UNIVERSIDAD DE LA REPÚBLICA - FACULTAD DE CIENCIAS

Tesina de grado: Licenciatura en Ciencias Biológicas.

ESTUDIO DEL USO DE CODONES EN ARCHAEA

Autor: LUCIA LEYTON

Tutores: Dr. HECTOR MUSTO – Dr. ANDRES IRIARTE

Montevideo, Uruguay. 2015

CONTENIDO.

RESUMEN.

El USO DE CODONES es el desvío no al azar de las frecuencias observadas de algunos codones sinónimos sobre otros. La expresión de los genes es el proceso molecular central de las células, debe llevarse a cabo en forma precisa, sensible y eficiente. Una característica conocida del código genético es su redundancia. La redundancia en el código genético ofrece un nivel extra de regulación en la producción de proteínas mediante el uso de codones sinónimos (UCS) mientras se mantiene la misma secuencia de aminoácidos. De esta forma se ha observado que el sesgo en el ucs desempeña un papel muy importante en la regulación de la expresión génica. La selección a favor del UCS promueve una traducción eficiente y tiene efectos locales en genes específicos y efectos globales en la adaptabilidad ("fitness") del organismo.

Las Archaeas constituyen un grupo de procariotas metabólicamente diverso siendo reconocido como una unidad taxonómica mayor; sin embargo, la mayor parte de la información que se dispone sobre el uso de codones sinónimos ha sido analizada en organismos pertenecientes al Dominio Bacteria. En esta tesina se plantea realizar el estudio del UCS en las Archaeas a través de herramientas bioinformáticas, dado que los trabajos publicados en este Dominio son escasos y además muchos de sus representantes viven en condiciones "extremas" (por ejemplo altas temperaturas, pH extremos, alta salinidad) y son, por tanto, un excelente modelo para testar la hipótesis de la influencia del nicho ecológico entre los factores causantes de sesgos en el uso de codones sinónimos.

PALABRAS CLAVE.

Codones; codón óptimo; sesgo; expresión; selección; inercia filogenética.

INTRODUCCIÓN

Las Archaeas constituyen un grupo de procariotas metabólicamente diverso siendo reconocido como una unidad taxonómica mayor, es decir uno de los "tres Dominios de la vida" (Woese, 1987; Woese, 1990). Forman un importante componente de la biósfera en conjunto y están involucrados en ciclos bio-geoquímicos globales (Pace, 1997; Gribaldo & Brochier-Armanet, 2006; Allen, *et al.*, 2009; Erwin *et al.*, 2014). A diferencia de lo que comúnmente se cree, no sólo los encontramos en ambientes extremos. Su origen se remonta a unos 2.7 a 3.4 mil millones de años, según estudios de presencia de metano y reducción de azufre de origen biológico, pero puede ser aún más remoto si se utilizan otros métodos de datación y calibración (Gao & Gupta, 2007).

Estos organismos fueron descubiertos en 1969 cuando Thomas D. Brock realizó un estudio en las fuentes termales del Yellowstone National Park de Wyoming y descubrió que *Termophillus aquaticus* efectivamente crece a más de 70°C. Este organismo es del cual se obtuvo posteriormente la enzima *Taq* polimerasa, fundamental en la PCR, constituyendo un hito en la biología molecular moderna (Madigan, 2000).

La relación filogenética con los otros dominios existentes -Bacteria y Eucarya- todavía está en debate porque estos organismos presentan características compartidas por ambos Dominios así como algunas propiedades únicas como la metanogénesis, y otras ligadas a la capacidad de vivir en ambientes extremos (Gribaldo & Brochier-Armanet, 2006).

* * *

Estudios realizados indican que el sesgo en el uso de codones sinónimos desempeña un papel muy importante en la regulación de la expresión génica. Sin embargo, la mayor parte de la información de que se dispone se obtuvo de análisis de organismos pertenecientes al Dominio Bacteria. En esta tesina se plantea analizar este problema en las Archaeas, dado que los trabajos publicados en este Dominio son escasos y además muchos de sus representantes viven en condiciones "extremas" (por ejemplo altas temperaturas, pH extremos, alta salinidad) y son, por tanto, un excelente modelo para testar la hipótesis de la influencia del nicho ecológico entre los factores causantes de sesgos en el uso de codones sinónimos. Entre otros aportes, el conocimiento del sesgo en el uso de codones en las especies provee de una herramienta útil de análisis filogenético porque la divergencia en el uso de codones está correlacionada con las distancias evolutivas entre las especies (Grantham *et al.* 1981; Knight, *et al.* 2001).

Características generales del grupo de estudio

Las Archaeas o arquebacterias son organismos procariotas que presentan en general un único cromosoma circular, carecen de orgánulos y en su morfología general se asemejan a las bacterias. Como éstas, el grupo inicia la traducción mediante la secuencia de Shine-Dalgarno y presenta mRNAs policistrónicos (Gribaldo & Brochier-Armanet, 2006).

En contraposición, los componentes de la maquinaria de procesamiento de la información guarda mayor homología con Eucarya, por ejemplo el aparato de replicación del DNA (Lindas & Bernander, 2013). Inician la replicación del cromosoma a través de la proteína Orc1/Cdc6, estructural y funcionalmente similar a la proteína bacteriana DnaA (Bernander, 2003). La RNA polimerasa presenta siete o más subunidades en lugar de cuatro, como en bacterias, el primer aminoácido incorporado es la Metionina (y no la N-formil metionina) y requieren algunos factores de iniciación extra (eIF2, 2A, 2B y 5A) (Cavallier-Smith, 2002; Gribaldo & Brochier-Armanet, 2006). Los factores de elongación son más similares a los eucariotas durante la síntesis de proteínas y lo mismo sucede con algunas modificaciones en los tRNAs. Presentan histonas del core con plegamiento típico de su contraparte eucariota (Cavallier-Smith, 2002; Gao & Gupta, 2007). Además, la pared celular carece de peptidoglicanos (Gao & Gupta, 2007).

La estructura química de la única membrana celular es característica exclusiva del Dominio y está compuesta por fosfolípidos levógiros ramificados, con el grupo glicerol (G1P) unidos por enlaces éster (Cavallier-Smith, 2002; Matte-Tailliez *et al.*, 2002; Gao & Gupta, 2007), típicamente más estable que los enlaces éster con ácidos grasos unidos a G3P que encontramos en el resto de los seres vivos (Gribaldo & Brochier-Armanet, 2006). En el sistema de replicación se encuentra una subunidad característica de la RNA polimerasa propia del grupo de Archaea (Cavallier-Smith, 2002; Gao & Gupta, 2007). Algunas modificaciones en los tRNAs son exclusivas del grupo, como la adición de archaeosina en el d-loop y la ausencia de queuina (Cavallier-Smith, 2002). Por último presentan un eje flagelar compuesto de glicoproteínas relacionadas a la pilina, ácido insolubles, a diferencia de la flagelina soluble en ácido (Cavallier-Smith, 2002).

Hábitat y recursos

Usualmente son llamados extremófilos, porque la mayoría de las Archaeas conocidas viven en ambientes extremos, pero también están presentes en el suelo y en los océanos (Bintrim *et al.*, 1997; Erwin *et al.*, 2014). La diversidad de ambientes colonizados está todavía siendo descubierta y se cree que en realidad podrían encontrarse estos microorganismos siempre que exista agua.

Un ambiente extremo es un hábitat en el que alguna variable física o química difiere significativamente de la encontrada en hábitats que permiten la vida vegetal y animal (Madigan, 2000). Se plantea como límite de temperatura tolerable al desarrollo de vida los 150°C porque a temperaturas mayores se pierde estabilidad a nivel molecular (Madigan & Mars 1997; Madigan 2000). En este sentido, las arquebacterias están siendo muy estudiadas principalmente debido a sus moléculas y actividad enzimática particular, resistentes a condiciones extremas por lo que se pueden emplear en la industria alimentaria, la textil o la metalúrgica (Madigan & Mars 1997).

Sobrevivir en condiciones extremas implica en algunos casos tener tasas de crecimiento óptimas a temperaturas mayores a 45°C en el caso de los llamados termófilos o mayores a 80°C en el caso de los hipertermófilos. Por otro lado, los organismos psicrófilos viven en ambientes temporalmente o permanentemente congelados como en el Lago Fryxell en la Antártida (Karr *et al.* 2006), sedimentos marinos en Alaska, el mar Báltico, el permafrost Ártico o lagos de agua fresca en Suiza (Allen, *et al.* 2009). Por su parte las enzimas generadas por los psicrófilos son útiles para trabajar en el procesamiento de comida que se realiza a bajas temperaturas o en la industria de los perfumes porque disminuye la evaporación de los componentes de las fragancias y desarrollo de detergentes para lavado en frío (Madigan & Mars 1997).

