



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**ESTUDIO DE LAS RELACIONES EVOLUTIVAS
ENTRE ARCHAEAS, BACTERIAS Y
EUCARIONTES, UNA PERSPECTIVA DESDE
LA GÉNOMICA COMPARADA**

**T E S I S
QUE PARA OBTENER EL TÍTULO DE:
B I O L O G O**

**PRESENTA:
JONATHAN ALVAREZ ALVARADO**

**DIRECTOR DE TESIS:
DR. LUIS JOSÉ DELAYE ARREDONDO**



2009



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

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

FACULTAD DE CIENCIAS
Secretaría General
División de Estudios Profesionales

Votos Aprobatorios

Por este medio hacemos de su conocimiento que hemos revisado el trabajo escrito titulado:

Estudio de las relaciones evolutivas entre Archaeas, Bacterias y Eucariontes, una perspectiva desde la genómica comparada

realizado por Alvarez Alvarado Jonathan con número de cuenta 0-9902520-7 quien ha decidido titularse mediante la opción de tesis en la licenciatura en Biología. Dicho trabajo cuenta con nuestro voto aprobatorio.

Propietario Dr. Arturo Carlos II Becerra Bracho

Propietario Dr. Luis Felipe Jiménez García

Propietario Tutor Dr. Luis José Delaye Arredondo

Suplente M. en C. Alfonso José Vilchis Peluyera

Suplente Biól. Tobías Portillo Bobadilla

Atentamente,

“POR MI RAZA HABLARÁ EL ESPÍRITU”

Ciudad Universitaria, D. F., a 01 de diciembre de 2008

EL COORDINADOR DEL COMITÉ ACADÉMICO DE LA LICENCIATURA EN BIOLOGÍA

DR. PEDRO GARCÍA BARRERA

FACULTAD DE CIENCIAS



UNIDAD DE ENSEÑANZA
DE BIOLOGÍA

Señor sinodal: antes de firmar este documento, solicite al estudiante que le muestre la versión digital de su trabajo y verifique que la misma incluya todas las observaciones y correcciones que usted hizo sobre el mismo.

DARWIN, 200 AÑOS DE SU NACIMIENTO, 200 AÑOS DE LA PUBLICACIÓN
DE LA “*Filosofía Zoológica*” DE J. B. LAMARCK,
A 150 AÑOS DE LA PUBLICACIÓN DE
“*El origen de las Especies*”



*No hemos obtenido ninguna explicación científica sobre
la visión ordinaria de que cada una de las especies
han sido creadas independientemente.
Charles Darwin, 1959.*

*No es el más fuerte de las especies el que sobrevive,
tampoco es el más inteligente el que sobrevive.
Es aquel que es más adaptable al cambio.
Charles Darwin, 1959.*



Agradecimientos.

A mis padres por su comprensión y gran apoyo incondicional así como por los principios y valores inculcados.

A mis hermanos por su apoyo.

A mis abuelos que sin duda fueron y son parte fundamental de mi constante formación.

A mi familia que directa e indirectamente son parte de este trabajo.

A Alina Katia por ayudarme a creer en mí y por todo su apoyo.

A todos mis amigos que siempre están en todos los momentos.

A Luis José Delayo Arredondo, mi tutor, mi más sincero agradecimiento por el respaldo, el espacio brindado, así como su gran apoyo en la asesoría, a pesar de la distancia y los imprevistos surgidos durante el desarrollo de este trabajo.

A Arturo Becerra por su comprensión y aportaciones a este trabajo y su gran sentido de vocación hacia la docencia.

Al Dr. Arturo Becerra, Dr. Luis Felipe Jiménez, Dr. Luis Delayo, M. en C. Alfonso Vilchis y al Biól. Tobías Portillo quienes formaron parte del comité de titulación, a quien agradezco sus comentarios, sugerencias y observaciones para mejorar este trabajo.

Al Dr. Antonio Lazcano por su inspiración y amor a la ciencia, y su gran vocación por compartir el conocimiento, quien es un modelo a seguir en el campo de la ciencia. Gracias por permitir la realización de este trabajo en su laboratorio.

A mi universidad, la UNAM, nuestra Máxima casa de estudios, ya que funge como pilar en el desarrollo de México. Por la formación recibida en ella.

"La ignorancia genera confianza más frecuentemente que el conocimiento, son aquellos que saben poco, y no esos que saben más, quienes tan positivamente afirman que este o aquel problema nunca será resuelto por la ciencia".

Charles Darwin, 1959

CONTENIDO

1. Resumen

2. Introducción

2.1 Filogenias universales y el Último Ancestro Común Universal

2.2 Transferencia horizontal de genes

2.3 Genoma mínimo

2.4 Reconstrucción del complemento genético del último ancestro común

2.5 Análisis de comparación de genomas; herramientas bioinformáticas

2.6 Planteamiento del problema

2.7 objetivos

3. Material y métodos

3.1 Identificación de secuencias homólogas conservadas.

3.2 Análisis funcional de genes conservados de acuerdo a las categorías del KEGG.

3.3 Interpretación de los patrones de conservación de los genes.

4. Resultados

5. Discusión y conclusiones

6. Literatura citada

7. Apéndice

1. Resumen

Este trabajo es un intento para comprender mejor las relaciones evolutivas entre las Archaeas, las Bacterias y los Eucariontes estudiando las similitudes y diferencias genéticas y funcionales entre estos tres grupos. Utilizando BLAST se identificaron los genes conservados entre dos de los tres linajes, así como la categoría funcional a la cual estos genes pertenecen (de acuerdo al KEGG Kyoto Encyclopedia of Genes and Genomes). Más precisamente, los análisis que se realizaron fueron los siguientes: (1) los genes conservados en Archaeas y Bacterias excluyendo a los Eucariontes; (2) los genes conservados en Bacterias y Eucariontes excluyendo a las Archaeas; y (3) los genes conservados que comparten Archaeas y Eucariontes excluyendo a las Bacterias.

Los resultados sugieren que los genes conservados en Archaeas y Bacterias excluyendo a los Eucariontes, identificados que en su mayoría del tipo operacional (genes metabólicos) serán en principio el resultado de la transferencia horizontal entre estos grupos, o bien genes que se encontraban en el cen ancestro, y que posiblemente se perdieron en la evolución temprana del linaje eucarionte y probablemente se retuvieron entre Archaeas y Bacterias debido a que son importantes para el estilo de vida procarionte. Para el segundo análisis, los genes conservados en Bacterias y Eucariontes con exclusión de Archaeas, identificamos en su mayoría son genes operacionales de varias clases funcionales y esto nos sugiere que probablemente el que compartan estos genes se deba a la transferencia de genes por el proceso de endosimbiosis, de los ancestros de los organelos (como las mitocondrias y cloroplastos) al núcleo-citoplasma eucarionte. En el tercer análisis los genes conservados por Archaeas y Eucariontes excluyendo a las Bacterias, encontramos que la mayoría son genes que están relacionados con la replicación y reparación del ADN, así como la traducción (genes informacionales), lo que nos sugiere que estos genes se originaron un poco antes del ancestro de este grupo.

Los estudios sugieren que los eucariontes son una mezcla de genes de procariontes; los bacterianos que en su mayoría son genes operacionales y los genes informacionales por parte de las Archaeas. En cuanto a Bacterias y Archaeas pudo ser un intercambio horizontal de genes entre ambos linajes, o en su caso por herencia del cen ancestro.

2. INTRODUCCIÓN.

Darwin afirmó en 1859 “es probable que todos los seres vivos que hay en la Tierra desciendan de un mismo ancestro”, sin embargo no dio muchos detalles sobre cuales serían las características de ese ancestro (Darwin, 1859) .

Las similitudes bioquímicas y genéticas básicas de los seres vivos reveladas en los últimos 50 años de investigación nos sugieren un origen común para todos los organismos, tal y como había intuido Darwin.

El uso de secuencias como marcadores moleculares ha permitido resolver controversias evolutivas y taxonómicas, difíciles de aclarar entre especies con las comparaciones morfológicas o los estudios paleontológicos.

2.1 Filogenias universales y el Último Ancestro Común Universal

Desde los inicios de los estudios biológicos, el sueño de muchos evolucionistas ha sido lograr clasificar a toda la diversidad biológica jerárquicamente en un gran árbol universal (Doolittle, 1999). Anteriormente no había sido posible hacer dicho árbol, debido a que hasta mediados del siglo XX la biología evolutiva dependía principalmente de comparaciones morfológicas y del estudio del registro fósil para hacer inferencias del pasado, lo que implica que para especies fenotípicamente muy distintas no había forma de establecer relaciones evolutivas basadas en caracteres compartidos.

Probablemente la primer filogenia universal fue realizada por Haeckel en 1866 (fig. 1). En dicha filogenia, Haeckel propone la existencia de un tercer reino denominado protista, en donde agrupa a todos los seres vivos que no pertenecen ni a las plantas ni a los animales.

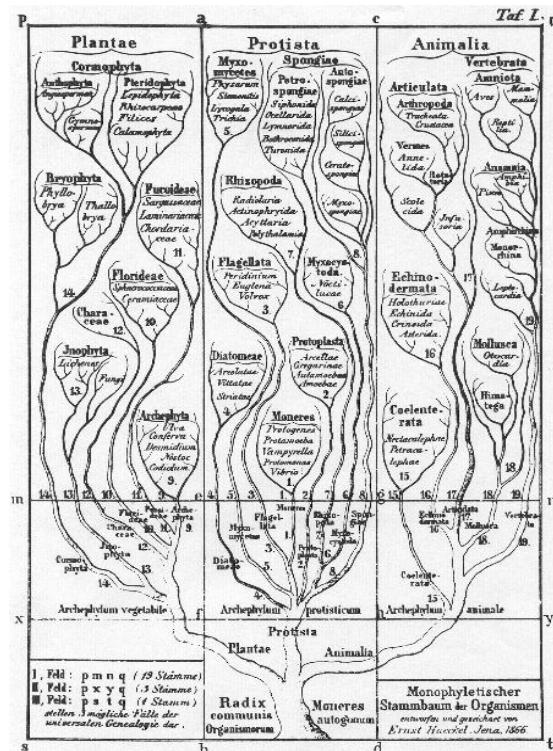


Fig. 1 Filogenia universal propuesta por Ernst Haeckel(1866)

Posteriormente, Edourd Chatton(1938) diferenció dos grandes tipos de células. Uno de estos grupos, los eucariontes, se caracterizan por presentar arreglos intramembranales que constituyen organelos y un núcleo-citoplasma. El otro grupo, que carecía de núcleo, se le denominó procarionte.

Más adelante, Lynn Margulis (1970) en donde se proponen argumentos convincentes para suponer que los organelos eucariontes evolucionaron mediante un proceso de endosimbiosis.

A partir de este momento es que se conoce el esquema tradicional de clasificación de los seres vivos, en el que se les agrupa en cinco reinos y éstos a la vez en dos grandes grupos: los procariontes y los eucariontes.

Con el descubrimiento de la estructura del DNA y con las primeras secuencias de aminoácidos y nucleótidos se desarrolló la disciplina de la evolución molecular. A partir del trabajo de Zuckerkandl y Pauling (1965), mostraron que, mediante la comparación de secuencias de nucleótidos y aminoácidos, se podía obtener más información con respecto a la evolución de las secuencias y de la evolución de los

seres vivos respaldando caracteres, funciones y eventos de evolución que no son detectables por caracteres fenotípicos. Este estudio de secuencias nos permitió estudiar las probables características de organismos ancestrales e incluso de eventos de evolución temprana de la vida. De esta forma se generaron las primeras reconstrucciones filogenéticas basadas en secuencias de genes.

Durante cerca de diez años este enfoque permitió no solamente comparar proteínas como las hemoglobinas, el citocromo C, las ferredoxinas y otras más, sino también construir árboles evolutivos que podían incluir organismos tan distintos entre sí como las bacterias, los hongos y los mamíferos marinos, lo cual hubiera sido imposible con los criterios morfológicos tradicionales.

Debido a la universalidad de algunos genes (como el rRNA entre otros) 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.

Basados en la metodología de las filogenias moleculares, a partir de la comparación de secuencias completas de 16S/18S de la subunidad pequeña del RNA ribosomal (cuyas siglas en ingles se abrevian SSU rRNA) provenientes de diversos organismos tanto procariontes como eucariontes, Woese y Fox (1977) encontraron que de acuerdo a esta molécula, los seres vivos actuales se dividen en tres grandes linajes (denominados dominios); Archaeas, Bacterias (ambos procariontes) y el Nucleocitoplasma Eucarionte, fig. 2.

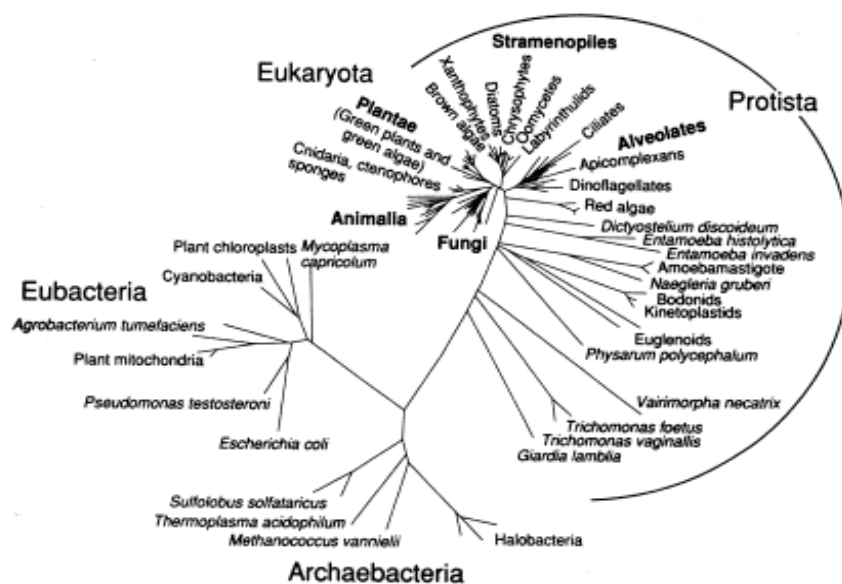


Fig. 2. Filogenia universal, basada en la SSU rRNA (Woese y Fox 1977)

Con esto se demostró que la mayoría de la diversidad de los seres vivos se encuentra a nivel microbiano, esto es, dos de los tres linajes (Bacterias y Archaeas) están compuestos exclusivamente por procariontes, mientras que gran parte del linaje eucarionte está compuesto por organismos unicelulares muy divergentes a las plantas y los animales.

El dominio Bacteria, está representado por procariontes, generalmente mesófilicos, en donde se presentan una amplia gama de metabolismos y adaptaciones a diferentes ambientes y tipos de vida; el dominio Archaea, separa a dos grandes grupos de procariontes: Crenarchaeota conformado por procariontes hipertermofílicos y especies relacionadas filogenéticamente que presentan algunos rasgos comunes con los eucariontes, y los Euryarchaeota, que incluye a los procariontes metanógenos y halofílicos. El tercer dominio es el Eucarya, que incluye a todos los organismos con células nucleadas (Fig. 3).

