



UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO

FACULTAD DE CIENCIAS

INDICE DE HOMOGENEIDAD DEL DNA COMO
HERRAMIENTA PARA EL ANALISIS EVOLUTIVO DEL
GENOMA

T E S I S
QUE PARA OBTENER EL TITULO DE
B I O L O G A
P R E S E N T A :
DANIELA SOSA PEREDO

DIRECTOR DE TESIS: BIOL. LUIS JOSE DELAYE ARREDONDO



2005



m. 340443

FACULTAD DE CIENCIAS
SECCION ESCOLAR



Universidad Nacional
Autónoma de México

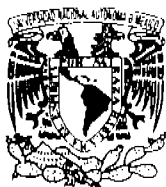


UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

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.



SECRETARÍA DE EDUCACIÓN PÚBLICA
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recepcional.

NOMBRE: Daniela Sosa Peredo

FECHA: 28-enero-05

FIRMA: Daniela Sosa P.

ACT. MAURICIO AGUILAR GONZÁLEZ
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo escrito:

Índice de Homogeneidad del DNA como herramienta para el
análisis evolutivo del genoma.
realizado por Daniela Sosa Peredo

con número de cuenta 9653266-9 , quien cubrió los créditos de la carrera de:

Biología

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis
Propietario

Biol. Luis José Delaye Arredondo

Propietario

Dr. Pedro Eduardo Miramontes Vidal

Propietario

Dr. Germinal Cocho Gil

Suplente

M. en C. Ana María Velasco Velasco

Suplente

M. en C. Sara Ernestina Islas Graciano

Consejo Departamental de Biología FACULTAD DE CIENCIAS

M. en C. Juan Manuel Rodríguez Chávez



UNIDAD DE INVESTIGACIÓN
DE BIOLOGÍA

*A mi mamá, a mi abuela, a Luis F.
a Diego R., y a Rodrigo H.
con todo mi cariño*

Agradecimientos

Indiscutiblemente agradezco a Toño por haber sido la principal fuente de inspiración para indagar en estos terrenos, y por tener abiertas las puertas de su laboratorio en todo momento y enseñarme sin quererlo que hay una vida entera por descubrir...

A Luis Delaye, mi director, por guiarme y darme toda la paciencia y confianza que solo él puede tener, además de ofrecerme este bonito proyecto y convidarme de su maravillosa curiosidad por estos temas y sin duda por ser un buen amigo.

A Luis Fernández no solo por darme el apoyo estadístico en este trabajo si no por todos los años de conocerlo, que en mi vida han significado demasiado, por contar con él incondicionalmente, por ser más que un padre, en verdad muchas gracias por todo. Te quiero muchísimo.

A Pedro Miramontes por ser otra fuente de inspiración bien importante en este trabajo, por tener toda la disposición de apoyarme y por su increíble atención y amistad.

A Germinal Cocho, por abrir estos espacios, por su atención y apoyo.

A Sara, Ana e Irma por compartir conmigo este tiempo en el laboratorio, por todas las buenas pláticas y consejos y porque las admiro y las estimo mucho, gracias por su tiempo y ayuda.

A mi Tía Rosalía por estar conmigo y compartir momentos muy agradables, por sus consejos, por escucharme, por entenderme y sobre todo por confiar en mí, por su gran cariño y por estar cerca de mí.

A mis primas Natalia y Lucía por toda la vida, el amor y la amistad tan fuerte que seguiremos compartiendo.

A mi Tío Julio por ser una maravilla de tío y amigo, a mi Tía Ruth, a Dora a Pato y a Fausto por todo su cariño y apoyo...

A Manuel Fernández, Ade (mamá), Ani Campa, Itza y Héctor por todo el cariño y porque de todos he aprendido mucho...

A Meche, mi angelita, por todo lo que me ayudo a encontrar, lo más valioso de todo lo que he conocido, y por su sincera amistad y a Liz por su gran ayuda y amistad...
A Javier de la Fuente por su sincera amistad y agradable compañía

A mis amigas las más hermosas del mundo, Yssel, Adela, Tana, Claudia, Gaby, Miroslava y Oli por que cada una representa una parte muy importante de mi vida y con cada una tengo muy buenos recuerdos, con los que me quedare toda la vida... Gracias por su apoyo... Las adoro.

Ade gracias por tu cariño, comprensión, apoyo y por los comportamientos propios de una hermana mayor.

Yssel, gracias por todo lo que compartimos y estos años en la fac tan buenos contigo y tu verdadera y gran amistad.

A Rodrigo, Diego, Ezequiel, y Ernesto mis buenos amigos y compañeros por darme tanta alegría y compartir tantos años juntos aunque ahora la distancia y otros caminos nos separen...

A Alejandro Áviles por ser en verdad un gran amigo y una gran persona, gracias por todo...

A Pancho, a Diego y a Rodrigo, por todo el inmenso amor que les tengo y la felicidad que me dieron...

A Israel por llegar en el momento indicado a mi vida por convidarme de su vida, por darme tanto ánimo, apoyo y alegría ...y por el tiempo que nos falte compartir...

A mi abuela Engracia, por todo su amor y su cariño el más fuerte y grande que he recibido, por cuidarme tantos años, por ser mi confidente, mi amiga, y enseñarme tantas y tantas cosas, por apoyarme y escucharme siempre...

Agradezco finalmente a mi mamá, porque sin ella nada de esto hubiera podido ser, por todo el amor que le tengo, que es mi principal motor, por ser la mejor mujer y por ello mi tan especial mamá. Gracias por todos los hermosos años juntas y por todo lo bueno que compartimos y compartiremos siempre, por tu infinito apoyo y sobre todo por ser mi mejor amiga y por todo tu amor.

Índice

<i>Resumen</i>	2
1. Introducción	
1.2 El DNA y los genes	3
1.3 La era de la genómica	6
1.4 El árbol universal de la vida y la transferencia horizontal	7
1.5 El proceso de evolución molecular	14
1.6 El Índice de homogeneidad (IDH)	16
2. Planteamiento del problema	20
3. Objetivo e Hipótesis	21
4. Metodología	22
5. Resultados	28
6. Discusión y Conclusiones	36
<i>Apéndice 1</i>	40
<i>Apéndice 2</i>	44
<i>Apéndice 3</i>	50
<i>Apéndice 4</i>	52
<i>Referencias</i>	53

Resumen

De acuerdo a las características y las propiedades físicoquímicas y termodinámicas del DNA, este presenta una conformación estructural que ha sido poco considerada en el campo de la biología evolutiva. En este trabajo se utilizó un algoritmo de distribución binaria (Índice de homogeneidad del DNA, IDH) propuesto por Miramontes et al, 1995 como herramienta para analizar algunos aspectos evolutivos de genomas. Este índice es un método que tomando en cuenta parámetros físicos y termodinámicos del DNA que afectan su estructura nos da una representación de su conformación estructural, mediante el análisis de genomas completamente secuenciados se ha demostrado que cada especie tiene una conformación particular que se representa gráficamente en un espacio tridimensional Ω , a lo que le llamamos firma del genoma.

Con esto se trataron de abordar algunos problemas de la evolución molecular, puesto que se observó este particular comportamiento del genoma creemos que debe haber una relación evolutiva con esta característica, también se trató de ver si se podía reconocer con esta herramienta un importante problema para la reconstrucción de la historia filogenética, que es el fenómeno de la transferencia horizontal.

Para abordar el primer problema se desarrolló una metodología que trata de buscar una relación entre la distancia filogenética de un gen obtenida por alineación de secuencias y la distancia de la posición de ese gen en la firma estructural al centro de la misma.

Para el segundo problema planteado se graficó, utilizando el IDH, el genoma completo de dos especies, *Giardia intestinalis* y *Rickettsia prowasekii* y dos genes homólogos entre ellas, que en *G. intestinalis* había sido identificado como producto de transporte horizontal.

Se encontró que para el primer problema, si existía una relación entre la distancia evolutiva y la estructural, lo que sugiere que la conformación estructural es debida a un sesgo mutacional, aunque se propone hacer más experimentos utilizando otros ejemplos de genes, y otros análisis estadísticos adecuados para cada experimento.

En el caso de la transferencia horizontal, por lo menos en el ejemplo utilizado no se distinguió claramente la ubicación del gen transferido horizontalmente en las representaciones gráficas, por lo que sugiere que los genes transferidos horizontalmente adquirirán la firma del genoma.

Introducción

1.1 El DNA y los genes

El DNA es la molécula que almacena la información genética, en los sistemas biológicos la información que especifica la estructura primaria de una proteína reside ahí.

Los ácidos nucleicos están formados por un azúcar, una base débil y al menos un grupo fosforilo, hay dos clases generales de nucleótidos los ribonucleótidos, en los cuales el azúcar es la ribosa, y los desoxirribonucleótidos, en los cuales el azúcar es la 2-desoxirribosa. (Horton *et al* 1981)

Las bases que se encuentran en los nucleótidos son pirimidina o purinas. La pirimidina es un compuesto heterocíclico que contiene cuatro átomos de carbono y dos de nitrógeno. La purina es una estructura bicíclica que tiene una pirimidina fusionada a un anillo de imidazol. En el DNA las pirimidinas son timina (T) y citosina (C) y las purinas adenina (A) y guanina (G).

El DNA consiste en dos cadenas antiparalelas de polímeros "lineales" de residuos de desoxirribonucleótidos enlazados por fosfodiésteres 3'-5'. Cada base forma puentes de hidrógeno con una base de la cadena opuesta formando pares de bases en donde solo se encuentran pares de G con C y T con A. Las dos cadenas se enrollan una alrededor de la otra para formar una estructura helicoidal. (Horton *et al*, 1981)

La interacción de pares de bases adyacentes acerca los pares de bases y crea un interior hidrofóbico que da como resultado la torsión del esqueleto de azúcar-fosfato para formar una hélice. Estas interacciones de acomodamiento estabilizan la doble hélice. Debido a la forma en que se apilan los pares de bases para formar una doble hélice, la hélice tiene dos surcos de anchura desigual.

La doble hélice se estabiliza por una variedad de fuerzas que son suficientes para mantener la estructura pero también permiten una flexibilidad de la misma. Mientras los enlaces covalentes definen las estructuras primarias de las macromoléculas las fuerzas débiles dominan el plegamiento de ellas, por ejemplo, los efectos de fuerzas hidrofóbicas mantienen a las purinas y pirimidinas al

interior de la hélice. Los pares de bases apilados forman contactos de van der Waals que en suma proporcionan una fuente importante de estabilidad. Los puentes de hidrógeno entre los pares de bases son una significativa fuerza estabilizadora. Y las interacciones electrostáticas entre grupos fosfodiester negativos generan una potencial inestabilidad que se mejora por las interacciones electrostáticas entre los grupos fosfodiester y los cationes. (Horton *et al*, 1981).

Los genes son secuencias del DNA o RNA genómico que codifican a una función específica, existen tres tipos de genes reconocidos actualmente, genes que codifican para proteínas, que son transcritos en RNA y luego traducidos a proteínas, genes específicamente de RNA que son únicamente transcritos y genes regulatorios. (Li, 1997)

Las mutaciones que ocurren en los genes son errores en la replicación de DNA o en la reparación de DNA, una mutación puede afectar a un nucleótido o a varios nucleótidos adyacentes. La mutación puede ser clasificada por el tipo de cambios causados dentro del evento mutacional: i) sustitución, donde se reemplaza a un nucleótido por otro, ii) recombinación, en donde se incluye el crossing-over, y la conversión, iii) la delección en donde se pierden uno o más nucleótidos y iv) la inserción donde se agrega uno o más nucleótidos. La sustitución de un nucleótido puede ser sinónima o silenciosa, cuando el cambio de nucleótido no afecta la codificación para el aminoácido o no sinónima cuando el cambio afecta al aminoácido y por tanto a la proteína. (Li, 1997) En el Apéndice 4 se muestra una tabla que ejemplifica y explica los tipos de mutaciones (Ridley, 1993).

Los cambios en la secuencia de nucleótidos en el tiempo son el proceso básico de evolución molecular. (Li, 1997) Para reconstruir la historia evolutiva se analizan los cambios en las secuencias de genes homólogos, genes que guardan relación en caracteres y que han divergido de un carácter común ancestral (Fitch, 2000)

Los genes homólogos pueden subdividirse en tres tipos de genes:

Genes Ortólogos, son genes en diferentes especies que evolucionaron de un gen ancestral común por especiación. (Roman

et al., 1997) Esto da origen a un grupo de secuencias cuyas verdaderas filogenias son exactamente las mismas a las verdaderas filogenias de los organismos de los cuales fueron obtenidas. (Fitch, 2000)

Genes Parálogos, cuando los genes homólogos son el resultado de una duplicación al interior del genoma. Tales genes pueden descender y divergir mientras existan lado por lado en el mismo linaje. (Fitch, 2000)

Genes Xenólogos o de transferencia horizontal, son genes homólogos que se transfieren de un organismo a otro. Este fenómeno puede darse por diversos mecanismos como en el caso de bacterias por conjugación, o por vehículos de intercambio genético como plásmidos, transposones, integrones, entre otros (Roy, 1999).

La evidencia de estos casos hasta ahora solo se puede inferir por observaciones en: anomalías en el contenido de GC, los genes transferidos horizontalmente generalmente no tienen el patrón general de los otros genes, la evidencia es mas clara cuando se trata de bacterias Gram + o Gram -, aunque no todas las anomalías en contenido de GC están relacionadas con transferencia horizontal; otra prueba esta en el uso de codones, cuando se tienen de 10 a 20 genes secuenciados ya es posible hacer una tabla del uso de codones que no solo refleja contenidos de GC también la relativa abundancia de tRNA en el organismo, los genes especialmente aquellos de alta expresión de proteínas, tienden a conformar los patrones de uso de codones y las excepciones nos pueden indicar recientes adquisiciones de transferencia horizontal; otro manera de detectar la transferencia horizontal es mediante los árboles filogenéticos, un alto grado de similaridad de un gen con organismos genéticamente muy distantes nos puede indicar transferencia horizontal (Roy, 1999).