Otros hábitat extremos involucran diferentes pH; así existen organismos en ambientes ácidos con pH menor a 5 (los llamamos acidófilos) como en estanques o geiser con gases sulfurosos, áreas contaminadas por drenajes mineros o en el estómago de un mamífero. Las archaeas de ambientes alcalinos, son alcalófilos, y viven a un pH mayor a 9. Los encontramos en lagos de soda de Egipto, en el desierto *Rift Valley* del este de Africa, en "slag dumps" producidos por la actividad metalúrgica o en el *Octopus Spring* del *Yellowstone National Park*, entre otros sitios.

Algunas extremo-enzimas alcalinas son usadas en la industria de los detergentes para lavandería, tratamiento de desechos, industria del papel, entre otras aplicaciones. Por su parte, las enzimas provenientes de los organismos acidófilos son usadas en la industria alimentaria. Es interesante notar que algunos acidófilos, como *F. acidophillum* son responsables de la contaminación de los drenajes mineros y cursos de agua cercanos porque oxidan hierro de la pirita (FeS_2) que hay en el drenaje, liberando ácido sulfúrico (Madigan & Mars 1997).

Por último los halófilos crecen en medios salinos e hipersalinos (mayor a 3M) como en el *Great Salt Lake* de Estados Unidos, el Mar Muerto en Oriente Medio o los lagos salados de la Antártida; así como en los depósitos artificiales de sal.

Relaciones evolutivas

Al constatar la presencia de moléculas constitutivas similares en todos los seres vivos y patrones similares de crecimiento y propagación, se asume que todos los seres vivientes están emparentados entre sí y derivaron de un ancestro común universal (Gupta, 1998b). Sin embargo diferentes explicaciones para las relaciones de parentesco entre los dominios de la vida tienen argumentos a favor y en contra, y hasta el presente sigue estando en debate (Gribaldo & Brochier-Armanet, 2006) según se sigan diferentes modelos evolutivos.

Woese estableció los tres dominios de la vida según estudios en base a SSU (rRNA 16S), los cuales habrían divergido en tres líneas evolutivas independientes desde el LUCA (de las siglas en inglés de "last ultimate common ancestor") universal o desde una comunidad primordial de células primitivas que evolucionaron como una unidad al principio y posteriormente se diferenciaron en líneas primarias más especializadas, las cuales habrían dado lugar a las tres líneas evolutivas ya mencionadas (Woese, 1998).

Asumiendo que Archaea es un único superdominio, en un modelo evolutivo se plantea que Bacteria derivó directamente del LUCA universal mientras que Archaea y Eucarya compartieron un último ancestro común más reciente que el LUCA universal y son por tanto linajes hermanos. La evidencia fósil indica que el grupo "Neomura" formado por los taxa hermanos arquebacterias y eucariotas no son ancestrales, por el contrario este grupo derivaría de la evolución de las "posibacterias" (Bacterias Gram positivas) (Cavallier-Smith, 2002). Esto requiere un episodio de aceleración evolutiva importante en la rama que lleva a Archaea, fenómeno apoyado por la ausencia de registros fósiles de organismos intermedios (en contraposición con la diversidad existente de organismos fósiles procariotas y eucariotas que precedieron y siguieron a este evento evolutivo) (Gupta, 1998b). Esta aceleración en teoría podría estar relacionada a dos fuerzas selectivas principales; genes de resistencia a antibióticos y medio ambiente fuertemente selectivos por sensibilidad al oxígeno de los organismos cuando la atmósfera cambió de anaeróbica a aeróbica y selección positiva para termofilia (Gupta, 1998b; Cavallier-Smith, 2002).

Este modelo implica que los estados de los caracteres presentes en bacteria son los ancestrales y el resto derivados. Por el contrario, se puede plantear que los estados de los caracteres que comparten Eucarya y Archaea son ancestrales o que los tres dominios equidistan filogenéticamente entre sí o que ninguno de los dominios existentes tienen hoy en día estados ancestrales conservados, implicando que

ninguno puede ser rastreado por análisis molecular hasta el LUCA universal (Gribaldo & Brochier-Armanet, 2006).

En otro modelo evolutivo se considera que Archaea y Eucarya serían bacterias modificadas (mosaicos genéticos en lugar de Transferencia Genética Horizontal [TGH] y simbiogénesis) con evolución vertical (Cavalier-Smith, 2002) o quimeras con aportes genéticos de varios grupos por TGH o fusiones celulares (Gribaldo & Brochier-Armanet, 2006; Gupta, 1998b). La dirección del cambio evolutivo va de bacteria (más antiguo) hacia archaea (más reciente) (Cavallier-Smith, 2002). La raíz del árbol de los procariotas estaría ubicada entre las arquebacterias y las bacterias gram-positivas, por lo que estos organismos serían ancestrales con respecto a las bacterias gram-negativas y los eucariotas (Gupta, 1998b). Estudios de filogenia de retroides en arquebacterias y bacterias, muestran ancestría común entre los dos dominios (Rest & Mindell, 2003).

En el debate se cuestiona incluso la categoría taxonómica de Dominio para Archaea (Gupta, 1998a; Gupta, 1998b; Cavallier-Smith, 2002), clasificación que es la mayormente aceptada en la comunidad científica. En estos trabajos se sostiene que las arquebacterias están muy relacionadas filogenéticamente a las bacterias Gram-positivas, organismos con los que comparten principalmente la ausencia de una segunda membrana celular (son monodermos) y los diferencia de las bacterias Gram-negativas, procariotas con doble membrana celular (Gupta, 1998b; Gupta, 1998c; Cavalier-Smith, 2002).

Como Dominio, Archaea es un grupo cuyas relaciones filogenéticas también están en discusión. Filogenias realizadas en base a ARNr 16S (Matte-Tailliez *et al.*, 2002) y en base a análisis genómicos por comparación de los mecanismos involucrados en el ciclo celular, replicación y traducción del ADN (She *et al.* 2001), reconocen como *Phylum* a Crenarchaeota y Euryarchaeota (Gao & Gupta, 2007). Estos son los dos *Phyla* que aparecen en el Bergey´s Manual of systematic bacteriology (www.springerlink.com/content/978-0-387-21609-6). Esta división se corrobora además si se compara la forma de los ribosomas en Crenarchaeota y eucariotas que presentan características compartidas (Henderson *et al.*, 1984 & Lake *et al.*, 1984) originando el grupo de los "eocytos". Lake divide a Archaea en base a las diferencias en la estructura tridimensional del ribosoma según estudios de microscopía electrónica. Los ribosomas aparecen con una proyección quasimétrica en una subunidad mayor y una proyección asimétrica en la subunidad menor, un espacio entre las subunidades que conforman la proyección quasimétrica, al cual le llaman "gap eocítico", y esto resulta en la forma característica de estos ribosomas (Lake *et al.*, 1984). A Crenarchea por esta razón se propone renombrarlo como EOCYTA ("dawn + cell"), ya que tendría un ancestro común con las células eucariotas (Lake *et al*., 1984; Gribaldo & Brochier-Armanet, 2006; Gupta, 1998b).

Algunos autores mencionan además otros dos *Phylum*: Nanoarchaeota y Korarchaeota (Gribaldo & Brochier-Armanet, 2006). En base al análisis filogenómico de proteínas ORFans (o "signature

proteins") Gao y Gupta (Gao & Gupta, 2007) demostraron que todas las archaeas metanogénicas forman un grupo monofilético y comparten un ancestro común con *Archaeoglobus* y que el dominio Archaea forma un grupo polifilético.

En el GenBank se agrupan las especies dentro de 5 *Phyla*: Crenarchaeota, Euryarchaeota, Korarchaeota (Elkins, *et al.*, 2008), Nanorchaeota (Huber, *et al.*, 2003) y Thaumarchaeota (Brocher-Armanet, *et al.*, 2008). En este trabajo se tomará como referencia la taxonomía que aparece en el sitio GenBank, es decir agrupando a las especies en 5 *Phyla*. De cualquier manera es importante tener en cuenta que la mayoría de las especies cultivables y bien investigadas pertenecen a Euryarchaeota y Crenarchaeota. El hecho de que se conozcan pocas especies dificulta el esclarecimiento de la filogenia, por tanto, a medida que más organismos sean conocidos y secuenciados, mejor será nuestra comprensión de la historia evolutiva del grupo.

En el grupo de los Crenarchaeota encontramos los órdenes correspondientes a Acidilobales, Desulfurococcales, Sulfolobales, Thermoproteales y otros aún no clasificados entre los que se destaca el MGI ("Marin Group I"). Euryarchaeota, en base a diferentes características fisiológicas y metabólicas, se subdivide en cinco subgrupos más; lo que de acuerdo al método de estudio filogenético cambia en algunos casos las relaciones de parentesco entre los distintos organismos: los metanógenos, los thermococcales, los halobacteriales, los thermoplasmatales y los archaeoglobales. Los organismos metanógenos están representados por los órdenes Methanopyrales, Methanobacteriales, Methanococcales, Methanomicrobiales y Methanosarcinales, estos últimos son llamados metanógenos de clase 2 (los primeros son los metanógenos de clase 1) (Gao & Gupta, 2007).