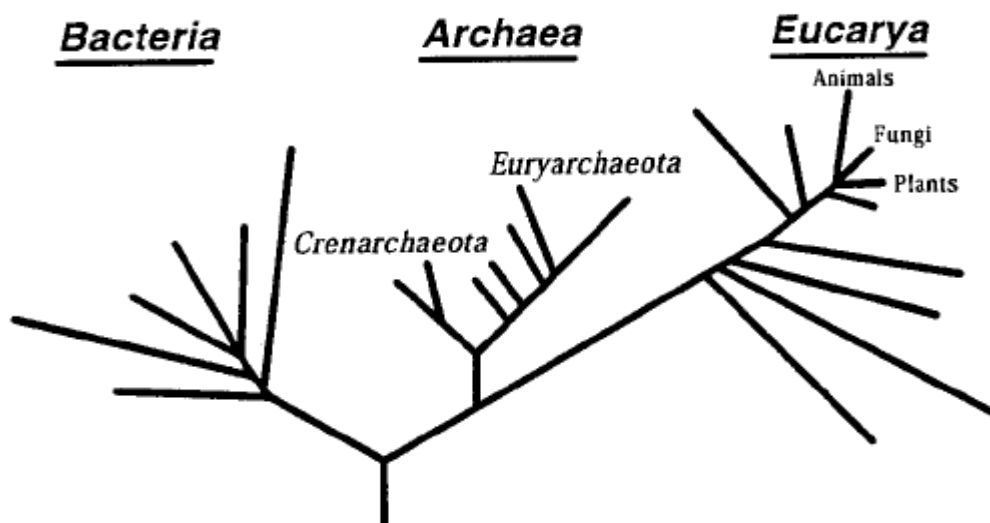


Fig. 3. Clasificación general de los tres dominios (Woese, 2000)

Además la filogenia de 16/18 SSU rRNA sirvió como base a partir de la cual construir una clasificación natural universal (Woese, et al, 1990)

Con la filogenia universal propuesta por Woese y Fox (1977), viene implícita la hipótesis de una especie ancestral a partir de la cual todos los seres vivos

descendemos. Inicialmente Woese y Fox denominaron *progenote* a la entidad biológica que se encuentra en la base del árbol universal y es una entidad hipotética primitiva en la que el fenotipo aun no se encontraba diferenciado del genotipo. Años más tarde, Woese (1983, 1987) continuó desarrollando su hipótesis y propuso que el *progenote* era un sistema donde el material hereditario estaba constituido por moléculas fragmentadas de RNA que aun no estaban integradas en un solo polímero genético.

No todos aceptaron la posibilidad de que el último ancestro común fuese, en efecto, un *progenote*. A partir del análisis de secuencias de tRNA de los tres linajes celulares, Fitch y Upper (1987) sugirieron que el ancestro común a éstos ya poseía un código genético equivalente al de las células contemporáneas y propusieron que el árbol del rRNA se trifurcaba no a partir de un *progenote* sino de un organismo complejo al que denominaron *cenancestro*, que literalmente quiere decir “último ancestro común” y que no necesariamente implica una entidad con características primitivas como aquéllas sugeridas por el *progenote*. Por otra parte, la comparación de las secuencias homólogas comunes a organismos de los tres linajes permitió proponer que el ancestro común del árbol de rRNA era, en realidad, una célula procarionte dotada de los mismos rasgos biológicos de una bacteria contemporánea (Lazcano et al., 1992; Lazcano, 1995).

Es posible describir las características biológicas del último ancestro común como el conjunto de aquellas características comunes y homólogas a todos los seres vivos y que se han heredado verticalmente desde el ancestro a sus descendientes, mas aquellas características que estuvieron presentes en el ancestro universal pero que se perdieron en uno o más linajes, menos aquellas características homólogas que son comunes a todos los seres vivos debido a que han sido heredadas horizontalmente entre linajes. Por lo anterior, el genoma del último ancestro común sería aquel conjunto de genes homólogos a los tres linajes celulares, que han sido heredados verticalmente desde el ancestro universal (Fig. 4).

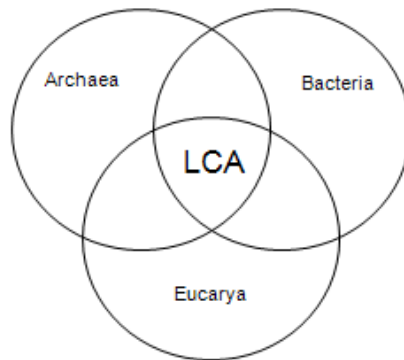


Fig. 4. Caracterización del Último Ancestro Común (LCA). (Lazcano et al., 1992).

Esta metodología para reconstruir el último ancestro común no es del todo infalible ya que su precisión dependerá de una serie de factores tales como el sesgo en el muestreo de especies con los cuales hagamos la reconstrucción, las pérdidas polifiléticas secundarias tanto de genes individuales como de rutas metabólicas completas, las sustituciones no ortólogas de genes y la intensidad del fenómeno de transferencia horizontal entre linajes, (Becerra et al. 1997). Por lo tanto, la caracterización del ancestro común universal como una célula procarionte compleja sugiere la existencia de una fase de evolución biológica previa a la trifurcación.

Las filogenias universales con raíz se hicieron utilizando moléculas parálogas como grupo externo; esto fue realizado por primera vez en los trabajos de Iwabe *et al.*, (1989) usando los factores de elongación (EF-G, EF-Tu) y por; Gogarten *et al.*, (1989) empleando las subunidades α y β de las ATP sintetasas. Ambos grupos llegaron independientemente a la conclusión de que la raíz universal se encuentra localizada en la rama de las bacterias, ya que a partir de moléculas a los que se les adjudica una antigüedad mayor a la del cenancestro es posible reconocer una relación estable entre los tres dominios.

Esto no solamente implica que la naturaleza del último ancestro común tendrá las características similares a las bacterias actuales, sino que también sugiere que una porción importante del nucleocitoplasma eucarionte ha sido heredado a partir de las Archaeas.

2.2 Transferencia horizontal de genes

Con la filogenia universal de Woese, el origen endosimbiótico de mitocondrias y plástidos, la raíz del árbol universal en el linaje bacteriano, la historia evolutiva de la vida en la tierra parecía estar resuelta en gran parte. El árbol evolutivo propuesto por Woese representaría la ancestría-descendencia entre todas las especies actuales. Dicho árbol evolutivo estaría dominado por la herencia vertical y las especies se clasificarían de forma jerárquica de acuerdo al grado de proximidad entre ellas. La molécula de 16S rRNA, por sus características de universalidad, tasa de evolución lenta y conservación de función, parecía ser el marcador ideal para trazar la evolución a nivel orgánico y construir el árbol universal de la vida.

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). Con las primeras secuencias completas de genomas se pensó que las filogenias de la mayoría de los genes reflejarían el esquema evolutivo propuesto por Woese y Fox (1977), con la excepción de los genes transferidos de forma horizontal entre los linajes. Sin embargo, los análisis reflejaron una historia distinta.

Las distintas filogenias moleculares mostraron que diferentes grupos de genes homólogos de un mismo grupo de organismos presentan árboles filogenéticos con topologías distintas a las propuestas por Woese y Fox (1977) y Brown *et al.*, (1997). (Fig. 5).

Este fenómeno de topologías distintas fue atribuido al fenómeno de la transferencia horizontal de genes, entre especies. De hecho se sugirió que la transferencia horizontal de genes entre diferentes linajes orgánicos parecía ser mucho más común de lo que hasta entonces se había creído.

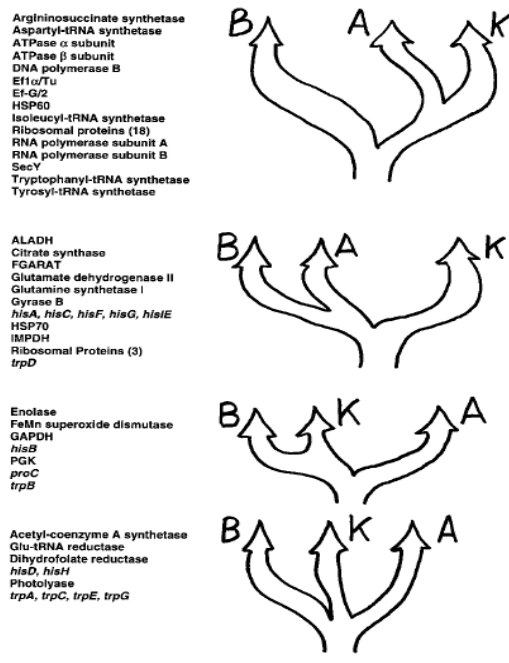


Fig. 5. Topologías de filogenias obtenidas utilizando distintos marcadores moleculares (Brown y Doolittle, 1997)

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). Por lo que tanto la idea de la existencia de linajes orgánico definidos cada uno por los linajes de genes que evolucionan “juntos” en un mismo genoma y que se pueden representar con un solo marcador molecular pareció desvanecerse y en su lugar se sugirió un esquema en donde la transferencia génica horizontal dominaba la estructura del árbol universal a gran escala. Fig. 6. Implicando que no era posible reconstruir el complemento genético del último ancestro común a partir de la comparación de genomas.

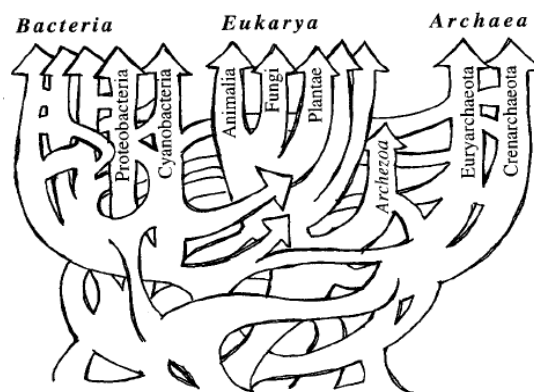


Fig. 6. Filogenia universal propuesta por Doolittle (1999)

Dicha situación llevo a sugerir que el último ancestro común no era una entidad discreta, sino una comunidad de células que evolucionaron como una unidad biológica y que el ancestro universal estaba constituido por una población de *progenotes* con sistemas de procesamiento de información muy poco precisos en donde la dinámica evolutiva estaba regida principalmente por una alta tasa de mutación y de transferencia horizontal y conforme evolucionaron estructuras biológicas más complejas y precisas, tanto la tasa de mutación como la transferencia horizontal disminuyo y la dinámica evolutiva comenzó a parecerse al de las células actuales.(Woese, 1998).

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, (Tekaia et al. 1999), así como una filogenia universal derivada de concatenar una serie de genes universalmente conservados (Brown, et al, 2001). También se ha sugerido que distintos linajes celulares han estado sujetos a distintas tasas de transferencia horizontal (Zhaxybayeva et al, 2004). Dichos análisis rescatan la existencia de los tres grandes grupos filogenéticos (Archaea, Bacteria y Eucarya) y parece indicar que la transferencia horizontal entre linajes no ha sido tan elevada como para borrar del todo el pasado “organísmico”.

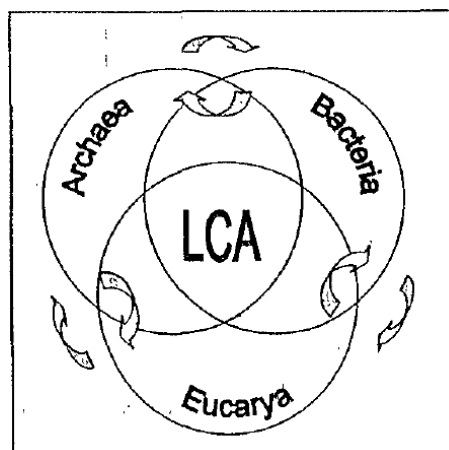


Fig. 7. Complemento genético del LCA, donde se muestra con flechas la transferencia horizontal de genes entre los tres dominios celulares

En este sentido sigue siendo un debate la cantidad de transferencia horizontal que ha existido entre los linajes celulares.

2.3 Genoma mínimo

Todos los organismos vivos están formados por células y cada una de ellas almacena, en su genoma toda la información necesaria para su correcto funcionamiento, el número de genes presentes en el genoma de cada especie puede ser muy variable y depende en parte de la complejidad del organismo, pero sólo en parte, de aquí deriva la paradoja del valor “C”. De hecho se puede decir que hasta el organismo más simple de los organismos unicelulares que habitan en nuestro planeta presenta un grado sorprendente de complejidad. Lo importante es preguntarse si la complejidad se atribuye a algo necesario para que se pueda desarrollar la vida, o si por el contrario, la vida puede ser posible con un número menor de componentes moleculares. Con estos planteamientos y la secuenciación de los primeros genomas completos, se realizó el primer trabajo de la aproximación al genoma mínimo mediante genómica comparada, y se remonta a 1996, cuando los dos primeros genomas bacterianos completos, correspondientes a *Haemophilus influenzae* y *Mycoplasma genitalium* estuvieron disponibles (Mushegian y Koonin, 1996)(Fig.8) .

El “genoma mínimo” es el acercamiento para estimar el número más pequeño de elementos genéticos suficientes para construir un tipo de organismo celular moderno de vida libre (Mushegian,1999). Lo que sugiere es que el conjunto de genes compartidos entre ambos genomas representa el número mínimo de genes necesarios para mantener una célula viva.

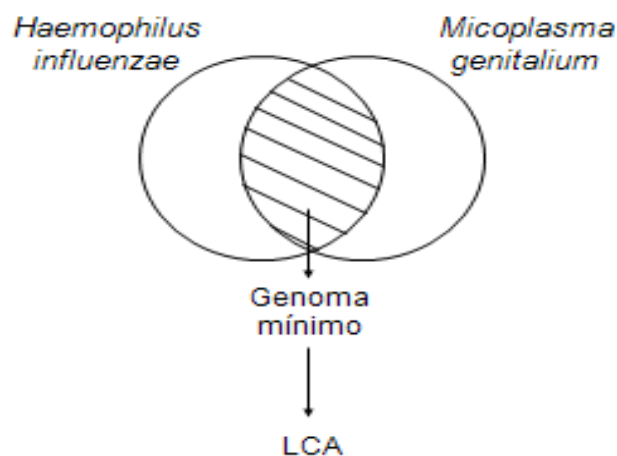


Fig. 8. Representación del genoma mínimo Mushegian y Koonin, (1996)

Con este trabajo encontraron que el conjunto de genes que comparten y formarían el genoma mínimo es de 256. Las dos bacterias utilizadas en la comparación están separadas por 1.500 millones de años de evolución, por lo que los genes conservados a través de tan grande distancia filogenética y tras el proceso de reducción genómica, fueron considerados muy buenos candidatos para ser esenciales, por lo tanto ser ideales para el genoma mínimo. Sin embargo un estudio con mutagénesis en *M. genitalium* permitió comprobar que muchos de los genes incluidos en el conjunto mínimo propuesto no eran esenciales (Hutchison III, et al.,1999). Además se ha visto que si se aumenta el número de genomas usados en la comparación, el conjunto de genes compartidos se reduce considerablemente.

Por lo anterior se llega a la conclusión de que el trabajo de Mushegian y Koonin es solo o parcialmente correcto ya que el conjunto de genes que representan el genoma mínimo es relativo al sistema parásito hospedero que se estudie y por lo tanto a la historia evolutiva de los organismos a partir de los cuales se hizo la comparación genómica. En segundo lugar, sugirieron que este conjunto de genes podía ser representativo del genoma del último ancestro común. Como en el conjunto de genes homólogos entre las dos especies faltaban una serie de proteínas clave en la síntesis del ADN llegaron a la conclusión de que el último ancestro común poseía un genoma de RNA. Conclusión que fue prematura debido a que en su estudio se utilizaron organismos de tipo parásito, por lo que lo más probable es que estos organismos hayan perdido genes y capacidades metabólicas debido a su naturaleza (Becerra et al, 1997).

A pesar de la indudable utilidad de la genómica comparada para tratar de definir un conjunto de genes esenciales, este tipo de análisis tiende a subestimar el número de genes que deben componer un genoma mínimo.

2.4 Reconstrucción del complemento genético del Último Ancestro Común

El estudio del origen y la evolución temprana de la vida ha sido un problema histórico. Se requiere de múltiples aproximaciones para tratar de reconstruir los eventos que comenzaron en la sopa prebiótica y que finalmente desembocaron en los primeros seres vivos.

Es decir, la comparación de las diferencias y similitudes que existen entre los tres linajes permite, en principio, conocer no sólo la relación evolutiva que guardan entre ellos, sino también las características de su ancestro, al que Fitch y Upper (Fitch y Upper, 1987) designó como *cenancestro* (del vocablo griego ancestro común).

