AE001781.1 AE001781	Thermotoga maritima	35	8.81

Cj1422

qi 5757614 qb U34622.1 CJU34622	Campylobacter jejuni	78	4e-233
qi 22990303 ref NZ AAK01000010.1	Enterococcus faecium	40	0.201
qi 21326777 dbj AB079602.1	Escherichia coli	35	0.422
qi 2688731 qb AE001178.1 AE001178	Borrelia burgdorferi	38	0.981
qi 13274354 qb AF302465.1 AF302465	Pasteurella multocida	37	1.81
qi 19712932 qb AE010475.1	Fusobacterium nucleatum	36	2.51
qi 16413128 emb AL596166.1	Listeria innocua	36	2.51
qi 290670 qb L13457.1 ERWAEP	Erwinia carotovora	36	3.51
qi 15981524 emb AJ414158.1	Yersinia pestis	30	3.52
qi 5199106 qb AF048749.1 AF048749	Bacteroides fragilis	35	6.61

Cj1423

qi 16357458 qb AF324836.2	Aneurinibacillus thermoaerophilus	146	1e-535
qi 15026105 qb AE007802.1 AE007802	Clostridium acetobutylicum	98	2e-424
qi 23321111 qb AF461769.1	Yersinia pseudotuberculosis	83	6e-385
qi 13446679 qb AF285636.1	Burkholderia mallei	121	5e-293
qi 20517022 qb AE013152.1	Thermoanaerobacter tengcongensis	69	2e-223
qi 23020895 ref NZ AANJ01000016.1	Clostridium thermocellum	75	5e-222
qi 2983301 qb AE000704.1 AE000704	Aquifex aeolicus	83	7e-212
qi 22971976 ref NZ AAH01000309.1	Chloroflexus aurantiacus	70	1e-204
qi 23015319 ref NZ AAP01003853.1	Magnetospirillum magnetotacticum	77	5e-203
qi 23011453 ref NZ AAA001000022.1	Trichodesmium erythraeum	76	4e-192
qi 14026358 dbj AP003010.2	Mesorhizobium loti	69	5e-172

Cj1424

qi 16357458 qb AF324836.2	Aneurinibacillus thermoaerophilus	314	9e-853
qi 13446679 qb AF285636.1	Burkholderia mallei	242	4e-632
qi 23321093 qb AF461768.1	Yersinia pseudotuberculosis	193	2e-613
qi 15026105 qb AE007802.1 AE007802	Clostridium acetobutylicum	172	5e-563
qi 23473702 ref NZ AAB101000002.1	Desulfovibrio desulfuricans	194	1e-481
qi 22966016 ref NZ AAG01000002.1	Rhodospirillum rubrum	191	1e-473
qi 2313982 qb AE000596.1 AE000596	Helicobacter pylori	188	1e-462
qi 22960729 ref NZ AAAF01000001.1	Rhodopseudomonas palustris	183	3e-452
qi 23053851 ref NZ AAAS01000003.1	Geobacter metallireducens	181	2e-441
qi 6968444 emb AL139077.2 CJ11168X4	Campylobacter jejuni	178	1e-433

Cj1425

qi 13446679 qb AF285636.1	Burkholderia mallei	156	1e-753
qi 16357458 qb AF324836.2	Aneurinibacillus thermoaerophilus	169	2e-684
qi 23321093 qb AF461768.1	Yersinia pseudotuberculosis	133	2e-654
qi 3261520 emb AL021926.1 MTV031	Mycobacterium tuberculosis	108	2e-463
qi 15026105 qb AE007802.1 AE007802	Clostridium acetobutylicum	86	6e-363
qi 23000469 ref NZ AAN01000197.1	Magnetococcus sp.	61	3e-173
qi 22972492 ref NZ AAH01000343.1	Chloroflexus aurantiacus	61	6e-172
qi 23016764 ref NZ AAP01003880.1	Magnetospirillum magnetotacticum	55	1e-153
qi 1652844 dbj D90909.1	Synechocystis sp.	53	2e-112
qi 16413394 emb AL596167.1	Listeria innocua	35	0.0482

Cj1426

qi 2277756 dbj AP004600.1	Oceanobacillus iheyensis	37	0.751
qi 2688623 qb AE001170.1 AE001170	Borrelia burgdorferi	36	1.41
qi 14025398 dbj AP003007.2	Mesorhizobium loti	32	2.52
qi 3844742 qb U39696.1 U39696	Mycoplasma genitalium	31	3.52
qi 18146728 dbj AP003190.2	Clostridium perfringens	35	3.71
qi 15023285 qb AE007558.1 AE007558	Clostridium acetobutylicum	30	5.02
qi 22972902 ref NZ AAH01000365.1	Chloroflexus aurantiacus	34	6.91