Uso de codones: generalidades

La expresión de los genes es uno de los procesos moleculares centrales de las células. Los organismos invierten energía, biomoléculas e información para llevar a cabo el proceso de expresión genética, optimizando la eficiencia, sensibilidad y precisión del mismo. Se entiende por eficiencia al resultado del rendimiento del proceso en relación al costo que conlleva realizarlo. La precisión es la probabilidad de que una proteína traducida no tenga errores y se corresponda con la secuencia presente en la secuencia del gen codificante, sumado a la probabilidad de que la proteína se pliegue correctamente en la célula (Gingold & Pilpel, 2011). La fidelidad de la transmisión del mensaje es una propiedad fundamental de la transferencia de información genética, la precisión final depende de la combinación de las tasas de errores de todos los procesos que están involucrados (De Koning *et al.*, 2010).

La información almacenada en el ADN está codificada en forma de trinucleótidos (codones), donde cada codón se corresponde con un aminoácido (AA). En el código genético canónico, existen 64 codones diferentes. De estos codones, tres son señales de STOP y los 61 restantes codifican para 20 aminoácidos diferentes. Así cada triplete de nucleótidos equivale a un AA y algunos aminoácidos están codificados por más de un triplete, esto último se conoce como degeneramiento del código genético (Bulmer, 1988). De esta manera el código es redundante pero no ambiguo. La redundancia del código genético ofrece un nivel extra de regulación en la producción de proteínas mediante el uso de codones sinónimos (UCS) mientras mantiene la misma secuencia de aminoácidos (Canarozzi *et al.*, 2010; Supek *et al.*, 2010; Gingold & Pilpel, 2011).

La codificación es universal. El código genético clásico es encontrado en la mayoría de genomas nucleares y mitocondriales pero algunos genomas tienen pequeñas variaciones (Musto *et al.*, 2001; Marquez *et al.* 2005). En la naturaleza se encuentran algunas variantes en la codificación para los aminoácidos, originando códigos genéticos alternativos. De todos modos, cualquiera de estos códigos resiste los errores mejor que los errores que presentarían códigos genéticos al azar para un amplio rango de diferentes propiedades de aminoácidos y modelos de generación de códigos al azar (Marquez *et al*., 2005).

Hay dos tipos de mutaciones puntuales en genes codificantes para proteínas: las mutaciones sinónimas (MS) y las mutaciones no sinónimas (MNS). Las MNS provocan cambios en la secuencia de aminoácidos que conllevan diferentes consecuencias para la proteína codificada según la mutación en particular. Por otro lado las MS no alteran la secuencia de la proteína y provocan que aparezcan distintos codones sinónimos para un mismo aminoácido. Tempranamente se observó que las sustituciones silenciosas que ocurren como consecuencia de fijaciones de MS, son más frecuentes que las sustituciones no sinónimas (Kimura, 1977), esto llevó a pensar que las mutaciones sinónimas eran neutrales, es decir, no eran objeto de selección (Nei 2005).

Sin embargo se ha demostrado que un uso adecuado y eficiente de codones puede reducir las mutaciones sin sentido o de cambio de sentido, así como el plegamiento incorrecto de las proteínas (Warnecke & Hurst, 2011). La precisión de la traducción, proceso afectado por el uso de codones sinónimos, aumenta la eficiencia porque permite que el ribosoma no se detenga innecesariamente mientras se "desplaza" por el mRNA (Sharp *et al.*, 2010). El plegamiento inapropiado de las proteínas puede interferir en la función de las mismas, llevar a su agregación o generar disrupción de la integridad de la membrana, que también puede resultar en disfunción celular, lo que se traduce en un mayor gasto energético para la célula (Gingold & Pilpel, 2011).

A pesar de la vasta investigación realizada tanto en bacterias como en modelos eucariotas, poco es lo que se conoce de los mecanismos de procesamiento de la información en las arqueobacterias (Emery & Sharp, 2011), y este conocimiento, sin dudas, jugará un rol importante en futuras investigaciones para elucidar detalles de los sistemas que controlan la fidelidad del procesamiento de la información (De Koning *et al.*, 2010), ya que mediante el estudio del sesgo en el UCS se puede conocer un aspecto de la regulación en el flujo de la información en este Dominio. Los patrones de tendencias en el UCS entre las especies es útil para la reconstrucción molecular filogenética y clarificar los roles relativos de la evolución neutral y la selección natural en las mismas y de la composición de los genomas (Knight *et al.*, 2001). Asimismo, este conocimiento será sin duda importante por sus aplicaciones biotecnológicas, fundamentalmente para expresar proteínas útiles en medios "no habituales".

El sesgo en el UCS varía entre los organismos y la variación tiende a ser mayor cuanto mayor sea la distancia taxonómica que separa a los organismos. Al mismo tiempo, hay una correlación que indica que las estrategias de UCS utilizadas tienden a ser conservadas en la evolución (Ikemura, 1985; Retchless & Lawrence, 2011).

Se ha observado que en diferentes organismos hay preferencias de uso de algunos de los tripletes sinónimos con respecto a los otros (Grantham, 1981; Musto *et al.* 2005; Sharp *et al.* 2005; Plotkin & Kudla, 2011) y que este mismo fenómeno también se observa a nivel inter-especie, si comparamos la preferencia de uso de codones a nivel de genes tanto en el tipo de codones elegidos como en la frecuencia de uso (Sharp, 1993; Sharp *et al.* 2005). Debido a las diferentes estrategias de uso de codones entre especies Grantham y colaboradores en 1980 postularon la "genome hypothesis" porque cada genoma tiene una estrategia de codificación (Ikemura, 1985; Novoa & Ribas de Pouplana, 2012) y esta estrategia es especie específica (Sharp & Li, 1987). Así, el sesgo en el uso de codones hace referencia a las diferentes frecuencias de uso de los codones sinónimos encontrados en el mRNA y se puede comparar a nivel de genes dentro un genoma (dentro de un organismo o intragenómico) o entre distintos genomas (entre diferentes organismos o intergenómico). Warnecke y colegas hablan de que el genoma presenta una anatomía y arquitectura particular en cada especie, que habría evolucionado para reducir la tasa de ocurrencia de errores o disminuir los efectos deletéreos en caso de que ya se haya producido alguno

(Warnecke *et al.* 2011). De esta manera el genoma es una unidad en su conjunto sujeta a selección, en vez de los genes individuales (Ikemura, 1985; Medrano Soto, 2005), y donde cada gen tiende a seguir el patrón de sesgo de UCS del genoma al que pertenece. Por anatomía génica se refiere a la composición y estructura de los genes y a sus productos. La arquitectura génica es la organización de los genes en el genoma (Warnecke *et al.*, 2011). Los sesgos mutacionales también varían entre los diferentes genes incluso en un mismo organismo. Si se comparan las hebras líder y retrasada del ADN se observa que los genes presentes en la hebra líder son usualmente más ricos en G+T (Sharp *et al.*, 2005). La composición de un genoma es el resultado de sesgos mutacionales (lo cual sería un proceso selectivamente "neutral" o bien la selección natural juega el rol más importante, llevando a la fijación de nucleótidos en forma no azarosa y cambiando, con el tiempo, la frecuencia de bases (Bernardi, 1993) (ver más abajo).

En consecuencia, cada genoma tiene una composición de bases diferente y se manifiesta en el contenido G+C. Las relaciones filogenéticas se reflejan en esta medida, ya que organismos cercanos muestran una composición de bases similar (Sueoka, 1962). A nivel inter-especie, la composición de bases es muy importante porque influye directamente en el uso de codones (Sharp *et al.* 2005; Behura & Severson, 2013). El contenido G+C del genoma explica más del 25% de la variación del UCS entre genomas (Lynn *et al*., 2002). Por otra parte, el contenido G+C genómico en procariotas, varía desde menos de 25% a más de 75% (Sueoka, 1962; Muto & Osawa, 1987; Musto *et al.*, 2004). En Archaea obtuvimos una variación de entre el 29% al 69% (datos no publicados).

La variación en el contenido en G+C de los genomas en procariotas se puede explicar desde un punto de vista neutralista (Sueoka, 1962) o desde un modelo seleccionista (Naya *et al.*, 2002; Novoa & Ribas de Pouplana, 2012). El enfoque neutralista propone que las variaciones observadas son el producto de la deriva genética de las mutaciones, siendo las mismas selectivamente neutras. El modelo seleccionista asume que el contenido G+C es una forma de adaptación a condiciones ambientales y/o fisiológicas de los organismos.

El contenido medio en G+C de las posiciones sinónimas varía también mucho entre especies y el valor de cada una está siempre en dirección la composición genómica en su conjunto y favorece o no la fijación de pares G+C (Sueoka, 1962; Bernardi, 1993) y es característico de cada genoma. En procariotas el principal determinante del UCS son los patrones de sustitución nucleotídica (Supek *et al.*, 2010).