Como ya hemos mencionado anteriormente los genes del cenancestro estarían definidos por el conjunto de secuencias presentes en la intersección de los conjuntos que representan los genomas de las Archaea, las Bacteria y los Eucarya. Sin embargo, en la práctica ésta reconstrucción se ha visto limitada por (a) el hecho de que los genomas secuenciados representan solo una fracción de la biodiversidad; (b) los distintos niveles de conservación de los genes, que distan mucho de ser los mismos para todas las secuencias; (c) los problemas para anotar, es decir, identificar a las secuencias presentes en las bases de datos; (d) la presencia de secuencias altamente conservadas cuya función es aún completamente desconocida por no disponer de datos experimentales; (e) la pérdida polifilética, es decir, independiente, de secuencias, funciones y rutas metabólicas que han sufrido diversos organismos, sobre todo parásitos y simbioses; y (f) el transporte horizontal de genes que, como lo ha demostrado la comparación de muchos genomas celulares completamente secuenciados, en algunos casos puede llegar a traspasar las fronteras que separan a los grandes dominios.

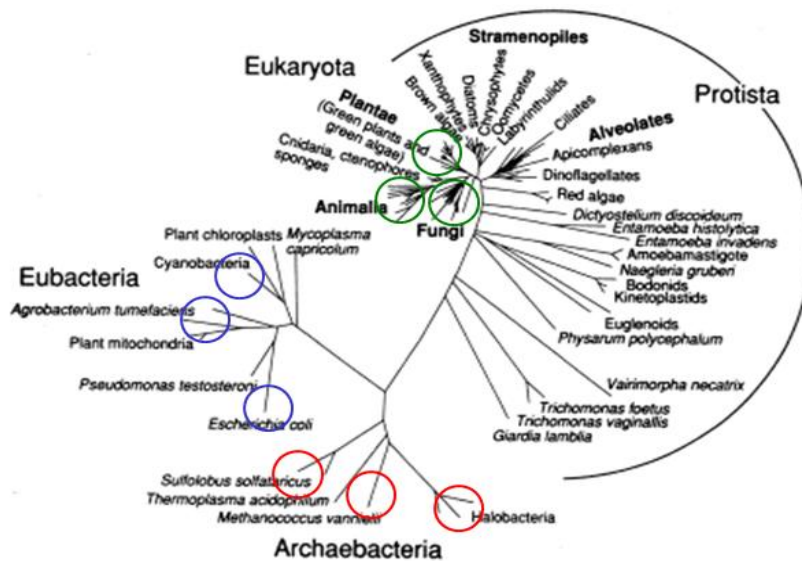
En el trabajo de Delaye y colaboradores (2005) realizaron un estudio para reconstruir el complemento genético del último ancestro común, basados en las características antes mencionadas. Se utilizaron organismos no parásitos ni endosimbiontes obligados.

El trabajo consistió en buscar los genes universalmente conservados, mediante el análisis de una base de datos de una muestra de veinte secuencias de genomas completos, tomados de los tres dominios celulares. En la siguiente tabla 1, se muestran los organismos utilizados, su clasificación y un comentario de la biología de cada especie.

Tabla 1: Organismos utilizados en el trabajo de Delaye et al (2005). (A =Archaeas, B = Bacterias y E =Eucariontes)

	Specie	Phylum/Class/Order	Biology (comment)
A	<i>Archaeoglobus fulgidus</i>	Euryarchaeota/Archaeoglobi/Archaeoglobales	Hyperthermophilic, metabolizes sulfur
	<i>Halobacterium sp.</i>	Euryarchaeota/Halobacteria/Halobacteriale	Halophile
	<i>Methanococcus jannaschii</i>	Euryarchaeota/Methanococci/Methanococcales	Autotrophic hyperthermophilic
	<i>Pyrococcus horikoshii</i>	Euryarchaeota/Thermococci/Thermococcales	Hyperthermophilic (98°C)
	<i>Thermoplasma acidophilum</i>	Euryarchaeota/Thermoplasmata/Thermoplasmatales	Extremophile (59°C and pH 2)
	<i>Aeropyrum pernix</i>	Crenarchaeota/Thermoprotei/Desulfurococcales	Hyperthermophilic (100°C)
	<i>Pyrobaculum aerophilum</i>	Crenarchaeota/Thermoprotei/Thermoproteales	Hyperthermophilic (104°C)
	<i>Sulfolobus solfataricus</i>	Crenarchaeota/Thermoprotei/Sulfolobales	Sulfur-rich, acidic environments (80°C)
B	<i>T. tengcongensis</i>	Firmicutes/Clostridia/Thermoanaerobacteriales	Hyperthermophilic (80°C)
	<i>Streptococcus pneumoniae</i>	Firmicutes/Bacilli/Lactobacillales	Human parasite
	<i>Bacillus subtilis</i>	Firmicutes/Bacilli/Bacillales	Gram-positive
	<i>Synechocystis sp.</i>	Cyanobacteria/Chroococcales	Photosynthesis
	<i>Fusobacterium nucleatum</i>	Fusobacteria/Fusobacteria (class)/Fusobacteriales	Human parasite
	<i>Escherichia coli K-12</i>	Proteobacteria/Gammaproteobacteria/Enterobacteriales	Found in the intestines of animals
	<i>Deinococcus radiodurans</i>	Deinococcus-Thermus/Deinococci/Deinococcales	Radiation-resistant
	<i>Aquifex aeolicus</i>	Aquificae/Aquificae (class)/Aquificales	Chemolithoautotroph (95°C)
E	<i>Arabidopsis thaliana</i>	Viridiplantae/Streptophyta	Photosynthesis
	<i>Caenorhabditis elegans</i>	Metazoa/Nematoda/Rhabditida	Animal
	<i>Saccharomyces cerevisiae</i>	Ascomycota/Saccharomycetes/Saccharomycetales	Aerobic/anaerobic metabolism
	<i>Schizosaccharomyces pombe</i>	Ascomycota/Schizosaccharomycetes/Schizosaccharomycetales	Fungi

Las especies seleccionadas fueron tomadas de acuerdo a la filogenia de Woese y Fox que se muestran en la figura 8: donde se tomo una muestra representativa para los tres dominios.



(Woese y Fox, 1977)

Fig.9. Filogenia universal donde se muestran algunos de los grupos de donde se seleccionaron las especies para el estudio.

El trabajo se realizó utilizando búsquedas de BLAST de una sola vía para identificar aquellos genes universalmente conservados. Los genes identificados provienen en principio de etapas tempranas de la evolución de la vida en la Tierra y por lo tanto es el complemento del último ancestro común. Además se identificaron los dominios conservados entre los grupos de genes homólogos altamente conservados para las especies; *E. coli*, *M. jannashii* y *S. cerevisiae*

Con este estudio encontraron que los genes universalmente conservados están relacionados principalmente al metabolismo del RNA, es decir encontraron que más del 80 % de los dominios conservados corresponden a proteínas que interactúan directamente con el RNA, lo que sugiere que previo a los sistemas celulares basados en DNA/RNA y proteínas, existió una etapa en donde las células estaban basadas en RNA y proteínas, Figura 9. Además con los genes universalmente conservados identificados aquí, sugieren la existencia de un LCA similar en complejidad al de las células actuales (Delaye et al, 2005).

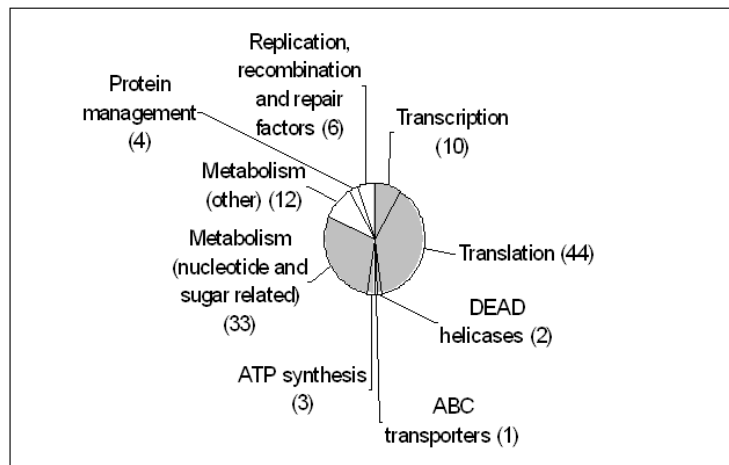


Fig. 10. Secuencias y dominios de proteínas altamente conservadas relacionadas con el metabolismo del RNA, encontradas en el complemento genético del último ancestro común. (Delaye et al, 2005).

2.5 Análisis de comparación de genomas; herramientas bioinformaticas

Cuando obtenemos la secuencia de un genoma vemos poco más que una o varias largas cadenas de letras. Allí está la información que buscamos. El análisis de

genomas se refiere a la tarea de *entender* qué *dice* la secuencia de un genoma: básicamente, qué genes contiene, dónde se encuentran, y qué función realizan las proteínas que son codificadas por ellos (Rouze *et al.*, 1999). A lo largo de la evolución, los cambios que sufren los genomas y sus genes, están sometidos a la presión de la selección natural, por lo que según las ventajas o inconvenientes que los cambios acarreen, dependerá que lleguen a las siguientes generaciones, o lo que es lo mismo, que los podamos observar en la actualidad.

La reconstrucción de estadios ancestrales ha adquirido una perspectiva totalmente novedosa con la disponibilidad, a partir de 1995, de un número creciente de genomas celulares completamente secuenciados. Como habían afirmado desde 1965 Zuckerkandl y Pauling, la historia evolutiva de un organismo está contenida en su genoma pero a menudo esta información es difícil de interpretar, debido a una serie de fenómenos biológicos entre los que se encuentra la falta de preservación de la estructura primaria de los genes.

Para lograr comprender como se ha analizado la relación entre los dominios celulares es necesario aclarar algunos conceptos en el estudio de las secuencias. Los caracteres se pueden definir como *homólogos* cuando se les reconoce un pasado evolutivo común, originándose de un mismo ancestro (Fitch, 2000).

Otros términos frecuentemente se usan para especificar el tipo de homología existente entre dos secuencias o genes son los de *ortólogos* y *parálogos* (Fitch, 1970; Tatusov *et al.*, 1997). Hablamos de ortólogos cuando los genes homólogos han divergido a partir de una especiación. Este tipo de homólogos suelen conservar una función común. Por otra parte, hablamos de parálogos cuando nos referimos a genes de la misma especie o de especies distintas que han surgido de un proceso de duplicación. Otro término son los genes xenólogos, que son secuencias homólogas compartidas entre dos o más organismos y que resultan de eventos de transferencia horizontal.

Por lo que el principal interés de la genómica comparada es buscar secuencias homólogas para después poder realizar el estudio que se desea, como son la búsqueda de genes o la predicción de la función y la estructura de las proteínas, etc.

Existen numerosos métodos para encontrar secuencias parecidas entre la enorme cantidad de secuencias de las bases de datos. Estos métodos se basan en modelos estadísticos para determinar cuándo estos parecidos se deben a que ambas proteínas comparten un mismo origen y cuándo se deben a parecidos al azar. Uno de los primeros métodos y de los más usados es el BLAST.

BLAST (Altschul et al., 1997). Este método realiza de forma muy rápida una búsqueda de secuencias parecidas en las bases de datos. Para cuantificar los parecidos se determina una puntuación del alineamiento entre las dos secuencias. Esta puntuación se obtiene consultando una matriz de sustitución en la que está representado, mediante un valor numérico, la frecuencia con que se observan los posibles cambios entre aminoácidos o la frecuencia con que éstos se conservan. En el caso de BLOSUM (Henikoff, 1992), la matriz se construye analizando alineamientos múltiples y contando la frecuencia con que se observan las posibles sustituciones de aminoácidos, a partir de estas frecuencias y de las frecuencias esperadas se calculan *log-odds* de los que se derivan las puntuaciones. BLAST aplica un marco estadístico (basado en un modelo aleatorio que describe cómo se distribuyen las puntuaciones de parecidos al azar y qué parámetros afectan a esta distribución) para determinar cuán significativa es una determinada puntuación dadas las características de la secuencia problema, de la base de datos y de la matriz de sustitución, y proporciona un *e-value* (valor esperado) que indica, para cada puntuación, cuántas veces esperaríamos que por azar apareciese esa determinada puntuación o una mejor en la base de datos utilizada (Karlin y Altschul, 1993). De forma que si una puntuación tiene asignado un *e-value* de 1, quiere decir que por azar esperamos encontrar un evento parecido con al menos esa puntuación en la base de datos.

El método de BLAST resulta muy útil para conocer de forma rápida cuáles son los homólogos cercanos de una secuencia y con esto proceder a nuestros siguientes análisis como son identificación de genes, clasificación, estructura de proteínas así como la evolución de las mismas e incluso del organismo.

Este tipo de análisis *in silico* nos dan mucha información, pero estos estudios dependen de los algoritmos que se utilicen, de la anotación del gen, proteína o genoma así como de la muestra con que se realice el estudio y demás variables que pueden variar o sesgar el estudio.

2.6 Planteamiento

Si bien se han realizado diversos análisis para estudiar el complemento genético del cen ancestro, se han hecho relativamente pocos estudios para identificar los genes que se comparten por dos de los linajes celulares con la exclusión del tercero. Se ha sugerido que la raíz del árbol universal se encuentra en las Bacterias. Por lo tanto, los genes compartidos entre Archaeas y Eucariontes serán el resultado de herencia vertical, mientras que los genes compartidos por Eucariontes y Bacterias serán el resultado de transferencia horizontal de los ancestros de los organelos al nucleocitoplasma eucarionte.

2.7 Objetivos

- Comparar genomas de los tres dominios celulares para identificar genes conservados entre los linajes Archaea-Bacteria con la exclusión de Eucariontes, Archaea-Eucarionte con la exclusión de Bacterias, Eucarionte-Bacteria con la exclusión de Archaeas.
- Identificar las funciones de dichos genes.

3. Material y método

Se utilizaron secuencias de aminoácidos en formato FASTA, se tomó una muestra de nueve genomas completos de los tres dominios celulares (Archaea, Bacteria y Eucariote).

Se seleccionaron genomas representativos, tres Eucariontes, tres de Bacterias y tres de Archaeas todos ellos de organismos no endosimbióticos ni parásitos obligados.

Bacterias:

eco	<i>Escherichia coli</i>
syn	<i>Synechocystis</i> sp.
bsu	<i>Bacillus subtilis</i>

Archaeas:

mja *Methanococcus jannaschii*
pfu *Pyrococcus furiosus*
pai *Pyrobaculum aerophilum*

Eucariontes:

sce *Saccharomyces cerevisiae*
ath *Arabidopsis thaliana*
cel *Caenorhabditis elegans*

Los genomas fueron tomados de la base de datos del Kyoto Encyclopedia of Genes and Genomes KEGG.

<ftp://ftp.genome.jp/pub/kegg/genomes/>

3.1 Identificación de secuencias homólogas (genes) conservadas

La base de datos de las secuencias de genomas fue sometido a un análisis de BLAST para obtener las secuencias homólogas de las intersecciones de los dominios. Para obtener las secuencias conservadas se realizaron tres análisis. (Fig. 10).

- 1) Secuencias homólogas que comparten Archaeas y Bacterias pero no Eucariontes. Este análisis consistió que existiera cuando menos un Hit en uno de los genomas, de los dominios Archaeas y Bacterias pero que no exista ningún homólogo en ningún genoma de Eucariontes.
- 2) Secuencias homólogas que comparten Archaeas y Eucariontes pero no Bacterias. Este análisis consistió que existiera cuando menos un Hit en uno de los genomas, de los dominios Archaeas y Eucariontes pero que no exista ningún homólogo en ningún genoma de Bacteria.
- 3) Secuencias homólogas que comparten Bacterias y Eucariontes pero no Archaeas. Este análisis consistió que existiera cuando menos un Hit en uno de los genomas de los dominios Bacterias y Eucariontes pero que no exista ningún homólogo en ningún genoma de Archaeas.