TESIS CON
FALLA DE ORIGEN

gi 1653083 dbj D90911.1	Synechocystis sp.	34	6.91
gi 10038693 dbj AP001118.1	Buchnera sp.	34	6.91
gi 15625506 gb AF410775.1	Staphylococcus aureus	29	8.32

Cj1427

gi 1653836 dbj D90917.1	Synechocystis sp.	431	e-1193
gi 23000469 ref NZ AAN01000197.1	Magnetococcus sp.	427	e-1183
gi 24195192 gb AE011336.1	Leptospira interrogans	53	5e-253
gi 23121882 ref NZ AAW01000001.1	Prochlorococcus marinus	67	1e-143
gi 22973511 ref NZ AAAH01000590.1	Chloroflexus aurantiacus	69	1e-143
gi 14600131 gb AF387640.1	Coxiella burnetii	55	2e-122
gi 10175792 dbj AP001518.1	Bacillus halodurans	66	9e-112
gi 20515651 gb AE013035.1	Thermoanaerobacter aurantiacus	53	7e-102
gi 23017216 ref NZ AAAQ01000018.1	Thermobifida fusca	44	9e-102
gi 22958382 ref NZ AAAE01000126.1	Rhodobacter sphaeroides	44	9e-104

Cj1428

gi 23053086 ref NZ AAAS01000001.1	Geobacter metallireducens	157	4e-894
gi 18152896 gb AF285774.2 AF285774	Bacteroides fragilis	168	7e-855
gi 24195192 gb AE011336.1	Leptospira interrogans	137	3e-844
gi 15025171 gb AE007718.1 AE007718	Clostridium acetobutylicum	140	4e-835
gi 23474173 ref NZ AABI01000003.1	Desulfovibrio desulfuricans	148	2e-784
gi 22989230 ref NZ AAAC01000305.1	Burkholderia fungorum	141	2e-754
gi 13362012 dbj AP002559.1	Escherichia coli	138	3e-754
gi 24052459 gb AE015227.1	Shigella flexneri	140	8e-754
gi 15980975 emb AJ414155.1	Yersinia pestis	133	1e-744
gi 16503247 emb AL627273.1	Salmonella enterica	141	5e-744

Cj1429

gi 15023994 gb AE007624.1 AE007624	Clostridium acetobutylicum	26	0.0654
gi 22972264 ref NZ AAAH01000329.1	Chloroflexus aurantiacus	40	0.121
gi 16410540 emb AL591978.1	Listeria monocytogenes	28	2.03
gi 22997362 ref NZ AAAM01000050.1	Xylella fastidiosa	35	4.11
gi 22776250 dbj AP004595.1	Oceanobacillus ihyensensis	34	5.61
gi 7288585 gb AF222438.1 AF222438	Borrelia garinii	34	5.61
gi 7381164 gb AF135373.1 AF135373	Acinetobacter calcoaceticus	34	5.61
gi 216840 dbj D13209.1 PRMBLAP	Proteus mirabilis	34	5.61
gi 15980810 emb AJ414154.1	Yersinia pestis	25	7.22
gi 22992693 ref NZ AAAK01000269.1	Enterococcus faecium	32	7.22

Cj1430

gi 23015319 ref NZ AAP01003853.1	Magnetospirillum magnetotacticum	98	1e-262
gi 24052435 gb AE015225.1	Shigella flexneri	61	2e-203
gi 9957844 gb AF279621.1 AF279621	Salmonella enterica	61	2e-183
gi 2828669 gb AF038816.1 AF038816	Serratia marcescens	52	3e-183
gi 23043478 ref NZ AAAU01000084.1	Trichodesmium erythraeum	58	3e-163
gi 24417728 gb AF529080.1	Escherichia coli	59	4e-163
gi 13518091 gb AY028370.1	Burkholderia mallei	50	7e-163
gi 22988102 ref NZ AAAC01000292.1	Burkholderia fungorum	55	2e-153
gi 22979310 ref NZ AAAI01000349.1	Ralstonia metallidurans	51	1e-143
gi 16611721 gb AF402312.1 AF402312	Shigella boydii	59	2e-143