Si a nivel de los tripletes sinónimos las terceras posiciones fueran selectivamente neutras, deberían usarse todos los codones sinónimos en igual frecuencia para un aminoácido particular. Cuando se observa que las frecuencias en el UCS no son equivalentes entre los codones sinónimos para un aminoácido, entonces alguna fuerza evolutiva debe estar desviando significativamente estas frecuencias. Así el UCS podría ser explicado por una combinación de sesgos mutacionales, deriva genética y selección natural (Bulmer 1988; Behura & Severson, 2013). Esta es la postura que se acepta actualmente y se conoce como modelo MUTACIÓN-SELECCIÓN-DERIVA (Bulmer, 1991; Thiele *et al* 2012). El grado en que varía el

sesgo en el UCS entre las distintas especies depende de varios factores: los codones serían seleccionados en base a su efecto en la eficiencia y precisión traduccional y por tanto en su efecto sobre las tasas de crecimiento en las células (Ran & Higgs, 2012). La composición nucleotídica presente en un organismo es un balance entre fuerzas modeladoras, en aquellos organismos con estilos de vida fuertemente selectivos, la selección aún está actuando y ésta es más fuerte que las otras fuerzas modeladoras, entonces, se evidencia el sesgo por el uso de codones vigente. En organismos donde actualmente la presión selectiva es nula o baja se observaría un producto vestigial de una selección pasada que se va deteriorando por mutaciones y/o deriva génica (Iriarte *et al.* 2011).

La teoría de MUTACIÓN-SELECCIÓN-DERIVA (MSD) es la teoría más unificadora para explicar el fenómeno del sesgo en el uso de codones (Sharp, 1993). Implica que los patrones observados en el UCS en una población finita, son producto del balance entre la selección actuando a favor de codones óptimos para un aminoácido y las fuerzas de mutación y deriva genética al azar (Rocha, 2004a; Sharp *et al.* 2005), que conducen a la persistencia de codones no óptimos (Bulmer, 1991) cuando la selección modeladora no constituye la fuerza dominante. Estas fuerzas evolutivas afectan a todos los genes pero al aumentar el nivel de expresión, se hace más evidente el efecto de la selección ya que en estos casos es lo suficientemente fuerte para constituir la fuerza evolutiva dominante (Lynn *et al.* 2002; Sharp *et al.* 2005). La existencia de codones mayores u óptimos en los genes de alta expresión señalan la influencia de la selección traduccional en la elección de los tripletes (Wang *et al.*, 2011).

Otra teoría es la de EXPRESIÓN-REGULACIÓN (ER) la cual explicaría el sesgo observado en los genes de baja expresión (Bulmer, 1991). Mientras algunos codones son usados preferentemente en genes de alta expresión, otros codones “raros” se observan en frecuencias menores a las esperadas en ausencia de fuerzas evolutivas. Esta teoría plantea que los codones raros están implicados en mecanismos de regulación de la expresión para mantener una expresión baja (Bulmer, 1991). Se han identificado agrupamientos de codones raros al inicio de genes codificantes para proteínas ribosomales de alta expresión rplK, rplJ y rpsM en *E. coli* (Ikemura, 1981b) y esto podría desempeñar un rol de señalización y regulación de estos genes (Novoa & Ribas de Pouplana, 2012).

Un aspecto a tener en cuenta en cualquier modelo, es que el coeficiente de selección para una mutación puntual, en un genoma con cientos de miles de sitios sinónimos variables, se espera que sea extremadamente pequeño. Por tanto, aunque el tamaño poblacional en bacterias sea grande, la estructura poblacional de las especies podría ser tal que redujera el tamaño efectivo poblacional al punto en el que la selección de codones sea menos efectiva. Otro aspecto que también afecta el sesgo en el UCS es la variación en el grado de recombinación y ligamiento lo que puede interferir en el proceso selectivo.

Uso de codones: reglas generales

En un grupo de codones sinónimos, son usados con mayor frecuencia aquellos reconocidos por los tRNAs más abundantes (Bulmer 1988; 1991; Kurland, 1991). En *E. coli* y levadura la elección de codones está determinada por la disponibilidad de tRNAs (Ikemura, 1985). A este fenómeno se lo ha llamado REGLA 1 DE IKEMURA Y OZEKI (Ikemura & Ozeki, 1983) y siempre es muy marcada en genes de alta expresión. Aunque las abundancias relativas de los tRNAs podrían ser las determinantes a corto plazo en el uso de codones, se ha sugerido que a través del curso de la evolución a largo plazo el cambio en las abundancias de tRNAs por sí mismas podrían evolucionar a corresponderse con los patrones genómicos de frecuencias de codones y nucleótidos; o sea que a largo plazo el UCS es seleccionado para correlacionarse con las abundancias relativas de tRNA isoaceptores y viceversa. En todo caso, hay fuerte evidencia de una co-adaptación de las frecuencias relativas de codones y sus respectivos anticodones en el genoma (Bulmer 1988, 1991).

Entre codones reconocidos por el mismo tRNA, se prefieren los que presentan apareamiento Watson-Crick (W-C) "tradicional" con el anticodón en la posición de balanceo (Bulmer, 1991; Kurland, 1991). La HIPÓTESIS DE BALANCEO ("WOOBLE HIPOTHESIS") indica que un mismo tRNA puede aparearse con varios codones por el balanceo en la tercera posición del codón (Roth, 2012). La presencia de bases modificadas también establece un patrón de UCS organismo-específico. El balanceo presenta algunas características de preferencias que también influyen en el UCS (Ikemura, 1985). La REGLA 2 nos indica que en caso de haber una uridina tiolada o 5-carboximetil uridina en la posición de balanceo del anticodón se prefiere un apareamiento con codones terminados en A por sobre los codones terminados en G. Si en la posición de balanceo del anticodón está presente una inosina, entonces se prefiere los codones terminados en U y C por sobre los codones terminados en A (REGLA 3). Esta observación evita apareamientos purina-purina erróneos. Los codones con pirimidinas en la primera y segunda posición (REGLA 4) preferirán equilibrar las fuerzas de interacción codón-anticodón con una C en la tercera posición del anticodón (Ikemura, 1985; Rocha, 2004a) (ver más adelante, MODELOS PREFERENCIA CODON-ANTICODON).

El sesgo es más evidente en genes de alta expresión (Bulmer, 1991; Hilterbrand *et al.*, 2012). En organismos modelo los genes altamente expresados muestran una composición de codones bien adaptada al pool de tRNAs y sus preferencias de codones sinónimos son consideradas como óptimas para maximizar la eficiencia de la traducción pues correlacionan bien con la concentración de tRNAs (Ikemura, 1981a, b; Bulmer, 1991). El sesgo en el UCS en los genes codificantes para las proteínas ribosomales, factores de elongación de la traducción y otras proteínas abundantes es muy pronunciado y está relacionado a las abundancias de los tRNA isoaceptores (Gouy & Gautier, 1982; Canarozzi *et al*., 2010; Von Mandach & Merkl, 2010). Sumado a lo anterior se confirmó que existe sesgo en el UCS para

la eficiencia traduccional en algunos grupos funcionales de genes (efectomas) relacionados con la síntesis y plegamiento de proteínas y con la producción de energía (Von Mandach & Merkl, 2010).

Se define codón óptimo a aquél que aparece en mayor frecuencia en forma estadísticamente significativa en genes altamente expresados (Sharp *et al.*, 1988), generalmente estimado en relación al resto de los genes del genoma. Estos codones serían reconocidos por los tRNAs más abundantes, presentando apareamiento W-C perfecto (Akashi, 2001; Rocha, 2004a; Henry & Sharp, 2007). Los genes de alta expresión usan en forma ampliamente mayoritaria los codones óptimos y presentan un sesgo extremo (Ikemura, 1981a, b, 1985). Algunos codones óptimos fueron coincidentes para *E. coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe y Drosophila melanogaster* y fueron llamados "codones universales óptimos" (Sharp & Devine, 1989). Estos codones son UUC, UAC, AUC, AAC, GAC y GGU.

Bajo la asunción de balance entre selección débil y deriva génica, la proporción esperada de codones preferidos fijados por selección es:

$$E(Fop) = 1 / (1 + (\mu / v)e^{-4Ns}),$$

donde μ y v son las tasas de mutación desde y hacia codones mayores, respectivamente y 4Ns es el coeficiente de selección actuando en los codones preferidos. La ecuación sugiere que los cambios en los parámetros evolutivos, como el tamaño poblacional o la tasa de mutación, pueden tener grandes efectos en la evolución de codones sinónimos aún en especies cercanamente relacionadas (Rocha, 2004a; Ingvarsson, 2008). Mientras que grandes cambios en codones preferidos son raros entre especies cercanas, cambios en parámetros evolutivos, en especial el tamaño efectivo poblacional, son relativamente comunes incluso en escalas de tiempo evolutivo corto (Ingvarsson, 2008).

Uso de codones: modelos de preferencias codón-anticodón.