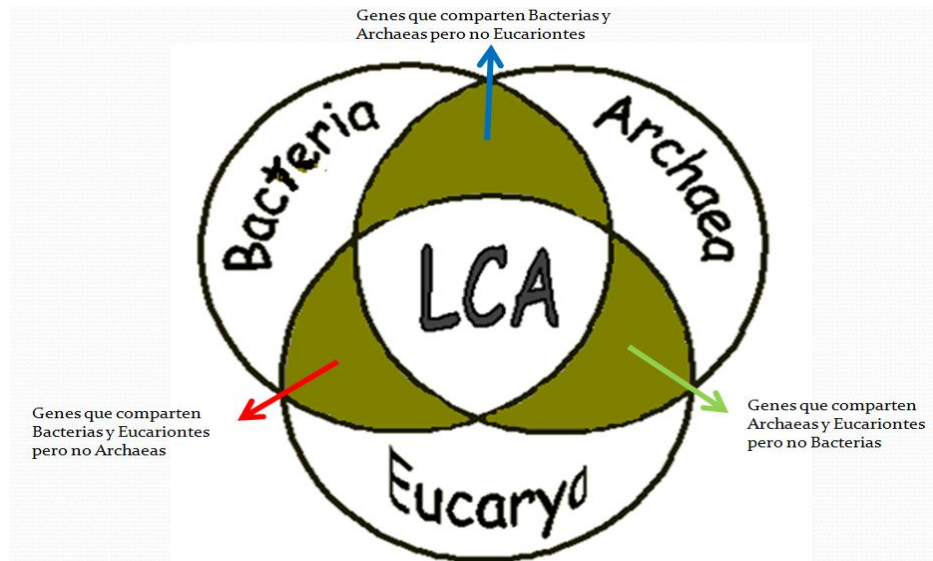


Fig. 11. Se indican las tres intersecciones de los análisis que se realizaron.

Los análisis de homología fueron realizados con el algoritmo de BLAST (Altschul, et al., 1997) de una sola vía y con un valor de “e” de 0.0001.

3.2 Análisis funcional de los genes conservados de acuerdo al KEGG

Con las secuencias homólogas conservados encontradas en los análisis anteriores se realizó un análisis funcional utilizando los organismos más estudiados de nuestra muestra de genomas, *E. coli* (Bacteria), *M. jannaschii* (Archaea) y *S. cerevisiae* (Eucarionte).

Dicho análisis consistió en clasificar funcionalmente los genes de *E. coli*, *M. jannaschii* y *S. cerevisiae* que se encuentran conservados en los distintos genomas de nuestra muestra, es decir;

- Genes de *E. coli* que se encuentran conservados en los genomas de Archaeas y Bacterias pero que no los presenta Eucariontes.
- Genes de *E. coli* que se encuentran conservados en los genomas de Eucariontes y Bacterias pero que no los presenta Archaeas.
- Genes de *M. jannaschii* que se encuentran conservados en los genomas de Archaeas y Bacterias pero que no los presenta Eucariontes.
- Genes de *M. jannaschii* que se encuentran conservados en los genomas de Archaeas y Eucariontes pero que no los presenta Bacterias.
- Genes de *S. cerevisiae* que se encuentran conservados en los genomas de Archaeas y Eucariontes pero que no los presenta Bacterias.

- Genes de *S. cerevisiae* que se encuentran conservados en los genomas de Bacterias y Eucariontes pero que no los presenta Archaeas.

Para realizar el análisis funcional se utilizó el programa GenEmerge, el cual se basa en la clasificación de los genes de acuerdo al Kyoto Encyclopedia of Genes and Genomes KEGG.

ftp://ftp.genome.jp/pub/kegg/pathway/map_title.tab

3.3 Interpretación de los genes conservados

De acuerdo a los genes conservados encontrados, interpretar que hay en términos funcionales en cada intersección de los dominios y a que se debe que se encuentre, si es por transferencia horizontal o transferencia vertical (ancestría común).

4. Resultados

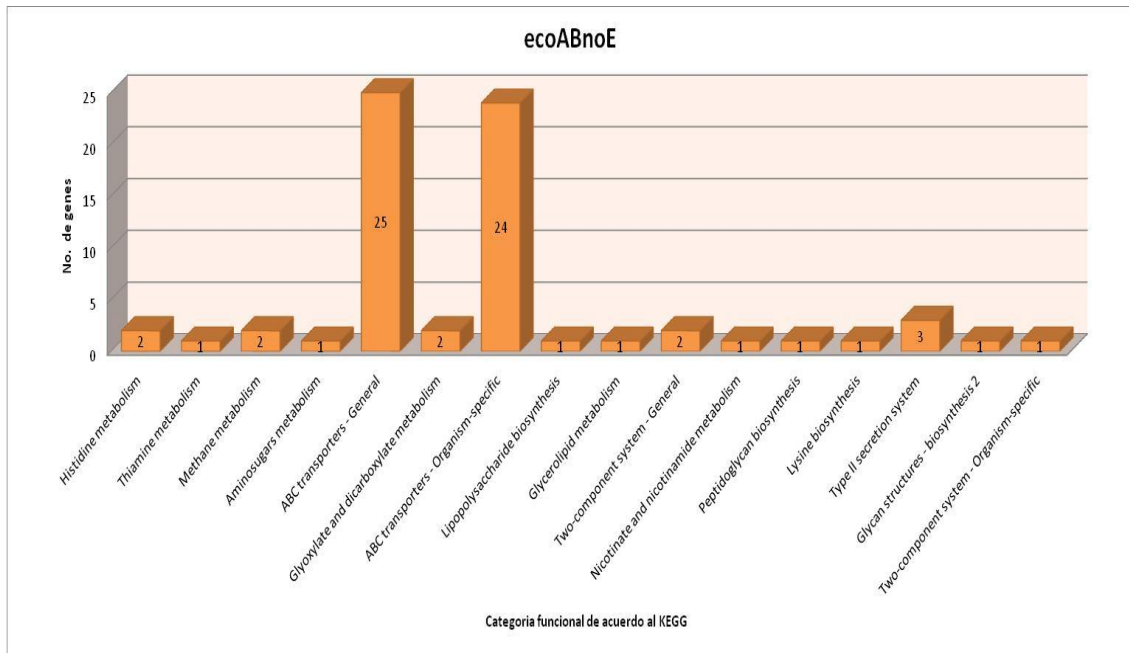
Genes de *E. coli* que se encuentran conservados en los genomas de Archaeas y Bacterias pero que no los presenta Eucariontes.

En este estudio del total de los genes de *E. coli* (4488 genes) se encontraron 77 genes conservados en los genomas de Archaeas y Bacterias. En la grafica 1, se muestran el número de genes así como a la categoría funcional a la que pertenecen de acuerdo al KEGG.

4. Resultados

Genes de *E. coli* que se encuentran conservados en los genomas de Archaeas y Bacterias pero que no los presenta Eucariontes.

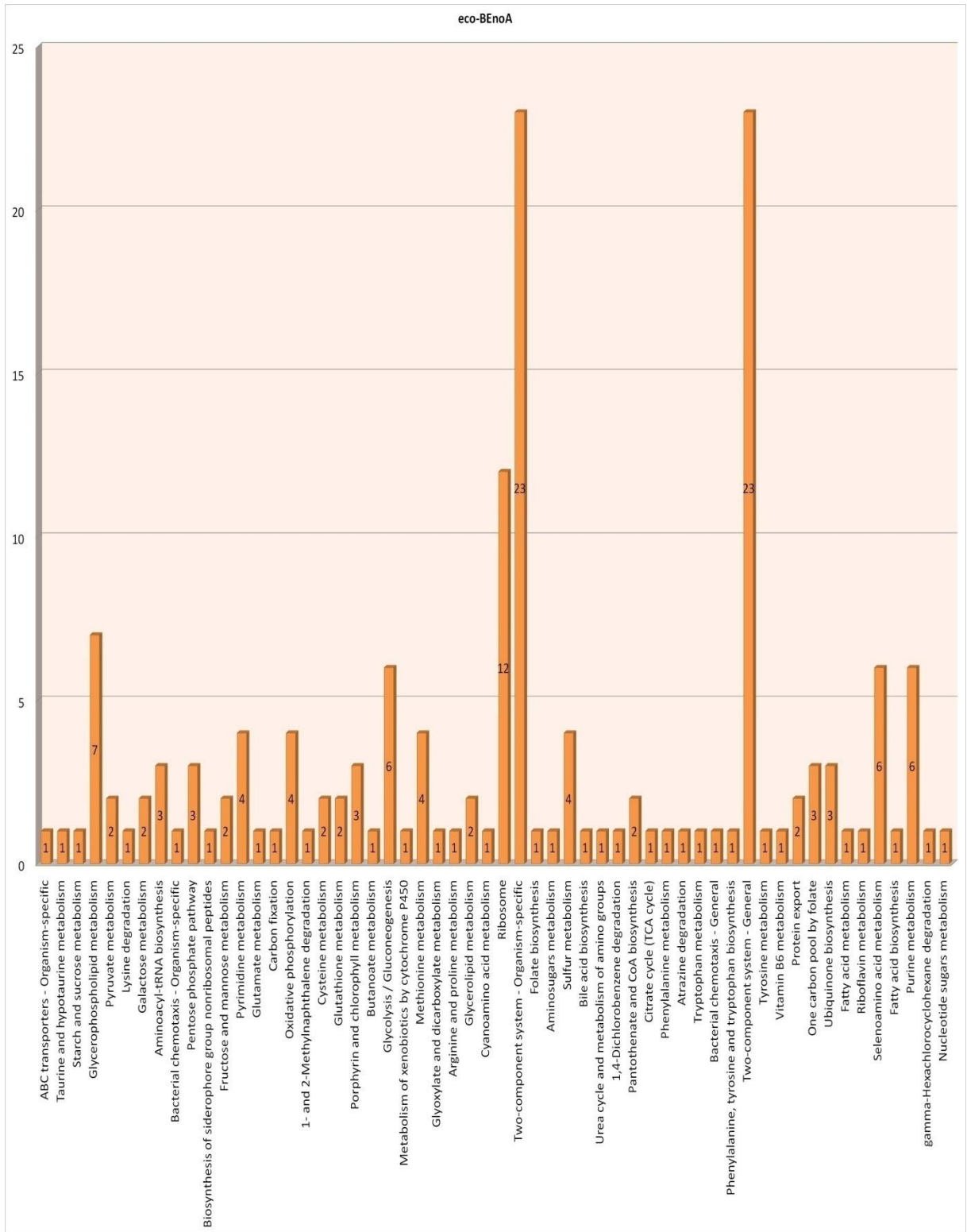
En este estudio del total de los genes de *E. coli* (4488 genes) se encontraron 77 genes conservados en los genomas de Archaeas y Bacterias. En la grafica 1, se muestran el número de genes así como a la categoría funcional a la que pertenecen de acuerdo al KEGG.



Grafica 1. Genes de *E. coli* conservados en Archaeas y Bacterias pero no en Eucariontes.

Genes de *E. coli* que se encuentran conservados en los genomas de Bacterias y Eucariontes pero que no los presenta Archaeas.

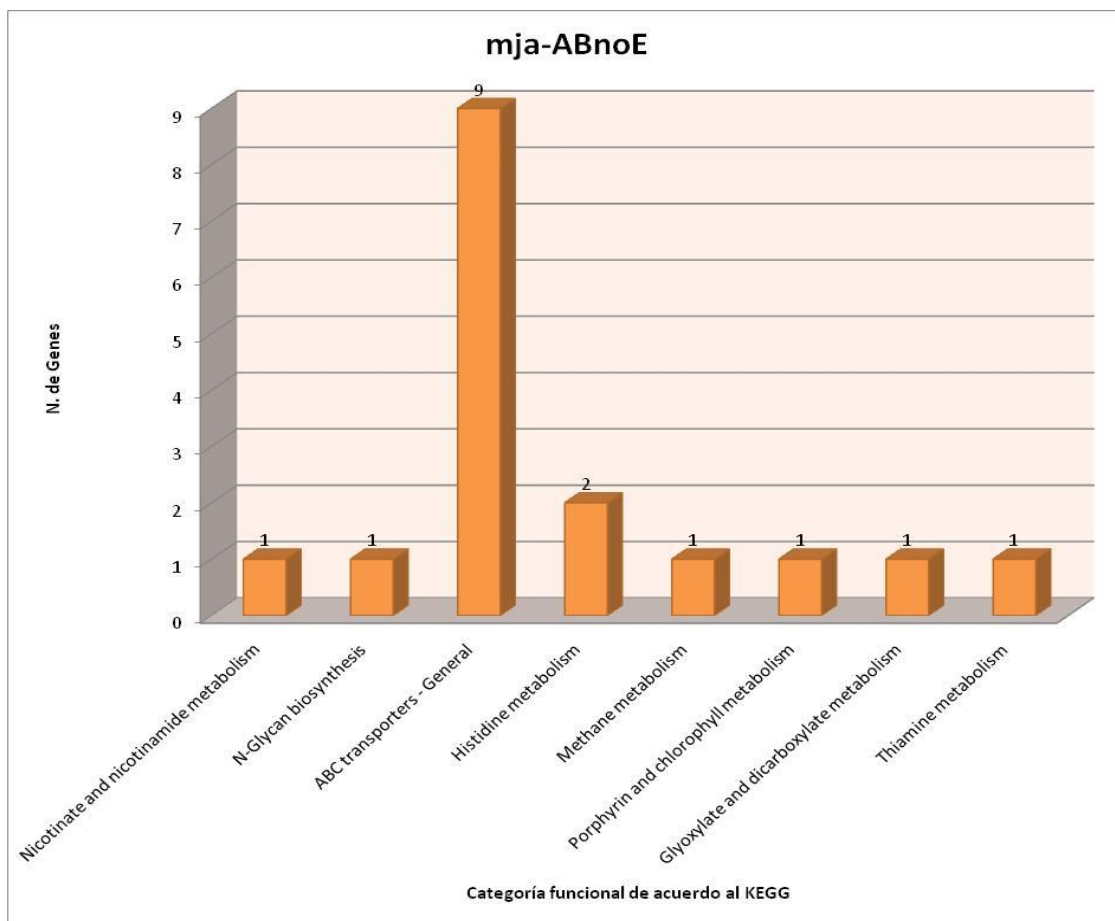
Del total de los genes de *E. coli* (4488 genes) para este caso se encontraron 220 genes conservados en los genomas de Bacterias y Eucariontes. En la grafica 2, se muestran el número de genes así como a la categoría funcional a la que pertenecen de acuerdo al KEGG.



Grafica 2. Genes de *E. coli* conservados en Bacterias y Eucariontes pero no Archaeas.

Genes de *M. jannaschii* que se encuentran conservados en los genomas de Archaeas y Bacterias pero que no los presenta Eucariontes.

