



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

Comparación entre un método filogenético y
uno composicional para la detección de
genes xenólogos.

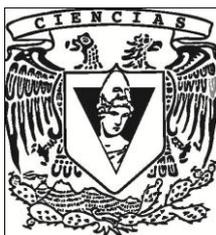
T E S I S

QUE PARA OBTENER EL TÍTULO DE:

B I Ó L O G O

P R E S E N T A :

OCTAVIO ABRAHAM MORENO GUILLÉN



DIRECTOR DE TESIS:

Dr. ARTURO CARLOS II BECERRA BRACHO

OCTUBRE 2011



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

DR. ISIDRO ÁVILA MARTÍNEZ
Director General
Dirección General de Administración Escolar
Presente

FACULTAD DE CIENCIAS
Secretaría General
División de Estudios Profesionales

Oficio FCIE/SG/DEP/0379/11

Autorización de examen profesional

Comunico a usted que la Dirección de esta Facultad ha autorizado a **Moreno Guillén Octavio Abraham**, con número de cuenta **3-0002007-4** y de la licenciatura en **Biología**, quien presenta la tesis

Comparación entre un método filogenético y uno composicional para la detección de genes xenólogos

a presentar su examen profesional el día **21 de octubre de 2011** a las **13:00 horas**, ante el siguiente jurado:

Presidente M. en C. Alfonso José Vilchis Pelayera

Vocal Dr. León Patricio Martínez Castilla

Secretario Tutor Dr. Arturo Carlos II Becerra Bracho

Suplente M. en C. Fabiola Ramírez Corona

Suplente M. en C. Sara Ernestina Islas Graciano

Atentamente,
"POR MI RAZA HABERÁ EL ESPÍRITU"
Ciudad Universitaria, D. F., 06 de octubre de 2011
EL JEFE DE LA DIVISIÓN DE ESTUDIOS PROFESIONALES

ACT. MIGUEL ÁNGEL GONZÁLEZ

Estimado(a) sinodal:

- Solamente debe firmar **una** hoja para acusar recibo del oficio citatorio que le entrega el(la) alumno(a) portador(a) de la presente; dicha hoja debe ser devuelta por el alumno.
- Le suplicamos presentarse 10 minutos antes de la hora fijada para la celebración del examen profesional, en la Sección Escolar de la Facultad de Ciencias.
- El oficio citatorio que el(la) alumno(a) le entrega sirve a usted como comprobante de participación como jurado.
- MAG/CZS/ygg

Hoja de Datos del Jurado.

1. Datos del alumno

Apellido paterno
Apellido materno
Nombres
Teléfono
U.N.A.M.
Facultad de Ciencias
Carrera
Número de cuenta

1. Datos del alumno

Moreno
Guillén
Octavio Abraham
56770325
U.N.A.M.
Facultad de Ciencias
Biología
300020074

2. Datos del tutor

Grado
Nombres
Apellido paterno
Apellido materno

2. Datos del tutor

Dr.
Arturo Carlos II
Becerra
Bracho

3. Datos del sinodal 1

Grado
Nombres
Apellido paterno
Apellido materno

3. Datos del sinodal 1

M en C.
Alfonso José
Vilchis
Peluyera

4. Datos del sinodal 2

Grado
Nombres
Apellido paterno
Apellido materno

4. Datos del sinodal 2

Dr.
León Patricio
Martínez
Castilla

5. Datos del sinodal 3

Grado
Nombres
Apellido paterno
Apellido materno

5. Datos del sinodal 3

M en C.
Fabiola
Ramírez
Corona

6. Datos del sinodal 4

Grado
Nombres
Apellido paterno
Apellido materno

6. Datos del sinodal 4

M en C.
Sara Ernestina
Islas
Graciano

7. Datos del trabajo escrito

Título

Comparación entre un método filogenético y uno composicional para la detección de genes xenólogos.

Número de páginas

Año

7. Datos del trabajo escrito

62

2011

Agradecimientos.

- ❖ A mis papás, Irma Guillén y Antonio Moreno por tenerme paciencia, ayudarme, patrocinarme y aguantarme de prángana por tanto tiempo.
- ❖ A mi tutor, Arturo Becerra, por recibirme, por las comidas que invitó y por la ayuda en las cuestiones más técnicas de éste trabajo.
- ❖ A Fabiola Ramírez Corona. Mi Sinodal, Maestra y amiga por permitirme trabajar en su laboratorio cada vez que lo necesité.
- ❖ A mis Sinodales quienes se tomaron un tiempo para revisar y corregir mi trabajo.
- ❖ Un agradecimiento muy especial a Vanessa Mariel Bustos Hernández por todo el cariño que le tengo y por los últimos meses, que en su compañía, han sido simplemente increíbles...
- ❖ A CONACYT por el apoyo brindado para la compra del equipo de cómputo en donde se realizó la mayor parte de este trabajo.
- ❖ A Verónica Cruz Hernández por las múltiples contribuciones, por su amistad y por tenerme paciencia.
- ❖ A mi perro, el burrito. La tesis de Maestría será sobre mi proyecto para clonarlo.
- ❖ A cualquier colado que sienta que contribuyó, pero a mi consideración no fue suficiente como para aparecer en esta honorable sección.

INDICE

1. Resumen	1
2. Introducción	2
2.1 Marco histórico e identificación del problema con la HGT	3
2.2 Mecanismos de intercambio genético	4
3. Antecedentes	5
3.1 Métodos para la detección e identificación de genes xenólogos	5
3.2 Comprobando la eficacia de otras metodologías	9
3.3 Consideraciones importantes con el problema de la HGT	11
4. Objetivos	15
5. Metodología	15
5.1 Testigos	17
5.2 Diagrama de Flujo	19
6. Resultados y Discusión	20
6.1 Comprobando los Resultados	23
6.2 Filogenias	24
7. Arboles Filogenéticos	29
7.1 Testigos	30
7.2 Casos Seguros de HGT	36
7.3 Casos en los que No hubo HGT	41
7.4 Casos en los que no es posible diferenciar entre ganancia (HGT) o pérdida de genes	46
8. Perspectivas	50
9. Conclusiones	51
10. Referencias	53

Resumen.

La transferencia horizontal de genes (por sus siglas en inglés, HGT), también conocida como transferencia lateral de genes (LGT), es la adquisición de un gen o genes entre una, dos o más especies que pueden o no estar filogenéticamente emparentadas. Es pues, un fenómeno capaz de cruzar barreras interespecies. En la actualidad, la HGT es un fenómeno muy reconocido y estudiado porque facilita eventos de innovación a nivel genoma que han modelado la historia evolutiva de una gran cantidad de organismos.

Aunque no todos los dominios han sido investigados con la misma consistencia y es notorio que la mayor cantidad de evidencia se encuentra concentrada en bacteria, este hecho plantea dudas respecto a la frecuencia e impacto del fenómeno en cada uno de las ramas del denominado árbol de la vida. Haciendo uso de diferentes metodologías, muchos estudios se han enfocando en la detección de genes xenólogos, que son el producto de la transferencia horizontal, en muchos organismos pertenecientes a diferentes clados y dominios. Mientras que el intrincado proceso por el cual los organismos comparten información de manera no vertical comienza a visualizarse, igualmente empezamos a desentrañar el verdadero alcance que la transferencia horizontal de genes tiene y ha tenido durante la evolución de los organismos.

Así, uno de los puntos más importantes en el estudio de la transferencia horizontal de genes es la identificación de los genes xenólogos (genes producto de la HGT), así como determinar la frecuencia de ocurrencia para éste fenómeno. Con este propósito, muchos esfuerzos se están realizando para acumular un conocimiento cada vez más completo sobre la HGT, y una parte importante de estos esfuerzos se concentra en el desarrollo de metodologías más rigurosas y altamente especializadas cuya finalidad es la identificación de este tipo genes. Lamentablemente, la realidad es que, a pesar de tales esfuerzos, aun resulta complicado identificar con precisión y confianza genes xenólogos. La identificación de este tipo de genes usualmente es dudosa y en ocasiones, los resultados son contradictorios (Becq, 2010), por lo que se vuelve muy complicado determinar si un gen es producto de HGT.

Como la HGT implica un medio de innovación a nivel genoma, entender el fenómeno nos permitiría tener una visión más clara de los procesos evolutivos de una gran cantidad de organismos. Es entonces que mientras metodologías progresivamente precisas y prácticas que explotan muchas características intrínsecas y extrínsecas de los genomas bajo análisis se siguen desarrollando, en este trabajo nos enfocamos en la metodología que pensamos tiene las mayores posibilidades de aportar la información más valiosa y confiable, el método filogenético, que consiste en realizar reconstrucciones filogenéticas completas para aquellos genes que se sospeche sean posibles xenólogos (Eisen, 2000).

Bajo este supuesto, ponemos a prueba la eficacia y precisión en la detección de genes producto de transferencias horizontales haciendo una comparación y validación de los genes reportados como xenólogos por un método composicional (metodología que se basa en la detección de genes xenólogos en base a características que los identifique como foráneos a nivel genoma [ver Métodos para la detección e identificación de genes xenólogos para una

referencia más completa]), haciendo uso del método filogenético como herramienta de búsqueda, detección e identificación, para un total de 15 organismos procariontes.

Nuestros resultados, basados en la construcción de filogenias usando los genes reportados como xenólogos por Cortéz y colaboradores (2005), para posteriormente comparar los resultados contra 5 filogenias testigo, indican que los genes identificados por Cortéz y colaboradores como posibles genes xenólogos tienen una gran posibilidad de serlo, sin embargo, en nuestro trabajo también apreciamos que el método filogenético no es la herramienta más apropiada para corroborar resultados cuando metodologías de tipo paramétricas (que únicamente toman en cuenta la información característica de cada genoma), fueron empleadas para obtenerlos, situación que ocurre frecuentemente, pero resulta ser una justificación inapropiada en la construcción de filogenias.

Mostramos, además, que restricciones importantes, como la falta de genes ortólogos (genes que se encuentran en especies diferentes pero evolucionaron de un ancestro común) para las comparaciones, secuencias contaminadas con genes parálogos (genes originados por duplicación en un mismo genoma, pero que aun podrían mantener una relación con los genes en otra especie), el tamaño de la muestra y la homogenización que experimentan los genes después de ser transferidos (término conocido en la literatura como “*amelioration*” [Lawrence y Ochman, 1997]) son obstáculos importantes en la construcción de filogenias confiables (con significado biológico plausible) que normalmente no son considerados en las metodologías paramétricas, y que por ende, descartan la construcción de filogenias como un método de control sobre los resultados obtenidos mediante este tipo de metodologías (paramétricas). En ciertas circunstancias, el método filogenético no siempre resulta adecuado para evaluar casos de transferencia horizontal.

Al final, concluimos que, dado que el fenómeno de la HGT es multifactorial, la detección de genes xenólogos necesita abordarse por distintos ángulos; y dado que experimentalmente ya se ha comprobado que lo mejor es emplear al menos dos metodologías que sean mutuamente compensatorias (las debilidades de una sean compensadas por las fortalezas de otra) a favor de detectar eventos de HGT de una forma precisa y poco susceptible a errores, nosotros proponemos que si bien el método filogenético es de gran utilidad, la correcta interpretación de los resultados obtenidos con éste es vital para no incurrir en errores cuestionables en el momento de postular explicaciones.

Introducción.

Con el advenimiento de las técnicas moleculares y con la consecuente capacidad de secuenciar y comparar genomas completos de una gran cantidad de organismos, se hizo posible construir filogenias más completas y confiables. Sin embargo, cuando un mayor número de filogenias fueron construidas y comparadas contra el árbol de la vida, se pudo apreciar que en la mayoría de los casos no hubo concordancia. Para explicar tales discordancias se invocaron procesos que ocurren naturalmente, como la recombinación ilegítima, la hibridación y la transferencia horizontal de genes (Woese, 2000).

Dado que este tipo de mecanismos facilitan la adaptación a ambientes altamente cambiantes en periodos de tiempo relativamente cortos, hay una gran cantidad de organismos que en repetidas ocasiones se han visto favorecidos; inclusive aquellos que no son tan promiscuos en el libre intercambio de material genético. El ejemplo más controvertido, debatido e investigado, es el descubrimiento de la transferencia horizontal de genes en eucariontes, que se ha confirmado intercambian activamente material genético con bacterias, e inclusive con otros eucariontes (Syvanen, 1985; Gogarten 2003; Boto, 2010; Holmgren, 2010; Andersson, 2005).

Marco histórico e identificación del problema con la HGT.

Con el famoso experimento de Frederick Griffith en 1928 se comenzó a descubrir mecanismos de intercambio genético que eventualmente llevarían a Joshua Lederberg y colaboradores al descubrimiento de la conjugación y la transducción, se apreció por primera vez el tipo de intercambios genéticos llevados a cabo entre bacterias diferentes. Sin embargo. Es importante resaltar que en aquellos experimentos el mecanismo de transmisión era lo que estaba siendo investigado (y lo que resultaba novedoso), más no las implicaciones del material transferido (fenómeno de la HGT en sí).

El estudio de la importancia e implicaciones de la HGT no fueron realmente considerados sino hasta finales de los años 50, cuando el problema de la resistencia a antibióticos por parte de las bacterias empezó a manifestarse como una preocupación mayor en el tratamiento de algunas enfermedades. La velocidad con la que las bacterias adquirían resistencia a un amplio espectro de antibióticos comenzó a generar la idea de que posiblemente existía una forma de intercambio genético que difería de la clásica transmisión vertical (herencia) y de la aparición azarosa de mutaciones ventajosas (Ohno, 1999; Ohta, 1989). La posible causa fue descrita por dos científicos japoneses Tomoichiro Akiba y Kunitaro Ochia cuando de forma incidental se observó que una bacteria adquiría resistencia a un antibiótico (al cual era inicialmente vulnerable) si esta se ponía en contacto con otra que ya poseía resistencia al mismo. La HGT interespecies había sido descubierta (Ochia, 1959). De inmediato, la adquisición de resistencia a un amplio espectro de antibióticos por parte de un número creciente de bacterias en un periodo de tiempo relativamente corto fue acreditado a un proceso no vertical de transmisión de genes (Syvanen, 1985). Entonces, el fenómeno conocido como transferencia horizontal de genes cobró gran importancia, al menos, inicialmente, por sus implicaciones dentro de un contexto médico y social (Davies, 1996).

Con el propósito de abordar el problema, se realizaron investigaciones con el objetivo de tener un mejor conocimiento del proceso. Al comienzo, los resultados sólo alcanzaban para reconocer que esta novedosa forma de transferir material genético sería de una gran relevancia e importancia en el estudio de la resistencia farmacológica entre una gran variedad de bacterias; sin embargo, en aquellos momentos no se podía predecir el impacto que el fenómeno tendría en muchos aspectos de las ciencias biológicas.

De los primeros investigadores que comenzaron los estudios más a fondo, tenemos a Michael Syvanen, quien escribió una revisión de lo que hasta el momento se sabía de la HGT (Syvanen, 1985). Aunque en algunos casos la evidencia no era tan sólida y en otros tantos escasa, Syvanen exitosamente

extrapoló muchas ideas en base a fenómenos que ya eran bien conocidos y que a fin de cuentas involucraban un intercambio de genes. Así logró postular, aunque fuese de forma indirecta, muchas de las ideas que hoy son bien conocidas y características de la transferencia horizontal de genes (Syvanen *et al* 1988). Syvanen propuso entonces que resultaba muy evidente que la HGT facilitada por una gran cantidad de medios de transmisión (intercambio de plásmidos [Koonin, *et al* 2009], transducción y transformación [Ochman, *et al* 2000]), había tenido una gran influencia en el modelado de las relaciones filogenéticas. Él asumió esto en base a que, “considerar el intercambio genético entre especies que se pensaría normalmente no llevan a cabo tales actividades, es la única forma de darle una explicación razonable y altamente probable a muchos hechos que hasta el momento han tenido intrigados a los biólogos” (Syvanen, 1985). Syvanen, además, fue el primero en reconocer y acertar en que el fenómeno tendría un alcance considerable aun en las relaciones filogenéticas más insospechadas, por tanto podría presentarse en dominios más allá de proteobacteria.

Aunque en principio la mayoría de las investigaciones y los respectivos hallazgos fueron hechos en organismos procariontes, con el paso del tiempo y la disponibilidad de una cantidad mayor de genomas secuenciados, se realizaron estudios en organismos de otros grupos filogenéticos. Así fue como se logró reunir evidencia de la transferencia horizontal de genes en hongos, plantas y animales, siendo en los insectos en donde se ha encontrado la evidencia más convincente de HGT en organismos Eucariontes (Woolfit *et al* 2009).

Naturalmente, la idea de que la transferencia horizontal pudiese ocurrir en organismos que pertenecen a otros dominios resalta la duda de si existen factores ya sean fisicoquímicos, genéticos o ambientales que restringen o facilitan el intercambio genético. Como parte fundamental para entender el fenómeno de la HGT de una forma integral, se comenzaron a investigar el o los factores que facilitan este tipo de intercambios.

Mecanismos de intercambio genético.

A pesar de que el proceso por el cual un organismo toma un segmento de DNA foráneo y lo incorpora a su genoma no es completamente entendido (Thomas y Nielsen, 2005), la mayoría del conocimiento adquirido viene de extrapolaciones del estudio de todas las formas conocidas en las que se transfieren genes entre diferentes organismos.

A pesar de que la visualización directa del fenómeno, en teoría, no supondría un gran reto tecnológico, no fue sino hasta el 2008 cuando Babic y colaboradores lograron visualizar el evento *in vivo* y seguir el proceso completo desde la toma del plásmido por la célula, la subsecuente integración al cromosoma y finalizando con la expresión del gen transferido en el genoma receptor (Babic *et al* 2008).

Uno de los primeros trabajos en hacer una contribución importante para tratar de establecer los mecanismos que rigen las transferencias genéticas fue el de Jain y colaboradores (1999) en donde se establecieron aquellos factores que influyen de forma determinante para que una transferencia genética se lleve o no a cabo. Su trabajo confirmó la suposición hecha anteriormente por otros autores (Ochman *et al* 2000; Gogarten *et al* 2002; Doolittle, 1999) de que:

1) No todos los genes son igualmente propensos de ser transferidos, y 2) Las diferencias en las tasas evolutivas de los diferentes genes no son suficientemente significativos como para determinar o influir en la frecuencia de transmisión (como por ejemplo se suponía que podría ser el caso con genes de evolución rápida [Molloy, 2005]).

Según la propuesta de Jain y colaboradores (1999) los genes operacionales que son transferidos exitosamente poseen funciones de mantenimiento (metabolismo, estructura, etc.) que no pertenecen a grandes complejos sujetos a un estricto control genético; es decir, son estructuras modulares (Woese Kandler y Wheelis, 1990). La adquisición de este tipo de genes permitiría ganar acceso a nuevos tipos de fuentes de energía y nutrientes que les facilitaría la adaptabilidad a nuevos ambientes (Beiko, Harlow y Ragan, 2005). Por el contrario, los genes informacionales (genes para transcripción, traducción, etc.) son genes usualmente componentes de una compleja maquinaria celular y no son susceptibles a un simple reemplazo por unidades siquiera similares (Jain 1999; Woese, 2000). Así que, según esta propuesta, la función es la que dicta la frecuencia en la transferencia de ciertos genes, y más importante aún, la entrada y permanencia dentro del genoma receptor. Si un gen adquirido confiere al organismo receptor algún tipo de ventaja adaptativa, este muy probablemente sea incorporado al genoma. Este tipo de eventos en los que la presión de selección es favorable, aceleran el intercambio de material genético (Ragan *et al* 2009); así que fenómenos como la resistencia a antibióticos y la adquisición de grupos completos de genes, como por ejemplo, las secuencias conocidas como islas de patogenicidad, que confieren a las células nuevas capacidades infecciosas, se cree fueron adquiridas de esta manera (Waack *et al* 2006).

Cuando los primeros genomas completos fueron accesibles en bases de datos públicas, se emplearon para tratar de encontrar evidencia filogenética de este fenómeno. En la mayoría de los casos, el propósito de tales esfuerzos era dar explicaciones a las cuestiones más básicas acerca de la HGT.