1.2 La era de la genómica

Nos encontramos en una época de pleno auge en el campo de la biología y la evolución molecular. La vitalidad de esta última disciplina se ha debido en gran medida a la acumulación de secuencias tanto de genes individuales como de genomas completos provenientes de una gran cantidad de organismos. A la fecha hay más de 200 genomas celulares, aproximadamente 80% de bacterias, 11% de eucariontes y 9% de arqueas completamente secuenciadas y disponibles en bases de datos públicas y en su conjunto representan una muestra de la diversidad genética de la biota. (Fig. 1)

Crecimiento del GenBank

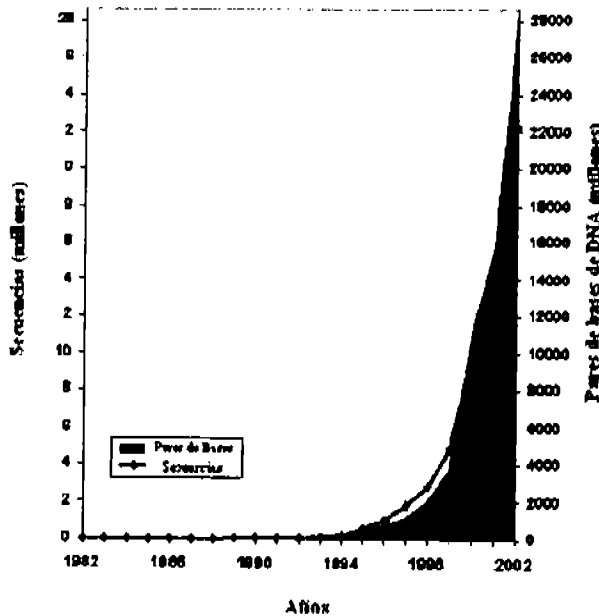


Figura 1. Número de secuencias acumuladas en GenBank

Esta información molecular ha modificado la comprensión del proceso evolutivo, así como de la historia de la vida en la Tierra, como se verá más adelante.

1.3 El árbol universal de la vida y la transferencia horizontal de genes

El sueño de muchos evolucionistas ha sido lograr clasificar a toda la diversidad biológica jerárquicamente en un gran árbol universal (Doolittle, 1999). Sin embargo anteriormente no había sido posible hacer dicho árbol debido a que para comparar y establecer relaciones filogenéticas se habían utilizado básicamente datos morfológicos, lo que implica que para especies fenotípicamente muy distintas no había forma de establecer relaciones evolutivas basadas en caracteres compartidos. Debido a la universalidad de algunos genes (como la chaperonina cpn60, el rRNA, etc.) su uso para establecer relaciones filogenéticas nos permite en principio hacer comparaciones entre todos los seres vivos lo que abre la posibilidad de hacer una filogenia universal.

Los datos morfológicos y la biología de los organismos son parámetros complementarios al árbol universal.

Woese a mediados de los años setenta analizando la molécula de 16S/18S de la subunidad pequeña del RNA ribosomal (cuyas siglas en inglés se abrevian SSU rRNA) propuso una filogenia universal que muestra una clara división entre tres grandes dominios: Archaea Eucarya y Bacteria. (Figura 2).

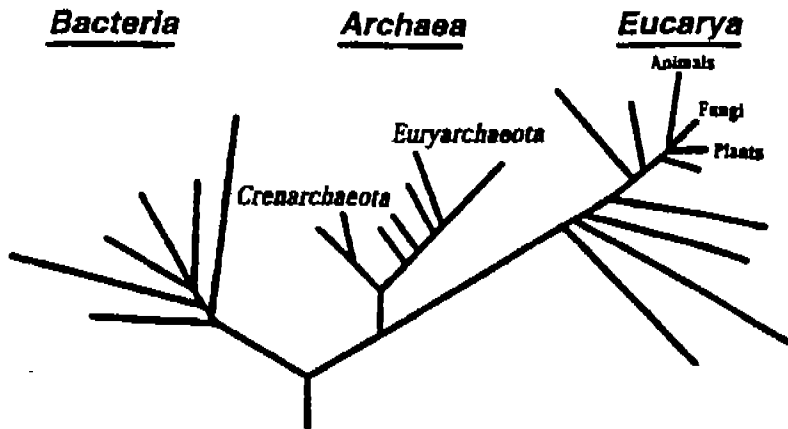


Figura 2. Clasificación general de los tres dominios, (Woese, 2000)

Inicialmente se consideró que el 16S/18S rRNA era un marcador molecular fiel de las relaciones filogenéticas entre las especies, debido a características tales como, estar presentes en todos los organismos celulares, tener una estructura universalmente conservada, además de ser una molécula esencial y fundamental en la funcionamiento celular con una tasa de cambio lenta. Dicha filogenia universal, fue modificada posteriormente para incluir el proceso de endosimbiosis Figura 3 (Doolittle, 1999) que origino a las mitocondrias y a los plástidos (Margulis, 1981).

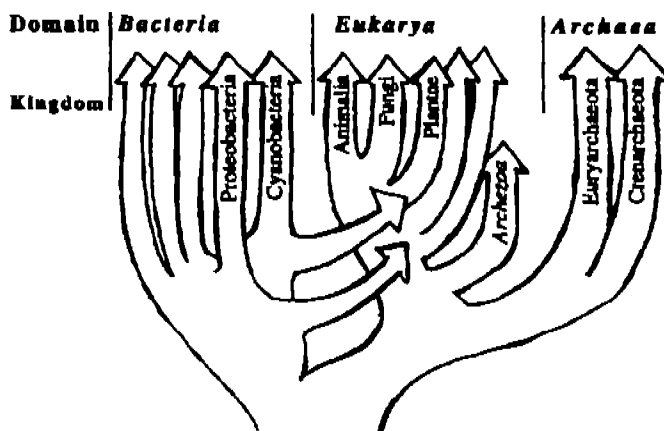


Figura 3. Representación del árbol filogenético con los tres dominios incorporando el fenómeno de endosimbiosis. (Doolittle, 1999)

Finalmente en 1989, de manera independiente Iwabe *et al* y Gogarten *et al*. utilizando dos grupos distintos de genes parálogos sugirieron que la raíz del árbol universal, se encontraba en la rama de las bacterias (Figura 4).

Evolution: Gogarten *et al.*

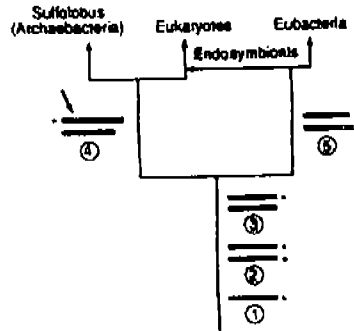


Figura 4a. Relaciones filogenéticas de arqueobacterias, eubacterias, eucariontes y organismos endosimbóticos derivados de la comparación de la subunidad de H⁺ - ATPasa. (Gogarten *et al*, 1989).

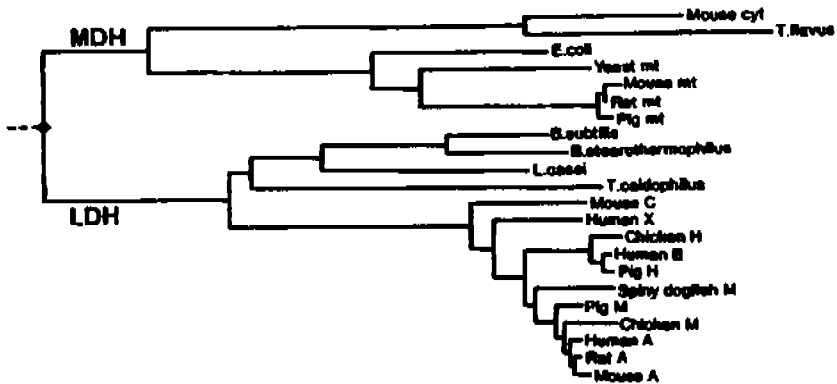


Figura 4b. Árbol filogenético inferido por la simultánea comparación de las enzimas maltato dehidrogenasa MDH y lactato dehidrogenasa LDH (Iwabe *et al*, 1989)

Con la filogenia universal de Woese, el origen endosimbótico de mitocondrias y plástidos, y la raíz del árbol universal en el linaje bacteriano, la historia evolutiva de la vida en la tierra parecía estar resuelta a gran escala.

Sin embargo, el optimismo en la filogenia del 16S/18S rRNA como marcador universal se cuestionó cuando se comenzaron a hacer otras filogenias universales basadas en otros genes

conservados en los tres linajes celulares y provenientes de genomas completamente secuenciados (Doolittle, 1999).

Las distintas filogenias moleculares mostraron que diferentes grupos de genes homólogos de un mismo grupo de organismos presentan árboles filogenéticos con topologías distintas (Brown *et al* ,1997). (Figura 5)

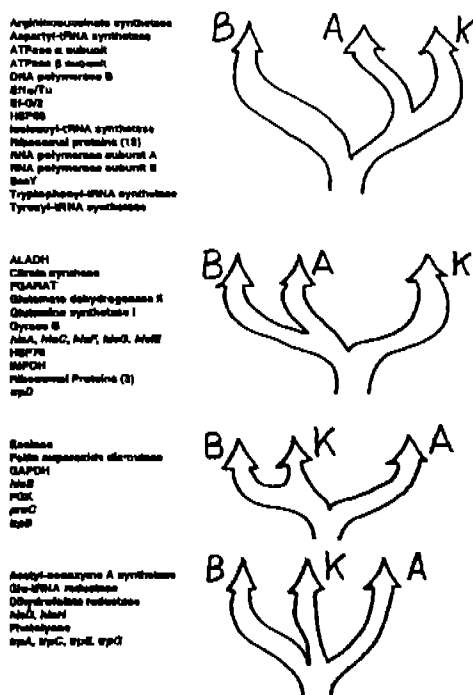


Figura 6. Alternativas de Topologías obtenidas al comparar genes parálogos presentes en los tres dominios. (Brown y Doolittle, 1997)

Este patrón fue atribuido al fenómeno de la transferencia horizontal de genes, entre especies. De acuerdo a Wolf *et al.* (2002) la transferencia horizontal representa una de las fuerzas principales del fenómeno evolutivo cuando menos en procariontes. Por ello se ha sugerido que para representar la historia evolutiva de una serie de linajes hay que considerar la historia de todos sus genes. (Doolittle, 1999). Esto hace que la historia de las ramas se anastomosen frecuentemente, (Figura 6).

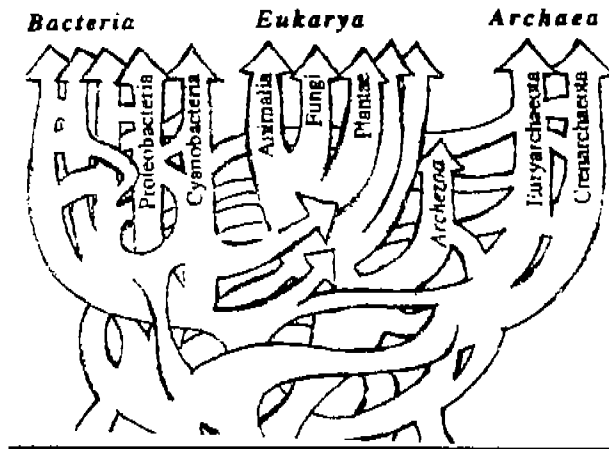


Figura 6. Árbol que representa la nueva propuesta de la historia evolutiva entre los tres dominios. (Doolittle W. 1999)

Sin embargo la intensidad del fenómeno de transferencia horizontal ha sido cuestionada ya que análisis hechos a partir de comparar el número de familias de genes que comparten los genomas completamente secuenciados, sugiere la existencia de los tres linajes celulares (Figura 7). (Tekaia *et al.* 1999). y la sugerencia de que muchos de los supuestos casos de transferencia horizontal, pueden ser debidos a errores metodológicos en el proceso de reconstrucción filogenética. Por ejemplo se ha observado que la concatenación de genes universalmente conservados tiene como resultado la obtención de filogenias robustas (Brown *et al.*, 2001) en donde se observan los tres dominios. (Figura 8). También se ha sugerido que distintos linajes celulares han estado sujetos a distintas tasas de transferencia horizontal (Zhaxybayeva *et al.*, 2004).

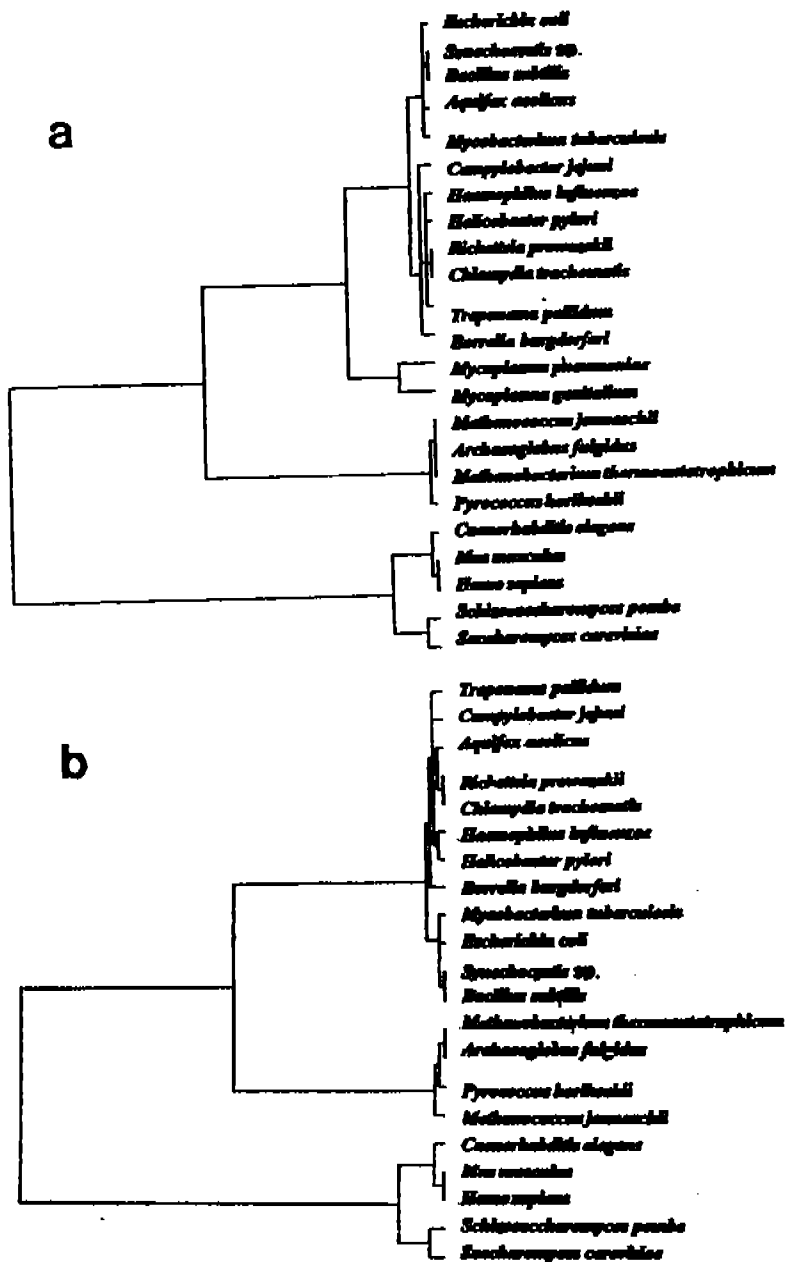


Figura 7. Árboles genómicos de los productos de los ORF en genomas completos (Tekala et al, 1999)