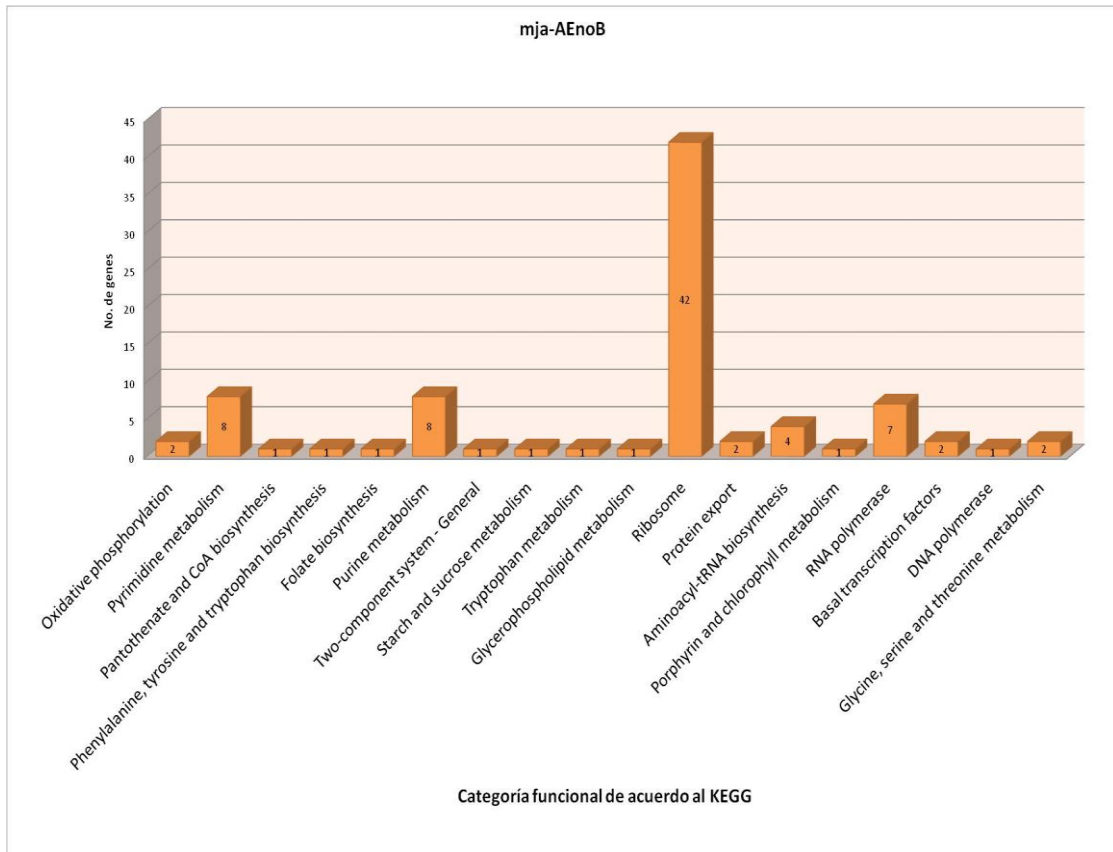
Del total de los genes de *M. jannaschii* (1830 genes) para este caso se encontraron 67 genes conservados en los genomas de Archaeas y Bacterias. En la grafica 3, se muestran el número de genes así como a la familia a la que pertenecen de acuerdo al KEGG.



Grafica 3. Genes de *M.jannaschii* conservados en Archaeas y Bacterias pero no Eucariontes.

Genes de *M. jannaschii* que se encuentran conservados en los genomas de Archaeas y Eucariontes pero que no los presenta Bacterias.

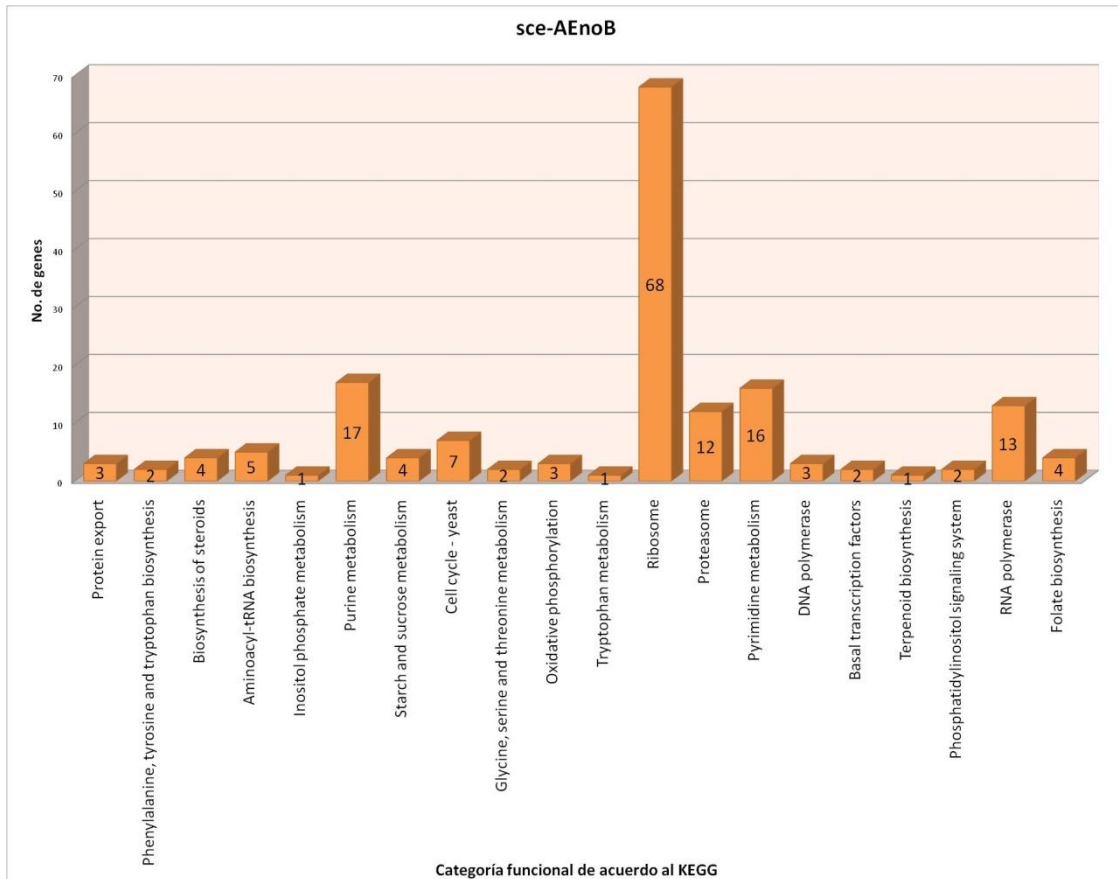
Del total de los genes de *M. jannaschii* (1830 genes) para este caso se encontraron 134 genes conservados en los genomas de Archaeas y Eucariontes. En la grafica 4, se muestran el número de genes así como a la familia a la que pertenecen de acuerdo al KEGG.



Grafica 4. Genes de *M.jannaschii* conservados en Archaeas y Eucariontes pero no Bacterias.

Genes de S. cerevisiae que se encuentran conservados en los genomas de Archaeas y Eucariontes pero que no los presenta Bacterias.

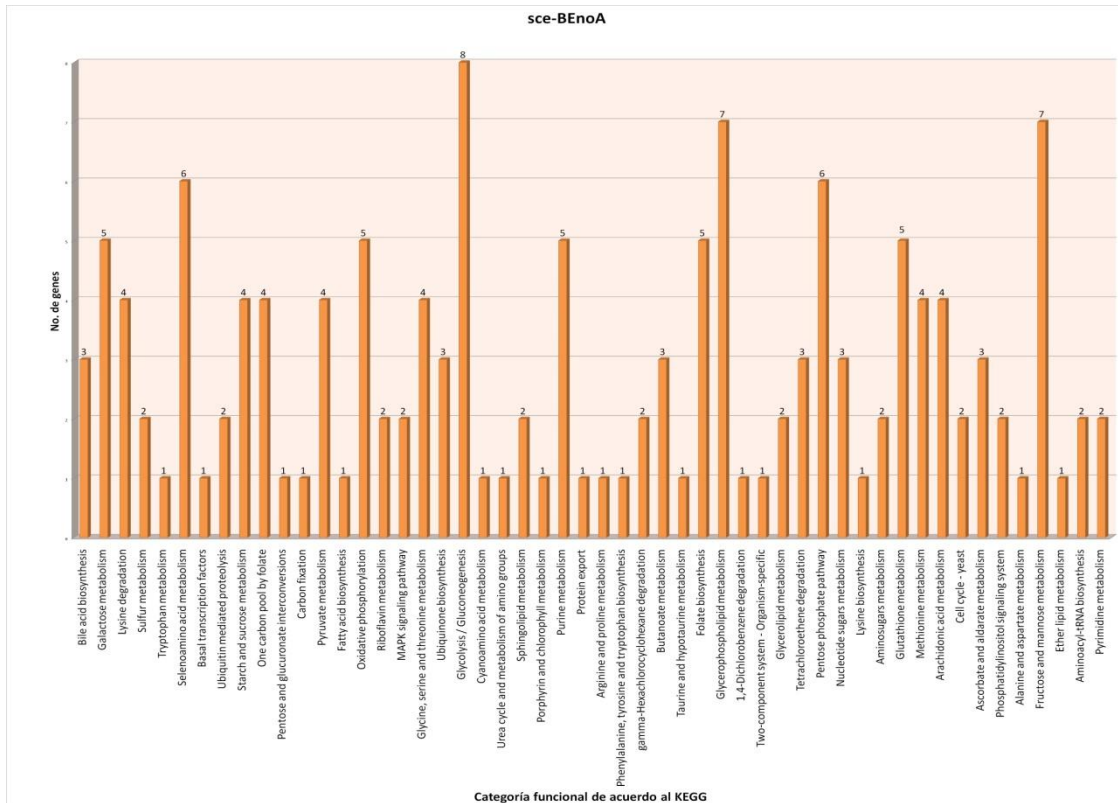
Del total de los genes de *S. cerevisiae* (6224 genes) para este caso se encontraron 211 genes conservados en los genomas de Archaeas y Eucariontes. En la grafica 5, se muestran el número de genes así como a la categoría funcional a la que pertenecen de acuerdo al KEGG.



Grafica 5. Genes de *S. cerevisiae* conservados en Archaeas y Eucariontes pero no Bacterias.

Genes de S. cerevisiae que se encuentran conservados en los genomas de Bacterias y Eucariontes pero que no los presenta Archaeas.

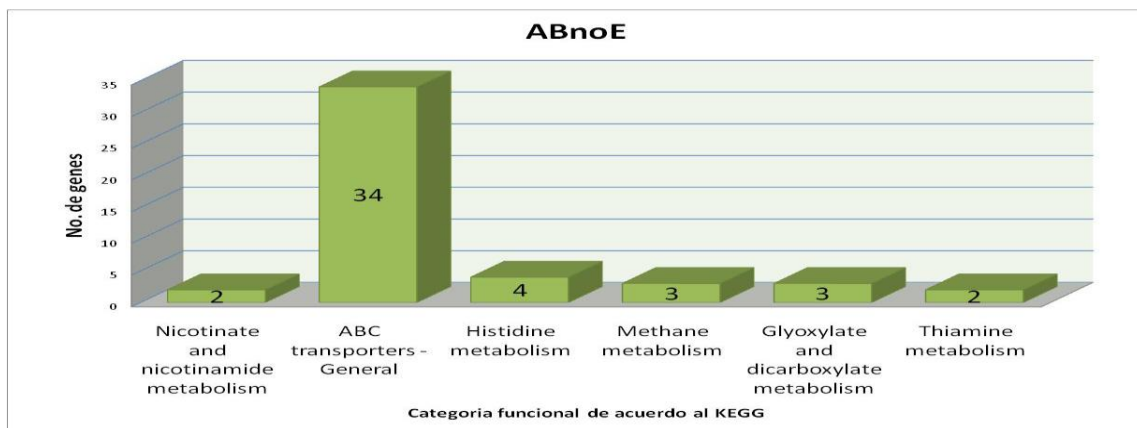
Del total de los genes de *S. cerevisiae* (6224 genes) para este caso se encontraron 305 genes conservados en los genomas de Bacterias y Eucariontes. En la grafica 6, se muestran el número de genes así como a la familia a la que pertenecen de acuerdo al KEGG.



Grafica 6. Genes de *S. cerevisiae* conservados en Bacterias y Eucariontes pero no Archaeas

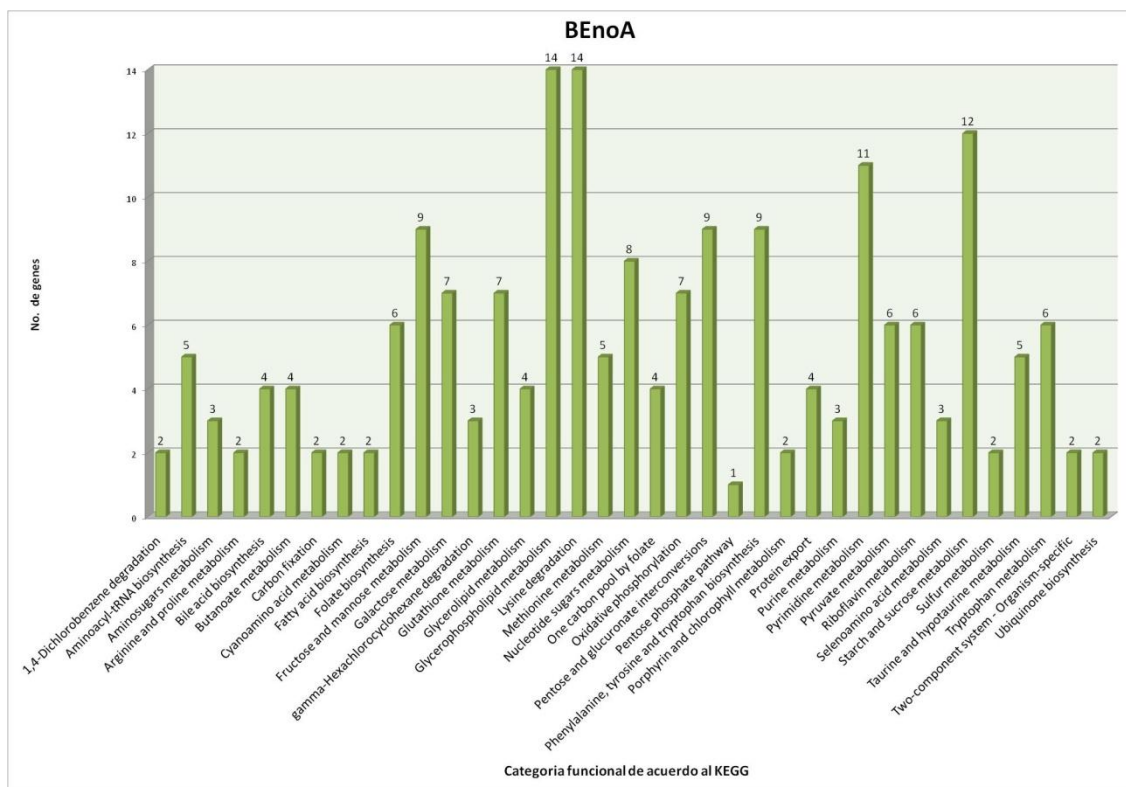
Para obtener un análisis más global de las gráficas anteriores, se analizaron y se hizo el resumen de la comparación de los genomas celulares, obteniendo tres gráficas finales donde se muestran los; Genes conservados en Archaeas y Bacterias excluyendo Eucariontes, Genes conservados en Bacteria y Eucariontes excluyendo Archaeas y Genes conservados en Archaeas y Eucariontes excluyendo Bacterias.

Genes que se encuentran conservados en genomas de Archaeas y Bacterias.



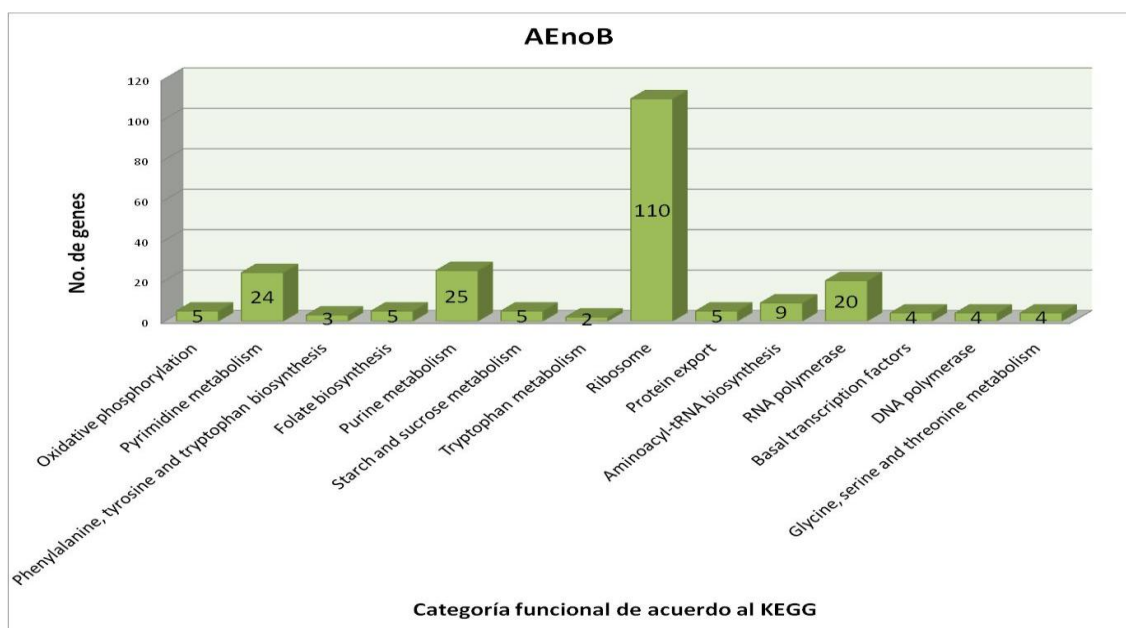
Grafica 7. Genes que se encuentra conservados en Arqueas y Bacterias pero no en Eucariontes.