En este sentido, los objetivos primarios eran: a) Detectar e identificar eventos de HGT. b) Detectar aquellos factores que aumentan la frecuencia de transmisión de determinados factores genéticos. c) Determinar qué tipo de consecuencias conllevan para las células receptoras realizar este tipo de intercambios genéticos, y d) Estimar la frecuencia con la que este tipo de intercambios ocurren (Ochman *et al* 2000).

Métodos para la detección e identificación de genes xenólogos.

Antes del artículo de Lawrence y Ochman (1997) en el que proponían una metodología completa para la detección de genes xenólogos y establecer un parámetro de la frecuencia de HGT en la evolución de las bacterias, muchos autores ya habían tratado de estimar esa frecuencia en base a extrapolaciones de otras investigaciones (Syvanen, 1985; Gogarten, 1993; Woese 1990). En ese tiempo la detección de eventos de HGT se hacía mediante la búsqueda de patrones intergénicos que indicaran la presencia de una secuencia transferida. Fue así que algunos autores propondrían que secuencias flanqueadas por integrasas de fago o regiones de integración con motivos, podrían ser consideradas como evidencias indirectas pero precisas de que genes rodeados por este tipo de secuencias anómalas seguramente serían el resultado de

HGT. La búsqueda de este tipo de patrones fueron los primeros intentos para correlacionar los eventos de HGT con firmas genómicas más simples de detectar y que a su vez fueran un reflejo o un indicador de HGT entre dos o más organismos.

Determinar la proporción de transferencias horizontales de genes es vital para poder comprender el impacto que este fenómeno tiene en la evolución de los organismos, y uno de los pasos más importantes encaminados en esa dirección es desarrollar un método/algoritmo de detección confiable.

Como veremos más adelante, algunos autores (Becq *et al* 2010; Gogarten 2002; Nakamura, 2004; Poptsova, 2007) señalan que la combinación de metodologías de detección es la mejor opción para hacer frente a este problema; sin embargo, tan sólo unos años atrás, la mayoría de las metodologías buscaban trabajar como un todo, sin ser dependientes una de otra.

Así pues, métodos como el contenido porcentual de G-C en un genoma, los patrones de utilización de condones y las frecuencias de di y trinucleótidos son metodologías de detección que eventualmente serían confirmadas como poco eficientes y que fallan de forma importante en la detección de transferencias antiguas y entre organismos muy cercanos (Cortéz *et al* 2005; Guindon y Perrière, 2001; Nakamura *et al* 2004; Deschavanne y Filipiski, 1995) condiciones bajo las cuales resulta imposible determinar el donante de las secuencias transferidas (Kanhare y Vingron 2009). Sin embargo, serían los primeros intentos serios para poder detectar y cuantificar la frecuencia con la ocurre la HGT.

En el año 2005, Cortéz y colaboradores hicieron una comparación de las metodologías hasta entonces más populares para detectar la transferencia horizontal de genes. Las distintas metodologías fueron categorizadas de acuerdo al parámetro de búsqueda. La primera, basado en contenido genómico, como el contenido porcentual de G-C de un genoma y el índice de adaptación de codones (CAI) (Smith, Feng y Doolittle, 1992). La segunda, basados en cercanía filogenética, como el método de la distribución de perfiles (DP) para cada gen dentro de un genoma. Y la tercer categoría, basada en metodologías cuyos resultados son sujetos a escrutinio matemático/estadístico; como los modelos basados en estadística Bayesiana y los basados en cadenas de Markov de primer orden.

El contenido de G-C y A-T de un genoma está determinado por mutaciones y presiones de selección característicos de cada especie (Daubin *et al* 2003a). Explotando esta cualidad yace detrás la idea de las metodologías basadas en la composición genómica como parámetro para determinar si un gen es o no xenólogo (Guindon *et al* 2001). Desafortunadamente este criterio se convirtió en más una excepción que una regla. Estudios posteriores mostraron que la distribución desigual en el contenido de G-C y A-T de los genomas, comprometían demasiado los resultados obtenidos por medio de éstas metodologías (Guindon *et al* 2001; Deschavanne y Filipiski, 1995). Por tanto, poco a poco, estas metodologías dejaron de emplearse.

Con el propósito de explicar las anomalías que presentaban algunos genes incorporados en los genomas de organismos receptores, Lawrence y Ochman (1997) postularon que los genes transferidos que originalmente poseían

características propias (como cierta preferencia por codones, mutaciones en 3ra posición, etc.), adoptan características del organismo receptor, lo que eventualmente les hace indistinguibles de los demás genes. De esta manera, ésta “adaptación” al genoma residente de un gen o una secuencia transferida nos impediría identificarlo como foráneo si ya ha pasado el tiempo suficiente para que adquiriera las características del genoma hospedero desde su integración a éste. Lawrence y Ochman (1997) denominaron a éste fenómeno como “*amelioration*”, sin embargo, en este trabajo usamos la palabra homogenización en su lugar.

En muchas ocasiones, los datos obtenidos han servido primordialmente sólo como una aproximación estadística; así que bajo tales circunstancias, un acercamiento comparativo, es decir, revisar la congruencia de los resultados obtenidos con la bibliografía que ya exista respecto a determinado tema, para así complementar y dar robustez a los resultados obtenidos, necesitaba adoptarse. De esta manera es más rápido y fácil determinar si los resultados entran en un rango de confianza aceptable y mejor aun, ampliar el espectro de detección del fenómeno para poder identificar eventos de transferencia que antes eran omitidos por completo, ya sea como resultado de artefactos de las metodologías o por ser erróneamente confundidos como falsos positivos.

Aunque la detección de HGT por variaciones genóticas resulta poco confiable, los resultados obtenidos por Cortéz *et al*, demostraron que cuando los genes analizados por esta metodología eran el resultado de transferencias entre especies muy divergentes o que cuando los mismos pertenecen a genomas cuyo contenido medio porcentual de G-C sea mucho menor a 50.8% (valor promedio en contenido de G-C para *E. coli* K12 MG1655) la capacidad de detección aumentaba dramáticamente. Aunque como Daubin y Perrière (2003a) sugieren, la interpretación que se le dé a los resultados obtenidos por medio de metodologías composicionales debe ser muy cuidadosa, ya que no se puede suponer que se trata de evidencia que corrobore la existencia de transferencias entre organismos distantes. El estudio detallado de la composición de los genomas y de las firmas genómicas nos ha permitido identificar la ahora reconocida heterogeneidad de los genomas, esto es que los genes pueden estar constituidos de forma diferente dependiendo de la posición que ocupen en el genoma (Daubin y Perrière, 2003b). Postulados recientes sugieren incluso que hay sustituciones intragénicas (dentro de un mismo gen) resultado de HGT (Ragan y Beiko 2009).

Aunque descubrimientos de ésta naturaleza ponen en duda la efectividad del método composicional, aun es posible hacer reconstrucciones filogenéticas confiables si cualquier factor de ruido es considerado y adecuadamente controlado.

Los métodos filogenéticos, a pesar de ser distorsionados por la presencia de HGT, son reconocidos como los mejores para detectar transferencias horizontales (Gogarten y Townsend, 2005). Si se hacen filogenias para genes individuales es fácil detectar cuando se presentan casos discordantes; el problema con estas metodologías es que requieren una gran cantidad de información de las especies analizadas y los estudios deben de ser a una escala considerable, pues durante las reconstrucciones filogenéticas la ausencia de especies cercanas compromete mucho los resultados y aumenta la posibilidad de que artulugios propios de las metodologías interfieran en ellos (Koski, Morton y Golding, 2001).

En el artículo de Cortéz y colaboradores se puso a prueba un método basado en reconstrucciones filogenéticas. El método de los perfiles distribucionales o DP; que funciona mediante la comparación filogenética de genomas cercanos, arrojó resultados similares a los obtenidos con el procedimiento de G-C y CAI, siendo la única ventaja notable de éste método su capacidad para descartar una cantidad mayor de falsos positivos. Una de las razones que podría ser causa para que estas metodologías tuviesen un nivel de detección comparable a las metodologías antes mencionadas, es que resulta complicado poder diferenciar entre ganancia (posibles transferencias horizontales) o pérdida de genes (Gogarten y Townsend, 2005; Smith, Feng y Doolittle, 1992); transferencias verticales mimetizadas (Kyrpides y Olsen, 1999) y que en muchas ocasiones, la falta de ortólogos apropiados para hacer las comparaciones (Poptsova y Gogarten, 2007) limitan el alcance de las metodologías filogenéticas. Y, aunque suele considerarse que lo más parsimonioso es lo más probable (Koski, Morton y Golding, 2001) en ocasiones esto influye hacia la detección de falsos positivos entre especies muy relacionadas.

A favor de rescatar las ventajas, Koski y colaboradores (2001), sugieren que metodologías basadas en caracteres composicionales podrían ser empleadas frecuentemente si son complementadas por otras que descarten la probabilidad de confundir falsos positivos como resultados reales. En su trabajo, Koski y colaboradores probó una metodología que combinaba similitud de secuencias y la correspondencia de los elementos de éstas con su posición en el cromosoma, lo que le evitaba detectar como transferencias, aquellos genes que, a pesar de poseer composiciones anómalas, hayan sido transferidos verticalmente (falsos negativos). Según la metodología de Koski y colaboradores, genes transferidos verticalmente se encontrarían en posiciones homólogas (dentro del cromosoma) en organismos estrechamente relacionados, mientras que los que sean resultado de HGT carecerán de una posición correspondiente (un ortólogo posicional) (Smith, Feng y Doolittle 1992).

Por supuesto, eventos como transposiciones, duplicaciones e inversiones complicarían tales comparaciones (Koski *et al* 2001); lo que eventualmente descartaría la verificación de datos a partir de métodos composicionales, por lo que la validación matemática/estadística de los resultados se volvió el paso natural en el desarrollo de metodologías de detección.

Así que, a partir de una caracterización de la información que se podía obtener de analizar el contenido de diferentes propiedades fisicoquímicas del genoma, se emplearon métodos probabilísticos basados en cálculos con estadística Bayesiana y cadenas de Markov, cuyas propiedades (ver más abajo) resultan ideales para lidiar con este tipo de problemas.

Entre los primeros en hacerlo estuvieron Nakamura y colaboradores en el año 2004; quienes desarrollaron una metodología para detectar transferencia horizontal de genes bajo un fuerte sustento probabilístico.

Empleando una categorización a partir del contenido génico de los organismos en estudio, Nakamura usó un algoritmo que se basa en inferencia Bayesiana y que originalmente fue concebido para detectar genes dentro de ORFS; así calcula la probabilidad media posterior (o índice de transferencia horizontal) para cada gen analizado. Sin embargo, Nakamura y colaboradores advierten que su método tiene la importante restricción de que podría detectar

únicamente eventos de transferencia horizontal reciente, y por tanto el índice de transferencia podría estar subestimado. Esto ocurre porque para la estimación de los parámetros estadísticos fue necesario hacer una categorización de las anomalías composicionales de los genomas analizados (al igual que con el método de estimación porcentual de G-C), así que el resultado de esa categorización sería fuertemente influenciada si los genes detectados ya han adoptado las características de sus nuevos genomas residentes (Lawrence y Ochman, 1997).

En su artículo de comparación de metodologías, Cortéz y colaboradores hacen una evaluación entre la metodología basada en probabilidad Bayesiana y la que se basa en Markov. Dado que una de las propiedades más importantes de las cadenas de Markov es que el estado futuro sólo depende del estado inmediatamente anterior (Metropolis, 1987), es factible entender la capacidad de aplicación de estos algoritmos a los problemas de HGT. Con anterioridad ya se habían empleado las Cadenas de Markov de primer orden y las simulaciones de Montecarlo para enfrentar cuestiones similares. (Hayes y Borodovsky, 1998;). Aunque ambas presentan valores de detección similares, la metodología basada en Cadenas de Markov de primer orden, presentan un muy buen nivel de detección cuando las transferencias son entre organismos cercanos (que es uno de los principales problemas con las demás metodologías); y al mismo tiempo, presenta el mejor nivel de detección cuando las transferencias entre organismos filogenéticamente distantes (Cortéz *et al* 2005).

A pesar de que los modelos basados en cadenas de Markov son altamente eficientes, no son perfectos. Una de las más grandes restricciones que poseen, radica en la dependencia de la calidad de la información con la que se alimenta al modelo. Markov puede arrojar una gran cantidad de falsos positivos si no se realiza una concatenación (categorización de acuerdo a un criterio previamente establecido) de la información a procesar, lo cual dificulta los análisis en genomas completos. La segunda restricción más importante es que si el margen de estimación estadística se amplía ($p < 0.03 - 0.05$), los falsos positivos aumentan en la misma proporción -que Cortéz propone son detecciones de pseudogenes-. Según el trabajo de Liu y colaboradores los pseudogenes se encuentran en proporciones mayores a las esperadas en genomas procariontes, muchos de ellos siendo el resultado directo de transferencias horizontales fallidas (Liu *et al* 2004).

Comprobando la eficacia de otras metodologías.

Siguiendo un esquema similar, Becq y colaboradores (2010), recientemente llevaron a cabo un exhaustivo estudio comparativo entre metodologías de detección de genes xenólogos. Bajo el supuesto de que la mayoría de estas metodologías no han sido correctamente probadas antes de ser promovidos como metodologías de detección confiables; más de 20 de ellas fueron analizadas con el propósito de averiguar la eficacia de estas bajo un escenario ideal.

Utilizando genomas artificiales (genomas a los que les fue removida cualquier región atípica), cuya principal ventaja era que las secuencias estaban libres de genes homogeneizados (refiriéndonos al proceso descrito como “*amelioration*” por Lawrence y Ochman [1997]) así como de parálogos que ocasionaran ruido filogenético significativo.

En el estudio se tomaron en cuenta ciertos parámetros que pueden influir en la detección de transferencias, y de esta manera se optimizó el umbral de detección para cada una de las metodologías. Parámetros como el origen de la secuencia donada, la cantidad (número de donadores para un solo genoma), tamaño, influencia del genoma residente (si un mismo gen se detectaba siempre con la misma frecuencia sin importar que se encontrase en *E. coli* K12 o en *S. sonnei*) y finalmente, la resolución cuando se aplicaba a un genoma real, que en este caso fue el de *E. coli* K12 MG1655.

Como resultados importantes, ellos señalan que encontraron metodologías que ni en condiciones ideales presentan buen nivel de detección, así que la probabilidad de que se desempeñaran apropiadamente en casos reales es muy baja. Además, mostraron que muchas metodologías composicionales fallan en detectar DNA foráneo y otras, basadas en algoritmos matemáticos, son incapaces de diferenciar entre DNA nativo y adquirido.

Según Becq y colaboradores, evaluar transferencias con base en el uso de oligonucleótidos (que son aquellas que identifican los cambios en la dinámica de sustitución de oligonucleótidos) es la mejor opción, ya que este tipo de metodologías son menos sensibles a parámetros que afectan mucho a otras. La única susceptibilidad encontrada parece radicar en el tamaño de los posibles xenólogos, que debido a las desviaciones estadísticas producto de análisis en muestreos pequeños, pueden arrojar resultados erróneos. Proponen además, que la mejor forma de detectar HGT es emplear al menos dos metodologías cuyas ventajas contrarresten mutuamente las dificultades encontradas en determinado caso (Becq *et al* 2010).

Todas las metodologías examinadas hasta ahora comparten la necesidad de hacer una clasificación previa de la información con la que se trabaja; las diferencias radican en la forma en la que se analizan estas categorías de información. Los métodos composicionales apoyados por algún tipo de análisis de confianza resultan ser apropiadas en la detección de genes xenólogos. Este tipo de metodologías, que no se basan en la construcción de árboles filogenéticos para comprobar los resultados, tienen las ventajas de ser rápidas (dependiendo del nivel de robustez que se quiera en el estudio), sencillas, accesibles en tiempo de cómputo y confiables.

Sin embargo, la creciente tendencia a recurrir a las reconstrucciones filogenéticas refuerza la postura de la mayoría de los autores (Gogarten, Doolittle y Lawrence, 2002) que la reconocen como la metodología más confiable y precisa; aunque requiere de tiempo, un acervo informativo importante (información extensiva de los *taxa* a analizar) y en ocasiones resultan computacionalmente restrictivas si el estudio es a gran escala (Gogarten y Townsend, 2005; Lawrence y Ochman, 1997; Gogarten *et al* 2002; Cortéz *et al* 2005; Doolittle, 1999; Guindon y Perrière, 2001).

La mayoría de las restricciones con el método filogenético no suelen ser impedimentos importantes; sin embargo, es necesario destacar que algunas de las críticas hechas al método ni siquiera están directamente relacionados con el mismo (Becq, 2010). La primera y posiblemente la más compleja, hace referencia a que en las bases de datos actuales no se cuenta con la información necesaria (secuencias ortólogas) para inferir filogenias confiables; y la segunda radica en que las reconstrucciones filogenéticas usualmente arrojan más de una posible solución, así que en muchas ocasiones, la

selección del árbol verdadero o en su defecto, el que represente mejor lo que posiblemente sucedió es un problema complejo de resolver, no solo por la cantidad de trabajo que requiere, sino porque, a pesar del crecimiento exponencial de las bases de datos, aun hay carencia de genes ortólogos comparables en 50% de los genomas secuenciados (Poptsova, 2007).

Obstáculos como el anterior en ocasiones dificultan la detección de genes xenólogos y comprometen los resultados, por lo que la incongruencia filogenética (discordancia filogenética entre el gen y el agrupamiento que resulta del mismo) sigue siendo el indicador más confiable de que estamos ante un evento de HGT (Cortéz *et al* 2005; Keeling y Palmer, 2008); y aun cuando corroborarlo más allá de toda duda es una tarea compleja, en la mayoría de los casos la HGT es la forma más plausible y lógica de explicar las discordancias filogenéticas más contradictorias.

Consideraciones importantes con el problema de la HGT.

Hasta la identificación del fenómeno de la transferencia horizontal de genes, el proceso que se pensaba tenía un peso fundamental y modelador en las relaciones filogenéticas de todos los organismos conocidos hasta el momento era la herencia; la cual, implica una transmisión vertical de caracteres benéficos adquiridos. Cuando la HGT fue documentada, fue considerada como un posible mecanismo por el cual un organismo podía adaptarse de forma muy rápida a un entorno cambiante sin necesidad la dependencia de las mutaciones azarosas que son fijadas en la población a través de un largo periodo de tiempo. Así fue que los biólogos comenzaron a enfrentar un problema que rápidamente se convirtió en una de las más grandes cuestiones aún por resolver de la biología contemporánea.

A finales de los años 70, Carl Woese y George Fox construyeron el que posteriormente sería denominado como “el árbol de la vida”, con base en la subunidad pequeña del rRNA (16s rRNA) (Woese y Fox, 1977).

Cuando la comunidad científica comparó una gran cantidad de árboles filogenéticos utilizando diferentes genes como marcadores, fue evidente que en muchas ocasiones el resultado dependía de las moléculas/genes que se emplearan para construirlas (Rivera, 1998); pero el desconcierto fue aún mayor cuando se encontraron casos en los que la discordancia filogenética se presentaba aun en aquellas filogenias construidas con base en genes altamente conservados (Gogarten y Townsend, 2005; Woese, 2000).

Como resultado, la comunidad científica comenzó a preguntarse si la transferencia horizontal y fenómenos análogos (ver más adelante) no representarían un obstáculo para llevar a cabo futuras reconstrucciones filogenéticas, ya que hasta entonces, procesos reconocidos por alterar el resultado esperado de las mismas (como la evolución reticulada, el mutualismo y la recombinación ilegítima) se presentaban con una frecuencia tan baja que sólo en casos muy especiales resultaban de consideración (Woese, 2000). Se sabe que, por ejemplo, en el estudio de ciertos clados del reino *plantae*, la evolución reticulada y el mutualismo son fenómenos frecuentes cuya influencia ha resultado modeladora en el curso de la evolución de estos organismos (Woese, Kandler y Wheelis, 1990; Brown y Doolittle, 1997; Pennisi, 1998).