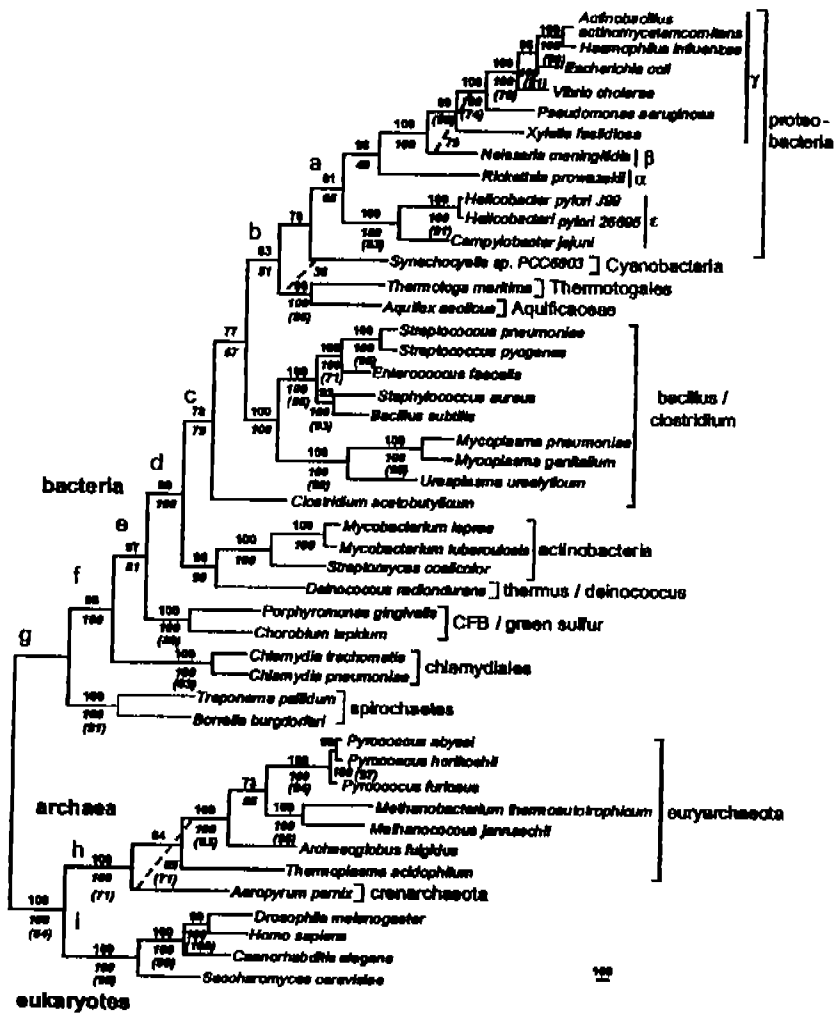


Figura 8. Árbol filogenético obtenido de la identificación y concatenación de 23 proteínas conservadas en todas las especies. (Brown *et al.*, 2001)

En este sentido sigue siendo un debate la cantidad de transferencia horizontal que ha existido entre los linajes de microorganismos. Es por ello que es necesario desarrollar nuevas metodologías para analizar las secuencias de DNA que nos permitan arrojar luz sobre el problema planteado, es decir, que tan común ha sido el fenómeno de transferencia horizontal y en que etapa de la evolución se hace más difícil dilucidarlo.

1.4 El proceso de evolución molecular

En su forma más reduccionista el proceso de evolución molecular consta de dos pasos. (Figura 9.)

- a). Se genera variación por mutación (inserción /delección, recombinación, conversión génica y mutación puntual etc.)
- b). Las fuerzas de selección natural o deriva génica determinan el destino de la variación. (Nei, 2000)

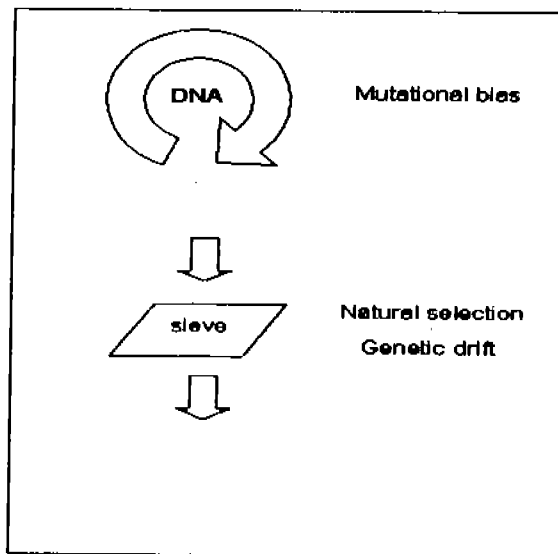


Figura 9. Esquema del proceso de las mutaciones del DNA

Las secuencias que están depositadas en las bases de datos son en última instancia el resultado de este proceso básico.

Desde los años 60 han existido dos modelos básicos que pretenden explicar el proceso de evolución molecular, el modelo seleccionista sugiere que la selección natural de mutaciones ventajosas es la fuerza más importante que dirige la sustitución alélica, y por otro lado el, modelo neutralista que sugiere que el proceso de sustitución esta gobernado principalmente por la deriva de mutaciones selectivamente equivalentes. (Page y Holmes, 1998)

Ambos modelos sugieren que existe tanto selección positiva, como negativa y variantes alélicas selectivamente equivalentes (neutras), sin embargo los modelos difieren en la proporción de las mutaciones que son neutrales o selectivamente ventajosas (Figura 10).

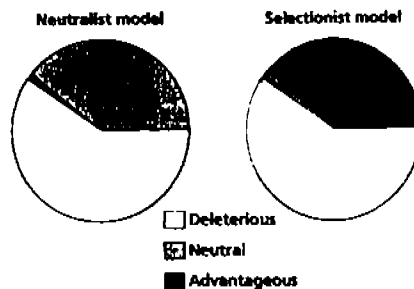


Figura 10. Proporciones de las mutaciones sugeridas por las dos escuelas. (Page y Holmes, 1998)

Si bien es claro que existen ejemplos tanto de evolución neutral como de evolución selectiva, el debate entre ambos modelos se centra en la relativa importancia (frecuencia) de ambos procesos. Tal vez, uno de los aspectos más importantes de la teoría neutral es que explica (en mayor o menor medida) ciertos aspectos de la evolución molecular que se mencionan a continuación: (i) existe una correlación entre la tasa de sustitución y el grado de restricción funcional de un gen, (ii) los patrones de composición de bases y el uso de codones reflejan presiones mutacionales, (iii) existe una tasa constante o un reloj molecular en la evolución de secuencias, y (iv) el nivel de variación genética entre especies es producida principalmente por el tamaño de la población y su tasa de mutación y es correlativa con el nivel de divergencia de secuencias entre especies. (Page y Holmes, 1998)

Es razonable concluir que tanto la selección natural como la deriva génica determinan el destino evolutivo de las mutaciones. Los neutralistas están probablemente correctos en creer que la mayoría de las sustituciones de nucleótidos son neutras en las regiones no codificantes de DNA y en sitios sinónimos. Muchos seleccionistas están también de acuerdo con esto. Sin embargo si vemos la dinámica de la evolución molecular dentro de pequeñas ventanas de tiempo en particular a sitios no sinónimos, podemos claramente identificar vestigios de la selección natural. Para una completa resolución del debate entre neutralistas y seleccionistas sin duda requerimos de métodos estadísticos para distinguir entre estos dos procesos, particularmente requerimos de aquellos métodos que incorporan más información sobre las relaciones filogenéticas de las secuencias en cuestión.

1.5 Índice de homogeneidad del DNA (IDH)

Las propiedades termodinámicas del DNA están estrechamente relacionadas con la estructura del DNA. (Chuprina *et al.* 1991; Dickerson 1992; Breslauer *et al.* 1986). La rugosidad, la lisura, (refiriéndose la rugosidad a una mayor varianza entre los ángulos estructurales de la molécula a diferencia de la lisura), la flexibilidad y la estabilidad del DNA pueden cambiar de acuerdo la composición de nucleótidos, la cual puede no ser uniforme a lo largo de la secuencia (Dickerson 1983), (Delcourt and Blake 1991). Tomando en cuenta estas condiciones Miramontes *et al.* propone un índice de homogeneidad (IDH) basado en un modelo de distribución binaria que cuantifica agregaciones de los nucleótidos presentes en la estructura primaria del DNA tomando en cuenta tres parámetros básicos. El primero mide la distribución local de enlaces energéticos dependiente de los enlaces fuertes CG (S) y débiles AT (W) en las bases apareadas a lo largo de la cadena. (Breslauer *et al.* 1996; Delcourt and Blake 1991). El siguiente parámetro toma en consideración es la distribución de purinas (R) pirimidinas (Y) a lo largo de una hebra de DNA, que influyen en la distancia adyacente entre fosfatos debido a los posibles choques estéricos cuando hay purinas o pirimidinas adyacentes (Calladine 1982) y por lo tanto esta relacionada con la tersura o rugosidad del DNA. El tercer parámetro agrupa a la adenina y citosina (M bases) por tener un grupo amino orientado hacia la ranura mayor de la doble hélice y a la guanina y timina (K bases) con un grupo cetona. (Figura 11.)


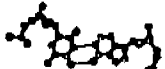
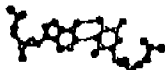
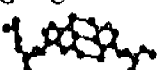
	dYR: R purina Y pirimidina	dWS: forma enlaces W débiles S fuertes	dMK: grupo orientado a la ranura mayor M amino K cetona
C - G 	R	S	K
T - A 	R	W	M
A - T 	Y	W	K
G - C 	Y	S	M

Figura 11. Esta figura muestra la diferencia de los Parámetros del IDH en cada nucleótido. Se hace referencia a los nucleótidos marcados en azul.

Los efectos de las agregaciones de *M* o *K* no han sido muy estudiados pero pueden tener importancia por la Interacción Iónica con otros metales, proteínas y con la misma cadena de DNA. (Marzilli and Kistenmacher 1977).

Una secuencia de DNA puede ser representada basándose en estas tres diferentes características con el IDH. Por ejemplo primero se le asigna un valor de 0 a las *Y* bases y de 1 a las *R* bases y repliando análogamente la sustitución para los casos *WS* y *MK*.

Definiendo al IDH como *d*, entonces:

$$d = \frac{(N_{00}N_{11}) - (N_{01}N_{10})}{(N_0N_1)}$$

donde N_{ij} es el número de ij dinucleótidos ($ij=00,10,01,11$) y N_0 y N_1 son el total de ceros y unos que hay en una secuencia y su producto normaliza el Índice en un intervalo de $[-1,1]$.

Este índice nos permitirá saber que tan agregados están los ceros y unos en la secuencia, si el término izquierdo del numerador se encuentra es dominante entonces D será mayor que 0 y si por otro lado la secuencia tiene mayor ceros y unos en alternancia entonces d será menor que 0.

Finalmente cuando el número de los cuatro diferentes dinucleótidos tengan aproximadamente el mismo valor entonces d tendrá un valor cerca de 0.

Con esto podemos representar cualquier secuencia de DNA en un espacio tridimensional Euclidiano el cual es definido por la identificación de cada secuencia con su vector (d_{YR}, d_{WS}, d_{MK}) . El codominio de este mapa es el paralelepípedo $\Omega = [-1, 1]^3$. Cualquier secuencia concebible de nucleótidos corresponde a un punto en este espacio Ω de la secuencia de DNA. (Figura 12).

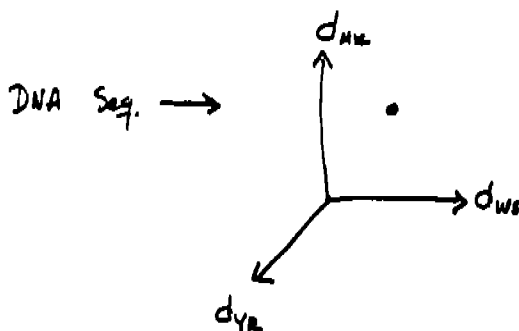


Figura 12. Espacio en Ω que ocupa una secuencia de DNA

Con el mapeo del IDH de los genes disponibles en aquella época (Miramontes *et al.* 1995), encontraron que los genes de cada especie se agregan en el espacio formando una nube de puntos. (Figura 13) También se encontró que parecía haber una diferencia significativa en el espacio ocupado por los genomas de bacterias, eucariontes (Miramontes *et al.* 1995).

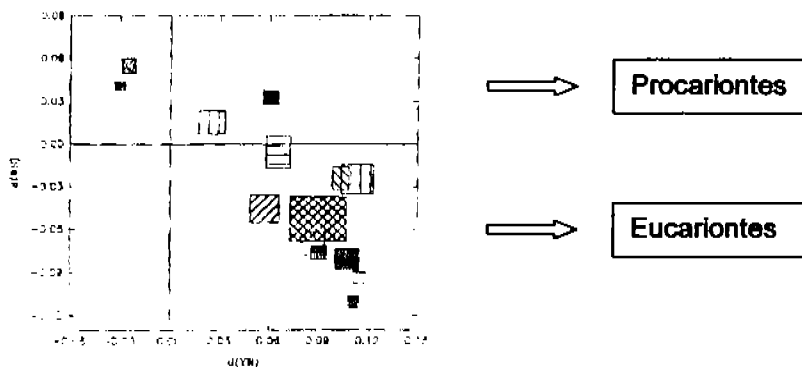


Figura 13. Representación de 13 organismos ubicados en los cuadros marcados en la gráfica. Cada cuadro representa aproximadamente el 95% de los genes analizados de distintas especies (tomada de Miramontes *et al.*, 1995).

Es significativo encontrar que cada genoma tiene un estilo de organización particular de acuerdo al IDH. En esta figura se analizan dos de los parámetros propuestos para el Índice de homogeneidad, (esto debido a que conociendo dos de los parámetros queda determinado el tercero) además de que los dos parámetros utilizados en esta representación tienen implicaciones concretas en la estructura y estabilidad de la cadena de DNA a diferencia del parámetro d_{MK} descrito anteriormente, con los genes secuenciados en esta figura 13 se muestra una separación entre genes procariontes y eucariontes, sin embargo no hay distinción entre los subgrupos de dichos dominios. Cuando se utilizan los tres parámetros como se muestra en la Fig. 14 se puede observar la dispersión de puntos con una muy significativa agregación de los genes en una firma, y claramente se ven las dos nubes de puntos de uno y otro genoma en el espacio Ω , las dos especies filogenéticamente alejadas mantienen su propio espacio conformacional. Este patrón se observa en todos los genomas hasta ahora secuenciados, dejando claro que estructuralmente el genoma mantiene un patrón definido en una firma o nube de puntos que se muestra de forma agregada.