El MODELO DE FRECUENCIAS define al codón más frecuente como aquel que se puede decodificar por la mayor cantidad de aa-tRNAs en la célula. Los codones reconocidos por el mismo tRNA deberían ser igualmente frecuentes, dejando de lado los sesgos mutacionales. Se favorecen codones que puedan ser leídos por la mayor cantidad de tRNAs con sus respectivos anticodones. El valor esperado bajo un UCS aleatorio es la suma del número de los codones más legibles (cuando hay más de uno) dividido por el número de codones de cada AA. La significancia de este modelo está dada por la razón de los codones observados / codones esperados (Rocha, 2004a).

En el MODELO DE INTERACCION PERFECTA ("PERFECT MATCH") se establece que el codón más frecuente debería tener la interacción codón:anticodón óptima, o sea que debería interaccionar con el anticodón más abundante en forma perfecta. Esto incrementaría la especificidad y sensibilidad del

ribosoma. Para este modelo se asume que la interacción perfecta es el apareamiento W-C y que los residuos modificados no cambian la interacción perfecta. Se recuerda que la base Inosina se corresponde con Uracilo y los nucleósidos modificados U y Q con A y C, respectivamente. La significancia de este modelo está dada por la suma del número de AA para el cual la interacción es perfecta con el anticodón más abundante dividida por el valor esperado (Rocha, 2004a).

El MODELO DE ESTABILIDAD sostiene que, como regla general, se deben evitar interacciones codón:anticodón demasiado fuertes o demasiado débiles, porque esto dificultaría el recambio o "turnover" de tRNAs en el ribosoma y porque podría llevar a errores de lectura y/o a altas tasas de rechazo incorrecto de aa-tRNAs por el ribosoma. Bajo este modelo los "mejores codones" tienen una fuerza de unión codón:anticodón intermedia; por lo tanto, la elección de las bases en las posiciones sinónimas (UCS) es condicionada por la bases que existen en las posiciones no-sinónimas. En otras palabras, los aminoácidos cuyos tripletes presentan bases fuertes (G o C) en la primera y segunda posición tendrían codones óptimos que tienen como 3era base una débil (A o U). En forma inversa, los aminoácidos codificados por codones que comienzan con bases débiles tendrían como óptimos aquellos que tengan una tercera base sinónima con enlace fuerte. El valor observado es la diferencia entre el número de AA que tienen el mayor número de codones que cumplen con el modelo contra aquellos que no lo cumplen (Rocha, 2004a).

Uso de codones: antecedentes

Algunos de los primeros estudios relacionados con el sesgo en el UCS datan de principios de los 80's y fueron dirigidos por Ikemura. En aquellos trabajos se determinó que en *E. coli*, *Salmonella typhimurium* y *S. cerevisiae* esta característica estaba correlacionada con la abundancia del tRNA isoaceptor (Ikemura, 1981a, b, 1982). El estudio de la distribución de codones óptimos y no óptimos en genes de *E. coli* y levadura sugiere fuertemente que la elección de codones está determinada por la concentración disponible de tRNAs y por la eficiencia del apareamiento codón-anticodón (Ikemura, 1981 a, 1985). En la misma época un estudio en *E. coli* demostró que en 83 genes la variación en el UCS era dependiente de los niveles de traducción y era muy marcado en el caso de los genes codificantes para proteínas abundantes (Gouy & Gautier, 1982). Estudios experimentales posteriores demostraron, en mutantes de *rel*A de *E. coli*, que la identidad del codón en sí mismo determina la frecuencia de aparición de errores de lectura (Precup & Parker, 1987).

En muchos organismos procariotas, como por ejemplo *E. coli*, *Methanococcoides burtonii* y *Methanococcus marispaludis* se ha observado que existe preferencia por algunos codones sinónimos (Grantham, 1981; Allen *et al.* 2009; Emery & Sharp, 2011). Este mismo fenómeno, el apartamiento del

uso aleatorio, también se observa a nivel intra-genómico, es decir si comparamos la preferencia de UCS entre genes, tanto en la identidad de los codones elegidos como en la frecuencia de su uso (Sharp, 2005; Canarozzi *et al.*, 2010; Emery & Sharp, 2011). Algunos otros ejemplos clásicos donde se observan correlaciones significativas entre el sesgo del UCS y el nivel de expresión se encuentran en *Physcomitrella patens* y en bacterias del género *Clostridium* (Musto *et al.*, 2003; Stenoien, 2005). Más recientemente se encontró evidencia de selección traduccional en 460 de 461 genomas microbianos examinados (Supek *et al.* 2010). También se observó selección en el uso de codones para la Clase Mollicutes y dado que éstos tienen un genoma sesgado hacia A+T se independiza el patrón observado de UCS con respecto al contenido en GC (Iriarte *et al.* 2011). En organismos multicelulares, también se evidencia el sesgo en el uso de codones. Se ha corroborado este fenómeno en *Caenorhabditis elegans, D. melanogaster, Zea mays, Arabidopsis thaliana, Oryza sativa* y en *Xenopus laevis* (Musto *et al.* 2005; Liu, 2012); con algunas características especiales en el caso de genomas compartimentalizados como los de mamíferos.

Entre otras posibles funciones se ha probado en genes *bla* y *omp*A en *E. coli* la modulación del tráfico de ribosomas en el mRNA que influye en la actividad RNAsa sobre el mRNA y por tanto, las concentraciones de mRNA celular (Thanaraj & Argos, 1996). En el caso de la atenuación de la transcripción del operón para *leucina*, la tasa de traducción de un grupo de codones de leucina decrece a medida que el complejo Leu-tRNA (Leu) aminoacetilado se vuelve poco abundante y no los codones que lo componen. De este modo es el tRNA poco abundante el que afecta la velocidad de la traducción (Bulmer, 1991). Se han publicado varios trabajos donde se relaciona el sesgo en el UCS con la regulación celular. Entre ellos, se ha encontrado que mediante el uso preferente de codones "raros" (siendo éstos los codones reconocidos por tRNAs de baja abundancia) se marca al mRNA para su degradación (Roche & Sauer, 1999). En los genes dnaG, lacI, trpR y araC de *E. coli* se observa un exceso de codones raros (Sharp & Li, 1987). Estos genes no tienen significativamente más codones raros que otros genes de *E. coli* expresados en forma moderada o baja. La presencia de codones raros, que conlleva un bajo CAI en estos genes, es resultado de una ausencia de selección negativa "purificadora" más que de una presencia de selección positiva (Sharp & Li, 1986; 1987).

El UCS también desempeña un rol en el plegamiento de secuencias donde estructuras de alfa-hélices son preferentemente codificadas por codones óptimos, en contraste con las secuencias enriquecidas en codones raros que resultan en estructuras de plegamiento lento como las láminas beta, loops, así como estructuras sin orden aparente (Thanaraj & Argos, 1996; Gingold & Pilpel, 2011). En las zonas de unión de módulos con plegamiento independiente existe un enriquecimiento en codones raros y el uso de estos codones posiblemente esté involucrado en la separación temporal de la traducción de los distintos dominios de una proteína (Gingold & Pilpel, 2011). Análisis en sustratos para la proteína chaperona GroEL en *E. coli* establecieron una interacción entre el UCS óptimos y la acción de GroEL,

indicando una correlación entre el sesgo en el UCS y la regulación en *cis* para el plegamiento de las proteínas (Warnecke & Hurst, 2011). Esta regulación se observó relacionada a secuencias internas a la región codificante tipo Shine-Dalgarno, que en el caso de *E. coli* y *B. subtilis* se identificaron con los codones menores GAG, GAG, AGG y GGG (Li *et al*, 2012). En dípteros e himenópteros se observaron secuencias específicas como reguladores con codones usados más frecuentemente en contextos de codones 3' del codón de inicio y 5' del codón stop (Behura & Severson, 2012).

Por último existe una correspondencia entre la variabilidad en el sesgo en el UCS con factores ambientales en bacterias mesofílicas y termofílicas (por ejemplo con la temperatura la correlación es de un 9.5% en Lobry & Chessel, 2003). Otros antecedentes en procariotas, han establecido correlación de la temperatura óptima de crecimiento y el contenido en GC (Musto *et al.,* 2004, 2006) y correlaciones con respecto al metabolismo de procesamiento del oxígeno y el contenido en GC (Naya *et al.*, 2002).