Como la teoría evolutiva clásica no contempla métodos de innovación genética que no sean resultado de la transmisión vertical entre ancestros y descendientes ni de procesos no graduales independientes de un intervalo temporal para fijarse en una población; se comenzaron a generar ciertas dudas respecto a que la teoría evolutiva clásica proveía de una explicación muy limitada para la evolución de los micro-organismos (Olendzenski y Gogarten 2009).

Antes del descubrimiento de las implicaciones y alcances de la HGT, la teoría evolutiva era una generalización del proceso evolutivo que se había estudiado en macro-organismos, por lo que evidentemente, la mayoría de los microorganismos y los mecanismos de innovación genética que en ellos resultaban recurrentes, no habían sido considerados.

A mediados del año 2000, la cantidad de genomas completos publicados estaba creciendo exponencialmente, y fue así que algunos genomas eucariontes estuvieron disponibles para analizarse. En algunas ocasiones con la intención de buscar HGT y en otras como meros incidentes, se comenzaba a obtener evidencia que corroboraba que la HGT ocurría en Eucariontes, aunque los protistas eran el ejemplo mejor documentado (Woolfit, 2009), pronto se reunirían evidencias sólidas para casos ocurridos en Plantas (Kim E. *et al* 2010), Hongos (Kondo N. 2002) e Insectos (Fitzpatrick *et al* 2008); lo que finalmente descartaría la idea de que la HGT es un fenómeno aislado que ocurre sólo en proteobacterias.

Toda esta serie de circunstancias dio pie a las proposiciones de que la HGT, principalmente en las ramas basales de la clasificación biológica y en las etapas en las que todas las células eran modulares (sus componentes podían ser intercambiados por otros sin sufrir efectos deletéreos) (Woese, 2000) era de tal magnitud, que muy probablemente resultaría complicado determinar con un buen nivel de confianza las cualidades genéticas de los ancestros de los organismos que vemos y conocemos hoy en día (Ver Figura 1). Es posible que las comunidades de células primigenias que dieron origen a todos los organismos actuales, no contuvieran todas las características que hoy consideramos como universales (ribosomas, RNA's mensajeros, DNA, etc.), y suponemos debieron estar presentes en *LUCA* (de las siglas en inglés, *Last Universal Common Ancestor*) (Kyrpides *et al* 1999; Becerra *et al* 2007; Lazcano y Forterre, 1999), pudieron de hecho, ser el resultado de una serie de transferencias horizontales ancestrales y/o de evolución convergente (Ver Figura 1) (Olendzenski y Gogarten, 2009).

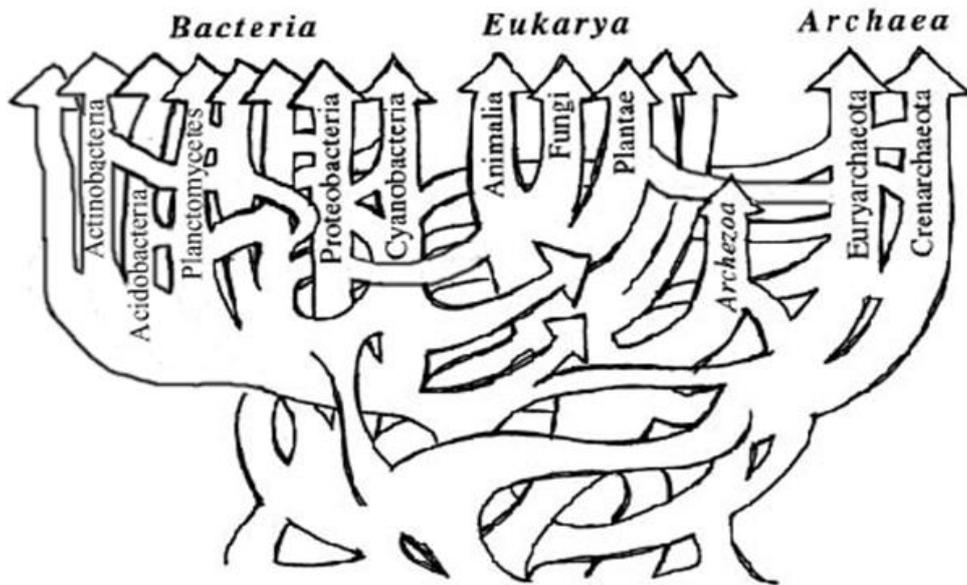


Figura 1. Esta imagen, modificada de Doolittle W.F. (1999), es una representación de los tres dominios. La gran cantidad de ramas simbolizan el intercambio genético (HGT) que hubo entre algunos de los miembros de estos dominios. Se cree que las células de las que descienden todos los organismos en la actualidad (representadas en la parte basal de la imagen) no eran entidades únicas como tal y por ende funcionaban como vectores en una red de HGT.

A finales de los años 90, W. Ford Doolittle y Peter J. Gogarten comenzaron a difundir fuertemente la idea de que el árbol de la vida era una interpretación conveniente pero demasiado simplista, irreal e impráctica para determinar las relaciones filogenéticas entre las especies (Ochman, 2000). Peter Gogarten postuló que los genomas se pueden considerar como un conjunto de mosaicos, como tal, diferentes partes del mismo cuentan diferentes historias evolutivas (Olendzenski y Gogarten, 2009), así que la HGT es la forma más factible de explicar las fuertes discordancias filogenéticas que se habían encontrado para una gran cantidad de moléculas y genes, entre ellas mismas y con el árbol de RNA (Gogarten, 2003).

Ambos proponen que una descripción gráfica más apropiada para la representación de las relaciones filogenéticas de los seres vivos sería la de una red o redes en la que los diferentes nodos reflejen de forma precisa las conexiones que existen o existieron entre ellos (Gogarten, 2003; Olendzenski y Gogarten, 2009); y aunque recientes esfuerzos por construir historias evolutivas reticuladas aún están en fases iniciales, hay opiniones y evidencias a favor de adoptar este tipo de construcciones filogenéticas, ya que nuevos algoritmos de han sido desarrollados para contrarrestar la dificultad que solía implicar la construcción una filogenia reticulada (Van Iersel L. *et al* 2010).

A pesar de ello, las representaciones reticuladas de evolución no han sido adoptadas porque la cantidad de estudios filogenéticos robustos que aportan evidencia contundente a favor de la viabilidad de las filogenias a pesar de la HGT se sigue incrementando (Lerat *et al* 2005), W. F. Doolittle propone que en el peor de los casos, resultaría imposible detectar con precisión las “huellas” dejadas por la HGT en los diferentes genomas procariontes, aun considerando que en algunos casos la influencia es más notoria que en otros (Doolittle 1999).

Aunque existen posturas que minimizan y en otras ocasiones exageran (Ver Figura 2) el papel de la transferencia horizontal en la evolución de los organismos (Kurland *et al* 2003); la mayoría de los autores concuerda en que existen tendencias evolutivas que resultan distinguibles (bajo cualquier frecuencia de HGT) y que podrían ser apropiadamente representadas por un árbol. La transmisión vertical es un mecanismo que al igual que la HGT, participa en la evolución de los organismos, y aunque es posible que la frecuencia de ambos sea radicalmente distinta, son procesos conjuntos y no excluyentes (Beiko *et al* 2005)

Con anterioridad ya hemos resaltado que es muy probable que las incongruencias en la identificación de los fenómenos de HGT se deba simplemente a la incapacidad que existe para detectarla con precisión y confianza (Koski *et al* 2001); por este motivo, corroborar que el método filogenético es la forma más imparcial en la identificación y detección de eventos de transferencia horizontal es el objetivo de este trabajo.

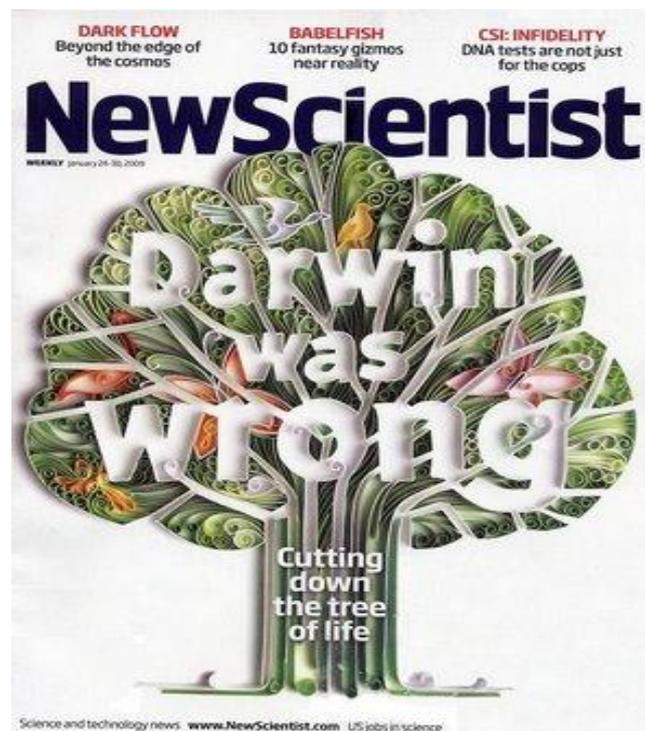


Figura 2. El descubrimiento de que ciertos patrones evolutivos no son apropiadamente representados por la metáfora del árbol de la vida, llevó a la declaración de que algunos conceptos fundamentalmente darwinistas dentro de la teoría evolutiva podrían estar equivocados. (Graham Lawton. "Darwin was wrong". 2009, Enero. New Scientist. Número 2692.)

Objetivos.

- 1) Comparar los resultados obtenidos en la identificación de eventos de HGT en el genoma de *Escherichia Coli K12 MG1655* por el método composicional en base a los obtenidos haciendo uso del método filogenético.
- 2) Determinar así, la eficiencia y particularidades de ambas metodologías en la detección de genes xenólogos.

Metodología.

En el año 2005, Cortéz y colaboradores publicaron en el International Journal for Computational Biology (*in silico Biology*) un estudio comparativo en el que hacían un análisis detallado del desempeño de las metodologías más populares hasta entonces disponibles para la detección de genes xenólogos.

Fue así que buscando e identificando genes resultado de HGT en el genoma de *Escherichia Coli K12 MG1655*, Cortéz y colaboradores reportaron 140 genes xenólogos. Nosotros, de los 140 genes reportados como xenólogos por Cortéz, extrajimos las secuencias para 109 de éstos genes. No seleccionamos los 140 porque algunos de esos genes estaban clasificados como hipotéticos de acuerdo a la clasificación del Kegg. Las secuencias de todos los genes fueron obtenidas del sitio FTP de la base de datos de la “Kyoto Encyclopedia of Genes and Genomes”, (<http://www.genome.jp/kegg/>).

Las secuencias de éstos genes fueron comparadas mediante un análisis por BLAST (con parámetros en *default* para trabajar con proteínas), programa bioinformático que busca similitud entre secuencias biológicas, de secuencias codificantes (proteínas) (Altschul *et al* 1990; Johnson *et al* 2008) contra los 15 organismos seleccionados para este estudio. Esto con el propósito de encontrar genes ortólogos, con los cuales se construirían filogenias para cada gen comparado.

Los 15 organismos seleccionados para este estudio pertenecen al Phylum Proteobacteria. 14 son de la Clase Gamma Proteobacteria y uno es de la Clase Epsilon Proteobacteria, Género: *Helicobacter*; *Helicobacter Pylori* 26695. (Ver Tabla 1). Estos organismos fueron seleccionados con base en la información filogenética de estos es confiable y a la existencia de estudios comparativos en los que se establece un parámetro para la frecuencia con la que intercambian secuencias de forma horizontal, por tanto, nuestros resultados pueden ser cuantitativamente comparados con los previamente reportados en caso de ser necesario (Guindon y Perrière, 2001; Nakamura, 2004; Liu, 2004; Wu, 2009).

Tabla 1. Los 15 organismos seleccionados para el estudio. Las secuencias codificantes de estos organismos así como la clasificación en base a 3 letras fueron obtenidas del KEGG.

Organismos	Código de 3 letras del Kegg
<i>Escherichia coli K-12 MG1655</i>	eco
<i>Escherichia coli O127:H6 E2348C_0177</i>	ecg
<i>Escherichia fergusonii ATCC 35469</i>	efe
<i>Salmonella subsp. Choleraesuis SC-B67</i>	sec

<i>Salmonella subsp. Typhimurium</i> LT2	stm
<i>Shigella flexneri</i> 301 (serotype 2a)	sfl
<i>Shigella sonnei</i> Ss046	ssn
<i>Enterobacter sakazakii</i> ATCC BAA-894*	esa
<i>Cronobacter turicensis</i> 3032	ctu
<i>Erwinia tasmaniensis</i> Et1/99	eta
<i>Yersinia pestis</i> CO92 (biovar Orientalis)	ype
<i>Vibrio cholerae</i> O1 N1696	vch
<i>Vibrio parahaemolyticus</i> 2210633	vpa
<i>Photobacterium profundum</i> SS9	ppr
<i>Helicobacter pylori</i> 26695	hpy

*En la base de datos del NCBI se agrupa dentro del género *Cronobacter*, por lo que su nombre cambia a *Cronobacter sakazakii*.

Con un BLAST bidireccional (Altschul *et al* 1990; Johnson *et al* 2008) de secuencias codificantes (proteínas) se hizo una búsqueda de genes ortólogos para los xenólogos reportados por Cortéz y colaboradores en los 15 organismos que seleccionamos. De acuerdo a parámetros cuantitativos del BLAST se decidió cuáles secuencias podrían ser ortólogas y por tanto fueron seleccionadas para la comparación; excepto para los 7 genes huérfanos que se identificarían. Algunas de las secuencias fueron curadas a mano (excluidas manualmente de secuencias que pudiesen introducir “ruido filogenético”) a fin de evitar la presencia de genes parálogos, que son inadecuados en la construcción de filogenias. Al mismo tiempo se evita sesgar la selección de las secuencias útiles que eventualmente se alinearían.

Haciendo uso de la paquetería de programas disponibles dentro del sitio de NCBI se empleó el programa PSI-BLAST (<http://www.ebi.ac.uk/Tools/sss/psiblast>) que se especializa en búsquedas de posibles genes homólogos en organismos filogenéticamente distantes (Altschul *et al* 1997), lo que facilitaría limpiar las secuencias obtenidas de parálogos que pudiesen introducir falsos negativos.

Ya que se contaba con todos los resultados del BLAST de proteínas y que las secuencias estaban libres de elementos que pudiesen introducir ruido, estas serían eventualmente alineadas usando *ClustalW* (Larkin *et al* 2007) y *Muscle* (Edgar, 2004). Ambos programas fueron empleados en sus parámetros *default*. Las alineaciones resultantes serían empleadas en la construcción de filogenias para los 15 organismos que seleccionamos. En esta labor se decidió utilizar la paquetería PHYLIP (ver 3.69) (Felsenstein, 2003) que es confiable, sencilla de implementar e incluye todos los programas necesarios para realizar el trabajo en un solo paquete. La construcción de los fenogramas se hizo con el método de Neighbor Joining (NJ). Se optó por usar el método de Neighbor Joining ya que es altamente eficiente, práctico, rápido, confiable, funciona bien cuando se trabaja con proteínas y la metodología no es susceptible a “desviarse” cuando se trabaja con un número pequeño *otus*.

Además, como una medida de confianza en los resultados, se aplicó un *Bootstrap* de 100 replicas en la construcción de cada fenograma.

El *bootstrapping* es un procedimiento muy empleado en la construcción de filogenias mediante el cual N número de caracteres de un conjunto de información son re-muestreados al azar. El resultado de esto es que algunos caracteres son eliminados y otros son duplicados. El *bootstrap* se emplea con frecuencia para fortalecer los análisis filogenéticos hechos con Neighbor Joining, aunque no es la única metodología a la que puede ser aplicada.

El *bootstrapping* fue realizado mediante el programa *seqboot* incluido en la paquetería de herramientas bioinformáticas PHYLIP (Felsenstein, 2003). Todos los parámetros de *seqboot* fueron mantenidos en sus valores *default*.

En la construcción de filogenias, métodos como el de Máxima Parsimonia (MP), y el de Máxima Verosimilitud (MV), gozan de mucha popularidad, aunque ambos tienen limitaciones que eventualmente nos hicieron optar por el método de Neighbor Joining.

El método de máxima parsimonia es usualmente reconocido como uno de los métodos que carecen de precisión cuando construyen el árbol verdadero o más parsimonioso. Las tasas de variabilidad de las secuencias suelen ser un factor que impacta de forma muy negativa este tipo de metodologías. En la actualidad métodos que funcionan bajo un supuesto de Parsimonia suelen ser el último recurso, en su defecto, metodologías basadas en inferencias estadísticas, como Máxima Verosimilitud, suelen ser empleadas con mayor frecuencia, ya que son más poderosas y consistentes. Sin embargo, en el caso de Máxima Verosimilitud, problemas en cuanto a la dificultad de implementación, la necesidad de considerar una mayor cantidad de parámetros en función de la cantidad de información, y que por ser un método exhaustivo, suele usarse para analizar nucleótidos, no para proteínas (por la cantidad de combinatorias que existen para 4 nucleótidos en comparación con 20 aminoácidos).

Testigos.

Para poder establecer un parámetro confiable en la detección de los genes xenólogos en las filogenias obtenidas, se buscó en la literatura publicada una filogenia completa para Gamma-Proteobacteria. La filogenia de Yarza P. y colaboradores (2008), que representa las relaciones filogenéticas de todos los organismos conocidos dentro de los Dominios Arquea y Bacteria se utilizó como parámetro en el momento de determinar posibles transferencias horizontales, sin embargo, dado el tamaño del árbol, la labor de comparación se complicaba, por lo que existía un gran riesgo de introducir un sesgo observacional importante y poco deseado.

Con la intención de evitar esta cuestión, se construyó el árbol filogenético de los 15 organismos para comparación. La secuencia del 16S rRNA de cada uno de los 15 organismos considerado en nuestro trabajo se obtuvo del sitio web "*The Ribosomal Data Project*" (Cole *et al* 2009).

El árbol de RNA que incluye únicamente a los 15 organismos que deseábamos comparar fue utilizado como un parámetro primario en nuestro estudio, sin embargo, se decidió que se construirían otros 5 árboles testigos que fueran conservados y ampliamente distribuidos en Proteobacteria (Tabla 2). El trabajo de Wu y colaboradores en el 2009 sirvió como una referencia en la selección de los genes, ya que ellos reportaron una lista de 36 genes como marcadores confiables para la reconstrucción de filogenias de Arquea y Bacteria. Las secuencias de éstos genes fueron obtenidas de la base de datos del GenBank dentro del sitio del NCBI, <http://www.ncbi.nlm.nih.gov/>.

Tabla 2. Los genes cuya filogenia serviría como referencia o testigo en las comparaciones visuales de los fenogramas.

Genes Testigo
<i>rRNA - 16S ribosomal RNA</i>
<i>frr - ribosome recycling factor</i>
<i>rplF - 50S ribosomal protein L6</i>
<i>rpsB - 30S ribosomal protein S2</i>
<i>dnaG - DNA primase</i>
<i>pyrG - CTP sintetase</i>

Las filogenias de los genes analizados se compararían visualmente con los testigos. Consideramos que un fenograma muestra un caso de transferencia horizontal cuando algún organismo esté fuera del nodo al que idealmente pertenece y en cambio se agrupe con organismos que filogenéticamente resulten más distantes. Si por ejemplo, *E. coli K12* está fuera del nodo que agrupa a las otras *E. colis* representantes del género *Escherichia* y en su lugar queda agrupada con una representante del género *Shigella*, como *S. Sonnei*, no consideraremos esto como un evento de HGT, ya que ambas bacterias son filogenéticamente cercanas. Pero si una *E. coli* queda agrupada con una representante del género *Vibrio*, como *V. Cholerae*, consideraremos éste como un caso de HGT basados en la distancia filogenética de ambas bacterias. Más adelante veremos con detalle las dificultades y peculiaridades de identificar un evento de HGT observando solamente un árbol filogenético.