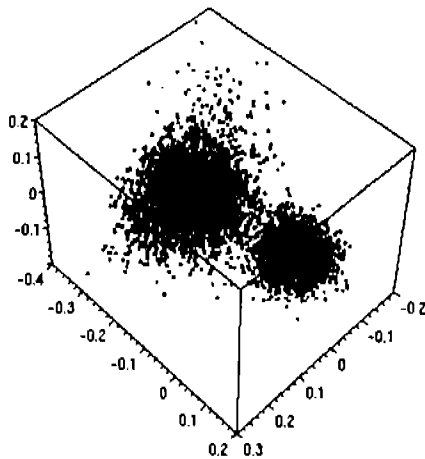


Figura 14. Representación de dos genomas *Acrobacter sp.* y *Rattus norvegicus* en el espacio Ω

2. Planteamiento del problema

De acuerdo a los tres índices IDH (d_{YR} , d_{WS} , d_{MK}), cada genoma tiene un estilo de organización propio. Pero, ¿cuál es la razón de este estilo de organización particular?, ¿cuál es el fenómeno evolutivo que esta detrás de este patrón?

Si el proceso de evolución molecular depende en última instancia de la dinámica entre la mutación (como fuente de variación genética), la deriva génica y la selección natural, que son las fuerzas que determinan el destino de la variación. Entonces, la firma estructural de cada genoma debe poder explicarse en términos de la interacción de estas tres fuerzas básicas.

Si la firma de cada genoma esta determinada por selección natural, entonces cabe esperar una correlación entre los parámetros estructurales y alguna característica ambiental. Por otro lado, si la firma estructural no esta determinada por selección, entonces es probable que se deba a un sesgo mutacional; ya que es posible que la deriva génica por si sola no pueda generar este patrón, en este caso, deberá existir una correlación entre la tasa de evolución y algún parámetro del IDH.

3. Objetivo e Hipótesis

Objetivo general: Estudiar el proceso de evolución molecular en el espacio Ω , el cual que representa los parámetros estudiados.

Objetivo particular: Estudiar si la firma de cada genoma en el espacio Ω esta determinada por un sesgo en la mutación.

Hipótesis: Si la firma estructural de un genoma en el espacio Ω está determinada por un sesgo en la tasa de mutación (μ), entonces las secuencias con una mayor tasa de evolución acumularán más mutaciones con la firma del genoma. De acuerdo a ello, hay dos predicciones que se deben cumplir,

- i) A mayor tasa de evolución molecular las secuencias van a estar más cercanas al centro de la nube de puntos del IDH. Es decir, van a tener más mutaciones con el "sello de la casa".
- ii) Los genes de transferencia horizontal adquirirán la firma del genoma.

4. Metodología

Para indagar las predicciones mencionadas en la hipótesis, se realizaron dos experimentos que se describen a continuación.

I) A mayor tasa de evolución molecular las secuencias van a estar más cercanas al centro de la nube de puntos del IDH.

Para poder estudiar la relación entre la tasa de evolución molecular de un gen y su firma estructural, se decidió comparar la variación en las tasas de evolución de un grupo de genes homólogos, con respecto a la variación que existe en la distancia en el espacio de la posición de cada gen con respecto al centro de cada uno de los genomas respectivos (Fig. 15)

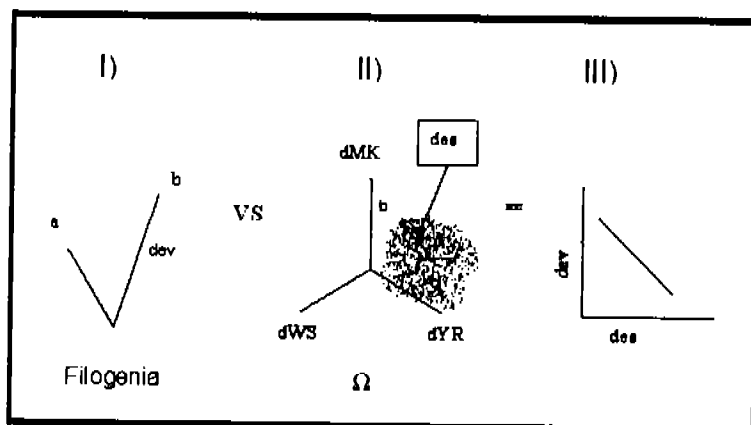
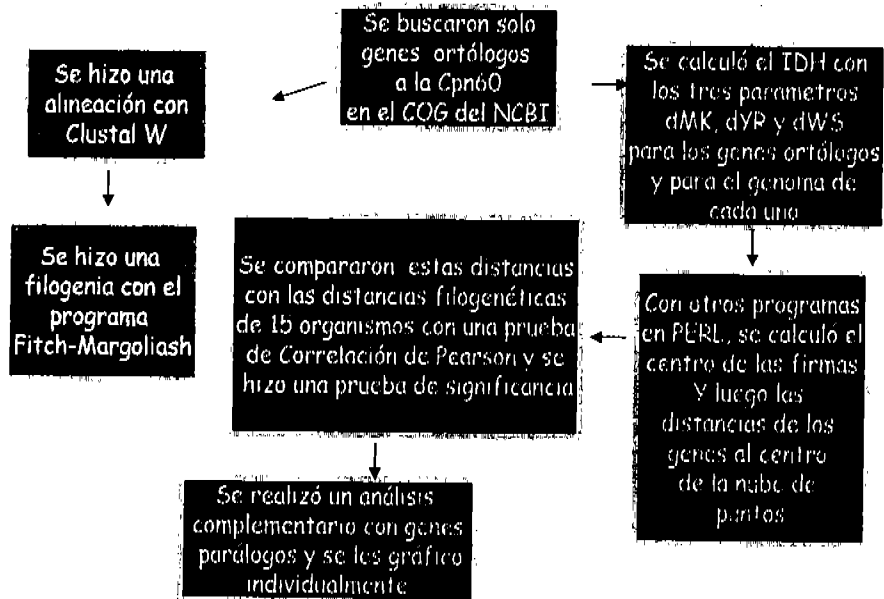


Fig. 15 Si la firma estructural de un gen está determinada por un sesgo en la mutación, entonces genes que han acumulado mutaciones más rápidamente que otros (I) tendrán ramas más largas en las filogenias y tendrán distancias más cortas al centro de la nube de puntos (II) por acumular más mutaciones con el "sello de la casa". La predicción es que habrá una relación proporcional entre ambas distancias (dev: distancia evolutiva; des: distancia estructural).

Se escogió al gen de la chaperonina con ID cpn60 de *Giardia Intestinalis* por ser un gen universalmente conservado y que presumiblemente tiene la misma antigüedad en todos los genomas, salvo un caso de transferencia horizontal (ver más adelante).

Esquema del desarrollo de la metodología



Como primer paso, se obtuvieron los genes ortólogos de la cpn60 de la base de datos del Cluster de grupos ortólogos COG depositada en www.ncbi.nlm.nih.gov/COG/, así como sus genomas completos, para el análisis estructural que se describe más adelante.

Tabla 1. Genes ortólogos de la cpn60 de *G. intestinalis*

ESPECIE	GEN	ESPECIE	GEN
<i>Archaeoglobus fulgidus</i>	AF2238	<i>Caulobacter crescentus</i>	CC0685
	AF1451	<i>Pseudomonas aeruginosa</i>	PA4385
<i>Saccharomyces cerevisiae</i>	YBR044	<i>Vibrio cholerae</i>	VCA082
	YLR259		VC2664
	YDR188	<i>Pasteurella multocida</i>	PM1107
	YJL111	<i>Haemophilus influenzae</i>	HI0543
	YIL142	<i>Buchnera aphidicola</i> APS	BU019
	YDR212	<i>Neisseria meningitidis</i> Z2491 (serogroup A)	NMA047
	YJL008	<i>Neisseria meningitidis</i> MC58 (serogroup B)	NMB197
	YJR064	<i>Xylella fastidiosa</i> 9a5c	XF0615
	YDL143	<i>Aquifex aeolicus</i>	aq_220
	YJL014	<i>Campylobacter jejuni</i>	Cj1221
<i>Mycoplasma genitalium</i>	MG392	<i>Helicobacter pylori</i> 399	jhp000
<i>Chlamydomytila pneumoniae</i> CWL029	CPn089	<i>Helicobacter pylori</i> 26695	HP0010
	CPn077	<i>Borrelia burgdorferi</i>	BB0649
	CPn013	<i>Treponema pallidum</i>	TP0030
	CT755	<i>Thermotoga maritima</i>	TM0506
<i>Chlamydia trachomatis</i>	CT604	<i>Halobacterium</i> sp. NRC-1	VNG209
	CT110		VNG222
	BH0562		<i>Thermoplasma acidophilum</i>
<i>Streptococcus pyogenes</i> SF370 (serotype M1)	SPy207	Ta0980	
<i>Synechocystis</i> sp. PCC6803	slr207	<i>Aeropyrum pernix</i>	APE207
	slr041		APE090
<i>Lactococcus lactis</i>	L19889	<i>Methanobacterium thermoautotrophicum</i>	MTH794
<i>Mycobacterium tuberculosis</i> H37Rv	Rv3417	<i>Methanococcus jannaschli</i>	MTH218
	Rv0440		MJ0999
<i>Deinococcus radiodurans</i>	DR0607	<i>Pyrococcus abyssi</i>	PAB234
		<i>Pyrococcus horikoshii</i>	PH0017
<i>Mesorhizobium loti</i>	mlt223	<i>Rickettsia prowazekii</i>	RP626
	mir239	<i>Giardia intestinalis</i> <i>Mycobacterium leprae</i>	gij2772932
	mlt820		ML0381
	mir934		ML0317
	mlt581		

Posteriormente, se realizó una alineación de los genes ortólogos utilizando el programa ClustaW (Thompson *et al*, 1994). A continuación, se elaboró una filogenia (con el programa Fitch-Margoliash), de la cual se obtuvieron las distancias evolutivas (Apéndice 1). De esta filogenia, utilizamos una sección del árbol (Fig. 16) en donde las relaciones filogenéticas de los genes corresponden a la biología de los organismos, para compararlos con los resultados de cuantificaciones en las variaciones locales con el IDH.

Para tener los datos trabajados con el análisis estructural, se obtuvo primero el IDH con un programa en PERL en los tres tipos de dicotomías descritas (d_{YR} , d_{WS} y d_{MK}) para los genes ortólogos mostrados en la Fig. 16, a fin de conocer su posición en el espacio dentro de un plano euclidiano (Ω). A los genomas de los organismos analizados se aplicó el mismo análisis, con el objeto de ubicar también su posición. Luego, con el mapeo de estos genomas completos se buscó un centro relativo en la nube de puntos con un programa en PERL.

Con otro programa realizado en PERL, se calculó la distancia de este centro a la posición del gen ortólogo, generando una matriz de datos que fue comparada con las distancias evolutivas (en términos relativos) que generó la filogenia de Fitch-Margoliash obtenida (Tabla 2).

Se calculó el coeficiente de correlación de Pearson, entre los resultados obtenidos con la filogenia de Fitch-Margoliash y las distancias estructurales obtenidas con el IDH (la Gráfica 1 muestra esta relación). Posteriormente, se realizó una prueba de significancia (prueba de hipótesis estadística) del coeficiente de correlación obtenido en la muestra.

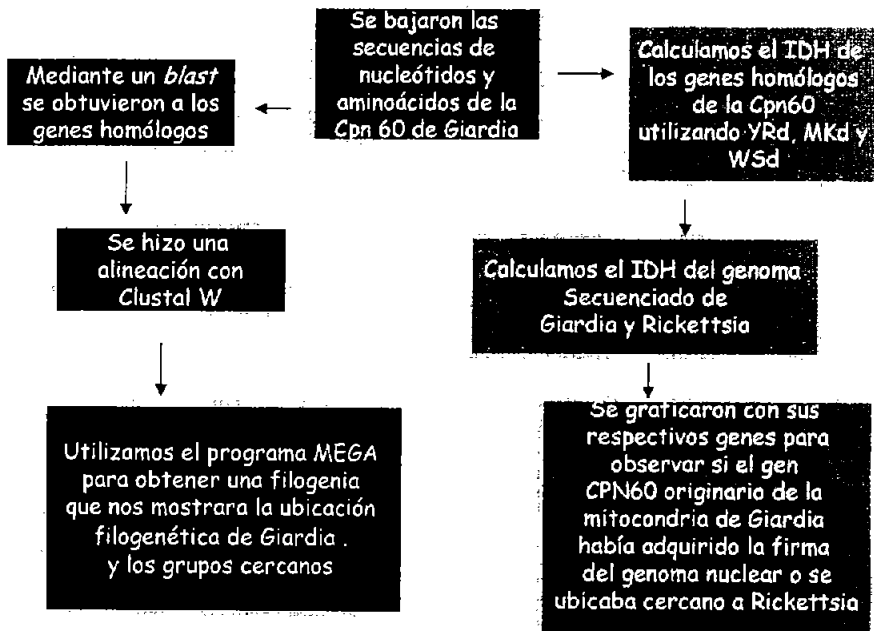
Se realizó un análisis complementario de la siguiente forma. De la misma filogenia de ortólogos a la Cpn 60 (Apéndice 1) se buscaron genes parálogos y se estudió la relación entre los datos obtenidos de las distancias trabajadas con el IDH, anteriormente mencionadas (Tabla 3).

Por último, se graficaron las distancias obtenidas en la filogenia de los doce genes parálogos, contra las distancias obtenidas con el IDH.

ii) Los genes de transferencia horizontal adquirirán la firma del genoma.

Utilizando el gen de la Cpn60 de *G. intestinalis*, reconocido como evidencia de que los diplomonados tuvieron un endosimbionte mitocondrial que transfirió este gen al núcleo antes de la pérdida secundaria del endosimbionte (Roger AJ, *et al.* 1998),

Esquema del desarrollo de la metodología



Se realizó un *blast* contra la base de datos del KEGG a fin de obtener todos los genes homólogos a dicha proteína, estos genes se muestran en la Tabla 4 del Apéndice 3.

De los genes obtenidos en el KEGG se tomaron las proteínas (en formato fasta) que codifican estos genes, y fueron alineados con el programa ClustalW. Posteriormente se realizó una filogenia con el método *Neighbor-Joining* en el programa Mega 2.1 Software (Kumar, *et al.* 2001). (Figura 17.) para mostrar el origen mitocondrial de dicho gen.

Se mapearon los genomas de *Rickettsia prowazekii* (por ser el genoma bacteriano de vida libre más cercano filogenéticamente a las mitocondrias) y *G. intestinalis* en el espacio Ω y se identificaron los dos genes homólogos de la cpn60 (gi|2772932| y RP626). A fin de presentar gráficamente la ubicación de los genes homólogos con respecto a la nube de puntos del resto del genoma (Gráfica 2).

5. Resultados

i) A mayor tasa de evolución molecular las secuencias van a estar más cercanas al centro de la nube de puntos del IDH.

Se obtuvo la Filogenia que se muestra en el apéndice 1 a partir de los genes ortólogos de la Cpn60, y de esa filogenia se utilizó la sección que se muestra en la Figura 16.