Cj1431

gi 6469928 gb AF130466.1 AF130466	Campylobacter jejuni	62	3e-182
gi 496601 emb X77617.1 ECKFAAF	Escherichia coli	40	7e-042
gi 13274354 gb AF302465.1 AF302465	Pasteurella multocida	45	0.0081
gi 18146729 dbj AP003191.2	Clostridium perfringens	39	0.481
gi 20502702 gb AF375657.1	Aeromonas hydrophila	38	0.901
gi 23130735 ref NZ AAX01000001.1	Prochlorococcus marinus	36	2.31
gi 24475413 dbj AB073701.1	Borrelia duttonii	29	3.82
gi 23135899 ref NZ AABE01000038.1	Cytophaga hutchinsonii	35	4.41
gi 2462696 emb Y12875.1 SSY12875	Staphylococcus saprophyticus	27	6.62
gi 10038987 dbj AP001119.1	Buchnera sp.	34	7.32

Cj1432

gi 471233 emb X78559.1 HISBCAL	Haemophilus influenzae	84	4e-334
gi 16419219 gb AE008729.1	Salmonella typhimurium LT2	97	7e-273
gi 17225478 gb AF327219.1	Helicobacter pylori CR35	50	3e-194

TESIS CON
 FALLA DE ORIGEN

qi 6469928 gb AF130466.1 AF130466	Campylobacter jejuni	50	2e-122
qi 2209207 gb AF004325.1 AF004325	Streptococcus pneumoniae	73	2e-112
qi 23114751 ref NZ AABR01000358.1	Desulfitobacterium hafniense	50	3e-062
qi 2983559 gb AE000722.1 AE000722	Aquifex aeolicus	53	5e-051
qi 13274354 gb AF302465.1 AF302465	Pasteurella multocida	50	3e-041
qi 18146725 dbj AP003187.2	Clostridium perfringens	46	0.0081
qi 2690041 gb AE000784.1 AE000784	Borrelia burgdorferi	40	0.0192

Cj1433

qi 23113434 ref NZ AABR01000288.1	Desulfitobacterium hafniense	123	7e-443
qi 5579417 gb U70376.2 U70376	Streptomyces flavopersicus	60	8e-122
qi 23128923 ref NZ AABC01000189.1	Nostoc punctiforme	68	2e-102
qi 150342 gb J03017.1 NGOPOR	Neisseria gonorrhoeae	62	2e-081
qi 23041985 ref NZ AAAU01000032.1	Trichodesmium erythraeum	60	1e-071
qi 483518 emb X79075.1 CBPUTOPP	Coxiella burnetii	59	2e-071
qi 4731413 gb AF145039.1 AF145039	Streptomyces spectabilis	46	5e-072
qi 22972264 ref NZ AAH01000329.1	Chloroflexus aurantiacus	56	2e-061
qi 12581462 gb AF105330.1 AF105330	Clostridium cellulovorans	55	4e-061
qi 2313137 gb AE000528.1 AE000528	Helicobacter pylori	50	1e-041

Cj1434

qi 1945641 emb Z94043.1 BSZ94043	Bacillus subtilis	123	1e-261
qi 13377447 gb AF316642.1 AF316642	Streptococcus pneumoniae	120	7e-261
qi 18146725 dbj AP003187.2	Clostridium perfringens	110	8e-232
qi 6601338 gb AF155804.1 AF155804	Streptococcus suis	109	1e-222
qi 13876769 gb AF355776.1 AF355776	Streptococcus agalactiae	107	5e-222
qi 12005996 gb AF267127.1	Lactobacillus delbrueckii	105	2e-211
qi 23002160 ref NZ AABR01000001.1	Lactobacillus gasserii	105	4e-211
qi 24636987 gb AF410175.1	Streptococcus thermophilus	104	4e-212
qi 15025159 gb AE0007717.1 AE0007717	Clostridium acetobutylicum	100	6e-201
qi 18146724 dbj AP003186.2	Clostridium perfringens	100	1e-191

Cj1435

qi 23029813 ref NZ AAAT01000036.1	Microbulbifer degradans	74	3e-121
qi 15979242 emb AJ414147.1	Yersinia pestis	51	6e-112
qi 9949378 gb AE004748.1	Pseudomonas aeruginosa	68	2e-101
qi 23471304 ref NZ AABH01000006.1	Pseudomonas syringae	68	3e-101
qi 21322764 dbj AP005274.1	Corynebacterium glutamicum	42	2e-092
qi 23472880 ref NZ AABH01000016.1	Pseudomonas syringae	60	5e-081
qi 23060917 ref NZ AABA01000137.1	Pseudomonas fluorescens	60	5e-081
qi 19656466 gb AE004269.1 AE004269	Vibrio cholerae	57	3e-071
qi 15023094 gb AE007539.1 AE007539	Clostridium acetobutylicum	38	9e-072
qi 3261507 emb AL021246.1 MTV008	Mycobacterium tuberculosis	36	4e-062