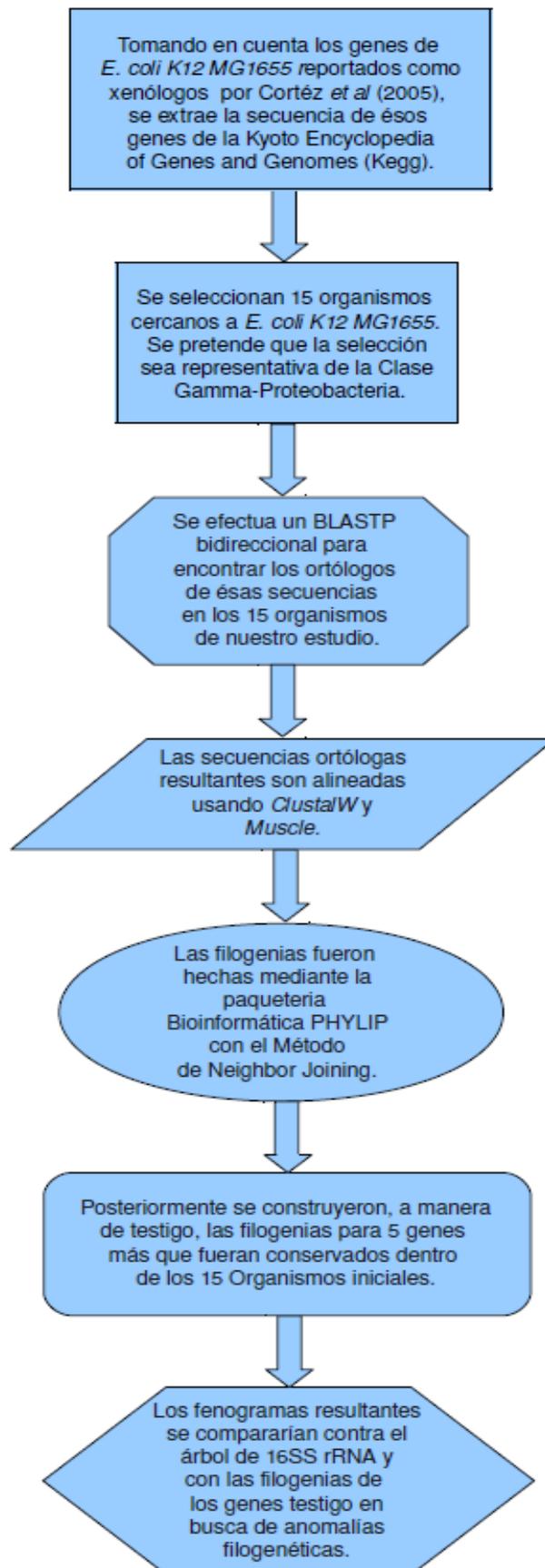
La construcción de los fenogramas fue hecha con la paquetería PHYLIP (Felsenstein, 2003), como ya antes lo habíamos mencionado, sin embargo, TreeView (Page, 1996) y FigTree (<http://tree.bio.ed.ac.uk/software/figtree>) fueron los dos programas empleados en la visualización de los fenogramas a partir del archivo de salida obtenido con PHYLIP y que servirían en la comparación directa de los mismos.

Dado que muchos de los procesos requeridos en la realización de este trabajo eran altamente repetitivos, se hicieron programas en el lenguaje de programación "PERL". Así, "*ancestor.pl*", "*clustero.pl*" y "*phylipero.cmd*", fueron empleados con tal propósito. El primero y el segundo fueron desarrollados sirviendo inicialmente en otros proyectos por otros investigadores dentro del laboratorio de Origen de la Vida de la Facultad de Ciencias. Para cumplir con nuestros objetivos ambos programas fueron ligeramente modificados. El último, es un "*batch file program*" que funciona en D.O.S. Fue escrito con ayuda de Antonio Moreno (PB).

Los tres programas y una descripción de ellos pueden ser encontrados en el sitio web: <http://microbio.fciencias.unam.mx/programas/>

A continuación se muestra un diagrama de flujo que resume en 7 etapas el procedimiento metodológico (Página 19).

Diagrama de Flujo del Procedimiento Metodológico.



Resultados y Discusión.

En el estudio comparativo de Cortéz y colaboradores en el 2005 se reportaron 140 genes como posibles xenólogos. Nosotros solo evaluamos 109 de esos genes como “total”, debido a que 31 de ellos correspondían a resultados hipotéticos que decidimos no considerar en el estudio.

Uno de los aspectos más importantes a notar es que nosotros sólo pudimos confirmar que de los 109 genes, 40 son en realidad xenólogos. En este sentido creemos que 3 factores jugaron un papel fundamental que influyeron en los resultados. A) Del total de posibles xenólogos reportados por Cortéz y colaboradores, nosotros no consideramos aquellos clasificados como hipotéticos o putativos en las bases de datos de donde extrajimos nuestras secuencias (Kegg y NCBI). Esto se decidió así porque es conocida la distorsión que las secuencias no caracterizadas pueden introducir en un estudio filogenético. B) Cortéz hizo una búsqueda de posibles xenólogos en todo el genoma de *E. coli* K12 MG1655, lo que posiblemente les permitió detectar un mayor número de genes xenólogos. En la literatura se han publicado trabajos que muestran que hasta el 15% del genoma de *E. coli* K12 podrían ser secuencias transferidas (Lawrence y Ochman, 1997), por lo que estos factores pudieron interferir mucho en la cantidad total de genes transferidos detectados en ambos trabajos (Ver Figura 3). C) Es posible que los resultados estén evidenciando que la sensibilidad de un método composicional y uno filogenético es muy diferente, que es precisamente uno de los objetivos de este trabajo.

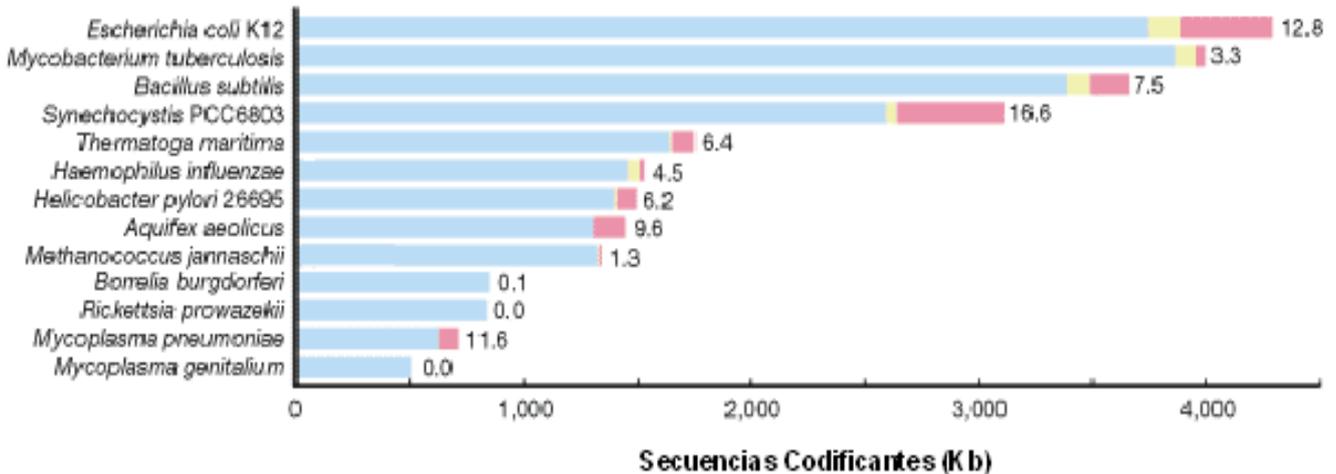


Figura 3. En esta gráfica, modificada de Ochman JH., Lawrence G. y Groisman E. (2000); se pueden observar las proporciones de secuencias codificantes foráneas correspondientes a cada organismo. La parte azul de las barras denota el DNA nativo, la parte amarilla simboliza DNA foráneo, como elementos móviles, transposones y bacteriófagos. Otro tipo de DNA foráneo que no entra en esta categoría se simboliza en rojo. A pesar de la dificultad de estimar la cantidad de genes xenólogos dentro de un determinado genoma, podemos considerar que esta gráfica muestra la “media” de HGT para algunas bacterias. Decimos la “media” porque algunos estudios estiman una proporción de transferencias considerablemente mayor al que se muestra en ésta gráfica.

Tabla 3. Desglose en categorías de los 109 genes analizados.

Categorías de Genes	Número de Genes
Xenólogos	43
NO Xenólogos	24
Posibles Xenólogos	35
Genes sin Ortólogos Asociados *	7
Total	109

*Originalmente considerados como Genes Huérfanos. Estos genes fueron inicialmente considerados así debido a que no tenían homólogos en ninguno de nuestros 15 organismos seleccionados. (Ver Tabla 5).

Desglosando la Tabla 3, la categoría “xenólogos” engloba aquellos genes que arrojan filogenias discordantes, por tanto, fueron considerados como genes xenólogos. A favor de nuestros resultados, la Tabla 4 evidencia aquellos genes que producen filogenias discordantes y que son congruentes con otros trabajos en los que se clasifican como genes de “mantenimiento”, los cuales se piensa se transfieren horizontalmente de forma muy activa, como genes de metabolismo secundarios, genes que codifican para distintas variantes de subunidades de algunas enzimas, genes que codifican para la conformación “L” (Levógira) de algunas moléculas o enzimas y genes para los que se piensa han sido transferidos desde la divergencia de ciertas familias de bacterias (como genes que codifican para las *acetil CoA* [Field *et al* 2000]) aparecen frecuentemente en la tabla. Además, genes que provienen de fagos, profagos y pseudogenes, que en la actualidad son ciertamente considerados como resultado de transferencias horizontales, también aparecen en dicha tabla. En porcentajes, la tabla 4 muestra que, considerados como xenólogos, tenemos en un 63% a genes de mantenimiento, 16% para genes que codifican proteínas que participan en metabolismo, primordialmente secundario, 14% para genes provenientes de bacteriófagos y 7% representando pseudogenes.

Tabla 4. Los 43 genes que resultaron ser xenólogos haciendo uso del método filogenético. Muchos de ellos provenientes de fagos, el resto, de genes frecuentemente reportados como transferidos (*acetil CoA*) y algunos más considerados como genes accesorios o de metabolismo secundario.

Etiqueta	Nombre del Gen	Descripción
<i>eco:b0227</i>	yafL, ECK0228, JW0217	Predicted lipoprotein and C40 family peptidase
<i>eco:b0261</i>	mmuM, ECK0263, JW0253, yagD	CP4-6 prophage. S-methylmethionine:homocysteine methyltransferase
<i>eco:b0273</i>	argF, arg5, argD, ECK0274, JW0266	Ornithine carbamoyltransferase 2, chain F; CP4-6 prophage
<i>eco:b0550</i>	rusA, ECK0541, JW0538, rus, ybcP	DLP12 prophage; endonuclease RUS
<i>eco:b0875</i>	aqpZ, bniP, ECK0866, JW0859	Aquaporin Z
<i>eco:b1338</i>	abgA, ECK1334, JW5205, ydaJ	P-aminobenzoyl-glutamate hydrolase, A subunit
<i>eco:b1363</i>	trkG, ECK1362, JW1358	Rac prophage; potassium transporter subunit
<i>eco:b1397</i>	paaJ, ECK1394, JW1392	3-oxoadipyl-CoA/3-oxo-5,6-dehydrosuberyl-CoA

eco:b1417	gapA	Glyceraldehyde-3-phosphate dehydrogenase
eco:b1442	ydcU, ECK1436, JW1437	Predicted spermidine/putrescine transporter subunit
eco:b1456	rhsE, ECK1450, JW1451	Pseudogene
eco:b1558	cspF, ECK1552, JW1550	Qin prophage; Cold Shock Protein
eco:b1776	ydjL, ECK1774, JW1765	Predicted oxidoreductase, Zn-dependent and NAD(P)-binding
eco:b2222	atoA, ECK2215, JW2216	Acetyl-CoA:acetoacetyl-CoA transferase, beta subunit
eco:b2223	atoE, ECK2216, JW2217	Short chain fatty acid transporter
eco:b2382	ypdC, ECK2378, JW2379	Predicted DNA-binding protein
eco:b2387	fryB, ECK2383, JW5389, ypdH	Predicted enzyme IIB component of PTS
eco:b2481	hyfA, ECK2477, JW2466, yffe	Hydrogenase 4, 4Fe-4S subunit
eco:b2483	hyfC, ECK2479, JW2468	Hydrogenase 4, membrane subunit
eco:b2488	hyfH, ECK2484, JW2473	Hydrogenase 4, Fe-S subunit
eco:b2492	focB, ECK2488, JW2477	Predicted formate transporter
eco:b2541	hcaB, ECK2538, JW2525, phdD, yfhX	2,3-dihydroxy-2,3-dihydrophenylpropionate dehydrogenase
eco:b2542	hcaD, ECK2539, hcaA4, JW2526, phdA, yfhY	Phenylpropionate dioxygenase, ferredoxin reductase subunit
eco:b2546	yphD, ECK2543 JW2530	Predicted sugar transporter subunit: membrane component of ABC superfamily
eco:b2639	arsB, arsF	Arsenical pump membrane protein
eco:b2715	ascF, ECK2710, JW5435, sac	Fused cellobiose/arbutin/salicin-specific PTS enzymes
eco:b2886	ygfS, ECK2881, JW5468	Predicted oxidoreductase, 4Fe-4S ferredoxin-type subunit
eco:b2934	cmtB, ECK2929, JW2901, tolM	Predicted mannitol-specific enzyme IIA component of phosphotransferase system
eco:b2978	glcF, ECK2972, gox, JW5486, yghL	Glycolate oxidase iron-sulfur subunit
eco:b2979	glcD, ECK2974, gox, JW2946, yghM	Glycolate oxidase subunit, FAD-linked
eco:b3063	ttdT, ECK3053, JW3035, ygcC, ygcE	L-tartrate/succinate antiporter
eco:b3111	tdcG, ECK3101, JW5520, yhaP, yhaQ	L-serine dehydratase 3
eco:b3112	tdcG, ECK3101, JW5520, yhaP, yhaQ	L-serine deaminase
eco:b3134	agaW, ECK3122, JW3103, yhaZ	Pseudogene
eco:b3135	agaA, ECK3123, JW5527, yraA	Pseudogene
eco:b3269	yhdX, ECK3256, JW5544	Predicted amino-acid transporter subunit
eco:b3476	nikA, ECK3460, hydC, hydD, JW3441	Nickel-binding, heme-binding periplasmic protein
eco:b3477	nikB, ECK3461, hydC, hydD, JW3442	Nickel transporter subunit
eco:b3479	nikD, ECK3463, hydC, hydD,	Nickel transporter subunit
eco:b3503	arsC, arsG, ECK3488	Arsenate reductase

<i>eco:b3709</i>	tnaB, ECK3702, JW5619, JW5622, tnaP, trpP	Tryptophan transporter of low affinity
<i>eco:b4289</i>	fecC, ECK4279, JW4249	KpLE2 phage-like element; iron-dictrate transporter subunit
<i>eco:b4318</i>	fimF, ECK4309, JW4281	Minor component of type 1 fimbriae

La categoría “NO xenólogos” (Tabla 3) engloba a aquellos genes para los que no se encontraron anomalías filogenéticas de consideración ni con el árbol de las especies ni con los producidos en base a los genes testigo, por tanto consideramos que no son producto de HGT.

Además, mostramos la categoría “Posibles xenólogos” (Ver Tabla 6); que son genes para los que existen evidencias (como el origen de las secuencias) de que podrían ser resultado de HGT; sin embargo, son extrapolaciones que se hicieron con base en la distancia filogenética que existe entre los organismos reflejados en los fenogramas y/o al conocimiento que se tiene del gen mismo. Desafortunadamente, el método filogenético no puede identificarlos como posibles genes transferidos debido a la falta de taxones para comparar. Confirmar que estos genes son efectivamente xenólogos requiere un acercamiento distinto al filogenético y al composicional (método con el cual originalmente fueron considerados como resultado de HGT). En este caso, emplear una combinación de metodologías mutuamente compensatorias sería lo más adecuado (Becq *et al* 2010).

La Tabla 3 muestra además la categoría “Genes sin Ortólogos Asociados”, inicialmente considerados como genes huérfanos; es decir, genes para los cuales no es posible hacer un análisis filogenético debido a que no se encontraron secuencias ortólogas para comparar, lo que nos impide hacer comparaciones y por ende construir fenogramas informativos. Emplear el método filogenético para determinar si estos genes son o no xenólogos es imposible debido a las restricciones del método en sí. Más adelante detallaremos esta parte.

Comprobando los Resultados.

Suponiendo que posiblemente estos 7 genes, considerados como posibles huérfanos, no tenían ortólogos en nuestros 15 organismos pero a lo mejor sí en otras clases dentro del dominio Bacteria, decidimos hacer una comprobación de nuestras sospechas.

Fue entonces que se realizó un segundo análisis que descartara errores por “*Ancestor.pl*” (el programa hecho en el lenguaje de PERL en el laboratorio de Origen de la Vida que automáticamente hacía todos los BLAST para cada una de las secuencias en comparación).

Con el objetivo de hacer una búsqueda más específica, se corrieron manualmente 7 BLAST correspondientes a cada gen huérfano identificado (véase tabla 5) haciendo comparaciones manuales en el *Kegg Orthology* y corriendo un BLAST de las secuencias codificantes. Haciendo uso de la base de datos del *Kegg Orthology* y del PSI-BLAST (Altschul *et al* 1997) nos aseguramos que las búsquedas efectuadas por BLAST sean más sensibles en la detección de únicamente genes ortólogos, disminuyendo así la identificación de falsos positivos, aún entre organismos distantes.

Tabla 5. Los 7 genes inicialmente considerados como genes huérfanos.

Etiqueta	Nombre del Gen	Descripción
<i>eco:b0349</i>	mhpC, ECK0346, JW0340	2-hydroxy-6-ketono-2,4-dienedioic acid hydrolase
<i>eco:b2035</i>	rfc, ECK2029, JW2020, wbbH, yefF	O-antigen polymerase
<i>eco:b2037</i>	rfbX, ECK2031, JW2022, wzx, wzxB	Predicted polisoprenol-linked O-antigen transporter
<i>eco:b2352</i>	gtrS, ECK2346, gtrIV, JW5382	Serotype-specific glucosyl transferase, CPS-53 prophage
<i>eco:b1139</i>	lit, ECK1125, JW1125	e14 prophage; cell death peptidase
<i>eco:b3329</i>	gspH, ECK3316, hofH, hopH, JW3291	Predicted general secretory pathway component, cryptic
<i>eco:b3334</i>	gspM, ECK3321, hopZ, JW5704	General secretory pathway component, cryptic

Los resultados de los BLAST hechos manualmente para las secuencias de la tabla 5 indicaron que los supuestos genes huérfanos en realidad sí poseen homólogos y ortólogos en una cantidad variable de organismos; sin embargo, hay dos casos que vale la pena detallar.

En primera instancia, tenemos el gen *eco: b1139*, que según la clasificación del *Kegg* es una peptidasa que induce apoptosis, probablemente resultado de una transferencia horizontal proveniente de un fago. En este caso, consideramos que *eco:b1139* tiene grandes posibilidades de ser un xenólogo basándonos en su origen, ya que las transferencias horizontales mediante transducción (en donde los fagos son los donantes de las secuencias) se han reportado y documentado con frecuencia (Canchaya *et al* 2003); aunque, con el método filogenético no podemos confirmar que estas secuencias son el resultado de HGT.

En otro caso, en las primeras comparaciones con BLAST (Altschul *et al* 1997) no se encontró evidencia de que el gen *eco:b3334* tuviera homólogos; pero en la posterior comprobación de los resultados descubrimos que en la base de datos del *Kegg Orthology* se pudo encontrar coincidencias en 3 de los organismos de nuestro estudio, específicamente en *ecg*, *efe*, *vch* y *vpa*. Estos podrían ser hallazgos importantes, ya que son congruentes con un gen de posible transporte horizontal que es componente primordial de sistemas generales de secreción, elementos fundamentales en bacterias patógenas, y que se ha demostrado se han transferido de forma activa a lo largo de Proteobacteria (Cascales *et al* 2003). A pesar de la evidencia que parece confirmar que *eco: b3334* es efectivamente un xenólogo, el método filogenético no es adecuado para responder dicha cuestión.