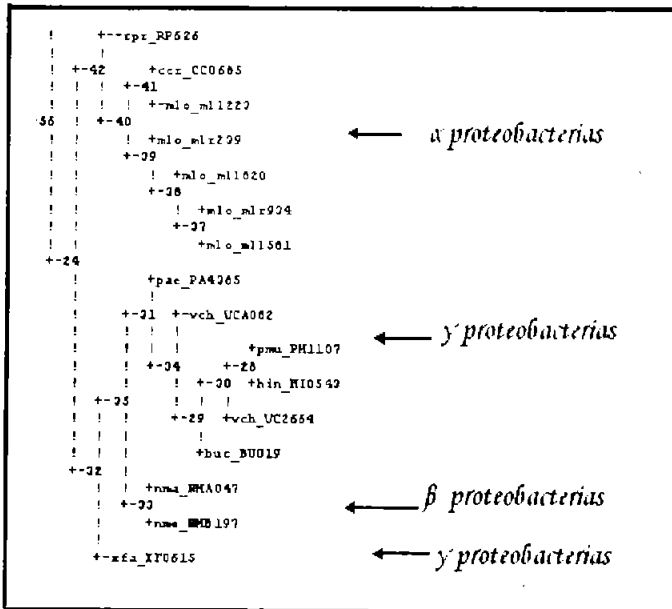


Figura 16. Rama de genes ortólogos obtenidos de la filogenia elaborada con el programa Fitch- Margoliash que se muestra en el Apéndice 1.

La filogenia que se muestra es una sección de la filogenia completa del Apéndice 1, del cual se tomó la rama de los 15 genes de organismos mostrados en la figura 16, para ser analizados con el programa para el IDH, al igual que los genomas completos a los que pertenecen dichos genes.

Una vez determinados los valores de los tres parámetros (d_{YR} , d_{WS} y d_{MK}), se obtuvo el valor del centro relativo de la firma de los genomas (nube de puntos) y se calculó la distancia del gen a dicho centro. Los valores de las distancias filogenéticas y los de las distancias estructurales con el IDH dentro del genoma de cada

organismo se muestran en la Tabla 2 y Figura 17. Cabe destacar que, para la muestra analizada, existe una relación inversa entre estas variables, acorde con lo planteado en la hipótesis. En efecto, el coeficiente de correlación (r de Pearson) es de -0.474.

Tabla 2. Distancias obtenidas con el IDH y distancias filogenéticas con el programa de Fitch-Margoliash

	Gen	des distancia estructural	dev distancia evolutiva
1	CC0685	0.10141285	0.19883
2	mlI2232	0.15377440	0.21338
3	mlr2394	0.12791874	0.19801
4	mlI8201	0.12500834	0.17964
5	mlr9342	0.12561703	0.19234
6	mlI5810	0.13211889	0.18825
7	PA4385	0.04277073	0.20117
8	VCA0820	0.05842995	0.26402
9	PM1107	0.12032831	0.20550
10	HI0543	0.10029905	0.21235
11	VC2664	0.06537636	0.20288
12	BU019	0.08265950	0.24660
13	NMA0473	0.14120810	0.20084
14	NMB1972	0.12533026	0.20229
15	XF0615	0.03815330	0.21780

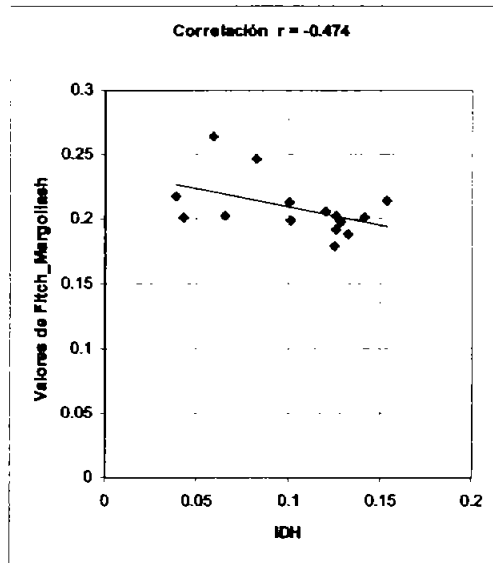


Figura 17. Relación entre las distancias estructural y evolutiva.

A continuación se describe el procedimiento seguido para la prueba de significancia estadística (prueba de hipótesis estadística) del coeficiente de correlación encontrado en la muestra.

La hipótesis nula (H_0) más atractiva de analizar, y que de hecho se desea rechazar, es que el coeficiente de correlación poblacional (ρ) sea igual a cero. Esto equivale a decir que no habría relación entre las 2 variables analizadas (en este caso, entre la distancia filogenética y la estructural) si se estudiara el universo de genes ortólogos; si esto fuese cierto, la correlación obtenida en la muestra de 15 genes se debería al azar (error de muestreo). La hipótesis alterna (H_1), que es la deseable de aceptar, se plantea en forma unidireccional negativa ($\rho < 0$), debido a que ya se conocen los resultados de muestra ($r = -0.474$). Así,

$$H_0 : \rho = 0$$

$$H_1 : \rho < 0$$

Como en este caso la H_0 establece que $\rho = 0$,¹ la distribución muestral es la "t" de Student con $n-2$ grados de libertad, donde "n" es el tamaño de la muestra. Por lo tanto, el estadístico de prueba, que no es otra cosa mas que el coeficiente de correlación estandarizado, está definido como

$$t = \frac{r - 0}{s_r} = \frac{r - 0}{\sqrt{(1 - r^2)/(n - 2)}} = \frac{r\sqrt{n - 2}}{\sqrt{1 - r^2}},$$

donde s_r es la desviación estándar de la distribución muestral de los coeficientes de correlación, que mide la variabilidad de los posibles valores que puede tener r si se tomaran muestras repetidas de tamaño n de una población con correlación igual a la planteada en la hipótesis nula ($\rho = 0$).

Por lo tanto, en forma estandarizada, el coeficiente de correlación encontrado en la muestra es igual a

$$t = \frac{(-0.474) - 0}{\sqrt{(1 - (-0.474)^2)/(15 - 2)}} = \frac{-0.474}{0.2442} = -1.94$$

¹ Cabe mencionar que si en la hipótesis nula se especificara cualquier otro valor de ρ que no fuese cero, la prueba es algo más complicada ya que $(r - \rho_0) / s_r$ no seguiría la distribución "t" de Student. Esto es más válido a medida que la correlación difiere mas de cero.

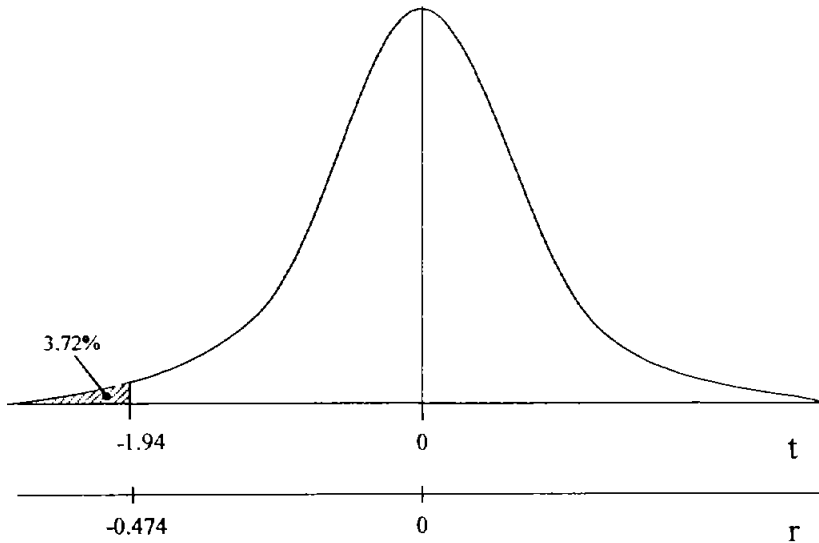


Figura 18. Prueba de significancia estadística del coef. de correlación de la muestra.

Como puede observarse en la Figura 18, si la H_0 fuese cierta ($\rho = 0$), existiría una probabilidad de sólo 3.72% de obtener una muestra con un coeficiente $r \leq -0.474$. Como esta probabilidad es muy pequeña (menor que el nivel de significancia α de 5%, unilateral), se rechaza la hipótesis nula y, en consecuencia, se acepta la hipótesis alterna ($\rho < 0$). Es decir, la distancia filogenética y la distancia estructural están correlacionadas en forma inversa de manera significativa. La posibilidad de que esta decisión sea incorrecta, el llamado error tipo I, es igual a la probabilidad antes citada.

Una vez probada la hipótesis con genes ortólogos correspondientes a 15 organismos, se realizó un análisis complementario con los genes parálogos de 12 organismos, que fueron marcados con verde en la Figura del Apéndice 1. La Tabla 3 muestra las distancias estructurales y evolutivas, donde puede apreciarse que en cada organismo existen dos o tres genes parálogos. En ocho de los organismos (marcados en rosa), la relación indica que a mayor distancia filogenética (distancia evolutiva) menor es la distancia del gen al centro de la nube de puntos del genoma al que pertenecen (distancia estructural). Estos

resultados son consistentes con el experimento anterior. Únicamente cuatro de los organismos analizados no cumplen con esta relación (indicados en verde). Además, se presenta un gráfico para cada organismo con sus genes parálogos (ver Gráficas A2-1 a A2-12 en el Apéndice 2)

Tabla 3. Distancias estructurales dentro del genoma de cada organismo y distancias filogenéticas entre ellos *

Genoma	Gen	des	dev
		Distancia estructural	Distancia evolutiva
NC_000917	AF1451	0.180704127	0.20083056
	AF2238	0.194101308	0.22940999
NC_000916	MTH218	0.121450572	0.28849694
	MTH794	0.12603042	0.24892463
NC_002578	Ta0980	0.087267199	0.30286754
	Ta1276	0.132413241	0.28365868
NC_002689	TVN1128	0.089853014	0.29569445
	TVN0507	0.073388104	0.30710634
NC_000854	APE0907	0.166671818	0.35211385
	APE2072	0.168638468	0.32247664
NC_001136	YDL143w	0.096193085	0.58969086
	YDR212w	0.114557831	0.57853339
NC_000962	Rv0440	0.0980409	0.23234263
	Rv3417c	0.091633733	0.37682749
NC_002677	ML0317	0.098339606	0.23959495
	ML0381	0.093189573	0.38220965
NC_000911	slI0416	0.148987729	0.30441343
	slr2076	0.162675684	0.25553454
NC_002678	MII5810	0.132118892	0.12822101
	MII8201	0.125008341	0.11898851
	MIIr2394	0.127918742	0.12816285
NC_000117	CT604	0.124502276	0.87020409
	CT755	0.183337955	0.85464746
	CT110	0.136168801	0.23503292
NC_000922	CPn0898	0.166327871	0.7533355
	CPn0777	0.160754249	0.51543506
	CPn0134	0.144912433	0.22918313

*Apéndice 1

ii) Los genes de transferencia horizontal adquirirán la firma del genoma.

Una de las características del grupo de los diplomonados al cual pertenece *G. Intestinalis*, es la ausencia de mitocondria. Se ha sugerido la pérdida secundaria de la mitocondria posterior a un evento de transporte horizontal del gen CPN60 del genoma mitocondrial al genoma nuclear de *G. Intestinalis*. (Roger *et al.*, 1998). Después de hacer el Blast de esta proteína contra los genomas completos secuenciados en el KEGG, mostrados en la Tabla 4 del Apéndice 3, realizamos una alineación con Clustal W y con el programa MEGA 2 obtuvimos el árbol filogenético que se muestra en la Figura 19 en donde podemos observar claramente los grupos filogenéticos de Archaeas, Eucariontes y Bacterias, y además observamos que el Cpn60 de *G. Intestinalis* se agrupa con los genes bacterianos y mitocondriales reflejando su origen mitocondrial.

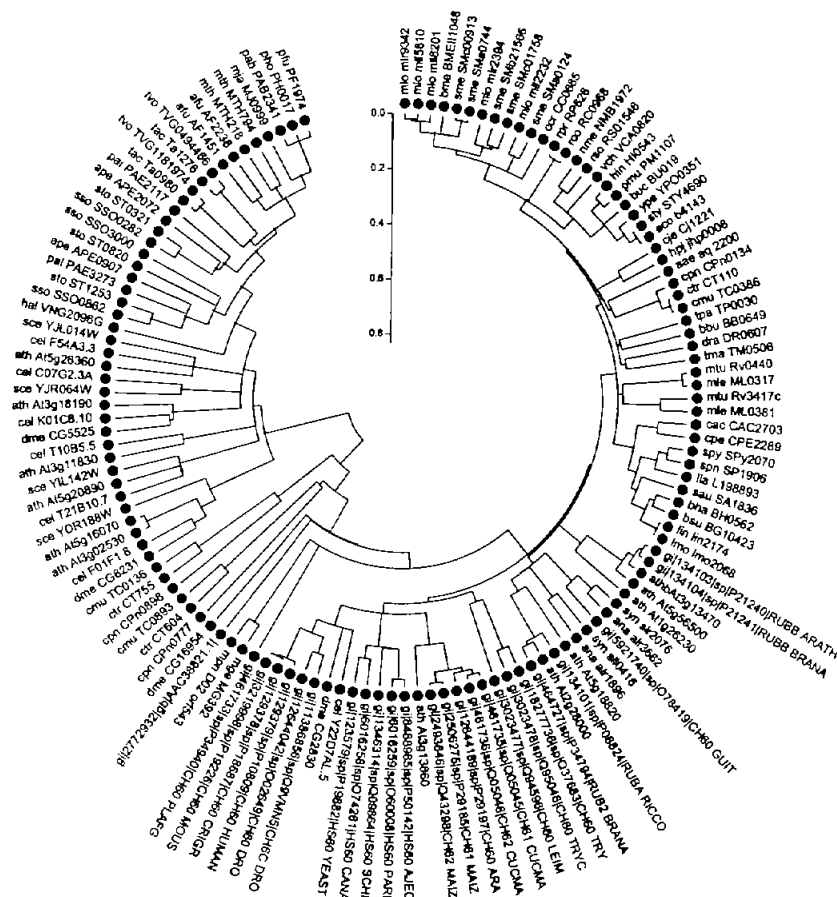


Figura 19. Árbol filogenético (NJ) construido en el MEGA

Se sabe por diversos análisis filogenéticos que *R. prowazekii* es el organismo más cercano al ancestro del endosimbionte mitocondrial.

Después de sacar el IDH con los tres parámetros (d_{WS} , d_{YR} , d_{MK}) para tener tanto la ubicación de los genomas de *G. Intestinalis* y *R. prowazekii*, como de sus genes homólogos del *cpn60* de *G. Intestinalis* y de *R. prowazekii*. Se realizó un análisis visual de la posición de los genes con respecto a los genomas en el espacio Ω para observar si el gen *cpn60* de *G. Intestinalis* había adquirido la firma del genoma Eucarionte y había perdido la firma del genoma mitocondrial. Pretendiendo ver que posición en el espacio Ω ocupaban estos genes con respecto a su genoma, se graficó el mapeo mostrado en la Figura 20.