Uso de codones: importancia

Consecuencias en la optimización en la transcripción y traducción. La selección traduccional es la selección asociada a la eficiencia y precisión en la traducción (Behura & Severson, 2013) y hace referencia a la optimización del proceso en sí mismo (Ran & Higgs, 2012) más que a la selección actuando sobre los productos producidos por la misma. La selección a favor del UCS promueve la traducción eficiente y tiene efectos locales en genes específicos y efectos globales en el "fitness" (adaptabilidad) del organismo (Wald *et al.* 2012; Novoa & Ribas de Pouplana, 2012). Los mecanismos que regulan los niveles de optimización de la traducción por medio del sesgo en el UCS permanecen desconocidos y se plantean distintas posibles explicaciones (Hershberg & Petrov, 2009; Novoa & Ribas de Pouplana, 2012; Behura & Severson, 2013). Bajo condiciones fisiológicas, un pequeño set de genes está encargado de la gran mayoría de la transcripción y traducción que tiene lugar en la célula. Este grupo incluye genes relacionados a la traducción, transcripción, al metabolismo energético y está bajo fuerte presión selectiva para la eficiencia traduccional (Rocha, 2004a). El sesgo en el UCS optimiza la velocidad de traducción en la célula porque minimiza el tiempo requerido para unir el aminoacil-tRNA correcto durante la síntesis de proteína (Sharp *et al.*, 2010). Una traducción más rápida aumenta el número libre de ribosomas en la célula (Kurland, 1991; Botzman & Margalit, 2011) y aumenta el número de mRNAs traducidos por ribosoma porque el ribosoma es liberado del mRNA más rápido. Algo también implícito es que al aumentar la velocidad de la traducción, la tasa total de producción de proteína puede obtenerse con menos ribosomas (Ran & Higgs, 2012). Todas estas formas de optimización son importantes porque el número de ribosomas es usualmente limitante (Kurland, 1991; Klumpp *et al*., 2012). Los ribosomas

requieren menos tiempo para incorporar un codón cuando usan los codones óptimos (Bulmer, 1991; Carlini, 2004).

Los codones reconocidos por tRNA "raros" requieren más tiempo para traducir porque el tiempo de espera antes de la unión del tRNA con el correspondiente codón es inversamente proporcional a su abundancia (Bulmer, 1987, 1988). En la mayoría de los genes, la velocidad de la traducción se reduce durante los primeros 30-50 codones formando una "rampa para la traducción", que es abundante en codones lentos y previene el embotellamientos de ribosomas (Novoa & Ribas de Pouplana, 2012; Klumpp *et al*. 2012). Se agrega un nivel de regulación extra en el mRNA por medio de sitios de pausa del ribosoma con secuencias internas tipo Shine-Dalgarno con codones específicos lentos. El 70% de las pausas fuertes en la traducción son debidas a estas secuencias y pueden además promover cambios en el marco de lectura (Li *et al*, 2012; Novoa & Ribas de Pouplana, 2012). En consecuencia el paso limitante de la tasa del ciclo de elongación de las cadenas polipeptídicas es la difusión del complejo ternario con el tRNA correcto ("cognate") (tRNA + EfTu + GTP) al sitio A del ribosoma (Kurland, 1991; Rocha, 2004 a). La selección en el ribosoma ocurre en dos etapas: una discriminación inicial del complejo ternario por el codón en el sitio A, en el que el tRNA "correcto" es más probable que sea aceptado que el tRNA no correcto, seguido del "proofreading" en el que el tRNA no correcto es más probable de que sea rechazado y liberado del sitio A que el tRNA correcto (Eyre-Walker, 1996; Bulmer, 1991).

Cuanto más rápida sea la tasa de elongación del ribosoma más eficientemente se utiliza la maquinaria de síntesis en la célula, lo cual incrementa su fitness global (Sharp *et al.*, 2010; Botzman & Margalit, 2011). El modelo más aceptado acerca de cómo la abundancia de tRNA afecta la tasa de elongación presupone que las abundancias de los tRNAs son suficientemente bajas para que los ribosomas no estén saturados con los tRNA correctos. El tiempo de rechazo de los tRNA no apropiados es despreciable pero el tiempo tomado por el arribo del primer tRNA apropiado en la proximidad del sito A es el factor limitante de la tasa de traducción. Este tiempo será inversamente proporcional a la abundancia absoluta del tRNA correcto (Bulmer, 1991).

El uso de codones mayores óptimos reduce el costo energético provocado por el "proofreading" rechazando tRNA no correctos. Se disminuye la síntesis de péptidos no funcionales, reduciendo las posibilidades de incorporación errónea y errores de procesividad (Akashi, 2001; Wald *et al*. 2012). Así es más favorable tener más tRNAs del mismo tipo porque permite una co-evolución del sesgo en el UCS en genes de alta expresión lo cual crea una fuerte demanda para sets pequeños de tRNAs (Rocha, 2004 a; Ran & Higgs, 2012). Esta co-evolución permite una mayor eficiencia en genes de alta expresión de forma que estos genes sobre representan los codones correspondientes a los tRNAs más abundantes.

Se ha relacionado el uso de codones específicos en el mRNA para la prevención de formación de horquillas o loops de anti-terminación, asegurando la continuidad apropiada en la transcripción (Li *et al*, 2012; Novoa & Ribas de Pouplana, 2012), así como también con el fin de evitar estructuras secundarias

en el sitio de inicio de la traducción para el ribosoma lo cual genera un reciclado eficiente de los ribosomas y un inicio adecuado de la traducción (Botzman & Margalit, 2011).

En proteínas traducidas a tasas similares, la probabilidad de incorporar aminoácidos en forma errónea aumenta cuanto más largo sea el gen. Por tanto, la selección para reducir incorporaciones erróneas será mayor cuanto más larga sea la secuencia (Eyre-Walker, 1996). La selección a nivel de la traducción favorece la reducción del tamaño de los genes de alta expresión y potencia la fidelidad de la síntesis proteica (Akashi, 2001). Se ha encontrado una correlación positiva entre el largo del gen y la variación en el sesgo de UCS en *E. coli* y en *S. cerevisiae*; tanto en proteínas ribosomales, como en proteínas multiméricas (Eyre-Walker, 1996; Akashi, 2001; Behura & Severson, 2013).

Mediciones de abundancias de mRNA pueden ser indicadores del sesgo del UCS. En el caso de los genes de levadura, el UCS está correlacionado positivamente con los niveles de abundancia de cada mRNA. (Akashi, 2001).

Consecuencias a nivel de las cadenas polipeptídicas. La degeneración en el código genético permite que el mRNA lleve información estructural que puede estar en un codón o en una región nucleotídica adyacente (Thanaraj & Argos, 1996), de manera que existiría una correlación entre propiedades físico-químicas como la hidrofobicidad de los AA codificados y la composición nucleotídica de los codones correspondientes para el plegamiento proteico, formación de láminas beta o hélices. Por tanto, una selección en el UCS sinónimos permite modular esta información estructural que no está en la secuencia aminoacídica (Thanaraj & Argos, 1996; Warnecke & Hurst, 2011).

El uso de codones presente para cada aminoácido, también influye en el plegamiento proteico y la velocidad de traducción (Thanaraj & Argos, 1996; Novoa & Ribas de Pouplana, 2012). De esta manera las proteínas con hélices alfa son codificadas preferentemente por regiones del mRNA con codones rápidos mientras que las láminas beta y "COILS" son preferentemente codificados por regiones lentas de mRNA (Thanaraj & Argos, 1996). Los codones raros se han encontrado en clusters en zonas de unión de cadenas polipeptídicas (Gingold y Pilpel, 2011). Con esto, se disminuye la generación de productos inactivos o tóxicos para la célula por la generación de polipéptidos con plegamiento incorrecto (Gingold & Pilpel, 2011; Waldman *et al*. 2011; Wald *et al*. 2012).

Consecuencias selectivas: factores ambientales y estilos de vida. El UCS sinónimos puede ser sujeto a selección natural y específicamente un factor ambiental (como la alta temperatura) puede favorecer algunos codones específicos en eubacteria y archaea (Lynn *et al*. 2002). Organismos con capacidad de vivir en ambientes múltiples o con amplia variabilidad metabólica, exhiben alto grado de sesgo en el UCS porque se favorecería la plasticidad de adaptarse eficientemente a diferentes medio-ambientes (Botzman & Margalit, 2011; Thiele *et al*., 2012) Diversos estudios muestran que parte de la

variación en el UCS está relacionada con la temperatura de crecimiento óptimo en genomas de procariotas mesófilos y termófilos, no estando relacionada con el contenido nucleotídico (Lynn *et al*., 2002; Singer & Hickey, 2003). La selección puede estar favoreciendo incrementar la estabilidad del mRNA a altas temperaturas. La estabilidad de la molécula de mRNA puede resultar en niveles incrementados de proteína traducida por molécula de mensajero. Así la estabilidad del mRNA puede ser sujeta a presión selectiva similar a la eficiencia traduccional. Los genomas de organismos termófilos son ricos en purinas en la hebra codificante y esta preferencia puede estar afectando tanto la estabilidad del mRNA así como la frecuencia de codones sinónimos dentro de estos genomas (Lynn *et al*., 2002).