Consideramos que de forma inicial no se encontraron secuencias homólogas para estos genes debido a que las búsquedas posteriores se hicieron con el *Kegg Orthology*; herramienta bioinformática de la base de datos del Kegg que emplea un algoritmo computacional más específico que el BLAST (Altschul *et*

al 1997). Lo que seguramente influyó de forma sustancial en los hallazgos posteriores (<http://www.genome.jp/kegg/ko.html>).

Como veremos más adelante, considerar los valores de BLAST como indicadores de HGT puede resultar en algo muy inconveniente (Eissen, 2000).

La mayoría de las metodologías recientes se enfocan en la detección bajo parámetros muy específicos, sin embargo, el método filogenético ha probado ser sumamente popular en la detección de eventos de transferencia horizontal, y no es para menos, ya que se ha confirmado que es el que presenta menos susceptibilidad a las condiciones de trabajo, con lo que se han podido hacer descubrimientos y confirmaciones que anteriormente solo resultaban de extrapolaciones lógicas de análisis más sencillos.

Uno de los ejemplos célebres y que evidencian la eficacia del método filogenético para inferir eventos de HGT aun bajo condiciones difíciles se presentó en el 2001.

Con el primer borrador del genoma humano completo por el consorcio privado, se anunció el descubrimiento de genes que se encontraban sólo en eucariontes vertebrados y cuyos homólogos estaban ampliamente distribuidos a lo largo de todo Proteobacteria. Como los estudios iniciales carecían de suficiente información comparativa (sin los controles positivos adecuados), se postuló que probablemente estábamos presenciando un evento de HGT ancestral entre genes de bacterias hacia los vertebrados que debió ocurrir durante el origen y diversificación de los eucariontes (Lander, 2001).

Dudando de tales conclusiones, Stanhope y colaboradores (2001) hizo una reconstrucción filogenética completa para los 28 genes sospechosos de ser resultado de HGT.

Los autores concluyen que ciertamente hay genes cuya historia evolutiva es difusa, por ello les permite ser considerados como transferencias entre bacterias y eucariontes vertebrados, pero esto sucede en el menor de los casos, ya que en general, fue factible construir la línea de descendencia para cada gen analizado. Parecía ser que los genes ya estaban presentes en el último ancestro más reciente de los vertebrados, mas no sería el caso para el ancestro de los invertebrados quien parecía haberlos perdido (Gogarten, 2003). Este hecho es de suma importancia, ya que ha vuelto a ocurrir que los resultados de BLAST se interpretan como posibles transferencias horizontales (Lander *et al* 2001). En nuestro trabajo, nos encontramos con 34 casos en el que el gen de *E. coli* K12 MG1655 comparado tiene menos de cuatro genes homólogos asociados. De éstos 34 genes, podemos postular que para 28 existe una gran probabilidad de que sean resultado de HGT. Factores ajenos al método, como la naturaleza del gen transferido y la distancia filogenética entre los organismos representados en el fenograma, nos permiten llegar a ésa conclusión. Sin embargo, como Stanhope y colaboradores (2001) apuntan, los resultados obtenidos por BLAST no pueden interpretarse como un indicador confiable de HGT, dado que la similitud no es un indicador de HGT. Sabemos que factores como el tamaño de las secuencias, la distancia filogenética entre ellas y las distintas tasas evolutivas que les afectan (Kanhere y Vingron 2009), pueden producir falsos positivos. Inclusive, Kurland y colaboradores (2003) propone que es posible que los resultados de BLAST reflejen el origen común de algunas secuencias, más no que éstas hayan sido transferidas.

Continuando con el desglose de nuestros resultados, encontramos que aproximadamente 35 genes tienen menos de 5 ortólogos asociados, por lo que debido a la falta de taxones para comparar no fue posible realizar una filogenia con ellos. A pesar de éste hecho, pudimos observar que algunos genes poseen ortólogos que se encuentran en organismos filogenéticamente distantes. Es factible que algunos de estos 35 genes sean efectivamente resultado de HGT, sin embargo, son sospechas apoyadas en otros factores, como la biología de los organismos involucrados y el origen de las secuencias transferidas.

En un intento posterior de comprobar las suposiciones de que algunos de éstos 35 genes son efectivamente xenólogos, se volvieron a efectuar búsquedas usando el *Kegg Orthology*. Aunque en principio los resultados indican que algunos de éstos genes pudieron estar presentes en el último ancestro de éstas bacterias para después perderse, duplicarse o experimentar un reemplazo de ortólogos; encontramos que en la mayoría de los casos los genes estaban plagados de secuencias parálogas, reconocidas en la literatura porque resultan inapropiadas para construir filogenias y por influenciar negativamente la búsqueda de HGT (Becq *et al* 2010).

Por ejemplo, en nuestros resultados encontramos que para el gen *eco:b1932* reportado en la base de datos del *Kegg* como una posible *aciltransferasa*, tres organismos pertenecientes al género *Escherichia*; *E. coli* K12, *E. coli* O127 y *E. fergusonii* aparecen en una filogenia con *V. parahaemolyticus*; además, en el fenograma es evidente que *E. fergusonii* y *V. parahaemolyticus* se agrupan juntas en un nodo, mientras que las otras dos bacterias representantes del género *Escherichia* quedan solas sin agrupar (ver el fenograma correspondiente, pag. 46). En nuestra opinión, éste podría ser un caso de transferencia horizontal, sin embargo, la falta de *otus*, en especial del género *Vibrio*, nos impide confirmar que el gen *eco:b1932* es un gen que sólo comparten cuatro organismos y por ende un posible xenólogo. El método filogenético es insuficiente para poder determinar si es un caso de HGT o no, y aunque ciertamente parece serlo, es bien conocido que hay factores que producen agrupaciones discordantes en filogenias pequeñas (y que no siempre se deben interpretar como transferencias horizontales).

En este sentido, uno de los más discutidos es el fenómeno de la atracción de ramas largas (Bergsten, 2005), en el que un muestreo incompleto de organismos cercanos entre sí favorece el agrupamiento de organismos filogenéticamente distantes. En otras ocasiones, se ha postulado que algunas agrupaciones discordantes no siempre reflejan HGT; en realidad podrían estar evidenciando que en los genomas hay secuencias con diferentes niveles de conservación, por lo que si se construye una filogenia con una secuencia de estas características podría resultar en agrupaciones discordantes que reflejan éste tipo de gradientes. El reemplazo de ortólogos, que es la sustitución de un gen residente por una versión adquirida que ocurre por recombinación homóloga o por pérdida secundaria del gen original, también genera topologías discordantes (Baptiste *et al* 2004).

Otra explicación podría recaer en el fenómeno descrito como “mosaicismo”, (distinto al término genético para referirse a las anomalías genotípicas de un individuo) que se refiere a la sustitución que ocurre cuando un gen transferido es incorporado parcialmente a la secuencia de un gen homólogo ya existente dentro de un genoma residente; por tanto, hay genes que pueden tener

diferentes historias evolutivas y por ende diferir en su origen. Así, si dos organismos son comparados mediante una secuencia que proviene del mismo origen, se agruparían, sin importar la distancia filogenética de los genomas portadores de la secuencia (Zhaxybayeva *et al* 2004; Hao *et al* 2010). Philippe y Forterre (1999) postulan que en ocasiones, aun cuando el método filogenético se aplica de forma correcta y se cuidan aquellos factores inherentes a la metodología que podrían ser causantes de discordancias filogenéticas, éstas no siempre funcionan correctamente.

Tabla 6. Los 35 genes posiblemente xenólogos según las búsquedas realizadas con BLAST pero que no pueden ser corroborados mediante la construcción de filogenias.

Etiqueta	Nombre del Gen	Descripción
<i>eco:b0259</i>	insH, ECK0261	IS5 transposase and trans-activator
<i>eco:b0263</i>	afuB, ECK0265, fbpB, JW0255	iron(III) transport system permease protein (pseudogen)
<i>eco:b0351</i>	mhpF, ECK0348, JW0342	acetaldehyde-CoA dehydrogenase II
<i>eco:b0352</i>	mhpE, ECK0349, JW0343	4-hydroxy-2-oxovalerate/4-hydroxy-2-oxopentanoic acid aldolase, class I
<i>eco:b0486</i>	ybaT, ECK0480, JW0475	Predicted Transporter
<i>eco:b0554</i>	essD, ECK0545, JW0543, ybcR	DLP12 prophage; predicted phage lysis protein
<i>eco:b0557</i>	borD, ECK0549, iss, JW0546, ybcU	DLP12 prophage; predicted lipoprotein
<i>eco:b0801</i>	ybiC, ECK0790, JW0786	Predicted Dehydrogenase
<i>eco:b1370</i>	insH, ECK0261	IS5 transposase and trans-activator
<i>eco:b1389</i>	paaB, ECK1386, JW1384, ynbF	Ring 1,2-phenylacetyl-CoA epoxidase possible subunit
<i>eco:b1391</i>	paaD, ECK1388, JW5217, ydbQ	Ring 1,2-phenylacetyl-CoA epoxidase subunit
<i>eco:b1396</i>	paaI, ECK1393, JW1391, ydbV	Thioesterase,
<i>eco:b1493</i>	gadB, ECK1487, JW1488	Glutamate decarboxylase B, PLP-dependent
<i>eco:b1566</i>	flxA, ECK1560, JW1558	Qin prophage; predicted protein
<i>eco:b1569</i>	dicC, ECK1563, ftsT, JW1561	Qin prophage; DNA-binding transcriptional regulator for DicB
<i>eco:b1575</i>	dicB, ECK1569, ftsT, JW1566	Qin prophage; cell division inhibition protein
<i>eco:b1597</i>	asr, ECK1592, JW5826	Acid shock-inducible periplasmic protein
<i>eco:b1932</i>	yedL, ECK1931, JW1917	Predicted acyltransferase
<i>eco:b2485</i>	hyfE, ECK2481, JW2470	Hydrogenase 4, membrane subunit
<i>eco:b2770</i>	ygcR, ECK2765, JW5441	Predicted flavoprotein
<i>eco:b2875</i>	yqeB, ECK2871, JW2843	Conserved protein with NAD(P)-binding Rossmann fold
<i>eco:b2931</i>	yggP, ECK2927, JW5477	Predicted dehydrogenase
<i>eco:b2976</i>	glcB, ECK2970, glc, JW2943	Malate synthase G
<i>eco:b2977</i>	glcG, ECK2971, JW2944, yghC	Conserved Protein

<i>eco:b3119</i>	tdcR, ECK3108, JW5525	DNA-binding transcriptional activator
<i>eco:b3129</i>	sohA, ECK3117, JW3098, prlF	Antitoxin of the SohA(PrIF)
<i>eco:b3510</i>	hdeA, ECK3494, JW3478, yhhC	Stress response protein acid-resistance protein
<i>eco:b3525</i>	yhjH, ECK3510, JW3493	Cyclic-di-GMP phosphodiesterase, FlhDC-regulated
<i>eco:b3622</i>	rfaL, ECK3612, JW3597, waaL	O-antigen ligase
<i>eco:b3623</i>	waaU, ECK3613, JW3598, rfaK	Lipopolysaccharide core biosynthesis
<i>eco:b3629</i>	rfaS, ECK3619, JW3604, waaS	Lipopolysaccharide core biosynthesis protein
<i>eco:b3645</i>	dinD, ECK3635, JW3620, pcsA	DNA-damage-inducible protein
<i>eco:b4225</i>	chpB, chpBK, ECK4221, JW4184	Toxin of the ChpB-ChpS toxin-antitoxin system
<i>eco:b4292</i>	fecR, ECK4282, JW4252	KpLE2 phage-like element; transmembrane signal transducer for ferric citrate transport
<i>eco:b4293</i>	fecI, ECK4283, JW4253	KpLE2 phage-like element; RNA polymerase, sigma 19 factor

Previamente habíamos mencionado que muchas de las críticas al método filogenético radican en que sólo es posible hacer filogenias con secuencias ortólogas. El problema es que la información con la que contamos en nuestras bases de datos, aunque cada vez más grandes y completas, aun carecen de la información apropiada para reconstruir la historia evolutiva de los organismos que habitan este planeta. Como Zhaxybayeva y Gogarten (2006) postulan, sólo se pueden utilizar genes ortólogos para reconstruir filogenias confiables y con significado biológico plausible. Este hecho, frecuentemente considerado como una deficiencia del método, en realidad debería de ser una crítica para la forma en la que construimos nuestras bases de datos, y es que a excepción de los 7 genes huérfanos, 97% de las secuencias que necesitábamos para construir nuestras filogenias estaban llenas de genes parálogos. La exclusión de este tipo de genes no fue tan laboriosa dado que el trabajo no fue realizado a gran escala, sin embargo, es necesario llevar a cabo la construcción de bases de datos que contengan solamente genes ortólogos, y en el caso de las ya existentes, especial cuidado necesita ponerse, ya que aun en estas que, supuestamente están libres de secuencias parálogas (como los COGS del Kegg) se encontraron evidencias de su presencia (datos sin publicar, experiencias compartidas dentro del laboratorio de Origen de la Vida, Facultad de Ciencias).

Filogenias.

Como parte central de los resultados, obtuvimos 86 fenogramas, cada uno correspondiente a cada gen analizado. Basándonos en la filogenia de rRNA de Yarsa y colaboradores (2008), en las filogenias de Gamma-Proteobacteria de (Kelly *et al* 2010) y en las filogenias de los 5 genes que usaríamos como testigos, evaluamos y comparamos las filogenias.

Durante la comparación de los fenogramas en busca de genes xenólogos nos encontramos con que, como otros autores ya lo han señalado, es notoria la

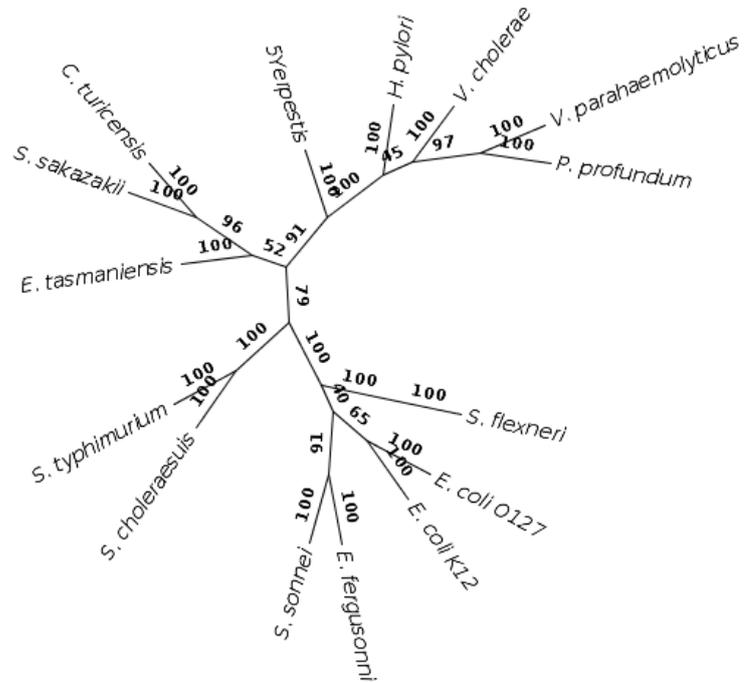
dificultad para distinguir entre ganancia o pérdida de genes cuando simplemente se observa un árbol filogenético. En este trabajo mostramos varios casos en los que no es posible determinar si el agrupamiento filogenético observado con base en si determinado gen es producto de una ganancia (posible HGT), una pérdida, a que el gen sea parte de lo que algunos llaman “genes accesorios” (que son aquellos que solamente se encuentran en muy pocos genomas de organismos relacionados [Lapierre y Gogarten 2009]), o a que el gen está reflejando distintas historias evolutivas, fenómeno conocido en la literatura como mosaicismo (Zhaxybayeva, Lapierre y Gogarten, 2004).

Los primeros 5 fenogramas deben de considerarse las filogenias control. En su construcción se utilizaron 5 genes conservados en nuestros organismos (ver Tabla 2). Posteriormente presentamos los fenogramas y una descripción de los casos representativos para genes que posiblemente son resultado de transferencias horizontales, así como casos que no reflejan evidencias de HGT. Al final mostramos las agrupaciones filogenéticas que no pueden resolverse concluyentemente haciendo uso del método filogenético. El principal problema con éstas filogenias es la falta de *otus* para comparar que nos indiquen que las agrupaciones resultantes no son resultado de artefactos o imperfecciones de la metodología *per se*. Bajo estas condiciones, ciertos fenómenos como la atracción de ramas largas (Bergsten, 2005), las pérdidas secundarias de genes en varios organismos cercanos a los que aun conservan las secuencias y las posibles diferencias en la tasa de divergencia de las secuencias comparadas (Eisen, 2000) son problemas que normalmente interfieren en las filogenias en donde no se maximiza el muestreo de organismos cercanos a los comparados.

(Filogenia Testigo).

Gen: *rRNA*

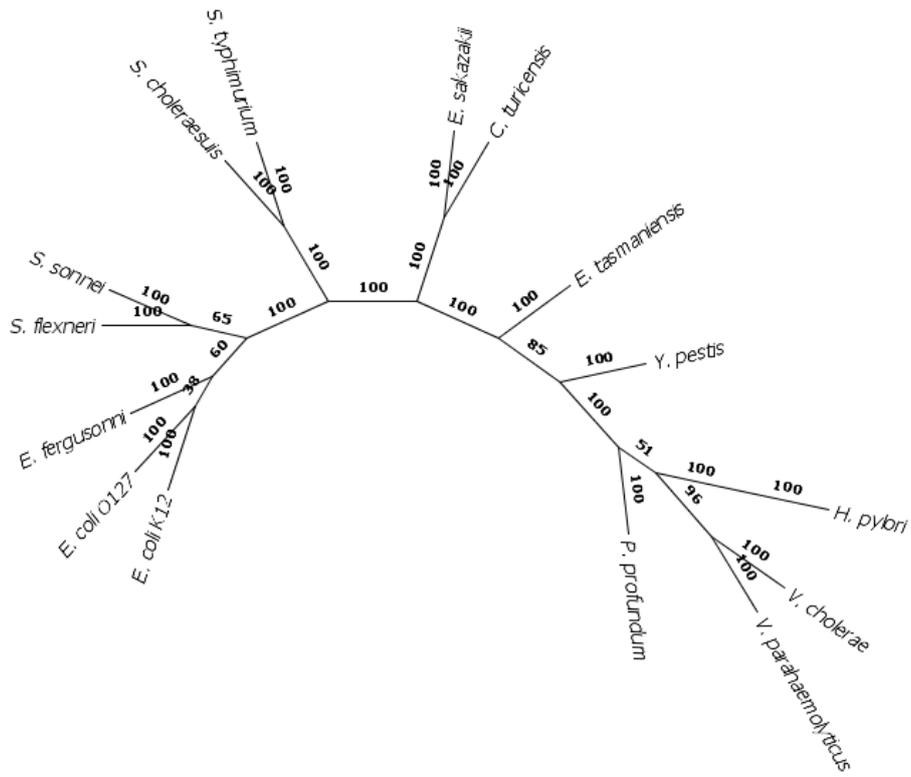
Definición: *rRNA - 16S ribosomal RNA*



(Filogenia Testigo).