Cj1436

qi 23026153 ref NZ AAAT01000001.1	Microbulbifer degradans	59	2e-344
qi 24413764 emb AL939112.1 SCO939112	Streptomyces coelicolor	60	2e-325
qi 2967841 gb AF052516.1 AF052516	Prevotella intermedia	45	1e-216
qi 20517144 gb AE013160.1	Thermoanaerobacter tengcongensis	45	3e-214
qi 18146728 dbj AP003190.2	Clostridium perfringens	58	1e-193
qi 24202250 gb AE011603.1	Leptospira interrogans	71	2e-163
qi 20804067 emb AL672114.1 ML0672114	Mesorhizobium loti	40	3e-165
qi 19714538 gb AE010604.1	Fusobacterium nucleatum	48	6e-163
qi 18146727 dbj AP003189.2	Clostridium perfringens	49	7e-123
qi 19623604 gb AF269708.1 AF269708	Staphylococcus epidermidis	46	2e-113

Cj1437

qi 23040634 ref NZ AAAU01000012.1	Trichodesmium erythraeum	118	3e-251
qi 20517144 gb AE013160.1	Thermoanaerobacter tengcongensis	75	2e-235
qi 23121882 ref NZ AAAW01000001.1	Prochlorococcus marinus	63	3e-233
qi 23126546 ref NZ AABC01000154.1	Nostoc punctiforme	106	8e-221
qi 23029675 ref NZ AAAT01000031.1	Microbulbifer degradans	101	2e-212
qi 17134031 dbj AP003598.1	Nostoc sp.	100	7e-201
qi 22293721 dbj AP005369.1	Thermosynechococcus elongatus	100	1e-191
qi 22775946 dbj AP004594.1	Oceanobacillus iheyensis	96	2e-181
qi 10038693 dbj AP001118.1	Buchnera sp.	47	3e-186
qi 14246388 dbj AP003360.2	Staphylococcus aureus	73	1e-172

TESIS CON
FALLA DE ORIGEN

Cj1438

gi 23022101 ref NZ_AAAJ01000078.1	Clostridium thermoecellum	65	3e-374
gi 2462110 emb Y11138.1 BCY11138	Bacillus cereus	66	3e-344
gi 24637512 gb AF454501.1	Streptococcus thermophilus	58	2e-295
gi 2635827 emb Z99121.1 BSUB0018	Bacillus subtilis	124	8e-271
gi 13377447 gb AF316642.1 AF316642	Streptococcus pneumoniae	122	4e-261
gi 18146725 dbj AP003187.2	Clostridium perfringens	110	1e-222
gi 6601338 gb AF155804.1 AF155804	Streptococcus suis	109	3e-222
gi 13876769 gb AF355776.1 AF355776	Streptococcus agalactiae	108	7e-222
gi 12005996 gb AF267127.1	Lactobacillus delbrueckii	105	3e-211

Cj1439

gi 6103223 emb AJ131984.1 SPN131984	Streptococcus pneumoniae	357	e-1514
gi 22218128 gb AF448502.1	Streptococcus thermophilus	357	e-1513
gi 1736708 dbj D90841.1	Escherichia coli	358	e-1507
gi 14578621 gb AF189282.1 AF189282	Bacteroides fragilis	350	e-1486
gi 23002160 ref NZ_AAB01000001.1	Lactobacillus gasseri	188	e-1255
gi 23037621 ref NZ_AAAZ01000006.1	Oenococcus oeni	182	e-1214
gi 23024822 ref NZ_AAAO01000045.1	Leuconostoc mesenteroides	176	e-1185
gi 23335957 ref NZ_AABF01000009.1	Bifidobacterium longum	104	1e-766
gi 557191 gb L31762.1 KPNRFRC	Klebsiella pneumoniae	116	8e-756
gi 17742413 gb AE009328.1 AE009328	Agrobacterium tumefaciens	121	2e-716

Cj1440

gi 6968444 emb AL139077.2 CJ11168X4	Campylobacter jejuni	204	e-1114
gi 16410252 emb AL591977.1	Listeria monocytogenes	77	1e-202
gi 13876769 gb AF355776.1 AF355776	Streptococcus agalactiae	81	2e-173
gi 23002160 ref NZ_AAB01000001.1	Lactobacillus gasseri	67	5e-172
gi 5123513 gb AF036485.2 AF036485	Lactococcus lactis	74	1e-162
gi 18146725 dbj AP003187.2	Clostridium perfringens	72	2e-164
gi 3818479 gb AF057294.1 AF057294	Streptococcus pneumoniae	75	4e-143
gi 6601338 gb AF155804.1 AF155804	Streptococcus suis	81	7e-142
gi 12005996 gb AF267127.1	Lactobacillus delbrueckii	66	4e-132
gi 22218112 gb AF434993.1	Streptococcus thermophilus	77	7e-132