Plasticidad frente a stress. Los patrones de UCS en *S. cerevisiae*, *S. pombe* y *C. elegans* cambian frente a situaciones múltiples de stress, y se observa una tendencia consistente donde se incrementan en el transcriptoma los codones "raros" que se corresponden con baja concentración de tRNAs y baja el grado de representación de los codones óptimos. En la levadura en condiciones de stress, el aumento es del 25% y en algunas situaciones de un 40% en relación a nivel del sesgo en el UCS de referencia. Por otro lado el patrón de variación del UCS está anti-correlacionado con experimentos de estrés y recuperación del estrés. Los cambios en el sesgo del UCS no son explicados por una necesidad de cambiar la composición aminoacídica del proteoma y tampoco pueden ser atribuídos a cambios en la disponibilidad de nucleótidos (Gingold *et al*., 2012).

OBJETIVOS:

Objetivo general: Estudiar el uso de codones sinónimos en organismos pertenecientes al superdominio Archaea.

Objetivos específicos:

1) Determinar la filogenia del superdominio Archaea.

2) Estudiar la relación existente entre los codones óptimos y las abundancias de tRNAs genómicos.

3) Determinar si los sesgos en el uso de codones pueden ser relacionados a alguna fuerza evolutiva.

   (a) Estudiar la influencia de la selección, si está actuando, en qué magnitud (cálculo de coeficientes de selección).

   (b) Analizar si se relaciona con alguna característica particular del ambiente (temperatura, disponibilidad de oxígeno, profundidad, pH, salinidad) en el que se encuentra la especie analizada.

4) Comparar patrones que se observan en los distintos grupos filogenéticos. Determinar si existen tendencias filogenéticas correspondientes a patrones de sesgo de UCS similares entre grupos. Determinar si son explicados por las fuerzas evolutivas vistas en (3) o son atribuibles a fenómenos de inercia propia de los grupos.

MATERIALES Y MÉTODOS:

4. Generación de base de datos.

Las secuencias de DNA codificante (datos ffn), las secuencias de DNA codificante para RNA 16S (datos frn) y las tablas de anotación de las proteínas codificadas (tablas ptt) para cada especie en Archaea (tabla 1) fueron obtenidas del GeneBank NCBI vía ftp ftp://ftp.ncbi.nih.gov/genomes/Bacteria a la fecha 27/03/2011.

5. Determinación de la filogenia.

Se alinearon las secuencias para RNA 16S obtenidas de los datos frn mediante MUSCLE 3.8 (Edgar, 2004) y se construyó un árbol filogenético por máxima verosimilitud (ML) en PHYML 3.0 tomando como modelo de sustitución nucleotídica HKY85 (Guindon & Gascuel, 2003) y se calculó el bootstrap para darle apoyo estadístico a los nodos.

6. Búsqueda de genes ortólogos.

Se obtuvo la secuencia de las proteínas de cada genoma y se realizó un BLASTp (Atschul *et al.*, 1990) para identificar las secuencias ortólogas entre todos los organismos analizados. Se usó el criterio de "best-reciprocal-hit" como estrategia de búsqueda donde se compara todas las combinaciones posibles de proteínas potencialmente ortólogas. De esta forma determinamos genes que divergieron a partir de un evento de especiación para dar robustez al análisis de selección en el uso de codones en la filogenia de Archaea y ésta no sea debida a particularidades de proteínas de cada especie en sí, duplicaciones genómicas o eventos de transferencia horizontal. Se alineó usando CLUSTALW (Thompson *et al*., 1994).