Gen: *rpsB*

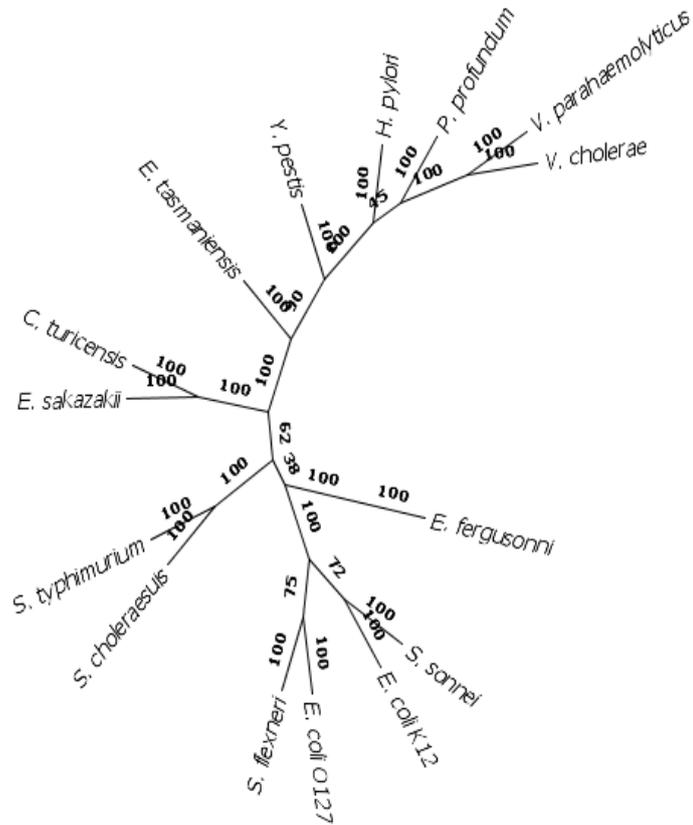
Definición: 30S ribosomal protein S2.



(Filogenia Testigo).

Gen: *rplF*

Definición: 50S ribosomal protein L6.

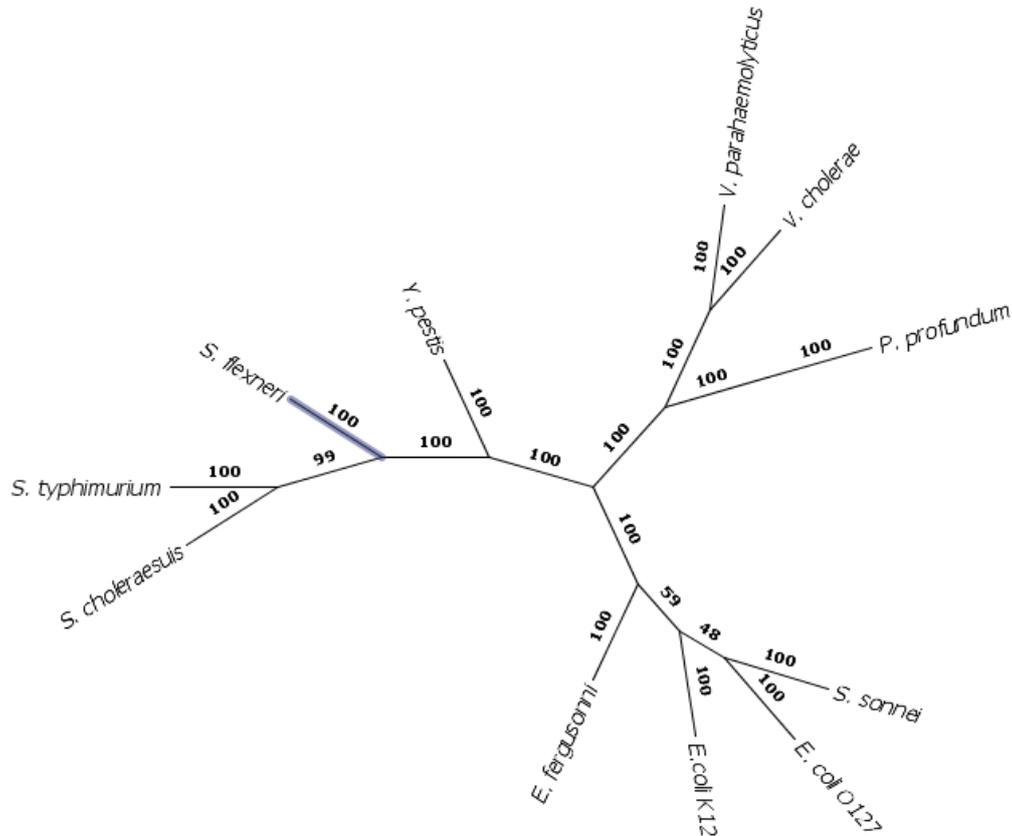


Casos Seguros de Transferencia Horizontal.

Gen: *eco:b2382*

Nombre del Gen: *ypdC*, *ECK2378*, *JW2379*

Definición: Predicted DNA-binding protein.



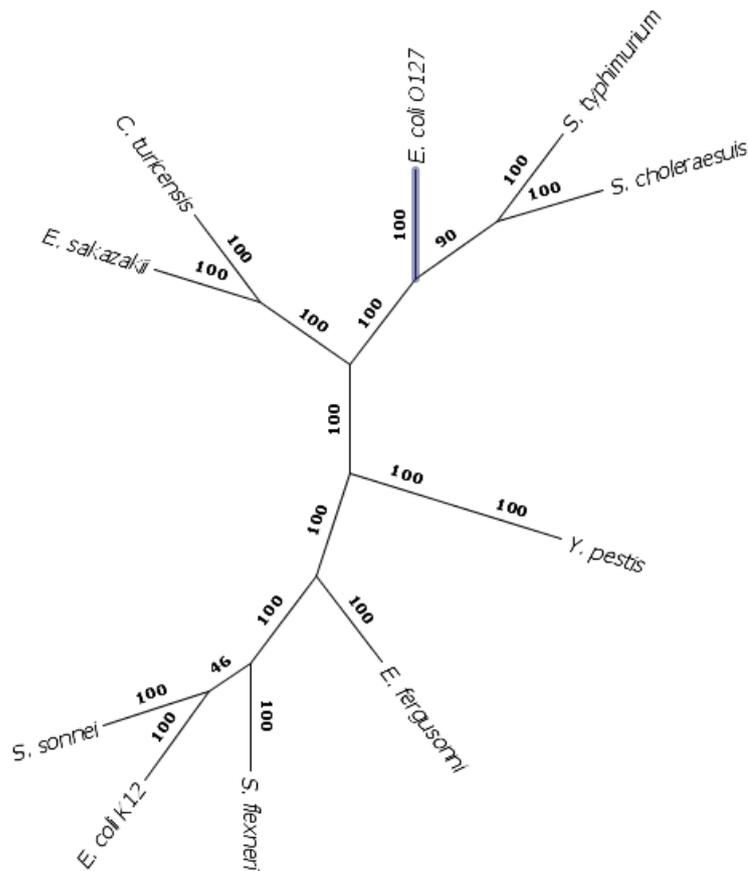
En este fenograma se representa la filogenia obtenida a partir del gen *eco:b2382*. Vemos que la topología del fenograma muestra 3 grandes nodos con representantes de los géneros *Escherichia*, *Salmonella* y *Vibrio*. Sin embargo, *S. flexneri* al no estar en el mismo nodo que *S. sonnei*, evidencia una discordancia filogenética. Consideramos que esto podría ser evidencia de un evento de transferencia horizontal. Interesantemente, dado que el gen *eco:b2382* parece estar participando en algún tipo de unión con el DNA, podemos considerar que el posible gen transferido resulta ser un gen informacional (Jain *et al* 1999); cuya frecuencia de HGT debería de ser baja, según se ha postulado, en base a que este tipo de genes poseen gran especificidad por su sustrato e interactúan con una gran cantidad de módulos celulares. A pesar de ello, consideramos que el gen *eco:b2382* es un xenólogo.

Casos Seguros de Transferencia Horizontal.

Gen: *eco:b2481*

Nombre del Gen: *hyfA*, *ECK2477*, *JW2466*, *yffE*

Definición: Hydrogenase 4, 4Fe-4S subunit.



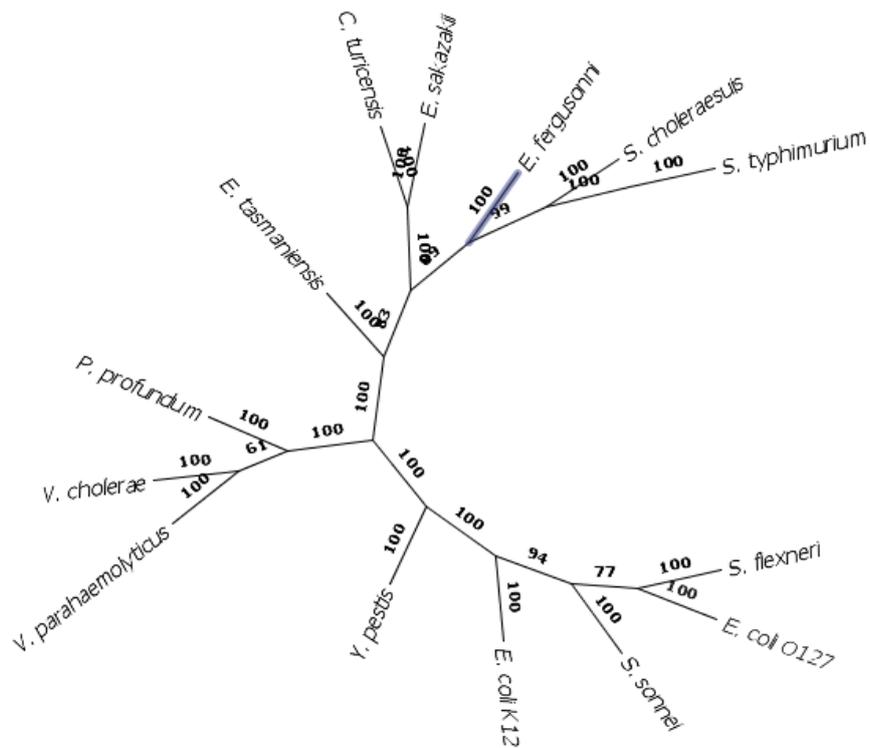
Para el caso del gen *eco:b2481*, consideramos que la discordancia filogenética es evidente. En el fenograma se muestra un nodo que agrupa a los representantes de los géneros *Escherichia* y *Shigella*. Sin embargo, *E. coli* O127 se agrupa en un nodo muy cercano a las representantes del género *Salmonella*. Que además, el valor de esta asociación sea apoyado por un excelente valor de Bootstrap (90%) nos indica que este fenograma posiblemente sea el reflejo de una transferencia horizontal.

Casos Seguros de Transferencia Horizontal.

Gen: *eco:b2546*

Nombre del Gen: *yphD*, *ECK2543*, *JW2530*

Definición: Predicted sugar transporter subunit: Membrane component of ABC superfamily.



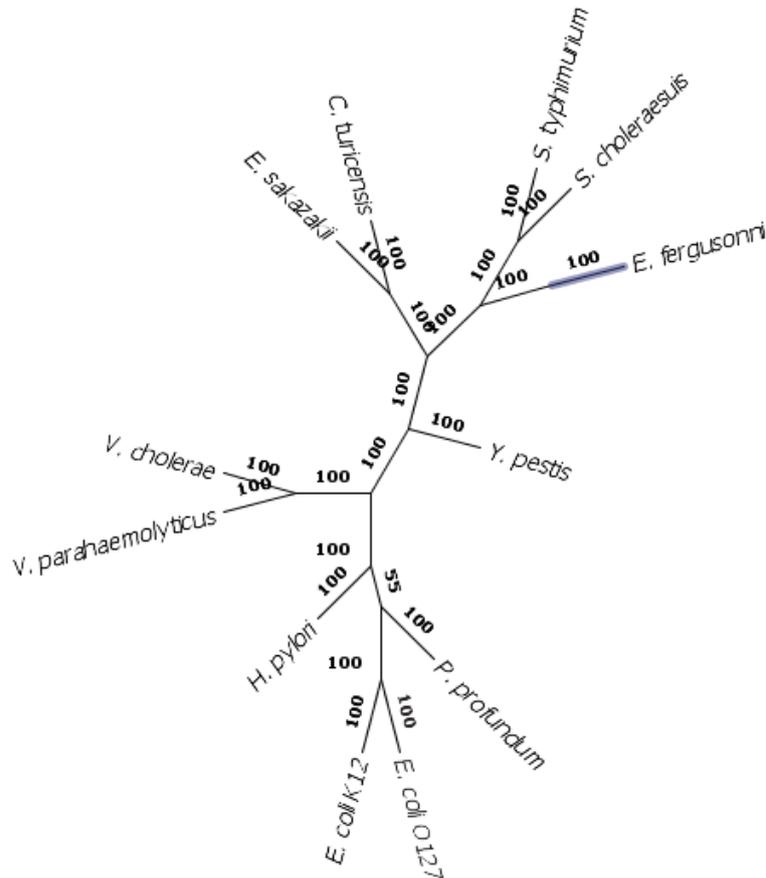
En este fenograma se muestra la agrupación resultante en base al gen *eco:b2546*. Aquí notamos que *Y. pestis* está muy cerca de algunos de los representantes de los géneros *Escherichia* y *Shigella*, aunque la naturaleza desordenada del fenograma nos indica que este hecho no resulta de mayor importancia. Sin embargo, consideramos que el verdadero caso de HGT está indicado por la posición de *E. fergusonii* entre organismos representantes del género *Salmonella* y *Cronobacter*.

Casos Seguros de Transferencia Horizontal.

Gen: *eco:b2979*

Nombre del Gen: *glcD*, *ECK2974*, *gox*, *JW2946*, *yghM*

Definición: Glycolate oxidase subunit, FAD-linked.



En este fenograma correspondiente a la agrupación resultante con base en el gen *eco:b2979* podemos observar que *Photobacterium profundum*, que en condiciones normales debería formar un nodo con los miembros del género *Vibrio*, queda con *E. coli* K12 y *E. coli* O127, quienes no se agrupan con *E. fergusonii*, bacteria del género *Escherichia* con la que mantienen una relación filogenética cercana.

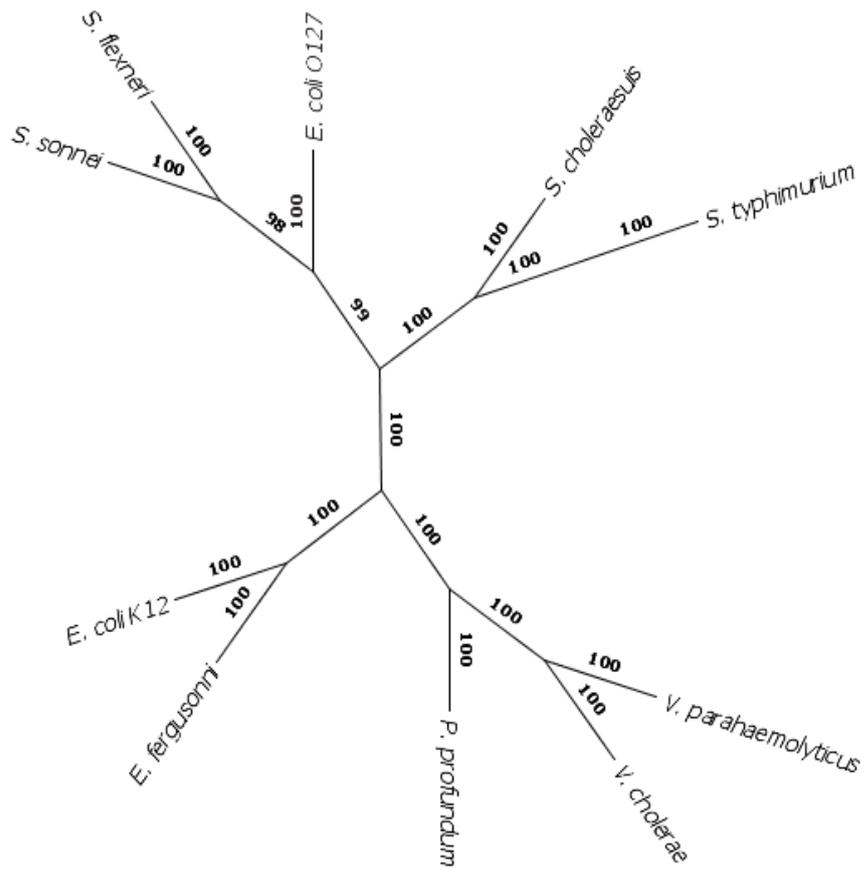
Además, es notoria la presencia de *Helicobacter pylori*, bacteria perteneciente al grupo de las Epsilon-Proteobacteria. Basándonos únicamente en la distancia filogenética, su presencia es injustificada. Consideramos que estas anomalías resultan indicadoras de un caso de HGT.

Casos en los que no hubo Transferencia Horizontal.

Gen: *eco:b1400*

Nombre del Gen: *paaY*, *ECK1397*, *JW1395*, *ydbZ*

Definición: Predicted hexapeptide repeat acetyltransferase.

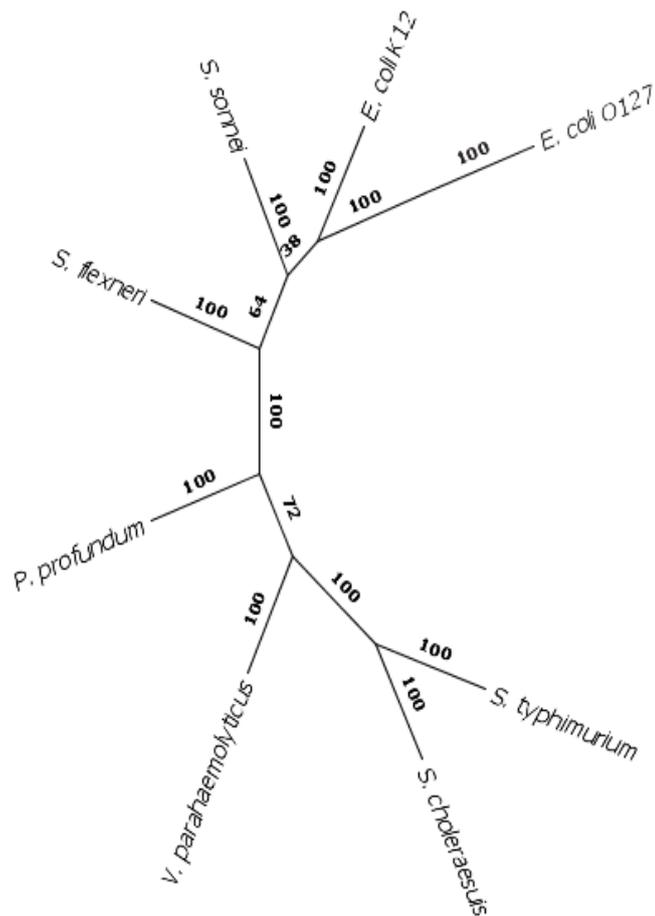


Casos en los que no hubo Transferencia Horizontal.

Gen: *eco:b2769*

Nombre del Gen: *ygcQ*, *ECK2764*, *JW5440*

Definición: Predicted flavoprotein.

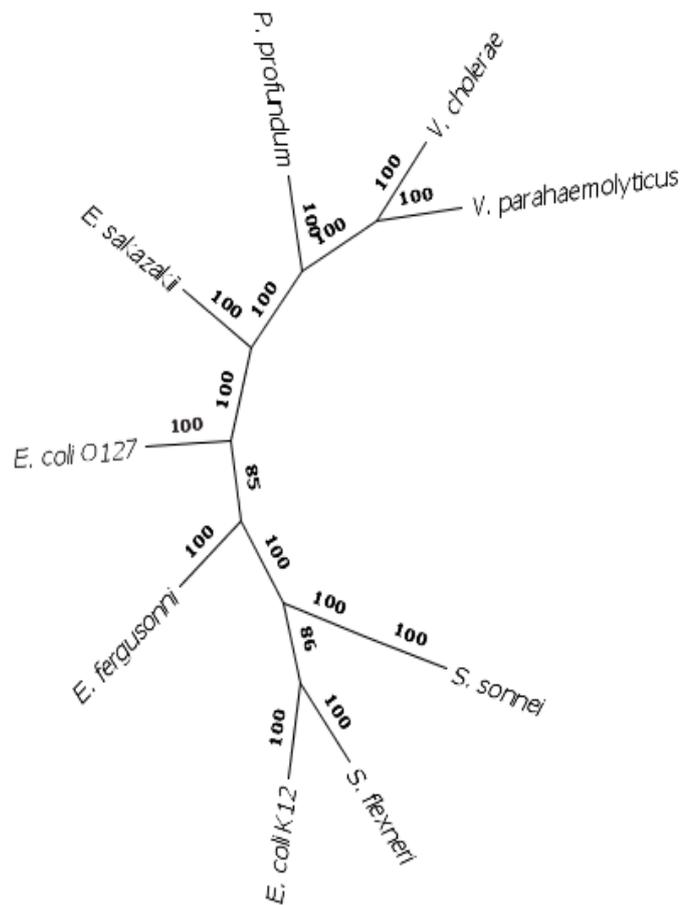


Casos en los que no hubo Transferencia Horizontal.