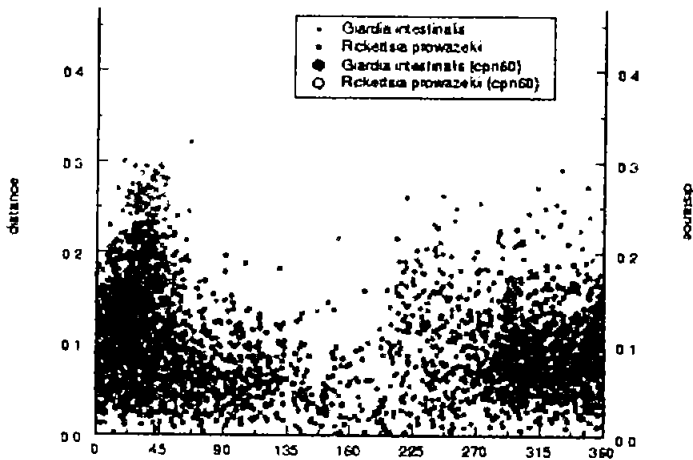


Figura 20

6. Discusión y Conclusiones

Existen mucho métodos de análisis para establecer relaciones filogenéticas, sin embargo la conformación estructural de la molécula del DNA no había sido contemplada como un carácter de análisis evolutivo hasta hace poco (Miramontes, 1995).

La importancia de esta trabajo radicó principalmente en tratar de establecer una relación entre el factor estructural, mediante el uso del IDH, y aspectos más tradicionales de la evolución molecular tales como la tasa de cambio de un gen.

Como se muestra en este trabajo cada genoma presenta una "firma estructural" característica de acuerdo al IDH, lo que representa una de las observaciones más relevantes.

En el enfoque actual del estudio de patrones espacio-temporales observados en diversos sistemas (biológicos, físicos, sociales...), se ha dejado a un lado la visión reduccionista, de ver por partes al sistema, para verlo como un todo, es decir, como un complejo en donde las partes pequeñas interactúan entre si operando con "reglas simples" para formar estos patrones, al parecer tan abundantes principalmente en sistemas biológicos. Estos patrones se conocen como sistemas complejos, que además se caracterizan también porque los cambios que en ellos pueda haber no son proporcionales a las causas que lo provocan.

Bajo este enfoque, muchos de los patrones que observamos pueden ser el resultado de procesos de autoorganización.

Los cambios que se dan en el sistema responden a estados llamados de criticalidad, en donde el sistema se encuentra oscilando fuera del equilibrio. Son los estados relacionados con el cambio del sistema, y en donde pueden existir interacciones de corto y largo alcance que repercutan en componentes vecinos o lejanos.

Para entender la dinámica de estos sistemas, es necesario describir las reglas de interacción y evolución que caracterizan la formación de estos patrones, y comprobar que son estas reglas las que llevan al estado crítico del sistema.

La firma estructural que presenta el genoma responde probablemente a un patrón espacio temporal que podría definirse como un sistema complejo. De ser cierta esta aseveración no se conocen a detalle las reglas que interaccionan dentro del genoma para formar dicho patrón, pero suponemos su existencia.

En este trabajo hemos estudiado al genoma mediante dos aproximaciones importantes (evolución molecular clásica y sistemas complejos). De esta forma, el origen de la "firma estructural" estaría determinada en parte por un sistema complejo (sesgo en el origen de la variación) y en parte por procesos evolutivos más ortodoxos tales como la selección natural y la deriva génica. Como sabemos el proceso de evolución molecular está determinado en última instancia por la mutación, la selección y la deriva génica. Los patrones que observamos hoy en día en los genes y genomas deben de poderse interpretar en términos de interacciones entre estas tres fuerzas evolutivas.

Las adaptaciones biológicas son el resultado del proceso de selección natural. Si la firma estructural de cada genoma está determinado por la selección entonces cabría encontrar alguna correlación entre dicha firma y algún parámetro ambiental. Dicha correlación hasta el momento no ha sido posible de demostrar (otros análisis no mostrados). Por otro lado, bajo un proceso gobernado solamente por la deriva génica, esperaríamos que cada gen, se moviera de forma aproximadamente aleatoria en el espacio Ω y no esperaríamos encontrar una firma estructural. En este trabajo sugerimos que la existencia de dicha firma es debida a un sesgo en la mutación. De ser cierta esta hipótesis entonces aquellos genes que tengan una tasa de evolución mayor deberán presentar una mayor acumulación de mutaciones con el sello de la casa y los genes que hayan sido adquiridos por transferencia horizontal adquirirán la firma del genoma. Los experimentos realizados en este trabajo parecen conformar dicha hipótesis.

También es necesario realizar más experimentos debido a que existen varios factores que pueden sesgar los resultados y con lo cual podemos llegar a conclusiones erróneas por ejemplo, el tiempo de permanencia de un gen en el genoma, también puede estar correlacionado con el número de mutaciones "de la casa" que va a acumular, variable que hemos intentado cancelar utilizando genes ortólogos del COG. Una predicción que sería interesante estudiar es comparar la relación entre las tasas de evolución de distintos genes (chaperoninas, hemoglobinas y fibropéptidos) con respecto a su posición en la nube de puntos de un genoma y observar si la predicción sugerida se cumple.

En principio este método sería capaz de detectar o sugerir genes de transferencia horizontal, sin embargo esto parece ser solo válido para genes transferidos recientemente. Haría falta estudiar si los genes alejados de la nube de puntos pertenecen a genes de transferencia horizontal.

El uso de las herramientas biomatemáticas en esta área nos ha permitido ampliar el panorama del comportamiento de las moléculas en el desarrollo de la historia evolutiva de las especies, y otra forma de ver el origen y aspectos de sistemas biológicos, este campo de desarrollo nos promete un mayor acercamiento para entender algunos aspectos importantes de la biología en muchas de sus ramas.

Por ahora en este trabajo podemos concluir que en las muestras de genes ortólogos y parálogos de la Cpn60, se encontró que a mayor tasa de evolución molecular menor es la distancia estructural al centro de la nube de puntos o mayores serán las mutaciones estructurales "de la firma".

Por esto podemos sugerir que existe una relación entre la distancia evolutiva y la estructural, y que entonces la conformación estructural del genoma es debida a un sesgo mutacional.

En los mapeos no se distingue al gen de transferencia horizontal, por lo menos en el ejemplo buscado.

Este análisis sugiere que los genes adquiridos por transferencia horizontal adquirirán mutaciones de la firma estructural del genoma.

Vale la pena analizar genes de transferencia horizontal tomando en cuenta el tiempo que han permanecido en el genoma.

Sugerimos utilizar métodos estadísticos más poderosos como máxima verosimilitud para estimar las tasas de evolución de los distintos genes.

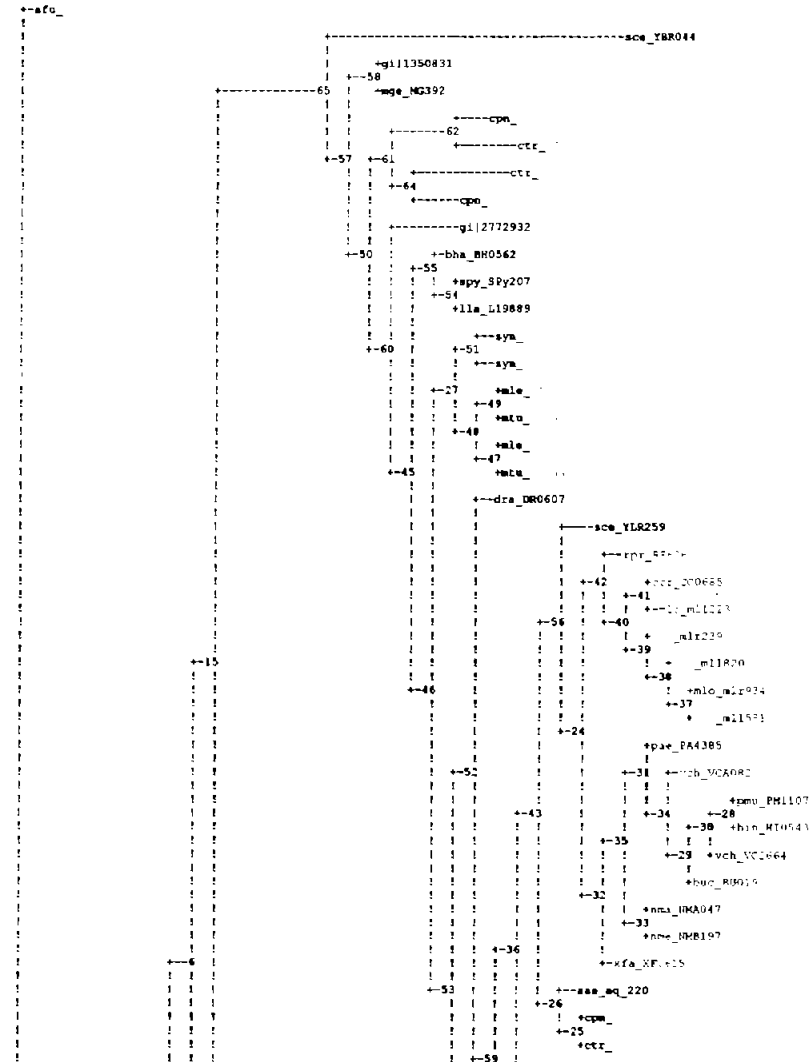
Apéndice 1

67 Populations

Fitch-Morgoliash method version 3.573c

$$\text{Sum of squares} = \frac{\sum_i \sum_j (\text{Obs} - \text{Exp})^2}{\sum_i \text{Obs}^2}$$

Negative branch lengths not allowed



remember: this is an unrooted tree!

Sum of squares = 10.63475

Average percent standard deviation = 4.90515

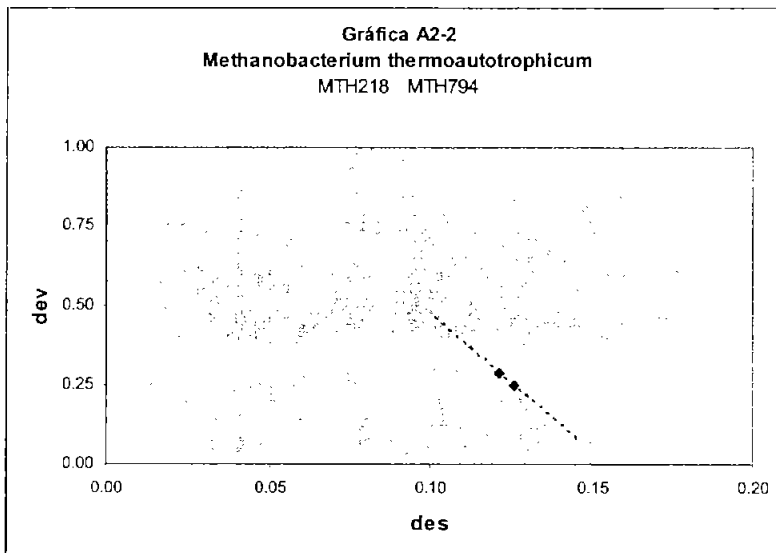
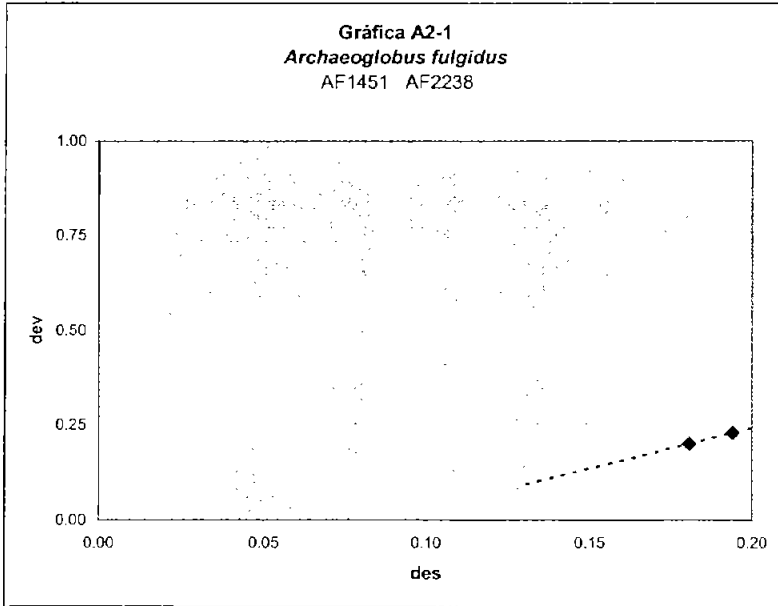
examined 12357 trees

Between	And	Length
-----	---	
1	afu_AF2238	0.14352
1	3	0.04553
3	7	0.02772
7	14	0.01948
14	11	0.02074
11	20	0.03049
20	10	0.01620
13	21	0.03566
21	6	0.01892
6	15	0.01194
15	85	1.26286
85	acc_YBR044	3.47347
55	57	0.21541
57	58	0.34362
58	gal1350831	0.07251
58	age_MG392	0.07882
57	50	0.07637
50	61	0.16229
61	62	0.72772
62	cpn_CPn089	0.45206
62	ctr_CT755	0.73859
61	64	0.25074
64	ctr_CT604	1.13693
64	cpn_CPn077	0.55170
50	60	0.03473
60	gal12772932	0.85834
60	45	0.02743
45	55	0.06405
55	bha_BR0562	0.13394
55	54	0.08154
54	spy_SPy207	0.08541
54	lla_L19889	0.10081
45	46	0.01909
46	27	0.02237
27	51	0.04674
51	syn_slr207	0.22303
51	syn_sl1041	0.23186
27	48	0.06619
48	49	0.26262
49	mle_ML0381	0.08623
49	mtu_Rv3417	0.09228
48	47	0.15853
47	mle_ML0317	0.02865
47	mtu_Rv0440	0.02444
46	53	0.00213
53	52	0.00598
52	dra_DR0607	0.24505
52	59	0.02126
59	36	0.01981
36	43	0.00382
43	56	0.01221
56	acc_YLR259	0.42283
56	24	0.01259
24	42	0.03112
42	rpr_RP626	0.24532
42	40	0.03933
40	41	0.00176
41	cor_C00685	0.12662
41	mlo_ml1223	0.14117
40	39	0.01197
39	mlo_mlr239	0.11559
39	38	0.08242
38	mlo_ml1820	0.01480
38	37	0.01270
37	mlo_mlr934	0.01495
37	mlo_ml1581	0.01071
24	32	0.06006
32	35	0.00955
35	31	0.02051
31	pae_PA4385	0.11095
31	34	0.00710
34	vch_VCA082	0.14670
34	29	0.02935
29	30	0.00472
30	28	0.05590
28	pen_PN1107	0.01821
28	hin_H10543	0.02506
30	vch_VC2664	0.07149
29	buc_BU019	0.11993
35	33	0.12854
33	naa_NMA047	0.00269
33	naa_NNB197	0.00414
32	xfa_XF0615	0.15774
43	26	0.02475
26	aaa_aa_220	0.24198
26	25	0.18532
25	cpn_CPn013	0.01997
25	ctr_CT110	0.04520
36	23	0.11651
23	cje_Cj1221	0.11516
23	22	0.14870
22	hpi_jbp000	0.00053
22	hpi_BPF010	0.00288