Genes que se encuentran conservados en genomas de Bacterias y Eucariontes.



Grafica 8. Genes que se encuentra conservados en Bacterias y Eucariontes pero no en Archaeas.

Genes que se encuentran conservados en genomas de Archaeas y Eucariontes.



Grafica 9. Genes que se encuentra conservados en Archaeas y Eucariontes pero no en Bacterias.

5. Discusión y conclusiones

1) Genes que se encuentran conservados en Archaeas y Bacterias con exclusión de Eucariontes.

Los genes conservados en Archaeas y Bacterias en principio serán el resultado de la transferencia horizontal entre estos grupos, o bien genes que se encontraban en el cenozoario, y que posiblemente se perdieron en la evolución temprana del linaje eucarionte y que se retuvieron entre Archaeas y Bacterias debido a que son importantes para el estilo de vida procarionte.

Para nuestro estudio, con la base de datos de genomas utilizados, encontramos que a pesar de que tanto bacterias como Archaeas son procariontes, los genes que comparten excluyendo a los eucariontes son relativamente pocos, por lo que esto nos da un indicio de la enorme distancia evolutiva que hay entre estos dos linajes de organismos.

En la grafica 7, identificamos que los genes conservados en estos grupos son totalmente metabólicos (genes operacionales) y en su mayoría son pertenecen a la familia del tipo ABC transportadores, que es una de las familias más comunes de transportadores, (por ejemplo, casi el 5% del genoma de *Escherichia coli* codifica componentes de transportadores ABC). Esta familia además posee un dominio ABC conservado muy antiguo que al parecer ya existía antes de la divergencia evolutiva entre procariontes y eucariontes (Davidson A.L, *et al*, 2008; Higgins, 2001). Se sabe que también hay ABC transportadores en Eucariontes (Davidson A.L, *et al*, 2008), sin embargo el hecho de que en nuestro análisis solo estén conservados ABC transportadores en Archaeas y Bacterias y no en Eucariontes se deba principalmente a la gran cantidad de ABC transportadores que existen y a su diversidad de funciones fisiológicas que realizan en este tipo de organismos (es decir, el grupo de ABC transportadores detectados debe de pertenecer a una sub-familia presente solo en procariontes). Ya que en procariontes los ABC transportadores tienen la principal función de transportar nutrientes al organismo así como factores de virulencia, debido a su estilo de vida, los sustratos a transportar entre procariontes debe de ser similar, en este caso.

Estudios previos con ABC transportadores apoyan la hipótesis de que los ABC transportadores ya se encontraban en el cenancestro o LUCA, el cual se heredaron a los procariontes y posteriormente las Archaeas perdieron algunos tipos de ABC transportadores. Los Eucariontes los adquirieron mediante la endosimbiosis y con el tiempo hubo modificaciones como duplicaciones, etc., que llevaron a su diferentes funciones en eucariontes (Davidson A.L, *et al*, 2008), lo que nos hace suponer que los genes de ABC transportadores entre los procariontes son muy similares por su forma de vida y en cambio en los eucariontes solo tendrían algunos dominios conservados como el dominio ABC.

En cuanto a los demás genes conservados que comparten los grupos procariontes son totalmente metabólicos, algunos de ellos muy particulares de metabolismos microbianos como lo son los genes del metabolismo del metano.

2) Genes que se encuentran conservados en Bacterias y Eucariontes con exclusión de las Archaeas.

Los genes conservados entre Bacterias y Eucariontes pertenecen a una gran cantidad de clases funcionales, la mayoría de ellas metabólicas (genes operacionales), probablemente se deba a la transferencia de genes de los ancestros de los organelos (como las mitocondria y cloroplastos) al núcleo-citoplasma eucarionte.

Se ha demostrado relación evolutiva cercana entre genes de Bacterias y Eucariontes. Análisis de genomas de Bacterias con genomas de Eucariontes muestran la presencia de homólogos de genes eucariontes implicados en las redes de transducción de señales y en varias rutas de muerte celular programada de plantas y animales. Esto podría indicar que al menos alguna de las complejas redes de señalización de eucariotas tiene sus equivalentes y predecesores evolutivos en bacterias. Mediante estudios de genómica comparada se han detectado una gran variedad de sistemas de transducción de señales en Bacterias y Archaeas que parecen ser el reflejo del estilo de vida del organismo (Koonin y Wolf ,2008).

Para nuestro análisis encontramos que los genes conservados que comparten estos grupos concuerdan con la hipótesis de la transferencia de genes por medio del ancestro de la mitocondria. Por ejemplo vemos en la grafica 8, que algunos de los

genes tienen que ver con características del metabolismo fotosintético (fijación del carbono, metabolismo de clorofila, porfirinas y riboflavinas, etc.) esto es porque tanto organismos procariontes como *Synechocystis* sp.(cianobacteria) y eucariontes como *Arabidopsis thaliana* (planta) realizan este proceso. Lo que apoya que la vía por la cual los Eucariontes obtuvieron esos genes es por la endosimbiosis. Estudios previos han demostrado que el 18% de los genes de *Arabidopsis thaliana* fueron originados en *Cyanobacteria* debido a que la endosimbiosis jugó un papel importante en el genoma de los Eucariontes fotosintéticos (Palenik, 2002).

Otros genes identificados en ambos grupos son los que tienen que ver con el proceso de la cadena de transporte de electrones, β -oxidación, metabolismo de la fructosa y manosa, síntesis de lípidos y metabolismo de sulfuro que concuerdan con el trabajo de Gabaldón y Huynen (2003) donde sugiere que estos genes son de origen proto-mitocondrial, por lo que fueron adquiridos por la endosimbiosis. Además encontramos genes relacionados con el ciclo del ácido cítrico y de las rutas pentosas fosfato que de acuerdo a Gabaldón y Huynen (2003) solo se recuperaron parcialmente y para el caso del metabolismo de aminoácidos y metabolismo de nucleótidos solo se tienen algunos genes aislados de estas vías metabólicas.

Los resultados nos parecen indicar que aunque existen varios genes que provinieron del ancestro de los organelos a los Eucariontes mediante endosimbiosis y transferencia horizontal, es una fracción mínima lo que demuestra que no son la única relación ancestral simbiótica.

3) Genes que se encuentran conservados en Archaea y Eucariontes con exclusión de las Bacterias.

Los genes que estén compartidos entre Archaeas y Eucariontes serán sinapomorfias, es decir que se originaron un poco antes del ancestro de este grupo, por tanto la transferencia será por herencia vertical, sin descartar que también haya transferencia horizontal donde se involucrarán procariontes como las Archaeas.

Las Archaeas y los Eucariontes comparten genes importantes que están implicados en la replicación y reparación del ADN, así como la traducción (genes informacionales), que son mucho menos propensos al proceso de transferencia horizontal de genes que los genes operacionales que codifican enzimas metabólicas,

sistemas de transporte y otras proteínas. El razonamiento subyacente es que estos genes codifican partes de complejas maquinarias moleculares que están fuertemente co-adaptadas y por tanto no pueden ser desplazadas fácilmente por genes ortólogos adquiridos de otros organismos. No obstante, hay muchos casos en que los genes informacionales como la aminoacil-tRNA sintetasa y proteínas ribosomales han sufrido eventos de transferencia horizontal de genes (Jain *et al.*, 1999).

Con respecto a nuestro análisis coincide con los resultados que se han mencionado en otros estudios donde encontramos que principalmente los genes conservados en Archaeas y Eucariontes excluyendo a las Bacterias son del tipo informacional; en su mayoría ribosomales, RNA ribosomales, DNA polimerasa, factores de transcripción, etc. Lo cual apoya la hipótesis de que las Archaeas tienen más parecido con Eucariontes que con Bacterias en cuanto a los procesos como la transcripción y traducción. Además apoya la filogenia de Woese, sobre que Archaeas y Eucariontes son grupos hermanos.

De acuerdo a nuestro análisis y a otros estudios en la siguiente figura, resumimos los patrones de conservación de los genes.

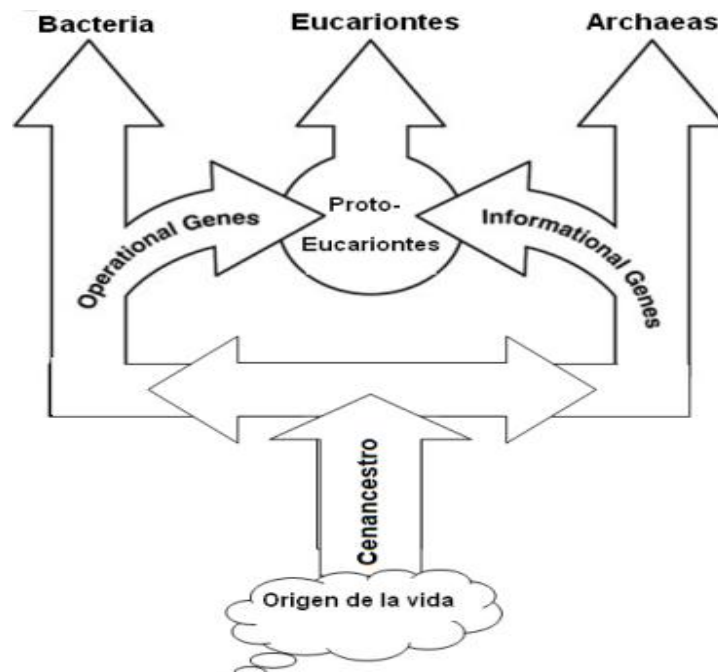


Fig. 12. Los estudios sugieren que los eucariontes son una mezcla de genes de procariontes; los bacterianos que en su mayoría son genes operacionales y los genes informacionales por parte de las Archaeas. En cuanto a Bacterias y Archaeas pudo ser un intercambio de genes entre ambos linajes, o en su caso por herencia del cenacestro. La herencia pudo ser por transferencia horizontal como herencia vertical. Modificado de Simonson *et al.*, 2005.

Sin duda la conservación de genes entre los distintos linajes celulares (Archaeas, Bacterias y Eucariontes) tiene un carácter central en la evolución, así como otros procesos que también son esenciales como es el caso de la transferencia vertical y la transferencia horizontal donde en estudios recientes se ha visto que ha rebasado las fronteras de los dominios filogenéticos.

Si bien en este trabajo se obtuvieron resultados que se pueden discutir y refinar, no podemos dejar de lado, que para poder obtener resultados satisfactorios el análisis depende de varios factores, entre los que destacan, la clasificación de los genes de acuerdo al KEGG (Kyoto Encyclopedia of Genes and Genomes), que muchas veces con seguridad son imprecisas e incompletas, por ejemplo, en varias ocasiones la homología no necesariamente nos dice que ese gen realiza la misma función en otro organismo, por lo que no siempre la anotación de los genomas y por tanto de los genes son del todo verdadero. Otro problema que nos encontramos al hacer este tipo de análisis es que aunque se tenga secuenciado el 100 % de genoma de algún organismo no siempre se tienen las anotaciones de los genes, aproximadamente entre el 60y 70% de los genes codificadores de proteínas se les asigna una función específica, y entre el 10% y 15% se les asigna una predicción funcional (Koonin y Wolf, 2008) y sin duda estos factores debilitan este tipo de análisis.

Otras variables que pueden afectar los resultados obtenidos son el tamaño y tipo de la base de datos, así como el método utilizado.

Por otro lado, este trabajo es un intento para comprender funcionalmente las similitudes y diferencias genéticas entre los principales linajes de organismos (es decir, Archaeas, Bacterias y Eucariontes).

No descartamos que esto es un indicio para realizar análisis con una mejor muestra y más exhaustivos con la ayuda de los avances en la tecnología, así como en el aumento de los genomas de organismos secuenciados.

6. Bibliografía

Altschul, S.F., T.L. Madden, A.A. Schäfer, J. Zhang, Z. Zhang, W. Miller y D.J. Lipman (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Reserch* **25**(17):3389-402.

Becerra, A., S. Islas, J.I. Leguina., E. Silva y A. Lazcano (1997) Polyphyletic gene losses can bias backtrack characterizations of the cenancestor. *Journal of molecular evolution* **45**(2):115-8.

Becerra A, Delaye L, Islas S, Lazcano A (2007) Very early stages of biological evolution related to the nature of the last common ancestor of the three major cell domains. *Annual Review of Ecology, Evolution, and Systematics. Annual reviews* **38**: 361-379

Brown, J.R. y W.F. Doolittle, (1997) Archaea and the prokaryote-to-eukaryote transition. *Microbiology and Molecular Biology Reviews* **61**: 456-502

Brown JR, Douady CJ, Italia MJ, Marshall WE, Stanhope MJ. (2001) Universal tres based on large combined protein sequence datasets. *Nat. Genet.* **28**:281–85

Brown, J.R, (2003) Ancient Horizontal Gene Transfer. *Nature Reviews Genetics.***4**:121-132

Doolittle, W.F. (1999) Phylogenetic classification and the universal tree, *Science* **284** (1999), p. 2124.

Chatton, E. (1938) Titres et travaux Scientifiques(1906-1937) de Edouard Cahtton. E. Sottano, Seté, France. Referenciado en Woese et al., 1990.

Davidson A.L., Dassa E., Orelle E., and Chen J. (2008) Structure, Function, and Evolution of Bacterial ATP-Binding Cassette Systems. *Microbiology and Molecular Biology REVIEWS* **72**: 317–364

Delaye, L., A. Becerra and A. Lazcano (2005) The last common ancestor: what's in a name? *Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life* **35**(6):537-54.

Delaye, L., Vázquez, H. and Lazcano, A.: (2001), The Cenancestor and its Contemporary Biological Relics: The Case of Nucleic Acid Polymerases, *First Steps in the Origin of Life in the Universe*, Kluwer Academic Publishers; 223–230

Doolittle, W.F. (1999) Lateral genomics, *Trends Cell Biology* **9** (1999), p. M5

FITCH W. M. (1970). Distinguishing homologous from analogous proteins. *Systematic Zoology* **19**: 99-113.

Fitch WM, Upper K. (1987) The phylogeny of tRNA sequences provides evidence of ambiguity reduction in the origin of the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* **52**:759–67

Fitch, W.M. (2000). Homology: a personal view on some of the problems. *Trends Genet.* **16**(5):227-31

Gabaldón T. and Huynen M. A. (2003) Reconstruction of the Proto-Mitochondrial Metabolism. *SCIENCE* **301**: 609

Gogarten, J. P., H. Kibak, P. Dittrich, L. Taiz, E.J. Bowman, B.J. Bowman, M.L. Manolson, J. Poole, T. Date, L. Oshima, L. Konishi, K. Denda y M. Yoshida (1989) Evolution of the vacuolar H⁺-ATPase, implications for the origin of eukaryotes. *Proceeding National Academy Sciences of the United States of America* **86**: 6661-6665

Haeckel, E. (1866) *Generelle Morphologie der Organismen. Allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von Charles Darwin reformierte Deszendenz-Theorie. Band I: Allgemeine Anatomie der Organismen.* Georg Reimer, Berlin.

Haeckel, E. (1868) *Natürliche Schöpfungsgeschichte.* Reimer, Berlin.