Gen: *eco:b3900*

Nombre del Gen: *frvA*, *ECK3893*, *JW3871*, *yiiK*

Definición: Predicted enzyme IIA component of Posphotransferase system.

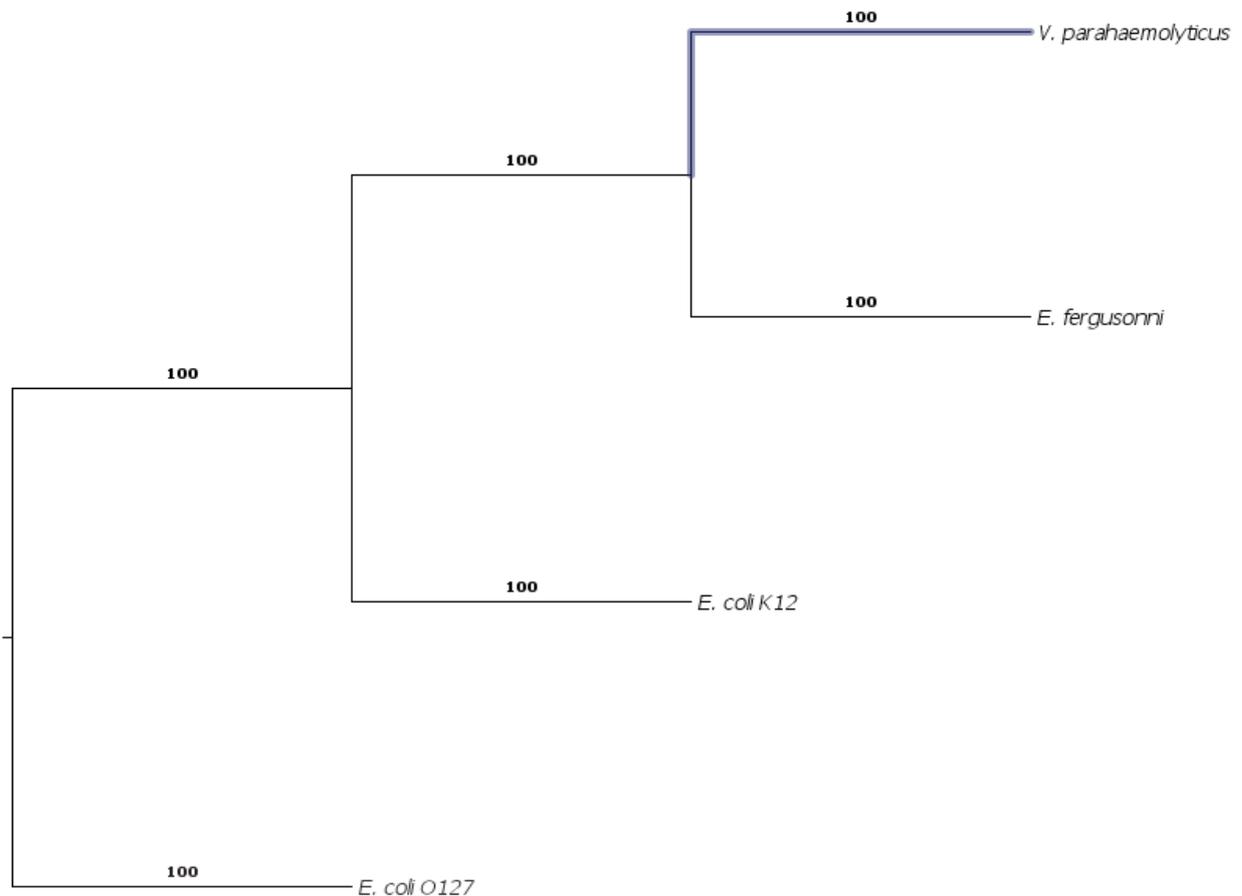


Casos en los que no es posible diferenciar entre ganancia (HGT) o pérdida de genes.

Gen: *eco:b1932*

Nombre del Gen: *yedL*, *ECK1931*, *JW1917*

Definición: Predicted acyltransferase.



En este caso, la agrupación resultante del gen *eco:b1932* se muestra en la forma clásica para la representación de filogenias debido a la ausencia de *otus* que permitan comparar un árbol sin raíz.

En esta filogenia tenemos representadas a las tres bacterias del género *Escherichia*. Interesantemente, la presencia de *V. parahaemolyticus* y más aun su agrupamiento con *E. fergusonii*, sugiere que este hecho podría considerarse como un evento de HGT. Sin embargo, la falta de *otus* nos impide, desde un punto de vista filogenético, considerar esto como un posible caso de HGT.

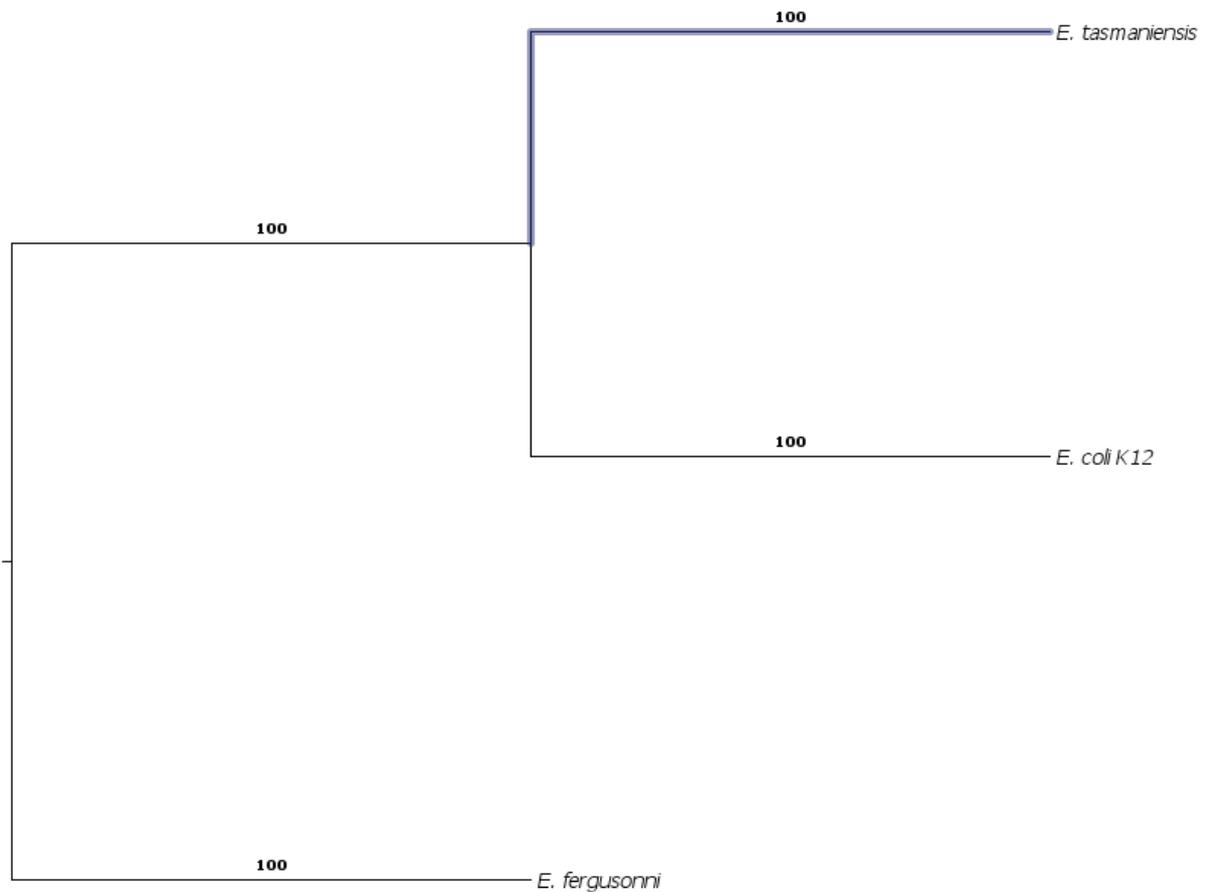
En estos casos, fenómenos como la atracción de ramas largas (Bergsten, 2005), las pérdidas secundarias de genes en varios organismos cercanos a los que aun conservan las secuencias y las posibles diferencias en la tasa de divergencia de las secuencias comparadas (Eisen, 2000) son problemas que normalmente interfieren en las filogenias pequeñas y generan agrupaciones filogenéticas cuestionables.

Casos en los que no es posible diferenciar entre ganancia (HGT) o pérdida de genes.

Gen: *eco:b0557*

Nombre del Gen: *borD*, *ECK0549*, *iss*, *JW0546*, *ybcU*

Definición: DLP12 prophage; predicted lipoprotein.



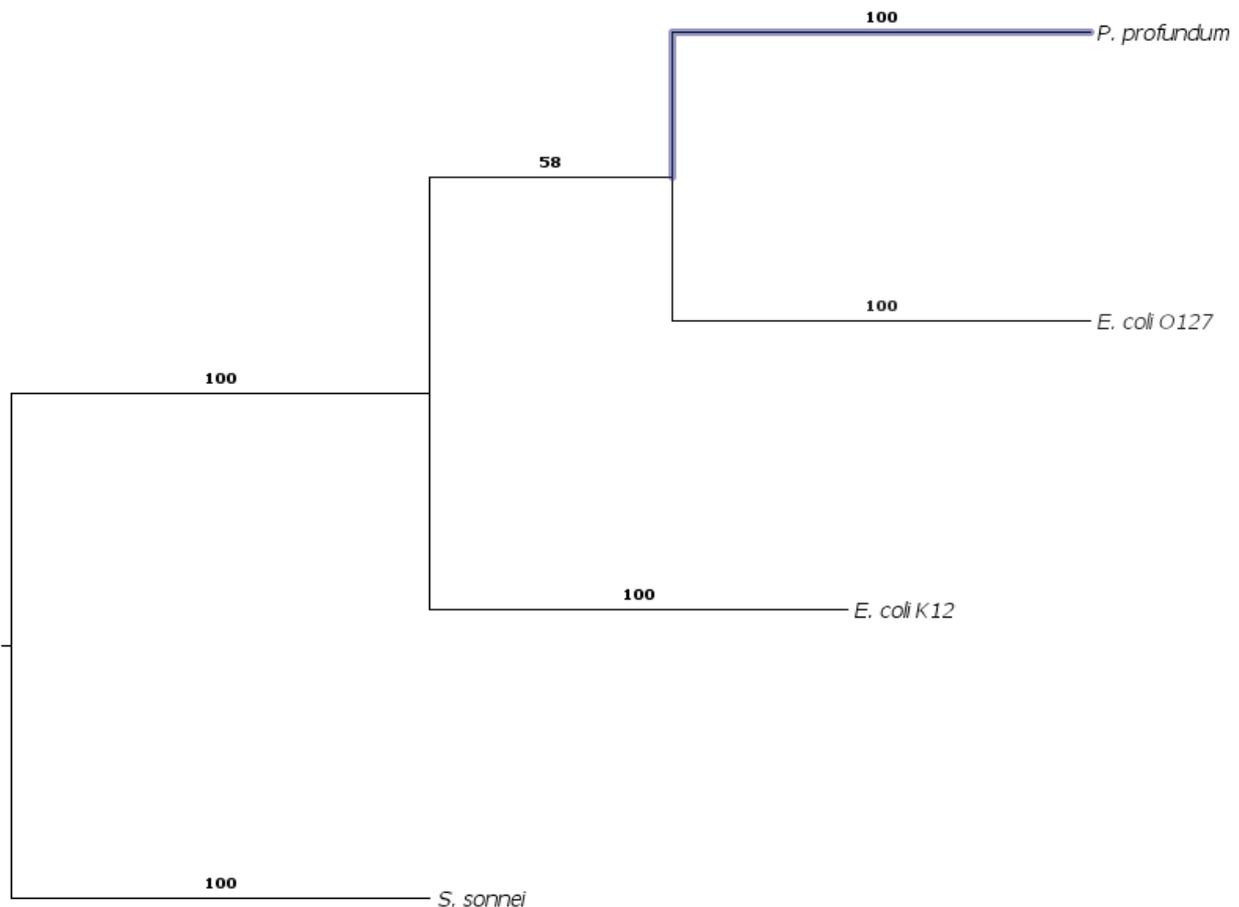
Para el fenograma hecho con base en el gen *eco:b0557* se presenta un caso muy similar al anterior, pero en esta ocasión entre *E. tasmaniensis* y *E. coli K12*. Sin embargo, por las mismas razones expuestas en el caso anterior, no es posible discernir, mediante el método filogenético, que el gen *eco:b0557* sea un xenólogo. Interesantemente, basándonos en la función y origen de la secuencia transferida, este parece ser un caso de transducción (HGT entre un fago y una bacteria), lo que apoyaría fuertemente un caso de HGT. En este caso, un acercamiento metodológico diferente al filogenético podría ser más que apropiado para resolver la duda.

Casos en los que no es posible diferenciar entre ganancia (HGT) o pérdida de genes.

Gen: *eco:b2875*

Nombre del Gen: *yqeB*, *ECK2871*, *JW2843*

Definición: Conserved protein with NAD(P)-binding Rossmann fold.



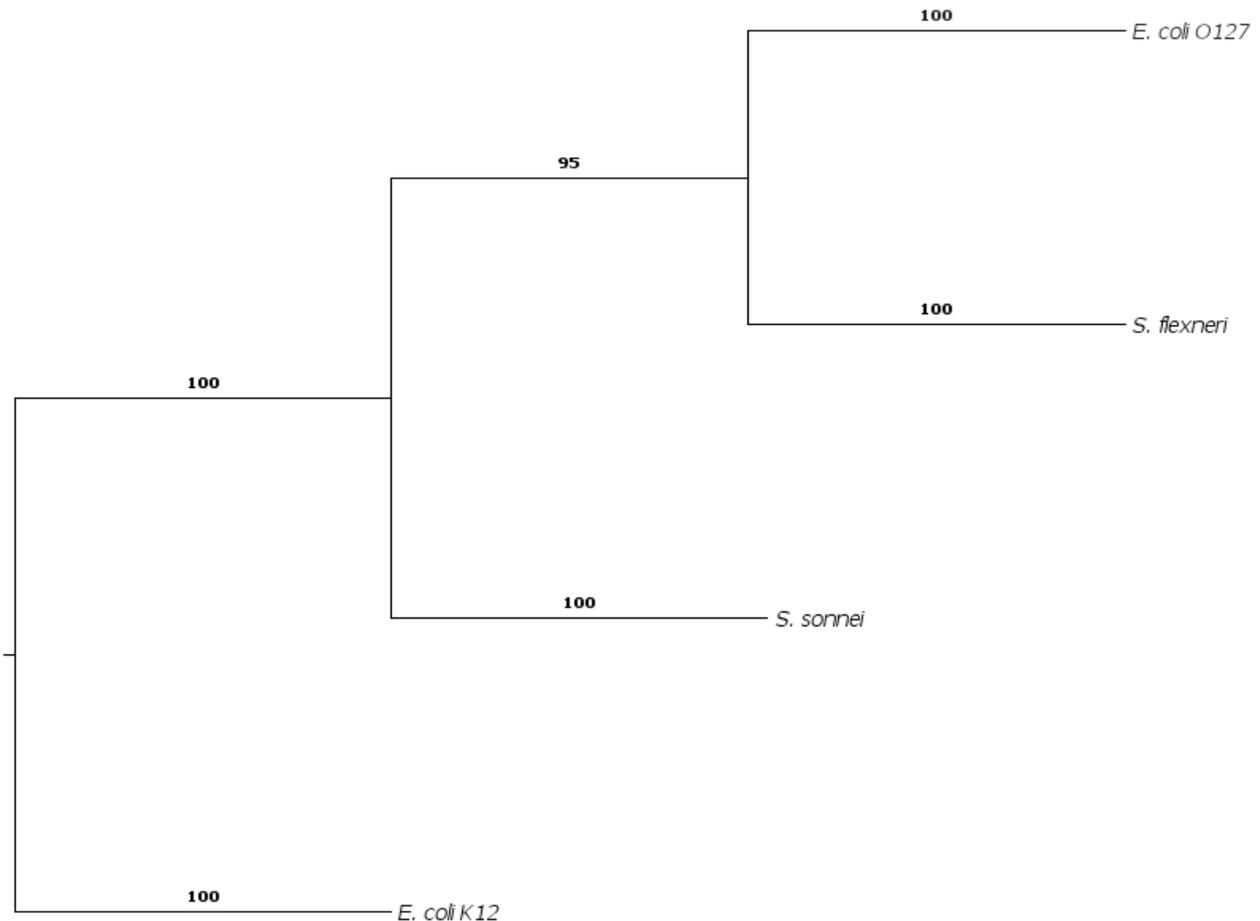
En este fenograma, basado en el gen *eco:b2875* de nuevo se expone un posible caso de HGT entre *E. coli O127* con una bacteria filogenéticamente lejana perteneciente al género *Vibrio*, *P. profundum*. En este caso, a diferencia del anterior, la función de la secuencia transferida no aporta información que de inmediato nos haga sospechar una transferencia horizontal, aunque la información biológica de los organismos involucrados ciertamente lo hace. Aquí, de nuevo, no es posible resolver, mediante la construcción de filogenias, que el gen *eco:b2875* sea un xenólogo. Una aproximación metodológica distinta a la filogenética (como una combinación de metodologías) facilitaría la diferenciación entre ganancia o pérdida de genes e incluso un posible reemplazo de ortólogos.

Casos en los que no es posible diferenciar entre ganancia (HGT) o pérdida de genes.

Gen: *eco:b2770*

Nombre del Gen: *ygcR*, *ECK2765*, *JW5441*

Definición: Predicted flavoprotein.



Este fenograma, en base al gen *eco:b2770*, es un ejemplo de lo que ciertamente resultan ser los casos menos apropiados para hacer inferencias filogenéticas. Aquí, se muestran cuatro *otus* filogenéticamente cercanos entre sí. Basados en ello, no podemos decir si estamos ante cuatro organismos que adquirieron un gen, es una pérdida masiva del mismo en el resto de Gamma-Proteobacteria o de nueva cuenta un reemplazo de ortólogos. La información biológica de los organismos o la función de la secuencia transferida ciertamente no nos ayuda a confirmar o refutar las sospechas. Corroborar que éste es un caso de HGT mediante la construcción de filogenias requeriría la construcción de una que incluya a todos los organismos dentro de Gamma-Proteobacteria, ciertamente una labor compleja y laboriosa.

Perspectivas...

Toda la investigación orientada en el desarrollo de metodologías para detectar xenólogos a lo largo de estos años parece indicar que no existe un único método que nos garantice la capacidad de lograrlo. La mejor opción y posiblemente la única para hacer frente a este tipo de problemas es emplear una combinación de métodos, lo que en principio facilitaría la obtención de resultados confiables. Sin embargo, la cantidad de problemas que la selección misma de los métodos implica, en ocasiones dificulta demasiado el proceso. En primera instancia, la misma elección del método paramétrico que mejor se ajuste al tipo de información que manejaremos no es una decisión sencilla y mucho menos trivial. Está claro, además, que realmente aun no se conocen muchos mecanismos por los que se llevan a cabo intercambios genéticos, y ya no sólo a nivel organismo, sino también a nivel celular. La reciente confirmación de la HGT intercelular y la posible correlación que existe de ésta con el desarrollo de nuevas funciones celulares dentro de un mismo organismo, podría ser una variante o al menos un mecanismo análogo, hasta ahora desconocido, de la HGT.