Cj1441

gi 12721075 gb AE006116.1 AE006116	Pasteurella multocida	412	e-1275
gi 25860996 gb U78086.1 ECU78086	Escherichia coli	387	e-1263
gi 10176109 dbj AP001519.1	Bacillus halodurans	394	e-1254
gi 4115688 dbj AB012957.1	Vibrio cholerae	390	e-1244
gi 22726324 gb AF416330.1	Ruegeria sp.	375	e-1244
gi 16420610 gb AE008792.1	Salmonella typhimurium LT2	383	e-1223
gi 23037621 ref NZ_AAAZ01000006.1	Oenococcus oeni	374	e-1216
gi 23026543 ref NZ_AAAO1000002.1	Microbulbifer degradans	383	e-1215
gi 21109951 gb AE012008.1	Xanthomonas axonopodis	370	e-1193
gi 22954051 ref NZ_AAAO1000001.1	Nitrosomonas europaea	374	e-1174

Cj1442

gi 15140405 emb AL603644.1 RME603644	Rhizobium meliloti	46	7e-112
gi 20502702 gb AF375657.1	Aeromonas hydrophila	48	2e-042
gi 24636987 gb AF410175.1	Streptococcus thermophilus	43	0.0191
gi 21203164 dbj AP004822.1	Staphylococcus aureus	40	0.171
gi 21903714 gb AF036961.2	Mycoplasma hominis	38	0.841
gi 19714865 gb AE010630.1	Fusobacterium nucleatum	38	0.841
gi 18152896 gb AF285774.2 AF285774	Bacteroides fragilis	36	2.21
gi 22954051 ref NZ_AAAO1000001.1	Nitrosomonas europaea	36	3.01
gi 16410252 emb AL591977.1	Listeria monocytogenes	36	3.01
gi 23138256 ref NZ_AABE01000099.1	Cytophaga hutchinsonii	31	3.42

Cj1443

gi 23467224 ref NZ_AABG01000006.1	Haemophilus somnus	261	3e-893
gi 12720774 gb AE006087.1 AE006087	Pasteurella multocida	261	4e-883
gi 1574529 gb U32841.1 U32841	Haemophilus influenzae	257	7e-882
gi 23000927 ref NZ_AAAO1000210.1	Magnetococcus sp.	274	1e-853
gi 22979688 ref NZ_AAAI01000356.1	Ralstonia metallidurans	309	5e-832
gi 23056119 ref NZ_AAAS01000024.1	Geobacter metallireducens	286	7e-813
gi 7380672 emb AL162758.2 NMA722491	Neisseria meningitidis	278	5e-783

TESIS CON
FALLA DE ORIGEN

gi 13363382 dbj AP002564.1	<i>Escherichia coli</i>	274	4e-742
gi 16504263 emb AL627278.1	<i>Salmonella enterica</i>	272	5e-742
gi 15981328 emb AJ414157.1	<i>Yersinia pestis</i>	267	7e-743

Cj1444

gi 20502702 gb AF375657.1	<i>Aeromonas hydrophila</i>	222	e-1124
gi 409137 gb MG4977.1 ECOKPSD	<i>Escherichia coli</i>	202	e-1064
gi 22979688 ref NZ AAA101000356.1	<i>Ralstonia metallidurans</i>	205	e-1044
gi 23055649 ref NZ AAAS01000015.1	<i>Geobacter metallireducens</i>	68	4e-213
gi 24349358 gb AE015757.1	<i>Shewanella oneidensis</i>	57	1e-194
gi 2983148 gb AE000693.1 AE000693	<i>Aquifex aeolicus</i>	76	3e-194
gi 3395597 emb X90547.1 VCOTNREGN	<i>Vibrio cholerae</i>	71	1e-132
gi 21108849 gb AE011898.1	<i>Xanthomonas axonopodis</i>	58	6e-071
gi 16755720 gb AF427012.1 AF427012	<i>Xanthomonas campestris</i>	53	2e-051
gi 22960729 ref NZ AAAP01000001.1	<i>Rhodopseudomonas palustres</i>	45	5e-052