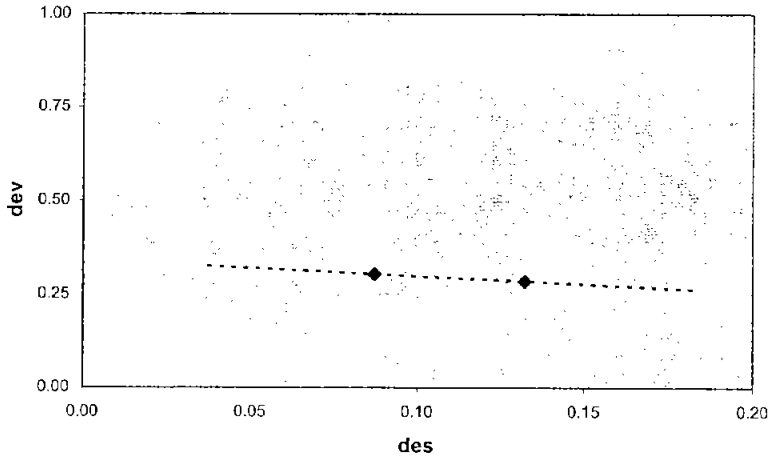
59	44	0.06450
44	hbu_RBG649	0.22912
44	tpa_TPO030	0.19958
53	tma_TMO506	0.23987
15	sca_YDR188	0.95147
6	hal_VNG209	0.37778
21	19	0.12118
19	sca_YJL111	0.69247
19	16	0.07601
18	sca_YIL142	0.72445
18	sca_YDK212	0.62766
13	hal_VNG222	0.30143
20	17	0.04346
17	63	0.05926
63	sca_YJL008	1.08931
63	16	0.12713
16	sca_YJRO64	0.59319
16	sca_YDL143	0.54376
17	sca_YJL014	0.68229
11	9	0.07177
9	10	0.22177
10	gll1354133	0.03068
10	tac_Ta1276	0.02421
9	8	0.22279
8	gll1354195	0.01947
8	tac-Ta0980	0.02931
14	12	0.09994
12	spe_APE207	0.25850
22	spe_APE090	0.32143
7	5	0.11115
5	mth_MTH794	0.07355
5	mth_MTH218	0.17786
3	4	0.03442
4	mja_MJ0999	0.18086
4	2	0.14381
2	peb_PAB234	0.01493
2	pho_PHO017	0.01250
1	afu_AFI451	0.14239

Figura 16. Filogenia Obtenida del análisis de los genes ortólogos, los genes marcada en rojo fueron escogida para el primer análisis de la metodología I porque las relaciones filogenéticas de esta rama corresponden a la biología de los organismos, y los genes marcados en verde son los genes parálogos que se utilizaron en el segundo análisis de la misma metodología.

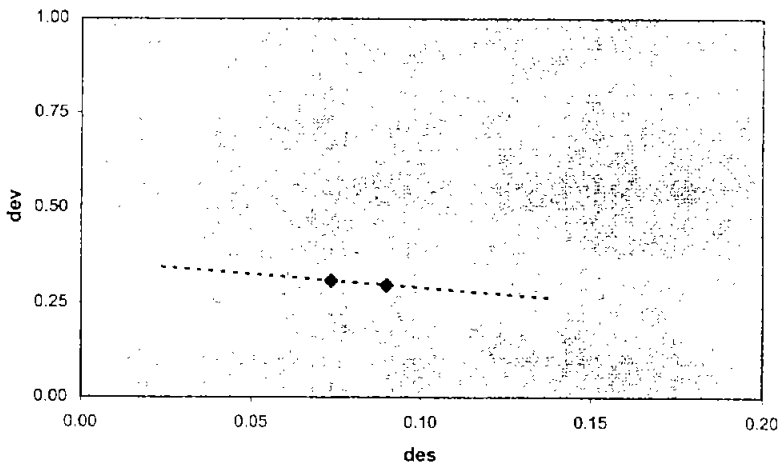
Apéndice 2



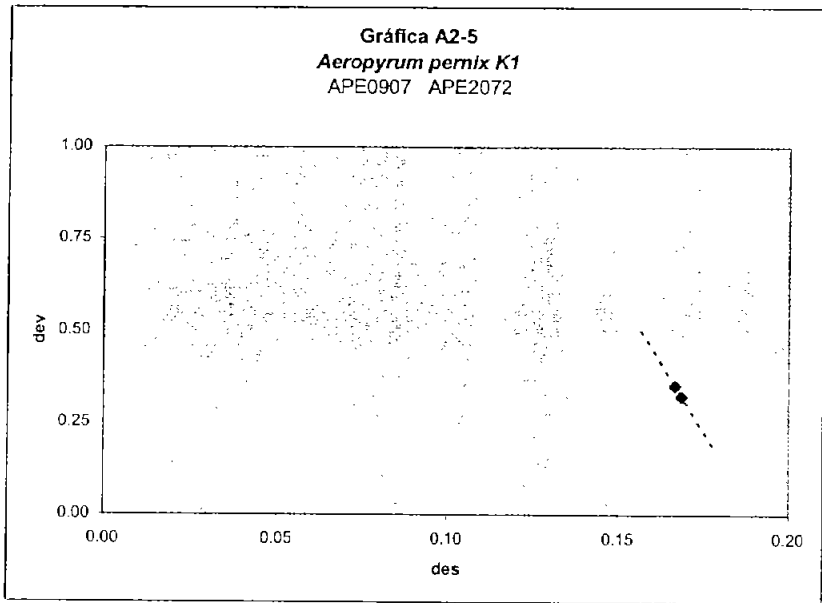
Gráfica A2-3
Thermoplasma acidophilum
Ta1276 Ta0980



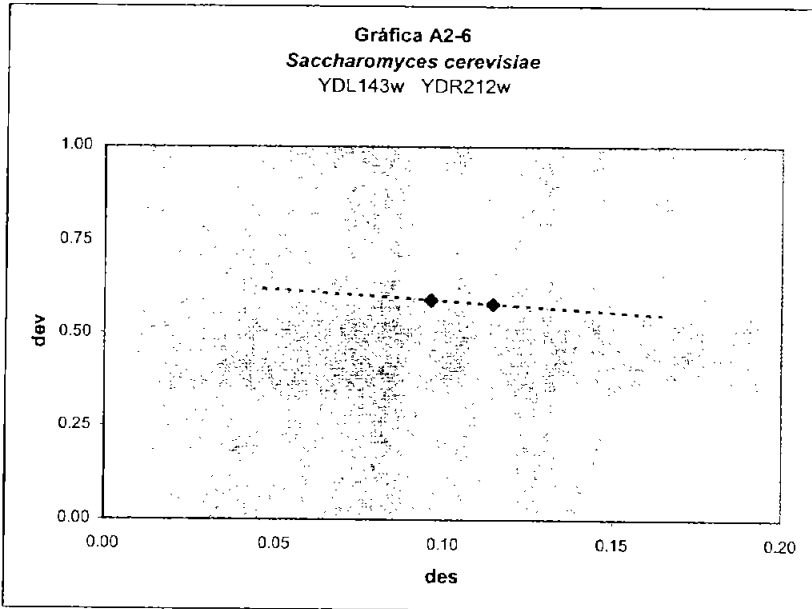
Gráfica A2-4
Thermoplasma volcanium
TVN1128 TVN0507

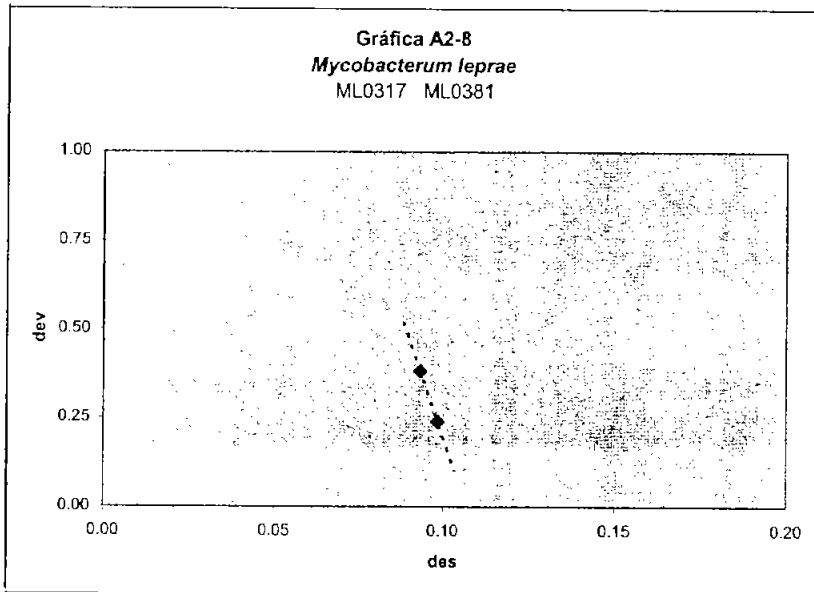
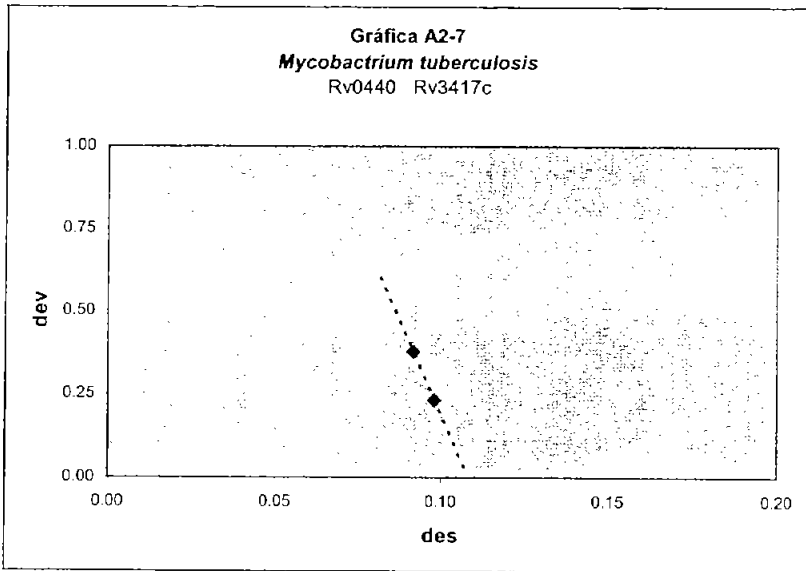