Higgins, C.F., (2001) ABC transporters: physiology, structure and mechanism – an overview. *Res. Microbiol.* **152** : 205–210

Hutchison C.A., S.N. Peterson, S.R. Gill, R.T. Cline, O. White, C.M. Fraser, H.O. Smith y J.C. Venter (1999) Global Transposon Mutagenesis and a Minimal Mycoplasma Genome. *Science*, **286**:, 2165–2169.

Iwabe, N., K. Kuma, M. Hasegawa, S. Osawa y T. Miyata (1989) Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proceedings of the National Academy of Sciences of the United States of America* **86**: 9355-9359.

Jain, R., Rivera, M. C., and Lake, J. A. (1999). Horizontal gene transfer among genomes: The complexity hypothesis, *Proc. Natl. Acad. Sci. USA* **96**: 3801–3806.

Karlin, S, y S.F. Altschul (1993). Applications and statistics for multiple high-scoring segments in molecular sequences. *Proc. Natl. Acad. Sci.* **90**:5873-7.

Koonin, E. V., K. S. Makarova, y L. Aravind (2001) Horizontal gene transfer in prokaryotes—quantification and classification. *Annu. Rev. Microbiol.* **55**:709-742

Koonin, E. V., and Yuri I. Wolf (2008) Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Research.* **36**: 6688–6719

Lazcano, A., G.E. Fox, y J. Oró (1992) Life before DNA: the origin and evolution of early Archean cells. In R. P. Mortlock (ed) *The Evolution of Metabolic Function* (CRC Press, Boca Raton), 237-295

Lazcano, A., E. Díaz-Villagómez; T. Mills y J. Oró (1995) On the levels of enzymatic substrate specificity: implications for the early evolution of metabolic pathways. *Advances in space research : the official journal of the Committee on Space Research (COSPAR)* **15**(3):345-56.

Lynn Margulis (1970) 'Origin of Eukaryotic Cells' Yale University Press

Mushegain A.R. (1999) The minimal genome concept. *Current Opinion in Genetics & Development.* **9**: 709-714

Mushegain A.R. y E.V. Koonin (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl Acad. Sci. USA*, **93**:, 10268–10273.

Palenik, B.(2002) The genomics of symbiosis: host keep the baby and the bath water. *Proc. Natl Acad. Sci. USA* **99**: 11996–11997.

Rouzé, P., N. Pavy y S. Rombauts (1999) Genome annotation: which tools do we have for it?. *Curr Opin Plant Biol* **2**, 90-5.

Simonson AB, Servin JA, Skophammer RG, Herbold CW, Rivera MC, Lake JA. (2005) Decoding the genomic tree of life. *Proc Natl Acad Sci U S A.* **102**; 6608-13.

Tatusov, R. L., E. V. Koonin y D. J. Lipman. 1997. A genomic perspective on protein families. *Science* **278**:631-637.

Tekaia, F., A. Lazcano y B. Dujon (1999a) The genomic tree as revealed from whole proteome comparisons. *Genome Research* **9**: 550-557

Tekaia, F., B. Dujon y A. Lazcano (1999b) Comparative genomics: products of the most conserved protein-encoding genes synthesize, degrade, or interact with RNA *Abstracts of the 12th International Conference on the Origin of Life & 9th ISSOL Meeting (San Diego, California, USA, July 11-16, 1999)* , Abstract c4.6, p. 53

Woese, C.R. y G.E. Fox (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA* **74**: 5088-5090

Woese C.R. (1983) *Evolution from Molecules to Men*, ed Bendall D S (Cambridge Univ. Press, Cambridge, U.K.) pp 209–233.

Woese,C.R. (1987) Bacterial evolution. *Microbiological Reviews* **51**: 221-271.

Woese,C.R., O. Kandler, y M.L. Wheelis (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America* **87**: 4576-4579.

Woese, C.R. (1998) The universal ancestor. *PNAS* **95**:6854-6859

Wolf, Y. I., Rogozin, I. B., Grishin, N. V. & Koonin, E. V.(2002) Genome trees and the tree of life. *Trends Genet.* **18**, 472–479.

Zhaxybayeva, O., P. Lapierre , y J.P. Gogarten (2004) Genome mosaicism and organismal linkages. *Trends Genet* **20**:254–260.

Zuckerkandl, E., y L. Pauling (1965). Molecules as documents of evolutionary history. *Journal of Theoretical Biology* **8**: 357-366

Sitios WEB.

Kyoto Encyclopedia of Genes and Genomes KEGG.

<ftp://ftp.genome.jp/pub/kegg/genomes/>

7. Apéndice

Analisis con GenEmerge.

1) Genes de *E.coli* conservados en Archaeas y Bacterias pero no en Eucariontes.

Gene	Pop	freq	Study	frac	Raw	es	score	Description	Contributing	genes																							
GeneMerge v1.2: esp_salida.xls																																	
Gene-Association File: esp_gene_map.tbl.f																																	
Description File: map_title.tbl																																	
Population File: population.e.coli.nuc																																	
Study File: study.e.coli.nuc																																	
340	0.002458906	nov-38	feb-77	0.014956701	0.21307217			Histidine metabolism	b2024	b2025																							
730	0.003787078	17/4488	ene-77	NA	NA			Thiamine metabolism	b0417																								
680	0.002873795	dic-38	feb-77	0.01793875	0.274510051			Methane metabolism	b4079	b3894																							
530	0.004456227	20/4488	ene-77	NA	NA			Aminosugars metabolism	b2225																								
2010	0.045900170	20/4488	25/77	1.205-01	1.32E+00			ABC transporters - General	b1711	b4208	b0153	b4209	b0590	b0589	b2726	b2727	b2728	b2424	b0067	b2423	b0784	b1124	b1125	b0857	b0856	b4033	b1443	b1442	b4032	b1211	b4230	b4290	b0592
630	0.007575757	34/4488	feb-77	0.114716801				1 Glyoxylate and dicarboxylate metabolism	b4079	b3894																							
2011	0.045379321	19/4488	24/77	2.295-01	3.67E+00			ABC transporters - Organism-specific	b1711	b4208	b0153	b4209	b0590	b0589	b2726	b2727	b2728	b2424	b0067	b2423	b0784	b1124	b1125	b0857	b0856	b4033	b1443	b1442	b4032	b1211	b4230	b4290	b0592
340	0.002208929	20/4488	ene-77	NA	NA			Lipopolysaccharide biosynthesis	b3628																								
530	0.004456227	20/4488	ene-77	NA	NA			Glycerolipid metabolism	b2945																								
2010	0.029108940	13/4488	feb-77	0.663789502				1 Two-component system - General	b2253	b0997																							
780	0.005792229	26/4488	ene-77	NA	NA			Nicotinate and nicotinamide metabolism	b0750																								
530	0.003565621	16/4488	ene-77	NA	NA			Peptidoglycan biosynthesis	b0807																								
330	0.003565621	16/4488	ene-77	NA	NA			Lysine biosynthesis	b2478																								
3090	0.005570492	25/4488	mar-77	0.006134478	0.136215588			Type II secretion system	b3326	b2950	b0107																						
1010	0.002896121	13/4488	ene-77	NA	NA			Glycan structures - biosynthesis 2	b3628																								
2012	0.019395029	8/4488	ene-77	NA	NA			Two-component system - Organism-specif	b0997																								
Total number of genes: 4488																																	
Total number of Study genes: 77																																	
Total number of Study gene GMRG terms (pop non-singlets): 16 (16)																																	
Genes with GMRG information: 41																																	
Genes with no GMRG information: 36																																	
These are:	b4290	b0298	b2059	b1022	b3615	b0599			b2712	b1497	b3800	b3781	b0444	b0661	b0835	b3350	b0047	b2777	b1029	b1175	b4389	b1507	b1598	b0894	b1072	b1272	b3857	b2385	b1053	b0427	b1065	b0045	b2781

2) Genes de *E. coli* conservados en Bacterias y eucariontes pero no en Archaeas

GeneMerge v1.2; eco_salida.xls

Gene Association File: eco_gene_map.tab.f
 Description File: map_title.tab
 Population File: population_e.coli.nuc
 Study File: study_e.coli.nuc

GMWG	Term	Pop. frac	Pop. frac	Study frac	Raw es	e-score	Description	Contributing genes
2011	0.04257932	151/4488	1/220	NA	NA	ABC transporters - Organism-specific	b1710	
430	0.00138696	jun-88 1/220	NA	NA	NA	Taurine and hypotaurine metabolism	b3447	
500	0.007130124	12/4488	1/220	NA	NA	Starch and sucrose metabolism	b3432	
564	0.00681042	27/4488	7/220	0.000236520	0.013014796	Glycerophospholipid metabolism	b0175 b1409 b1912 b3426 b1241 b3608 b3018	
620	0.00581105	43/4488	2/220	0.63986385		1 Pyruvate metabolism	b1241 b1651	
110	0.004010695	12/4488	1/220	NA	NA	Lysine degradation	b0727	
52	0.007130124	12/4488	2/220	0.470166070		1 Galactose metabolism	b0759 b3916	
570	0.005370492	25/4488	3/220	0.121174518		1 Aminoacyl-tRNA biosynthesis	b3384 b1657 b3288	
2011	0.00467044	21/4488	1/220	NA	NA	Bacterial chemotaxis - Organism-specific	b1382	
20	0.006684481	20/4488	3/220	0.179919791		1 Pentose phosphate pathway	b1852 b4094 b3916	
1053	0.00138696	jun-88 1/220	NA	NA	NA	Biosynthesis of siderophore group nonrib	b0596	
51	0.012620861	54/4488	2/220	0.761136320		1 Fructose and mannose metabolism	b3919 b3919	
240	0.011490129	50/4488	4/220	0.220263009		1 Pyrimidine metabolism	b3164 b2066 b2235 b2676	
251	0.006971088	31/4488	1/220	NA	NA	Glutamate metabolism	b1014	
710	0.005475952	34/4488	1/220	NA	NA	Carbon fixation	b3919	
190	0.005347591	41/4488	4/220	0.139175442		1 Oxidative phosphorylation	b0428 b3734 b3733 b3735	
624	0.001762521	ago-88 1/220	NA	NA	NA	1- and 2-Methylnaphthalene degradation	b1241	
272	0.003424249	15/4488	2/220	0.165496757		1 Cysteine metabolism	b2414 b2421	
480	0.002450980	nov-88 2/220	0.096318491			1 Glutathione metabolism	b1852 b3447	
860	0.005347591	24/4488	3/220	0.110425063		1 Porphyrin and chlorophyll metabolism	b0428 b0638 b3844	
650	0.005347591	41/4488	1/220	NA	NA	Butanoate metabolism	b1241	
10	0.008912659	40/4488	6/220	0.012278624	0.675225301	Glycolysis / Gluconeogenesis	b3919 b1241 b4395 b0755 b1779 b3916	
980	0.00138696	jun-88 1/220	NA	NA	NA	Metabolism of xenobiotics by cytochrome b	b1241	
271	0.00365062	16/4488	4/220	0.006414787	0.352811307	Methionine metabolism	b0261 b4019 b2542 b3288	
630	0.00576757	34/4488	1/220	NA	NA	Glyoxylate and dicarboxylate metabolism	b2579	
330	0.004563227	20/4488	1/220	NA	NA	Arginine and proline metabolism	b1014	
561	0.004563227	20/4488	2/220	0.256707366		1 Glycerolipid metabolism	b1241 b3018	
460	0.00133714	jul-88 1/220	NA	NA	NA	Cyanosulfonamide metabolism	b3447	
3010	0.01360452	79/4488	12/220	0.000294659	0.021706227	Ribosome	b3342 b3186 b0911 b3231 b0169 b2609 b3319 b3986 b3321 b3301 b3307 b3320	
2012	0.01360452	87/4488	23/220	5.72E+02	5.35E+04	Two-component system - Organism-speci	b2786 b2218 b3210 b1969 b2220 b1235 b4004 b3270 b3405 b0571 b2079 b0995 b4113 b2554 b3868 b1130 b4401 b0993 b3025 b1608 b4398 b0399 b3912	
790	0.002896613	13/4488	1/220	NA	NA	Folate biosynthesis	b3215	
530	0.004563227	20/4488	1/220	NA	NA	Aminosugars metabolism	b0676	
520	0.002896613	13/4488	4/220	0.002893125	0.155659185	Sulfur metabolism	b2414 b2421 b2750 b2764	
120	0.002201631	oct-88 1/220	NA	NA	NA	Bile acid biosynthesis	b1241	
220	0.00461672	29/4488	1/220	NA	NA	Urea cycle and metabolism of amino grou	b0343	
617	0.00689449	mar-88 1/220	NA	NA	1,4-Dichlorobenzene degradation	b3830		
770	0.004339311	19/4488	2/220	0.238196496		1 Pantotheate and CoA biosynthesis	b0103 b3974	
20	0.006338959	28/4488	1/220	NA	NA	Citrate cycle (TCA cycle)	b0727	
960	0.004339311	19/4488	1/220	NA	NA	Phenylalanine metabolism	b0549	
791	0.000223816	ene-88 1/220	NA	NA	NA	Atrazine degradation	b2559	
980	0.002896613	13/4488	1/220	NA	NA	Tryptophan metabolism	b3384	
2010	0.00370707	17/4488	1/220	NA	NA	Bacterial chemotaxis - General	b1382	
400	0.005791261	26/4488	1/220	NA	NA	Phenylalanine, tyrosine and tryptophan b	b1637	
2010	0.02618846	111/4488	23/220	5.48E+06	3.01E+09	Two-component system - General	b2786 b2218 b3210 b1969 b2220 b4004 b3270 b3405 b0571 b2079 b0995 b4113 b2554 b3868 b1382 b1130 b4401 b0993 b3025 b1608 b4398 b0399 b3912	
510	0.001342492	14/4488	1/220	NA	NA	Tyrosine metabolism	b1241	
750	0.00285547	sep-88 1/220	NA	NA	NA	Vitamin B6 metabolism	b2927	
3600	0.00365062	16/4488	2/220	0.183345069		1 Protein export	b0096 b3705	
670	0.002450980	nov-88 3/220	0.014308953	0.786992467		One carbon pool by folate	b4019 b3288 b4006	
130	0.006894491	30/4488	3/220	0.179919791		1 Ubiquinone biosynthesis	b4040 b0662 b2907	
71	0.003424249	15/4488	1/220	NA	NA	Fatty acid metabolism	b1241	
740	0.003424249	15/4488	1/220	NA	NA	Riboflavin metabolism	b0025	
450	0.00365062	16/4488	6/220	6.86E+09	0.093774139	Selenoamino acid metabolism	b2414 b2421 b2750 b2764 b2942 b3447	
61	0.002671796	dic-88 1/220	NA	NA	NA	Fatty acid biosynthesis	b1392	
230	0.017179679	76/4488	6/220	0.181830736		1 Purine metabolism	b3164 b2235 b2676 b2750 b3848 b4006	
361	0.000891263	abr-88 1/220	NA	NA	NA	gamma-Hexachlorocyclohexane degradat	b3830	
520	0.00365062	16/4488	1/220	NA	NA	Nucleotide sugars metabolism	b0759	

Total number of genes: 4488
 Total number of Study genes: 220
 Total number of Study gene GMWG terms (pop non-singletons): 56 (55)
 Genes with GMWG information: 98
 Genes with no GMWG information: 122

These are: b1055 b2566 b3183 b2086 b1749 b1206 b3781 b2582 b3407 b3887 b2733 b2592 b0882 b9488 b1856 b4392 b2742 b3885 b2963 b2958 b0211 b3913 b1133 b2960 b3414 b3012 b0207 b1781 b3081

3) Genes de *M.jannaschii* conservados en Archaeas y Bacterias pero no en Eucariones.