Cj1445

gi 397404 emb X74567.1 ECKPSCNST	<i>Escherichia coli</i>	181	1e-612
gi 22954051 ref NZ AAAY01000001.1	<i>Nitrosomonas europaea</i>	113	3e-493
gi 20502702 gb AF375657.1	<i>Aeromonas hydrophila</i>	156	6e-382
gi 13446679 gb AF285636.1	<i>Burkholderia mallei</i>	79	9e-333
gi 22979688 ref NZ AAA101000356.1	<i>Ralstonia metallidurans</i>	121	1e-313
gi 65233003 emb AJ245666.1 RME245666	<i>Rhizobium meliloti</i>	131	4e-291
gi 23132699 ref NZ AABD01000001.1	<i>Synechococcus sp.</i>	130	8e-291
gi 4809242 gb AF143906.1 AF143906	<i>Actinobacillus pleuropneumoniae</i>	79	8e-282
gi 426443 dbj D14156.1 STYVIAB	<i>Salmonella typhi</i>	117	4e-251
gi 13172665 gb AF316551.1 AF316551	<i>Citrobacter freundii</i>	117	5e-251

Cj1447

gi 41877 emb X53819.1 ECKRPSMT	<i>Escherichia coli</i>	157	3e-723
gi 20502702 gb AF375657.1	<i>Aeromonas hydrophila</i>	135	2e-602
gi 22979688 ref NZ AAA101000356.1	<i>Ralstonia metallidurans</i>	223	4e-572
gi 65233003 emb AJ245666.1 RME245666	<i>Rhizobium meliloti</i>	143	1e-562
gi 22726324 gb AF416330.1	<i>Ruegeria sp.</i>	145	3e-562
gi 22960082 ref NZ AAAP01000155.1	<i>Rhodobacter sphaeroides</i>	121	1e-522
gi 7710189 gb AF067175.2 AF067175	<i>Pasteurella multocida</i>	112	6e-512
gi 148866 gb IM19995.1 HEABEXA	<i>Haemophilus influenzae</i>	114	8e-502
gi 6424984 gb AF170495.1 AF170495	<i>Pasteurella haemolytica</i>	116	2e-492
gi 150249 gb IM57677.1 INGOCSPG	<i>Neisseria meningitidis</i>	114	4e-492

Cj1448

gi 14423109 gb AF322113.1 AF322113	<i>Campylobacter jejuni</i>	549	e-1554
gi 4877778 gb AF126952.1 AF126952	<i>Escherichia coli</i>	61	3e-243
gi 20502702 gb AF375657.1	<i>Aeromonas hydrophila</i>	59	1e-214
gi 22979688 ref NZ AAA101000356.1	<i>Ralstonia metallidurans</i>	68	2e-192
gi 22954051 ref NZ AAAY01000001.1	<i>Nitrosomonas europaea</i>	67	2e-182
gi 426443 dbj D14156.1 STYVIAB	<i>Salmonella typhi</i>	39	2e-052
gi 23108212 ref NZ AAAP01000129.1	<i>Novosphingobium aromaticivorans</i>	42	2e-052
gi 15140405 emb AL603644.1 RME603644	<i>Rhizobium meliloti</i>	43	8e-052
gi 13172665 gb AF316551.1 AF316551	<i>Citrobacter freundii</i>	38	2e-042
gi 23132699 ref NZ AABD01000001.1	<i>Synechococcus sp.</i>	46	0.0021

Cj1449

gi 15981328 emb AJ414157.1	<i>Yersinia pestis</i>	44	4e-082
gi 16504263 emb AL627278.1	<i>Salmonella enterica</i>	42	3e-062
gi 1789536 gb AE000396.1 AE000396	<i>Escherichia coli</i>	49	7e-051
gi 24053620 gb AE015331.1	<i>Shigella flexneri</i>	49	7e-051
gi 3036829 emb AJ000856.1 CJAJ0856	<i>Campylobacter jejuni</i>	47	2e-041
gi 12722388 gb AE006232.1 AE006232	<i>Pasteurella multocida</i>	45	9e-041
gi 18146726 dbj AP003188.2	<i>Clostridium perfringens</i>	39	0.0551
gi 23126256 ref NZ AABC01000148.1	<i>Nostoc punctiforme</i>	32	6.51
gi 23043910 ref NZ AAAP01000139.1	<i>Trichodesmium erythraeum</i>	32	6.51
gi 24347089 gb AE015581.1	<i>Shewanella oneidensis</i>	32	8.91