Gráfica A2-5
Aeropyrum pernix K1
APE0907 APE2072

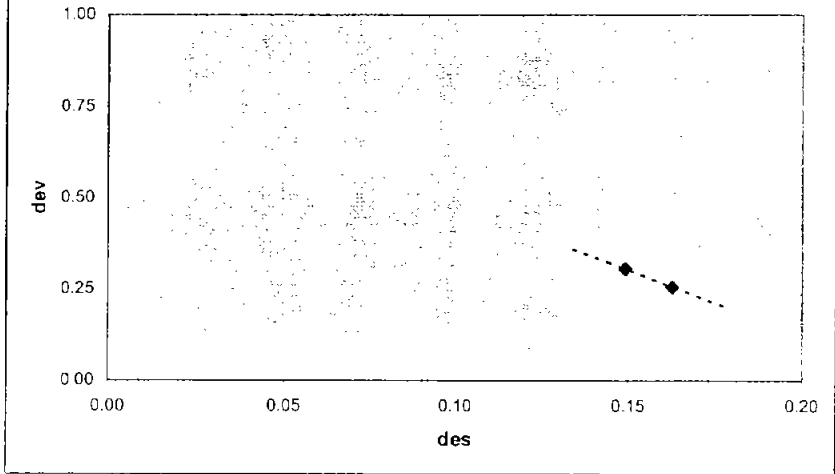


Gráfica A2-6
Saccharomyces cerevisiae
YDL143w YDR212w

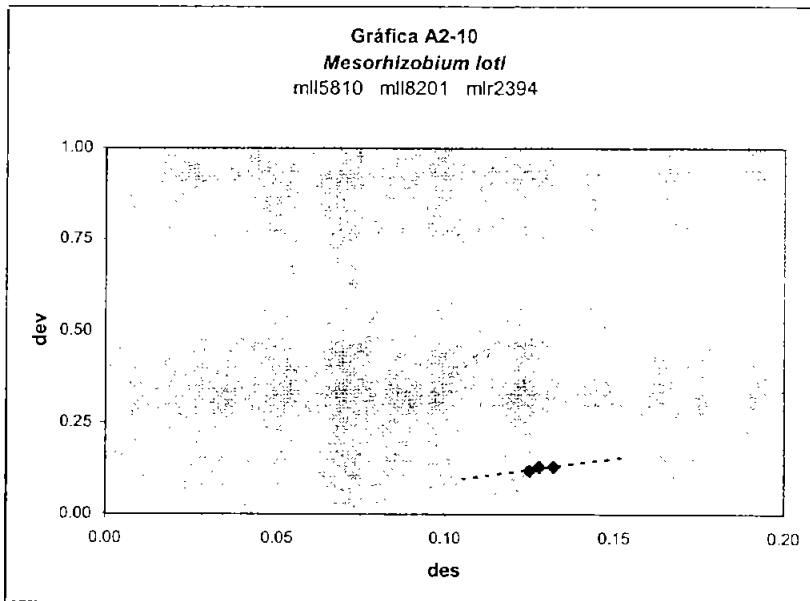


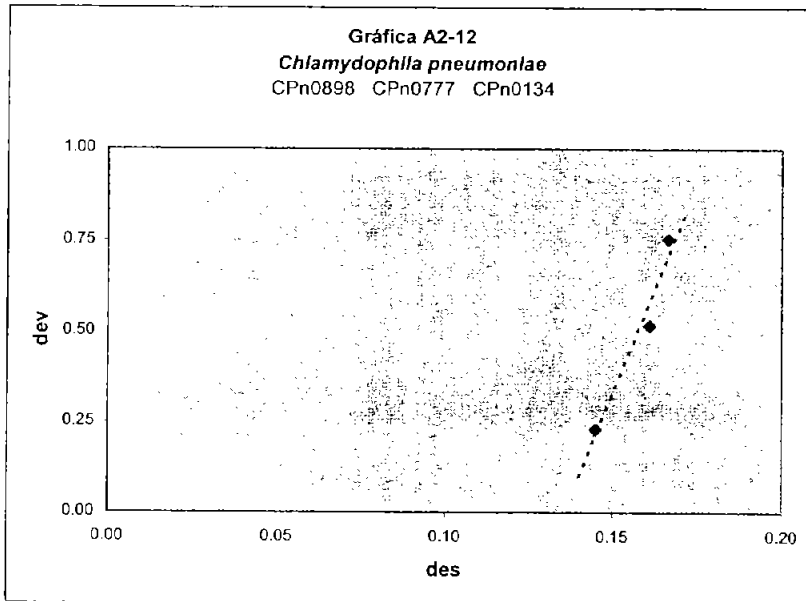
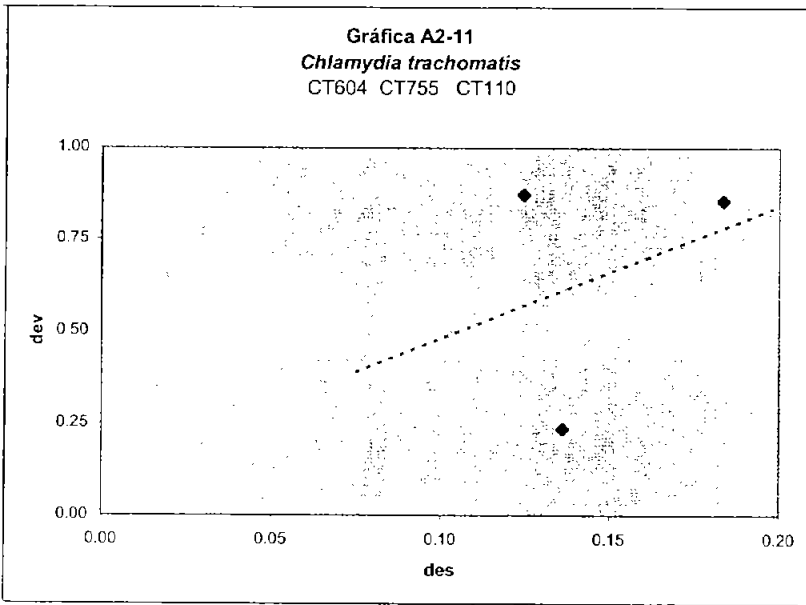


Gráfica A2-9
Synechocystis sp.
SI10416 SIr2076



Gráfica A2-10
Mesorhizobium loti
ml15810 ml18201 mlr2394





Apéndice 3.

Tabla 4 (1) de los genes homólogos a *G. intestinalis*

ESPECIE	GEN	ESPECIE	GEN
<i>Mesorhizobium loti</i>	mlr9342	<i>Mycobacterium tuberculosis</i> H37Rv	Rv0440
	mlI5810		Rv3417c
	mlI8201	<i>Clostridium acetobutylicum</i>	CAC2703
	mlr2394	<i>Mycobacterium leprae</i>	M10317
	mlI2232		ML0381
<i>Deinococcus radiodurans</i>	DR0607	<i>Streptococcus pyogenes</i> SF370 (serotype M1)	Spy2070
<i>Sinorhizobium meliloti</i>	SMc00913	<i>Streptococcus pneumoniae</i> TIGR4	SP1906
	SMA0744	<i>Lactococcus lactis</i>	L1198893
	SMb21566	<i>Staphylococcus aureus</i> N315 (MRSA)	SA1836
	SMc01758	<i>Bacillus halodurans</i>	BH0562
	SMA0124	<i>Bacillus subtilis</i>	BG10423
<i>Caulobacter crescentus</i>	CC0685	<i>Listeria monocytogenes</i> EGD	Imo2068
<i>Rickettsia prowazekii</i>	RP626	<i>Listeria innocua</i>	lin2174
<i>Rickettsia conorii</i>	RC0968	<i>Arabidopsis thaliana</i>	At3g13470
<i>Neisseria meningitidis</i> MC58 (serogroup B)	NMB1972		At5g56500
<i>Ralstonia solanacearum</i>	RS01546		At1g26230
<i>Vibrio cholerae</i>	VCA0820		At2g28000
<i>Haemophilus influenzae</i>	HI0543		At5g18820
<i>Pasteurella multocida</i>	PM1107		At3g02530
<i>Buchnera aphidicola</i> APS	BU019		At5g16070
<i>Yersinia pestis</i> CO92	YPO0351		At5g20890
<i>Salmonella typhi</i> CT18	STY4690		At3g11830
<i>Escherichia coli</i> K-12 MG1655	B4143		At3g18890
<i>Campylobacter jejuni</i>	Cj1221		At5g26360
<i>Helicobacter pylori</i> J99	jhp0008	<i>Synechocystis</i> sp. PCC6803	slr2076
<i>Aquifex aeolicus</i>	Aq2200		slI0416
<i>Chlamydomonas reinhardtii</i>	CPn0134	<i>Anabaena</i> sp. PCC7120	alr3662
	CPn0777		alr1896
	CPn0898	<i>Caenorhabditis elegans</i>	Y22D7.1L5
<i>Chlamydia trachomatis</i>	CT110		F01F1.8
	CT604		T21B10.7
	CT755		T10B5.5
<i>Chlamydia muridarum</i>	TC0386		KD1C8.10
	TC0893		C07G2.3A
	TC0136		F54A3.3
<i>Treponema pallidum</i>	TP0030	<i>Drosophila melanogaster</i>	CG2830
<i>Borrelia burgdorferi</i>	BB0607		CG1695
<i>Brucella melitensis</i>	BMEH1048		CG8231
<i>Thermotoga maritima</i>	TMO506		CG5525

Tabla 4 (2) de los genes homólogos a *G. intestinalis*

ESPECIE	GEN	ESPECIE	GENES
<i>Mycoplasma pneumoniae</i>	D02orf543	<i>Arabidopsis thaliana</i>	gi 134104 sp P21240 RUBB ARATH
<i>Saccharomyces cerevisiae</i>	YII142W YDR188W YJRO64W	<i>Guillardia theta</i> <i>Ricinus communis</i>	gi 12644189 sp P29197 CH60 ARA gi 5921740 sp 078419 CH60 GUIT gi 134101 sp P08824 RUBA RICCO
<i>Halobacterium sp. NRC-1</i>	YJL014W VNG2096G	<i>Brassica napus</i>	gi 464727 sp P34794 BRANA gi 134104 sp P21241 RUBB BRANA
<i>Sulfolobus solfataricus</i>	SSO0862 SSO3000 SSO0282	<i>Tripanosoma brucei brucei</i> <i>Leishmania major</i> <i>Cucurbita maxima</i>	gi 18277736 sp Q37683 CH160 TRY gi 3023478 sp Q94596 CH60 LEIM gi 461735 sp Q05045 CH61 CUCMA
<i>Sulfolobus tokodaii</i>	ST1253 ST0820 STO321	<i>Zea mays</i>	gi 141736 sp Q05046 CH162 CUCMA gi 2506275 sp P29185 CH161 MA1Z gi 2493646 sp Q34298 CH162 MA1Z
<i>Pyrobaculum aerophilum</i>	PAE0907 PAE2117	<i>Ajellomyces casulatus</i> <i>Paracoccidioides brasiliensis</i>	gi 8488985 sp P50142 HS60 AJEC gi 6016249 sp O60008 HS60 PARB
<i>Aeropyrum pernix</i>	APE2072	<i>Schizosaccharomyces pombe</i>	gi 134631 sp Q09864 HS60 SCHIP
<i>Thermoplasma acidophilum</i>	Ta0980 Ta1276	<i>Candida albicans</i> <i>Saccharomyces cerevisiae</i>	gi 6016258 sp O74261 HS60 CANA gi 123579 sp P19882 HS60 YEAST
<i>Thermoplasma volcanium</i>	TVG1181974 TVG0494466	<i>Drosophila melanogaster</i>	gi 11386856 sp Q9VMN5 CH60 DRO gi 12644042 sp O02649 CH160 DRO
<i>Archaeoglobus fulgidus</i>	AF1441 AF2238	<i>Homo sapiens</i> <i>Cricetulus griseus</i>	gi 129379 sp P10809 CH60 HUMAN gi 129378 sp P18687 CH160 CRIGR
<i>Methanobacterium thermoautotrophicum</i>	MTI1218 MTI1794	<i>Mus musculus</i> <i>Plasmodium falciparum</i>	gi 3219998 sp P19229 CH60 MOUS gi 461733 sp P34940 CH60 PLAFG
<i>Methanococcus jannaschii</i>	MJ0999	<i>Giardia intestinalis</i>	gi 2272932 gb AAC38821.1
<i>Pyrococcus abyssi</i>	PAB2341		
<i>Pyrococcus horikoshii</i>	PH0017		
<i>Pyrococcus furiosus</i>	PF1974		
<i>Clostridium perfringens</i>	CPE2289		
<i>Mycoplasma genitalium</i>	MG392		

Apéndice 4

Tipos de mutaciones

Secuencia original del DNA	Secuencia mutada/ cambio
ccg ctc gtc aac tag Gly-Glu-Gln-Leu-Ile	1) c→t ccg ctf gtc aac tag Gly-Glu-Gln-Leu-Ile
	2) a→g ccg ctc gtc agc tag Gly-Glu-Gln-Ser-Ile
	3) a→c ccg ctc gtc cac tag Gly-Glu-Gln-Val-Ile
	4) ↓ ccg cct cgt caa cta g Gly- Gly-Ala-Val-Asp
	5) g→a ccg ctc atc aac tag Gly-Glu-Alto

La mutación 1, corresponde a una mutación de tipo sinónima, en donde el cambio no afecta la información para el aminoácido; la mutación 2 es una transición, que son mutaciones en donde se cambian purinas por purinas o pirimidinas por pirimidinas, la mutación 3 es una transversión, son cambios de purinas por pirimidinas y viceversa; la mutación 4 es una inserción, provoca el cambio de marco de lectura; y la mutación 5 es una mutación de paro. (Ridley, 1993).

Referencias

Miramontes P, Medrano L, Cerpa C, Cedergren R, Ferbeyre G, Cocho G. (1995) Structural and thermodynamic properties of DNA uncover different evolutionary histories. *J Mol Evol*; **40**(6):698-704

Roger AJ, Svard SG, Tovar J, Clark CG, Smith MW, Gillin FD, Sogin ML. (1998) A mitochondrial-like *cha* peronin 60 gene in *Giardia lamblia*: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. *Proc Natl Acad Sci USA* **95**(1):229-34

Kumar, S. Tamura, K. Jakobsen, I.B., Nei, M. (2001) MEGA: molecular evolutionary genetics analysis software.

Doolittle W. Ford. (1999) Phylogenetic Classification and the Universal Tree. *Science* **284** : 2124-2128.

Nei, M., Kumar, S. (2000) *Molecular Evolution and Phylogenetics*, Oxford University Press. New York.

Li W-H (1997) *Molecular Evolution*, Sinauer Associates, Inc. Publishers, Massachusetts, U.S.A.

L. Margulis. *Symbiosis in cell evolution, life and its environment on the early earth*, Freeman, 1981

Gogarten JP, Kibak H, Dittrich P, Taiz L, Bowman EJ, Bowman BJ, manolson MF, Poole RJ, Date T, Oshima T (1989) Evolution of the vacuolar H⁺ - ATPase: implications for the origin of eucariotes, *Proc Natl Acad Sci USA*; **86**(17):661-5.

Iwabe N, Kuma K, Hasegawa M, Osawa S, y Miyata T (1989) Evolutionary relationships of archaeobacteria eubacteria and eukaryotes inferred from phylogenetic, *Proc Natl Acad Sci USA*; **86** (23): 9355-9.

Brown JR, Doudy CJ, Italia MJ, Marshall WE and Stanhope MJ (2001) Universal trees based on large combined protein sequence data sets, *Nat Genet.*; **28** (3) : 181- 5

Brown JR, Doolittle W. Ford (1997) Archea and the procarote- to- Eukariote Transition, *Microbiol. Mol. Bio Rev.*, **61** 456-502.

Page Roderic.DM, Holmes Edward C.(1998) *Molecular Evolution: A phylogenetic approach.*, Oxford; Madlen, MA: Blackwell Science. Paris, France.

Zaxybayeva O, Laperre P, Gogarten JP, (2004) Genome mosaicism and organismal lineages, *Trends Genet* **20**. 254-260

Yuri I. Wolf, Igor B. Rogozin, Nick V. Grishin and Eugene V. Konnin, (2002) Genome trees and the Tree of life. *TRENDS in Genetics* **18** No. 9

Fredj Tekai, Antonio Lazcano, and Bernard Dujon (1999) The Genomic Tree as Revealed from Whole Proteome Comparisons, *Genome Research*, **9**:550-557

Thompson, Higgins, and Gibson, 1994 ClustalW

Fitch W. and Margoliash E., 1967. Construction of phylogenetic trees. A method based on mutation distances as estimated from cytochrom C sequences is of general applicability. *Science* **155**: 279-284

National Center of bioinformatic, database, www.ncbi.nlm.nih.gov/COG/

Chuprina, VP, Lipanov AA, Fedoroff Oyu, Kim SG, Kihtar A, Reid BR, 1997 Sequence effects on local DNA topology, *Proc Natl Acad Sci USA* **88**:20 9087-91

Dickerson RE, DNA structure from A to Z, 1992 *Methods Enzymol* **211**: 67-111

Delcourt SG, Blake RD, Stracking energies in DNA, 1991; **266**(23):15160-9

Dickerson RE, 1983 Base sequence and helix structure variation in B and A DNA *J. Mol. Biol.* **166**(3):419-4

Breslauer KJ, Frank R, Blocker K, Mark LA, 1986 Predicting DNA duplex stability from the base sequence. *Proc Natl. Acad. Sci USA*, **83** (11): 3746-5

Breslauer KJ, 1995 Extracting thermodynamic data from equilibrium melting curves for oligonucleotide order-disorder transitions, *Methods Enzymol*, **255**:221-42

Calladine CR. 1982, Mechanics of sequence-dependent stracking of bases in B-DNA *J.Mol. Biol.* **161**(2): 343-52.

Marzilli LG, Kistenmacher TJ, Ross M, 1977, An extension of the role of O2 of cytosine residues in the binding of metal ions. Synthesis and structure of 1 methylcytosine *J. Am Chem. Soc.* **99**(8):2797-8.

Horton, Moran, Ochs, Rowm, Scrimgeour, 1981 *Bioquímica*, Prentice-hall Hispano america.

Walter m. Fitch, 2000, Homology a personal view on some of the problems, *Elsevier Science Ltd.* **16**: 227-231

Roy Paul H. 1999, Horizontal transfer of genes in bacteria. *Microbiology Today* **26**: 168-170.

Ridley M B, 1993 *Evolution*, Scientific publications.