GeneMerge v1.2; mja_salida.xls															
Gene Association File: mja_gene_map.tab.f															
Description File: map_title.tab															
Population File: population.m.jannaschii.nuc															
Study File: study.m.jannaschii.nuc															
GMRG_Term	Pop_freq	Pop_frac	Study_frac	Raw_es	e-score	Description	Contributing_genes								
760	0.003825137	7/1830	ene-67	NA	NA		1 MJ0407								
510	0.00273224	5/1830	ene-67	NA	NA	N-Glycan biosynthesis	MJ1113								
2010	0.017486339	32/1830	sep-67	9.86E-07	7.89E-06	ABC transporters - General	MJ0087	MJ0877m	MJ0876	MJ1013	MJ1014	MJ1015	MJ1368	MJ0085	MJ0878
340	0.003278689	6/1830	feb-67	0.018010928	0.144087424	Histidine metabolism	MJ0703	MJ0411							
680	0.007650273	14/1830	ene-67	NA	NA	Methane metabolism	MJ1353m								
860	0.01147541	21/1830	ene-67	NA	NA	Porphyrin and chlorophyll metabolism	MJ0391								
630	0.003278689	6/1830	ene-67	NA	NA	Glyoxylate and dicarboxylate metabolism	MJ1353m								
730	0.00273224	5/1830	ene-67	NA	NA	Thiamine metabolism	MJ0028								
Total number of genes: 1830															
Total number of Study genes: 67															
Total number of Study gene GMRG terms (pop non-singletons): 8 (8)															
Genes with GMRG information: 16															
Genes with no GMRG information: 51															
These are:	MJ0244	MJ1057	MJ1059	MJ1264	MJ0676	MJ1591m	MJ0640	MJ1416	MJ0712	MJ0713	MJ0900	MJ1287	MJ0781	MJ1288	MJ0907

4) Genes de *M. jannaschii* conservados en Archaeas y Eucariontes pero no en Bacterias

Gene	Term	Pop. freq.	Pop. freq.	Study freq.	Raw es	e-score	Description	Contributing genes
Gene Association File: m_j_gene_map.tab.f								
Description File: map_title.tab								
Population File: population.m_jannaschii.nuc								
Study File: study.m_jannaschii.nuc								
191	0.000196701	15	100	2	124	0.30163559	1 Oxidative phosphorylation	M0022 M0615
241	0.016393409	15	100	8	124	0.00989135	0.017804494 Pyrimidine metabolism	M0085 M0349 M0292 M0296 M0289 M0148 M0297 M0197
770	0.00235792	4	100	1	124	NA	NA	M0280
401	0.010302519	15	100	1	124	NA	NA	M0108
791	0.01493029	4	100	1	124	NA	NA	M0542
201	0.010302519	15	100	8	124	0.007724916	0.01939325 Purine metabolism	M0085 M0349 M0292 M0296 M0289 M0148 M0297 M0197
202	0.003276601	5	100	1	124	NA	NA	M0296
501	0.00473294	3	100	1	124	NA	NA	M0542
301	0.01030296	2	100	1	124	NA	NA	M0415
54	0.01030296	2	100	1	124	NA	NA	M0179
310	0.036920219	57	100	42	124	1.125-20	2,065-18 Ribosome	M0471 M0470 M0373 M0124 M0120 M0036 M0040 M0049 M0177 M0245 M0467 M0468 M0469.1 M0472 M0473 M0474 M0476 M0477 M0519 M0543 M0591 M0633 M0673 M0692 M0930 M0983 M0994 M1001
360	0.003276601	5	100	2	124	0.063684292	1 Protein export	M0291 M0470
570	0.010302519	15	100	4	124	0.044971940	0.094895704 Aminoacyl-tRNA biosynthesis	M0108 M0228 M0297 M0415
601	0.01147439	2	100	1	124	NA	NA	M0441
310	0.00653797	12	100	7	124	5.62E-01	0.00010101 RNA polymerase	M0349 M0292 M0296 M0289 M0148 M0297 M0197
312	0.01030296	2	100	2	124	0.005349570	0.093483027 Base transcription factors	M0507 M0702
330	0.001625944	3	100	1	124	NA	NA	M0085
261	0.00473294	3	100	2	124	0.111395998	1 G(lycine, serine and threonine) metabolism	M0228 M0197
Total number of genes: 133								
Total number of Study genes: 134								
Total number of Study gene GO term (pop non-singletons): 18 (18)								
Genes with GO information: 64								
Genes with no GO information: 70								
These are:	M0024	M0125	M0126	M0160	M0162	M0171		M0349 M0362 M0326 M0329 M0404 M0225 M0415 M0222 M0114 M0491 M0048 M0051 M0097 M0174 M0029 M0297 M0063 M0061 M0489 M0213 M0493 M0444 M0173 M0445 M0493 M0510 M0557 M0570

5) Genes de *S.cerevisiae* conservados en Archaeas y Eucariontes pero no en Bacterias.

GeneMerge v1.2; sce_salida.xls										
Gene Association File: sce_gene_map.tab.f										
Description File: map_title.tab										
Population File: population.s.cerevisiae.nuc										
Study File: study.s.cerevisiae.nuc										
GMRG_Term	Pop_freq	Pop_frac	Study_frac	Raw_es	e-score	Description	Contributing_genes			
3060	0.0014460154	sep-24	3/211	0.002773447	0.055468942	Protein export	YDR292C	YLR378C	YBR283C	
400	0.0032133671	20/6224	2/211	0.146024719		1 Phenylalanine, tyrosine and tryptophan biosynthesis	YLR060W	YGR165C		
100	0.001767352	nov-24	4/211	0.000351224	0.007024492	Biosynthesis of steroids	YJL167W	YLR450W	YML075C	YMR208W
970	0.005944730	37/6224	5/211	0.007688627	0.153772542	Aminoacyl-tRNA biosynthesis	YLR060W	YPR081C	YBR121C	YOL097C
562	0.004330046	27/6224	1/211	NA	NA	Inositol phosphate metabolism	YDR208W			
230	0.014299485	89/6224	17/211	4.75E+05	9.50E+06	Purine metabolism	YNL262W	YDL102W	YOR341W	YDL140C
500	0.007872750	49/6224	4/211	0.083308342		1 Starch and sucrose metabolism	YER172C	YGR271W	YGL251C	YER171W
4111	0.017512853	109/6224	7/211	0.076109538		1 Cell cycle - yeast	YJL074C	YPR019W	YEL032W	YGL201C
260	0.006908740	43/6224	2/211	0.431114863		1 Glycine, serine and threonine biosynthesis	YPR081C	YBR121C		
190	0.011086118	69/6224	3/211	0.416309950		1 Oxidative phosphorylation	YMR054W	YOR270C	YEL051W	
380	0.002892030	18/6224	1/211	NA	NA	Tryptophan metabolism	YOL097C			
3010	0.022172236	138/6224	68/211	3.85E+52	7.69E+51	Ribosome	YNL067W	YGL147C	YLR367W	YJL190C
3050	0.005141388	32/6224	12/211	2.10E+04	4.20E+05	Proteasome	YGR135W	YGR253C	YOL038W	YOR362C
240	0.010925449	68/6224	16/211	5.34E+04	1.07E+06	Pyrimidine metabolism	YNL262W	YDL102W	YOR341W	YDL140C
3030	0.003374035	21/6224	3/211	0.032567271	0.651345437	DNA polymerase	YNL262W	YDL102W	YIR008C	
3022	0.003695372	23/6224	2/211	0.182351877		1 Basal transcription factors	YER148W	YPR086W		
900	0.000803841	may-24	1/211	NA	NA	Terpenoid biosynthesis	YJL167W			
4070	0.004659383	29/6224	2/211	0.257845144		1 Phosphatidylinositol signaling	YFR019W	YDR208W		
3020	0.004498714	28/6224	13/211	1.29E+02	2.59E+03	RNA polymerase	YOR341W	YDL140C	YOR116C	YJL021W
790	0.002892030	18/6224	4/211	0.002701615	0.054032309	Folate biosynthesis	YER172C	YGR271W	YGL251C	YER171W
Total number of genes: 6224										
Total number of Study genes: 211										
Total number of Study gene GMRG terms (pop non-singletons): 20 (20)										
Genes with GMRG information: 127										
Genes with no GMRG information: 84										
These are: YBR118W YPR080W YDR172W YOR290C YLLO35W YPL022W YPL009C YOR005C YDL164C YDLO58W YPL174C YLR277C YLR115W YNR037C YBL091C YGL115W YAL021C YOR353C YPR016C YBL057C YPL237W										

6) Genes de *S.cerevisiae* conservados en Bacterias y Eucariontes pero no en Archaeas.

GeneMerge v1.2; sce_salida.xls										
Gene Association File: sce_gene_map.tab.f										
Description File: map_title.tab										
Population File: population.s.cerevisiae.nuc										
Study File: study.s.cerevisiae.nuc										
GMRG_Term	Pop_freq	Pop_frac	Study_frac	Raw_es	e-score	Description	Contributing_genes			
120	0.003534704	22/6224	3/305	0.090190846		1 Bile acid biosynthesis	YHR104W	YOR120W	YDR368W	
52	0.005141388	32/6224	5/305	0.018526621	0.944857693	Galactose metabolism	YHR104W	YOR120W	YDR368W	YGR240C YMR205C
310	0.003856041	24/6224	4/305	0.027646124		1 Lysine degradation	YHR104W	YOR120W	YDR368W	YGL154C
920	0.001928020	dic-24	2/305	0.114296940		1 Sulfur metabolism	YKL001C	YFR030W		
380	0.002892030	18/6224	1/305	NA	NA	Tryptophan metabolism	YDR268W			
450	0.003052699	19/6224	6/305	0.000207936	0.010604751	Selenoamino acid metabolism	YGR155W	YKL001C	YFR030W	YDR502C YLR180W YLR299W
3022	0.003695372	23/6224	1/305	NA	NA	Basal transcription factors	YBR198C			
4120	0.004820051	30/6224	2/305	0.436458289		1 Ubiquitin mediated proteolysis	YIL046W	YFL009W		
500	0.007872750	49/6224	4/305	0.217742697		1 Starch and sucrose metabolism	YGL163C	YDR038C	YDR040C	YEL011W
670	0.002410025	15/6224	4/305	0.005021192	0.256080836	One carbon pool by folate	YER183C	YBL013W	YLR028C	YMR120C
40	0.001124678	jul-24	1/305	NA	NA	Pentose and glucuronate interconver	YHR104W			
710	0.002892030	18/6224	1/305	NA	NA	Carbon fixation	YDR050C			
620	0.005302056	33/6224	4/305	0.075754678		1 Pyruvate metabolism	YDL174C	YHR104W	YML004C	YNL071W
61	0.000803341	may-24	1/305	NA	NA	Fatty acid biosynthesis	YKL182W			
190	0.011086118	69/6224	5/305	0.248678181		1 Oxidative phosphorylation	YPL172C	YBL099W	YKL192C	YBR039W YDR298C
740	0.002088688	13/6224	2/305	0.130877157		1 Riboflavin metabolism	YDR236C	YPR073C		
4010	0.008836760	55/6224	2/305	0.759525852		1 MAPK signaling pathway	YML004C	YIL147C		
260	0.006908740	43/6224	4/305	0.157740120		1 Glycine, serine and threonine metab	YHR104W	YOR120W	YDR368W	YGR155W
130	0.001124678	jul-24	3/305	0.003519595	0.179499349	Ubiquinone biosynthesis	YNR041C	YGR255C	YKL192C	
10	0.007712082	48/6224	8/305	0.002058091	0.104962658	Glycolysis / Gluconeogenesis	YDR050C	YKL152C	YJL052W	YJR009C YGR192C YGR240C YMR205C YNL071W
460	0.001767352	nov-24	1/305	NA	NA	Cyanoamino acid metabolism	YLR299W			
220	0.003052699	19/6224	1/305	NA	NA	Urea cycle and metabolism of amino	YOR323C			
600	0.002249357	14/6224	2/305	0.147962797		1 Sphingolipid metabolism	YLR260W	YOR171C		
860	0.002410025	15/6224	1/305	NA	NA	Porphyrin and chlorophyll metabolis	YPL172C			
230	0.014299485	89/6224	5/305	0.443564828		1 Purine metabolism	YJL026W	YKL001C	YDR454C	YLR028C YMR120C
3060	0.001446015	sep-24	1/305	NA	NA	Protein export	YER154W			
330	0.002249357	14/6224	1/305	NA	NA	Arginine and proline metabolism	YLR142W			
400	0.003213367	20/6224	1/305	NA	NA	Phenylalanine, tyrosine and tryptoph	YPL097W			
361	0.001606683	oct-24	2/305	0.083049018		1 gamma-Hexachlorocyclohexane degr	YDL086W	YPR073C		
650	0.004177377	26/6224	3/305	0.132340943		1 Butanoate metabolism	YHR104W	YOR120W	YDR368W	
430	0.000321336	feb-24	1/305	NA	NA	Taurine and hypotaurine metabolism	YLR299W			
790	0.002892030	18/6224	5/305	0.001381442	0.070453564	Folate biosynthesis	YGL163C	YDR038C	YDR040C	YMR113W YOR241W
564	0.004016709	25/6224	7/305	0.000141771	0.007230371	Glycerophospholipid metabolism	YBR029C	YDL142C	YPR113W	YIL155C YOL059W YDL022W YDL052C
627	0.000160668	ene-24	1/305	NA	NA	1,4-Dichlorobenzene degradation	YDL086W			
2021	0.000482005	mar-24	1/305	NA	NA	Two-component system - Organism-s	YIL147C			
561	0.003534704	22/6224	2/305	0.293614911		1 Glycerolipid metabolism	YHR104W	YDL052C		
625	0.001606683	oct-24	3/305	0.010811455	0.551384242	Tetrachloroethene degradation	YHR104W	YOR120W	YDR368W	
30	0.004177377	26/6224	6/305	0.001319354	0.067287062	Pentose phosphate pathway	YNL241C	YGR240C	YHR163W	YGR248W YNR034W
520	0.002249357	14/6224	3/305	0.028363314		1 Nucleotide sugars metabolism	YHR104W	YOR120W	YDR368W	
300	0.002410025	15/6224	1/305	NA	NA	Lysine biosynthesis	YGL154C			
530	0.002249357	14/6224	2/305	0.147962797		1 Aminosugars metabolism	YIL043C	YKL150W		
480	0.001767352	nov-24	5/305	9.90E+09	0.005048134	Glutathione metabolism	YBR244W	YIR037W	YKL026C	YNL241C YLR299W
271	0.002731362	17/6224	4/305	0.008102353	0.413220021	Methionine metabolism	YGR155W	YDR502C	YLR180W	YBL013W
590	0.000803341	may-24	4/305	2.72E+09	0.001387254	Arachidonic acid metabolism	YBR244W	YIR037W	YKL026C	YLR299W
4111	0.017512853	109/6224	2/305	0.973239013		1 Cell cycle - yeast	YIL046W	YFL009W		
53	0.002088688	13/6224	3/305	0.023105749		1 Ascorbate and aldarate metabolism	YHR104W	YOR120W	YDR368W	
4070	0.004659383	29/6224	2/305	0.419355749		1 Phosphatidylinositol signaling system	YBR029C	YPR113W		
252	0.005462724	34/6224	1/305	NA	NA	Alanine and aspartate metabolism	YNL071W			
51	0.004659383	29/6224	7/305	0.000388622	0.019819745	Fructose and mannose metabolism	YDR050C	YHR104W	YOR120W	YDR368W YJL155C YGR240C YMR205C
565	0.000803341	may-24	1/305	NA	NA	Ether lipid metabolism	YDL052C			
970	0.005944730	37/6224	3/305	0.271369605		1 Aminoacyl-tRNA biosynthesis	YDR268W	YPL097W	YBL013W	
240	0.010925449	68/6224	3/305	0.655078736		1 Pyrimidine metabolism	YPR062W	YNR012W	YJL026W	
Total number of genes: 6224										
Total number of Study genes: 305										
Total number of Study gene GMRG terms (pop non-singletons): 52 (51)										
Genes with GMRG information: 73										
Genes with no GMRG information: 232										
These are: YOR280C YMR222C YHR049W YCR057C YBR175W YJL112W YKL213C YKR036C YPL151C YGL099W YNR036C YJL096W YBL019W YPR003C										