TESIS CON
FALLA DE ORIGEN

BIBLIOGRAFÍA

1. Altschul S.F. et al., Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs, *Nucleic Acids Res.*, 1997, 25(17):3389-402
2. Anderson C.L., Karr T.L., Wolbachia: Evolutionary Novelty in a Rickettsial Bacteria, *BMC Evol Biol*, 2001, 1:10 <http://www.biomedcentral.com/1471-2148/1/10>
3. Brown E. W. et al., Three R's of Bacterial Evolution: How Replication, Repair and Recombination Frame the Origin of Species, *Environ Mol Mutagen*, 2001, 38:248-60
4. Copley S.D., Dhillon J.K., Lateral Gene Transfer and Parallel Evolution in the History of Glutathione Biosynthesis Genes, *Genome Biol*, 2002, 3(5):research0025.1-0025.16
5. Doolittle W.F., Phylogenetic Classification and the Universal Tree, *Science*, 1999, 284:2124-8
6. Espinosa-Aguirre J. et al., Characterization of the CYP isozyme profile induced by cyclohexanol. *Mutagen*, 1997, 12(3):159-62
7. Fagan T., Hastings J.W., Morse D., The Phylogeny of Glyceraldehyde-3-Phosphate Dehydrogenase Indicates Lateral Gene Transfer from Cryptomonads to Dinoflagellates, *J Mol Evol*, 1998, 47:633-9
8. Fickett J.W., Torney D.C., Wolf D.R., Base Compositional Structure of Genomes, *Genomics*, 1992, 13:1056-64
9. Goodwin C. S. et al., Transfer of *Campylobacter pylori* and *Campylobacter mustelae* to *Helicobacter* gen. nov. as *Helicobacter pylori* comb. nov. and *Helicobacter mustelae* comb. nov. respectively, *Int J Syst Bacteriol*, 39:397-405
10. Hacker J. et al., Pathogenicity Islands of Virulent Bacteria: Structure, Function and Impact on Microbial Evolution, *Mol Microbiol*, 1997, 23(6):1089-1097
11. Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 41:95-98.
12. He J., Fulco A.J., A Barbiturate-regulated Protein Binding to a Common Sequence in the Cytochrome P450 Genes of Rodents and Bacteria, *J Biol Chem*, 1991, 266(12):7864-69
13. Heatley R. V., The *Helicobacter pylori* Handbook, *Blackwell Science Ltd.* 1996.
14. Hillis D. M., Moritz C., Mable B. K., *Molecular Systematics*, Second Ed., *Sinauer Associates Inc.*, 1996
15. Hsiung-Li W., Graur D., *Fundamentals of Molecular Evolution*, *Sinauer Associates Inc.*, 1991
16. IARC Monographs, Infection with *Helicobacter pylori*, Vol. 61:188-240
17. Karlin S., Campbell A.M., Mrázek J., Comparative DNA Analysis Across Diverse Genomes, *Annu Rev Genet*, 1998, 32:185-225
18. Kyrpides, N.C., Olsen G.J. Archaeal and Bacterial Hyperthermophiles: Horizontal Gene Exchange or Common Ancestry?, *Trends in Genetics*, 1999, 15:298-9
19. Lawrence J.G., Roth J.R., Evolution of Coenzyme B12 Synthesis Among Enteric Bacteria: Evidence for Loss and Reacquisition of a Multigene Complex, *Genetics*, 1996, 142:11-24
20. Lawrence J.G., Ochman H., Reconciling the Many Faces of Lateral Gene Transfer, *Trends Microbiol*, 2002, 10(1) Research Update
21. Lawrence J.G., Roth J.R., Selfish Operons: Horizontal Transfer May Drive the Evolution of Gene Clusters, *Genetics*, 1996, 143:1843-60
22. Marshall B. J., Unidentified curved bacilli on gastric epithelium in active chronic gastritis (Letter to the Editor), *Lancet*, i, 1273-5
23. Morett E., Bork P., Evolution of New Protein Function: Recombinational Enhancer Fis Originated by Horizontal Gene Transfer from the Transcriptional Regulator NtrC, *FEBS Letters*, 1998, 433:108-112