Así pues, tenemos el trabajo de Holmgren (2010), quien propone que durante el desarrollo del cáncer, los cuerpos apoptóticos podrían ser un vehículo para la transferencia de DNA entre células moribundas y células sanas. Si esto resulta cierto, sería un mecanismo que satisfactoriamente explicaría la heterogeneidad de las células tumorales y la capacidad de estas para adquirir características que les confieran algún tipo de ventaja, como la resistencia a la quimioterapia.

En otro caso, trabajos separados por parte de Thomas M. y Nielsen M. en el año 2005, así como de Doerfler y colaboradores, sugieren que DNA presente en el torrente sanguíneo y el tracto intestinal de algunos animales puede persistir lo suficiente como para entrar en contacto con varios órganos del organismo (Thomas y Nielsen, 2005; Berndt, Meier y Wackernagel, 2003). Que la transformación por células de diferentes órganos pertenecientes al mismo individuo sea un proceso que pueda ocurrir naturalmente es una idea reciente y poco investigada. Hasta el momento resulta muy pronto anticipar y especular acerca de los alcances que fenómenos así tienen y tendrán en la evolución de cierto tipo de organismos, en especial de los multicelulares, usualmente expuestos a grandes cantidades de estrés ambiental y que bien podrían beneficiarse profundamente de mecanismos como éste.

Si bien no hay muchos casos descritos que confirmen la ocurrencia de tales mecanismos, el sólo hecho de que su existencia, al menos de forma teórica, permita dar una explicación razonable a ciertos fenómenos poco conocidos o estudiados, es suficiente para reformularnos de nueva cuenta el alcance que la HGT o fenómenos análogos podrían tener en aspectos en los que hace algunos años hubiese resultado imposible de anticipar. Entender de forma integral el papel de la HGT en la evolución de los organismos es primordial para enfrentar y eventualmente resolver las implicaciones médicas, sanitarias y sociales que este fenómeno está provocando, como la aparición de nuevas cepas bacterianas resistentes a los antibióticos más poderosos y el descubrimiento de que algunos eucariontes, como insectos y mamíferos (algunos roedores, por ejemplo) están desarrollando resistencia a plaguicidas que solían usarse como método de control.

Conclusiones.

En este trabajo hemos puesto a prueba el método filogenético para corroborar los resultados de una metodología igualmente popular en la identificación de genes xenólogos, es decir, el método composicional.

Durante la realización de nuestro trabajo, pudimos confirmar muchas de las ventajas y críticas que se le han hecho al método filogenético y que han sido frecuentemente citadas en la literatura. Con esta idea en mente, deseamos reforzar que el método filogenético es muy confiable para realizar detecciones e identificaciones de HGT; sin embargo, la dificultad para interpretar correctamente los resultados y que el nivel de practicidad es inversamente proporcional a la complejidad requerida para que el método funcione adecuadamente son restricciones importantes que se deben tomar en cuenta cuando se decida utilizar el método. Inclusive, en ocasiones, aun cuando la metodología sea correctamente aplicada, los resultados obtenidos podrían resultar cuestionables. Por lo tanto, considero que el método filogenético es indudablemente apto para enfrentar ciertas cuestiones, pero evidentemente, al igual que con cualquier otra metodología para la identificación de HGT, sus desventajas le impiden ser la mejor opción para enfrentar otros casos.

Parece ser que la consideración más importante en la detección e identificación de genes xenólogos y que aplica a todas las metodologías de análisis, es que la confiabilidad del resultado depende mucho de la correcta selección de la metodología más adecuada para resolver el problema al que nos estamos enfrentando (Becq *et al* 2010). El método filogenético no resulta universalmente infalible para resolver esta cuestión.

Es igualmente importante resaltar los problemas que existen con la información dentro de las bases de datos, que a pesar de ser cuantiosa, no resulta ser adecuada para llevar a cabo estudios filogenéticos, por lo que, en la actualidad, un paso importante en la realización de cualquier trabajo es categorizar la información con la que se cuenta de acuerdo al tipo de estudio que se llevará a cabo (Poptsova y Gogarten, 2007). Las limitaciones que encontramos en el método para poder identificar a aquellas filogenias incongruentes que posiblemente reflejen una transferencia horizontal siempre estuvieron basadas en la dificultad para encontrar información confiable para analizar, así la falta de *otus* para comparar y la ausencia de secuencias ortólogas; o en su defecto, la dificultad para diferenciar éstas de las parálogas, son dos de los problemas más grandes que se enfrentan (y que son intrínsecos a las bases de datos de donde se obtiene información y no al método en sí) para la construcción de filogenias.

A pesar de los muchos acercamientos metodológicos que se han desarrollado y que tienen por objetivo ser la herramienta de identificación de genes xenólogos más eficaz; durante la realización de este trabajo, me he dado cuenta de que la HGT, al ser un fenómeno multifactorial, se puede abordar de mejor manera cuando se combinan distintas herramientas bioinformáticas que permitan enfrentar el problema desde distintos ángulos. Sólo así resulta factible determinar con la mayor precisión posible el papel de la transferencia horizontal de genes en la evolución de los organismos.

Es por esto que la HGT es un fenómeno de gran importancia, sumamente complejo, no trivial y que está lejos de comprenderse.

Desde antes de que se conocieran las implicaciones de la HGT en la evolución, habían dudas respecto a que el efecto acumulativo de mutaciones puntuales simples bastaba para dar explicación al surgimiento de adaptaciones complejas (Goldschmidt, 1960). Recientemente, algunos autores han criticado a la Biología Evolutiva por no incorporar en la teoría general, mecanismos que generan diversidad e innovación genética sobre los que la evolución puede actuar y que podrían representar alternativas para la resolución de las cuestiones teóricas y prácticas que la Teoría Evolutiva Clásica no contempla ni resuelve (Feder, 2007).

La Transferencia Horizontal de Genes no pretende sustituir la visión clásica de la Evolución, que considera que las mutaciones benéficas fijadas y transferidas verticalmente en las poblaciones son el mecanismo primordial para generar innovación genética y Evolución. Simplemente, necesita reconocerse que aunque la fijación de mutaciones puntuales y la frecuencia de la HGT son fenómenos distintos que varían en la influencia que manifiestan a lo largo del “Árbol de la Vida”, sin duda alguna son fuerzas complementarias y no excluyentes (Brown, 2003).

La HGT, como parte del proceso evolutivo, es un concepto que requiere atención y su adecuada incorporación a la teoría evolutiva, y hasta que estas ideas y fenómenos dejen de considerarse como eventos aislados durante el curso de la Evolución y sean finalmente abordadas con el peso y el interés que en realidad se necesita, la teoría evolutiva estará incompleta (Feder, 2007).

Referencias.

- Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990). "Basic local alignment search tool". *J Mol Biol* 215 (3): 403–410
- Altschul S.F., Madden, T.L., Schäffer, A.A., *et al* (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Res.* 25:3389-3402.
- Andersson J. O. (2005). Lateral gene transfer in eukaryotes. *Cellular and Molecular Life Sciences.* 62, 1182-1197.
- Andersson J.O., Doolittle W.F y Nesbo C.L. (2001). Genomics. Are there bugs in our genome?, *Science* 292, pp. 1848–1850.
- Burke J.M y Arnold M.L, (2001). Genetics and the fitness of hybrids, *Annu. Rev. Genet.* 35 pp. 31–52.
- Babic A., Lindner A.B., Vulic M., Stewart E.J., Radman M. (2008). Direct visualization of horizontal gene transfer. *Science.* Mar 14; 319(5869):1533-6.
- Baptiste E., Boucher Y., Leigh J. Doolittle W.F. (2004). Phylogenetic reconstruction and lateral gene transfer. *Trends Microbiol. Sep;*12(9):406-11.
- Becerra A., Delaye L., Islas S. y Lazcano A. (2007). The very early stages of biological evolution and the nature of the last common ancestor of the three major cell domains. *Annual review of ecology evolution and systematics* 38: 361
- Boto L. (2010). Horizontal gene transfer in evolution: facts and challenges. *Proc Biol Sci.* Mar 22; 277(1683):819-27.
- Brown J. (2003) Ancient horizontal gene transfer. *Nat. Rev. Genet.* 4, 121–132.
- Brown J. y Doolittle W. F. (1997). *Microbiol Mol Biol Rev.* 61:456–502.
- Canchaya C., Fournous G., Chibani-Chennoufi S., Dillmann M.L y Brüssow H. (2003). Phage as agents of lateral gene transfer. *Current Opinion in Microbiology* Volume 6 Issue 4. August.
- Cascales E. y Christie P. J. (2003). The versatile bacterial type IV secretion systems. *Nature Rev. Microbiol.* 1, 137–149.
- Cole J.R., Q. Wang., E. Cardenas., J. Fish., B. Chai et al. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 37 (Database issue): D141-D145; doi: 10.1093/nar/gkn879
- Cortéz D., Delaye L., Lazcano A., Becerra A. (2005). Composition-based methods to identify horizontal gene transfer. *Methods Mol Biol.* 2009;532:215–25.
- Daniels S.B, Peterson K.R., Strausbaugh L.D., Kidwell M.G., Chovnick A. (1990). Evidence for horizontal transmission of the P transposable element between *Drosophila* species, *Genetics* 124, pp. 339–355.
- Davies J. (1996). Origins and evolution of antibiotic resistance. *Microbiology* 12; 9-16.
- Deschavanne P, y Filipski J. (1995). Correlation of GC content with replication timing and mechanisms in weakly expressed *E.coli* genes. *Nucleic Acids Res.* 25;23(8):1350-3.
- Doolittle W.F. (1999). Phylogenetic classification and the universal tree. *Science.* 25;284(5423):2124-9.
- Doolittle W.F. (1998). You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes, *Trends Genet.* 14, pp. 307–311.

- Edgar R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792–1797.
- Feder M.E. (2007). Evolvability of physiological and biochemical traits: evolutionary mechanisms including and beyond single-nucleotide mutation. *J Exp Biol.* 210(Pt 9):1653-60.
- Field J., Rosenthal B. y Samuelson J. (2000). Early lateral transfer of genes encoding malic enzyme, acetyl-CoA synthetase and alcohol dehydrogenases from anaerobic prokaryotes to *Entamoeba histolytica*. *Mol Microbiol.* Nov;38(3):446-55.
- Fitzpatrick D.A., Logue M. y Butler G. (2008). Evidence of recent interkingdom horizontal gene transfer between bacteria and *Candida parapsilosis*. *BMC Evol Biol.* 2008. PMID: PMC2459174
- Fraser J. S., Yu Z., Maxwell K. L., Davidson A. R. (2006). Ig-like domains on bacteriophages: a tale of promiscuity and deceit. *J. Mol. Biol.* 359, 496–507.
- Gogarten JP. y Hilario E. (1993). Horizontal transfer of ATPase genes—the tree of life becomes a net of life. *Biosystems.* 1993;31(2-3):111-9.
- Gogarten JP. (2003). Gene transfer: gene swapping craze reaches eukaryotes. *Curr Biol.* Jan 21;13(2):R53-4.
- Gogarten JP., Doolittle, W. F. & Lawrence, J. G. (2002). Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19, 2226–2238.
- Gogarten JP. y Townsend JP. (2005). Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol.* 3:679–687
- Goldschmidt R. B. (1960). *The Material Basis of Evolution.* Paterson, NJ: Pageant Books.
- Guindon S. y Perrière G. (2001). Intragenomic base content variation is a potential source of biases when searching for horizontally transferred genes. *Mol. Biol. Evol.* 18, 1838-1840.
- Hao W., Richardson A.O., Zheng Y. y Palmer J.D. (2010). Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion. *Proc. Natl. Acad. Sci. U.S.A.* 107, pp. 21576–21581.
- Hayes W. y Borodovsky M. (1998). How to interpret an anonymous bacterial genome: machine learning approach to gene identification. *Genome Res.* 8, 1154-1171.
- Holmgren L. (2010). Horizontal gene transfer: you are what you eat. *Biochem Biophys Res Commun.* May 21;396(1):147-51.
- Huson D.H. y Bryant D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23: 254–267.
- Jain R., Rivera MC., Lake JA. (1999). Horizontal Gene Transfer among genomes: the complexity hypothesis. *Proc Natl Acad sci USA.* 96(7):3801-3806
- Johnson M., Zaretskaya I., Raytselis Y., Merezuk Y., McGinnis S., y Madden T.L. (2008) "NCBI BLAST: a better web interface" *Nucleic Acids Res.* 36:W5-W9.
- Kanhere A. y Vingron M. (2009). Horizontal Gene Transfers in prokaryotes show differential preferences for metabolic and translational genes. *BMC Evol Biol.* 9:9.
- Keeling P.J., Palmer J.D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet.* 9(8):605–18. doi: 10.1038/nrg2386.
- Koonin E.V. (2009). Darwinian evolution in the light of genomics. *Nucleic Acids Res.* 37:1011–1034.

- Koonin E., Makarova K., Aravind I. (2001). Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol.* 55:709-742.
- Kondo N., Nikoh N., Ijichi N., Shimada M., Fukatsu T. (2002). Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect. *Proceedings of the National Academy of Sciences of the United States of America* 99, 14280-14285.
- Koski L. B., Morton R. A. y Golding G. B. (2001). Codon bias and base composition are poor indicators of horizontally transferred genes. *Mol. Biol. Evol.* 18, 404-412.
- Kurland C.G., Canback B., Berg O.G. (2003). Horizontal gene transfer: a critical view. *Proc Natl Acad Sci U S A.* 2003 Aug 19;100(17):9658-62. Epub Aug 5.
- Kyrpides N. y Olsen G. J. (1999). Archaeal and bacterial hyperthermophiles: horizontal gene exchange or common ancestry? *Trends Genet.* 15, 298-299.
- Kyrpides N, Overbeek R, Ouzounis C. (1999). Universal protein families and the functional content of the last universal common ancestor. *J. Mol. Evol.* 49:413–23
- Lander S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., Devon K., Dewar K., Doyle M. y FitzHugh W. et al., (2001). Initial sequencing and analysis of the human genome, *Nature* 409, pp. 860–921.
- Lapierre P. y Gogarten J. P. (2009). Estimating the size of the bacterial pan-genome. *Trends Genet.* 25, 107–110.
- Lawrence J. G, Hendrickson H. (2003). Lateral gene transfer: when will adolescence end?. *Mol Microbiol.* Nov;50(3):739-49.
- Lawrence J. G. y Ochman, H. (1997). Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* 44, 383-397.
- Lerat E., Daubin V., Ochman H., Moran N.A. (2005). Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol.* May;3(5):e130.
- Liu Y., Harrison P., Kunin V. y Gerstein M. (2004). Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. *Genome Biol.* 5, R64.
- Molloy S. (2005). Coordinated uptake. *Nature reviews Microbiology* 3, 667.
- Nakamura, Y., Itoh, T., Matsuda, H. y Gojobori, T. (2004). Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nature* 36, 760-766.
- Ochman H., Lawrence J. G. y Groisman E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299-304.
- Ochia, K. Yamanaka, T. Kimura, K. and Sawada, O. (1959). Inheritance of drug resistance (and its transfer) between *Shigella* strains and between *Shigella* and *E. coli* strains. *Nihon Iji Shimpo* 1861: p34
- Ohno S. (1999). Gene duplication and the uniqueness of vertebrate genome circa 1970-1999. *Semin Cell Dev Biol.* 10(5):517-22. Review.
- Ohta T. (1989). Role of gene duplication in evolution. *Genome.*;31:304–310.
- Olendzenski L. y Gogarten J.P. (2009). Evolution of genes and organisms: the tree/web of life in light of horizontal gene transfer. *Ann N Y Acad Sci.* Oct; 1178:137-45.
- Page R. D. M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* 12: 357-358.

- Pennisi E. (1998). Genome data shake tree of life. *Science*. 1;280(5364):672-4.
- Puigbò P., Wolf Y. y Koonin V. E. (2009). Search for a 'Tree of Life' in the thicket of the phylogenetic forest. *J Biol.* 8(6): 59.
- Philippe H. y Forterre P. (1999). The rooting of the universal tree of life is not reliable. *J Mol Evol.* 49:509-523.
- Poptsova M.S. y Gogarten J.P. (2007). The power of phylogenetic approaches to detect horizontally transferred genes. *BMC Evol Biol.* Mar 21;7:45.
- Ragan M.A., Harlow T.J, Beiko R.G. (2006). Do different surrogate methods detect lateral genetic transfer events of different relative ages?. *Trends Microbiol.* Jan;14(1):4-8.
- Rivera, M. C., Jain, R., Moore, J. E. y Lake, J. A. (1998). Genomic evidence for two functionally distinct gene classes. *Proc. Natl. Acad. Sci. USA* 95, 6239-6244.
- Sharp P.M. y Li W.H. (1987). The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* Feb 11;15(3):1281-95.
- Smith M.W, Feng D.F, Doolittle R.F. (1992). Evolution by acquisition: the case for horizontal gene transfers. *Trends Biochem Sci.* 17(12):489-93. Review.
- Stanhope M.J, Lupas A. Italia M.J., Koretke K.K., Volker c. y Brown J.R.. (2001). Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates, *Nature* 411, pp. 940–944.
- Syvanen M. (1985). Cross-species gene transfer: implications for a new theory of evolution. *J Theor Biol.* 21;112(2):333-43.
- Syvanen M. (1998). *Horizontal Gene Transfer*. Chapman & Hall. Kado Editors.
- Thomas C.M. y Nielsen K.M., (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol.* Sep; 3(9):711-21. Review.
- Van Iersel L., Kelk S., Rupp R., Huson D. (2010). Phylogenetic networks do not need to be complex: using fewer reticulations to represent conflicting clusters. *Bioinformatics.* Jun 15;26(12):i124-31.
- Waack S., Keller O., Asper R., Brodag T., Damm C., Fricke W.F., Surovcik K., Meinicke P., y Merkl R. (2006). Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics.* 16;7:142.
- Williams K.P., Gillespie J.J. et al. (2010). Phylogeny of gammaproteobacteria. *J Bacteriol.* May;192(9):2305-14. Epub.
- Whelan S., Liò P. y Goldman, N. (2001). Molecular phylogenetics: state-of-the-art methods for looking into the past. *Trends Genet.* 17, 262-272.
- Woese C. y Fox G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA.* Nov;74(11):5088–5090.
- Woese C., Kandler O., Wheelis M. (1990). "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.". *Proc Natl Acad Sci USA* 87 (12): 4576–9. doi:10.1073/pnas.87.12.4576. PMID 2112744.
- Woese C.R. (2000). Interpreting the universal phylogenetic tree. *Proc Natl Acad Sci USA.*
- Woolfit M., Iturbe-Ormaetxe I., McGraw E.A., O`Neil S.L. (2009). An ancient horizontal gene transfer between mosquito and the endosymbiotic bacterium *Wolbachia pipientis*. *Mol Biol Evol.* Feb;26(2):367-74.

Wu D., Hugenholtz P., Mavromatis K., Pukall R., Dalin E., Ivanova N.N., Kunin V., Goodwin L., Kyrpides N.C., Klenk H.P, Eisen J.A. et al (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature*. Dec 24;462(7276):1056-60.

Yarza P., Richter M., Peplies J., Euzéby J., Amann R., Scheleifer K.H., Ludwig W., Glockner F.O., Roselló-Móra. (2008). The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *R. Syst Appl Microbiol*. 241-50.

Zhaxybayeva O., Lapierre P., Gogarten JP. (2004). Genome mosaicism and organismal lineages. *Trends Genet*. May;20(5):254-60. Review.

Zhaxybayeva O., Gogarten JP., Charlebois J.L. et al. (2006). Phylogenetic analyses of cyanobacterial genomes: Quantification of horizontal gene transfer events. *Genome Res*. 16: 1099–1108.

Figtree. Software para visualizar los árboles filogenéticos.
(<http://tree.bio.ed.ac.uk/software/figtree/>).

Imágenes.

Graham Lawton. "Darwin was wrong". (2009, Enero). *New Scientist*. número 2692.
http://scienceblogs.com/pharyngula/upload/2009/01/ns_cover.jpeg

Ochman, H., Lawrence, J. G. and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299-304. (gráfica).

Doolittle WF. (1999). Phylogenetic classification and the universal tree. *Science*. Jun 25;284(5423):2124-9.