24. Mount W. D., *Bioinformatics: Sequence and Genome Analysis*, Cold Spring Harbor Laboratory Press, 2001.
25. Nelson K.E. et al., Evidence for Lateral Gene Transfer Between Archaea and Bacteria from Genome Sequence of *Thermotoga maritima*, *Nature*, 399:323-9
26. Nielsen K.M., Horizontal Gene Transfer – DNA in the Soil, AgBio View in Mindfully.org, 15May2001, <http://www.mindfully.org/GE/GE2/Horinzontal-Gene-Soil.htm>
27. Ochman H., Lateral and oblique gene transfer, *Curr Opin Genet Dev*, 2001, 11:616-9
28. Ochman H., Lawrence J.G., Groisman E.A., Lateral Gene Transfer and the Nature of Bacterial Innovation, *Nature*, 2000, 405:209-304
29. Ochman H., Lawrence J.G., Phylogenetics and the Amelioration of Bacterial Genomes in *Molecular Evolution and Population Genetics Section G*, 2627-37
30. Olsen G.J., Woese C.R., Overbeek R., The Winds of (Evolutionary) Change: Breathing New Life into Microbiology, *J Bacteriol*, 1994, 176(1):1-3
31. Omura T., Ishimura Y., Fujii-Kuriyama Y., *Cytochrome P-450*, Kodansha Ltd., Second edition, 1993
32. Parkhill J. et al, The Genome Sequence of the Food-Borne Pathogen *Campylobacter jejuni* Reveals Hypervariable Sequences, *Nature*, 2000, 403:665-8
33. Poupin P. et al., Cloning and Characterization of the Genes Encoding a Cytochrome P450 (PipA) Involved in Piperidine and Pyrrolidine Utilization and Its Regulatory Protein (PipR) in *Mycobacterium smegmatis* mc²155, *J Bacteriol*, 1999, June:3419-26
34. Poupin P., Truffaut N., Combourieu B., Besse P., Sancelme M., Veschambre H., Delort A.M., Degradation of Morpholine by an Environmental *Mycobacterium* Strain Involves a Cytochrome P-450, *Appl Environ Microbiol*, 1998, 64(1):159-65
35. Ragan M.A., Charlebois R.L., Distributional Profiles of Homologous Open Reading Frames Among Bacterial Phyla: Implications for Vertical and Lateral Transmission, *Int J Syst Evol Microbiol*, 2002, 52:777-87
36. Rodrigues-Lima F., Dupret J.M., In Silico Sequence Analysis of Arylamine N-acetyltransferases: Evidence for an Absence of Lateral Gene Transfer from Bacteria to Vertebrates and First Description of Paralogs in Bacteria, *Biochem Biophys Res Commun*, 2002, 293:783-92
37. Roelofs J., Van Haastert P.J.M., Brief Communications Genomics: Genes Lost During Evolution, *Nature*, 2001, 411:1013-4
38. Saitou N., Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Mol Biol Evol*, 1987, Jul 4(4):406-25
39. Schbath S, Prum B, de Turckheim E, Exceptional Motifs in Different Markov Chain Models for a Statistical Analysis of DNA Sequences, *J Comput Biol*, 1995, 2(3):417-37
40. Shahamat M. et al., Evaluation of Liquid Media for Growth of *Helicobacter pylori*, *J Clin Microbiol*, 1991, Dec:2835-37
41. Sharp P.M., Li W., The Codon Adaptation Index – A Measure of Directional Synonymous Codon Usage Bias, and its Potential Applications, *Nucleic Acids Res*, 1987, 15(3):1281-95
42. Shen Z., Feng Y., Fox J.G., Identification of Enterohepatic *Helicobacter* Species by Restriction Fragment-Length Polymorphism Analysis of the 16S rRNA Gene, *Helicobacter*, 2000, 5(3):121-8
43. Smith M. W., Feng D., Doolittle R. F., Evolution by acquisition: the case for horizontal gene transfers, *TIBS*, 1992, Dec:489-93
44. Stanhope M.J. et al., Phylogenetic Analyses Do Not Support Horizontal Gene Transfers from Bacteria to Vertebrates, *Nature*, 2001, 411:940-4
45. Sueoka N., Directional Mutation Pressure, Mutator Mutations, and Dynamics of Molecular Evolution, *J Mol Evol*, 1993, 37:137-53
46. Synanen M., Horizontal Gene Transfer: Evidence and Possible Consequences, *Annu Rev Genet*, 1994, 28:237-61

47. Thompson J. D., Higgins D. G., Gibson T. J., CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acid Res*, 1994, 22(22):4673-80
48. Tomb J. et al., The complete genome sequence of the gastric pathogen *Helicobacter pylori*, *Nature*, 1997, 388:539-47
49. Viale A.M., Arakaki A.K., Soncini F.C., Ferreyra R.G., Evolutionary Relationships Among Eubacterial Groups as Inferred from GroEL (chaperonin) Sequence Comparisons, *Int J Syst Bacteriol*, 1994, 44(3):527-33
50. Wada K., et al., Codon Usage Tabulated From the GenBank Genetic Sequence Data, *Nucleic Acids Res*, 1990, 18(Supp):2367-9
51. Warren J. R., Unidentified curved bacilli on gastric epithelium in active chronic gastritis (Letter to the Editor), *Lancet*, i, 1273
52. Wen L., Ruettinger R.T., Fulco A.J., Requirement for a 1-Kilobase 5'-Flanking Sequence for Barbiturate-inducible Expression of the Cytochrome P-450BM3 Gene in *Bacillus megaterium*, *J Biol Chem*, 1989, 264(19):10996-11003
53. Wood D. W. et al., The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58, *Science*, 2001, 294:2317-23