Nombre	Phyllum	Clase	Orden	ftp CODIGO	Cód.	N genes	TG (Mb)	%GC
Acidilobus saccharovorans 345-15 chromosome	Crenarchaeota	Thermoprotei	Acidilobales	NC_014374	asa	1499	1.5	58
Aeropyrum pernix K1	Crenarchaeota	Thermoprotei	Desulfurococcales	NC_000854	aac	1700	1.67	57
Desulfurococcus kamchatkensis 1221n	Crenarchaeota	Thermoprotei	Desulfurococcales	NC_011766	ber	1471	1.37	46
Hyperthermus butylicus DSM 5456	Crenarchaeota	Thermoprotei	Desulfurococcales	NC_008818	arf	1602	1.67	54
Ignicoccus hospitalis KIN4/I	Crenarchaeota	Thermoprotei	Desulfurococcales	NC_009776	avu	1434	1.3	57
Ignisphaera aggregans DSM 17230	Crenarchaeota	Thermoprotei	Desulfurococcales	NC_014471	iad	1930	1.88	36
Staphylothermus hellenicus DSM 12710	Crenarchaeota	Thermoprotei	Desulfurococcales	NC_014205	shd	1599	1.58	37
Staphylothermus marinus F1	Crenarchaeota	Thermoprotei	Desulfurococcales	NC_009033	aro	1570	1.57	36
Thermosphaera aggregans DSM 11486	Crenarchaeota	Thermoprotei	Desulfurococcales	NC_014160	bqh	1387	1.32	47
Metallosphaera sedula DSM 5348	Crenarchaeota	Thermoprotei	Sulfolobales	NC_009440	asy	2256	2.19	47
Sulfolobus acidocaldarius DSM 639	Crenarchaeota	Thermoprotei	Sulfolobales	NC_007181	ajb	2223	2.23	37
Sulfolobus islandicus L.D.8.5	Crenarchaeota	Thermoprotei	Sulfolobales	NC_013766	bnp	2916	2.75	36
Sulfolobus solfataricus P2	Crenarchaeota	Thermoprotei	Sulfolobales	NC_002754	abt	2977	2.99	36
Sulfolobus tokodaii str. 7	Crenarchaeota	Thermoprotei	Sulfolobales	NC_003106	adb	2825	2.69	34
Caldivirga maquilingensis IC-167	Crenarchaeota	Thermoprotei	Thermoproteales	NC_009954	awx	1963	2.08	44
Pyrobaculum aerophilum str. IM2	Crenarchaeota	Thermoprotei	Thermoproteales	NC_003364	adp	2603	2.22	52
Pyrobaculum arsenaticum DSM 13514	Crenarchaeota	Thermoprotei	Thermoproteales	NC_009376	aso	2299	2.12	56
Pyrobaculum calidifontis JCM 11548	Crenarchaeota	Thermoprotei	Thermoproteales	NC_009073	art	2149	2.01	58
Pyrobaculum islandicum DSM 4184	Crenarchaeota	Thermoprotei	Thermoproteales	NC_008701	aqi	1978	1.83	49
Thermofilum pendens Hrk 5	Crenarchaeota	Thermoprotei	Thermoproteales	NC_008698	aqf	1824	1.81	58
Thermoproteus neutrophilus V24Sta	Crenarchaeota	Thermoprotei	Thermoproteales	NC_010525	azo	1966	1.77	60
Vulcanisaeta distributa DSM 14429	Crenarchaeota	Thermoprotei	Thermoproteales	NC_014537	vdd	2493	2.37	46
Archaeoglobus fulgidus DSM 4304	Euryarchaeota	Archaeoglobi	Archaeoglobales	NC_000917	aam	2420	2.18	49
Archaeoglobus profundus DSM 5631	Euryarchaeota	Archaeoglobi	Archaeoglobales	NC_013741	bnk	1819	1.56	42
Ferroglobus placidus DSM 10642	Euryarchaeota	Archaeoglobi	Archaeoglobales	NC_013848	bnv	2480	2.2	45
Halalkalicoccus jeotgali B3 chromosome	Euryarchaeota	Halobacteria	Halobacteriales	NC_014297	hjb	3035	3.7	66
Haloarcula marismortui ATCC 43049 (II)	Euryarchaeota	Halobacteria	Halobacteriales	NC_006396	ahx	3412	4.27	63
Halobacterium salinarum R1	Euryarchaeota	Halobacteria	Halobacteriales	NC_010364	aym	2110	2.67	69
Halobacterium sp. NRC-1	Euryarchaeota	Halobacteria	Halobacteriales	NC_002607	abh	2075	2.57	68
Haloferax volcanii DS2 chromosome	Euryarchaeota	Halobacteria	Halobacteriales	NC_013967	boz	2945	4.01	67
Halomicrobium mukohataei DSM 12286	Euryarchaeota	Halobacteria	Halobacteriales	NC_013202	bla	3173	3.33	67
Haloquadratum walsbyi DSM 16790	Euryarchaeota	Halobacteria	Halobacteriales	NC_008212	ann	2610	3.18	49
Halorhabdus utahensis DSM 12940	Euryarchaeota	Halobacteria	Halobacteriales	NC_013158	bkk	2998	3.12	64
Halorubrum lacusprofundi ATCC 49239 (II)	Euryarchaeota	Halobacteria	Halobacteriales	NC_012028	bgb	3184	3.69	66
Haloterrigena turkmenica DSM 5511	Euryarchaeota	Halobacteria	Halobacteriales	NC_013743	bnl	3739	5.44	67
Natrialba magadii ATCC 43099	Euryarchaeota	Halobacteria	Halobacteriales	NC_013922	bok	3559	4.44	63
Natronomonas pharaonis DSM 2160	Euryarchaeota	Halobacteria	Halobacteriales	NC_007426	ajw	2659	2.75	64
Methanobrevibacter ruminantium M1 chromosome	Euryarchaeota	Methanobacteria	Methanobacteriales	NC_013790	bnr	2217	2.94	36
Methanobrevibacter smithii ATCC 35061	Euryarchaeota	Methanobacteria	Methanobacteriales	NC_009515	atz	1793	1.85	32
Methanosphaera stadtmanae DSM 3091	Euryarchaeota	Methanobacteria	Methanobacteriales	NC_007681	alg	1534	1.77	29
Methanothermobacter marburgensis str. Marburg	Euryarchaeota	Methanobacteria	Methanobacteriales	NC_014408	mmm	1753	1.64	50
Methanothermobacter thermautotrophicus str. Delta	Euryarchaeota	Methanobacteria	Methanobacteriales	NC_000916	aal	1873	1.75	51
Methanocaldococcus fervens AG86	Euryarchaeota	Methanococci	Methanococcales	NC_013156	bkj	1546	1.51	33
Methanocaldococcus infernus ME chromosome	Euryarchaeota	Methanococci	Methanococcales	NC_014122	bpx	1441	1.33	34
Methanocaldococcus sp. FS406-22 chromosome	Euryarchaeota	Methanococci	Methanococcales	NC_013887	bob	1804	1.77	33
Methanocaldococcus vulcanius M7	Euryarchaeota	Methanococci	Methanococcales	NC_013407	bls	1727	1.76	33
Methanococcus aeolicus Nankai-3	Euryarchaeota	Methanococci	Methanococcales	NC_009635	auo	1490	1.57	31
Methanococcus maripaludis C5	Euryarchaeota	Methanococci	Methanococcales	NC_009135	asc	1813	1.79	34
Methanococcus vannielii SB	Euryarchaeota	Methanococci	Methanococcales	NC_009634	aun	1678	1.72	32
Methanococcus voltae A3 chromosome	Euryarchaeota	Methanococci	Methanococcales	NC_014222	mva	1717	1.94	31
Methanocella paludicola SANAE	Euryarchaeota	Methanomicrobia	Methanocellales	NC_013665	bmz	3004	2.96	57
Uncultured methanogenic archaeon RC-I	Euryarchaeota	E.S.		NC_009464	ath	3085	3.18	56
Candidatus Methanoregula boonei 6A8	Euryarchaeota	Methanomicrobia	Methanomicrobiales	NC_009712	avk	2450	2.54	56
Candidatus Methanosphaerula palustris E1-9c	Euryarchaeota	Methanomicrobia	Methanomicrobiales	NC_011832	bez	2655	2.92	57
Methanocorpusculum labreanum Z	Euryarchaeota	Methanomicrobia	Methanomicrobiales	NC_008942	ark	1739	1.8	51
Methanoculleus marisnigri JR1	Euryarchaeota	Methanomicrobia	Methanomicrobiales	NC_009051	arq	2489	2.48	63
Methanoplanus petrolearius DSM 11571	Euryarchaeota	Methanomicrobia	Methanomicrobiales	NC_014507	mpd	2785	2.84	49
Methanospirillum hungatei JF-1	Euryarchaeota	Methanomicrobia	Methanomicrobiales	NC_007796	alv	3139	3.54	46
Methanococcoides burtonii DSM 6242	Euryarchaeota	Methanomicrobia	Methanosarcinales	NC_007955	amn	2273	2.58	42
Methanohalobium evestigatum Z-7303 chromosome	Euryarchaeota	Methanomicrobia	Methanosarcinales	NC_014253	mez	2148	2.41	38
Methanohalophilus mahii DSM 5219	Euryarchaeota	Methanomicrobia	Methanosarcinales	NC_014002	bpb	1987	2.01	44
Methanosaeta thermophila PT	Euryarchaeota	Methanomicrobia	Methanosarcinales	NC_008553	apk	1696	1.88	55
Methanosarcina acetivorans C2A	Euryarchaeota	Methanomicrobia	Methanosarcinales	NC_003552	adw	4540	5.75	45
Methanosarcina barkeri str. fusaro	Euryarchaeota	Methanomicrobia	Methanosarcinales	NC_007355	ajq	3606	4.87	42
Methanosarcina mazei Go1	Euryarchaeota	Methanomicrobia	Methanosarcinales	NC_003901	adz	3368	4.1	44
Methanopyrus kandleri AV19	Euryarchaeota	Methanopyri	Methanopyrales	NC_003551	adv	1687	1.69	61
Pyrococcus abyssi GE5	Euryarchaeota	Thermococci	Thermococcales	NC_000868	aad	1780	1.77	45
Pyrococcus furiosus DSM 3638	Euryarchaeota	Thermococci	Thermococcales	NC_003413	adr	2125	1.91	41
Pyrococcus horikoshii OT3	Euryarchaeota	Thermococci	Thermococcales	NC_000961	aar	1955	1.74	42
Thermococcus gammatolerans EJ3	Euryarchaeota	Thermococci	Thermococcales	NC_012804	biq	2156	2.05	54
Thermococcus onnurineus NA1	Euryarchaeota	Thermococci	Thermococcales	NC_011529	bdo	1975	1.85	52
Thermococcus sibiricus MM 739	Euryarchaeota	Thermococci	Thermococcales	NC_012883	bja	2035	1.85	41
Picrophilus torridus DSM 9790	Euryarchaeota	Thermoplasmata	Thermoplasmatales	NC_005877	agw	1535	1.55	37
Thermoplasma acidophilum DSM 1728	Euryarchaeota	Thermoplasmata	Thermoplasmatales	NC_002578	abg	1482	1.56	47
Thermoplasma volcanium GSS1	Euryarchaeota	Thermoplasmata	Thermoplasmatales	NC_002689	abo	1499	1.58	41
Aciduliprofundum boonei T469 chromosome	Euryarchaeota	un. Euryarchaeota	Aciduliprofundum	NC_013926	bol	1544	1.49	39
Candidatus Korarchaeum cryptofilum OPF8	Korarchaeota	cand. Korarchaeum		NC_010482	aza	1602	1.59	50
Nanoarchaeum equitans Kin4-M	Nanoarchaeota	Nanoarchaeum		NC_005213	agj	536	0.49	31
Nitrosopumilus maritimus SCM1	Thaumarchaeota	M A G 1	Nitrosopumilales	NC_010085	axj	1795	1.65	35

Tabla 1. Lista de especies y relaciones filogenéticas (Phyllum, Clase y Orden) que se muestran en esta tabla son las que se muestran en el GeneBank NCBI al 27/03/2011. **ftp CÓDIGO** se corresponde al código NCBI de los genomas completos para cada una de las especies. **Cód** en referencia al código generado para este trabajo para identificar a cada especie. **N genes** significa número de genes. **TG (Mb)** Tamaño genómico en Mb. **%GC** contenido GC calculado en el presente trabajo.

7. Estudio de uso de codones.

4.1) Análisis de correspondencia (COA).

El COA es un análisis estadístico multivariado que establece un sistema de asociaciones entre filas y columnas. Típicamente en estas matrices se encuentra información acerca de valores indicativos de uso de codones (como el RSCU) o de aminoácidos en las columnas matriciales, y los relaciona con los genes en estudio que colocamos en las filas de la matriz rectangular. Las filas y columnas de la matriz definen distancias que son correlacionadas a distancias euclidianas en una tabla de contingencia (Peden, 1999). El COA identifica tendencias en la variación de los datos incluso si la variación es continua y distribuye los genes en un eje según la tendencia identificada (Peden, 1999).

En definitiva el COA permite comparar proyecciones de puntos originalmente multidimensionales en dimensiones menores, visibles mediante gráficas una cantidad de genes en conjunto donde cada punto de la gráfica representa un gen con características definidas en la matriz multidimensional, quedando más cercanos aquellos con distancias similares y es posible establecer tendencias generales identificadas que explican la variación continua de los datos, según cómo se ubiquen en la gráfica.

Se analizó la existencia de patrones de agrupamiento a nivel de proteínas ribosomales y los resultados obtenidos mediante CODONW (Peden, 1999). En el análisis de la variabilidad interna en el UCS para cada especie vimos si hay una tendencia principal que muestre una posible selección traduccional en la variación del UCS, los genes altamente expresados se encuentran en uno de los extremos del eje principal (incluyen proteínas ribosomales) y los genes raramente expresados se encuentran en el otro extremo (proteínas reguladoras). Este clustering constituye un método sensitivo para observar UCS no al azar y ha sido probado en diferentes organismos: *E. coli* (Medigue *et al*., 1991), *B. subtilis* (Sharp *et al.*, 1990), *S. cerevisiae* (Sharp & Cowe, 1991), *D. melanogaster* (Shields *et al*., 1988), *Dictyostelium discoideum* (Sharp & Devine, 1989), *C. elegans* (Stenico *et al.*, 1994), *Rickettsia prowazwkii* (Andersson & Sharp, 1996), *Borrelia burgdorferi* (McInerney, 1998), *Mycobacterium tuberculosis* (Andersson & Sharp, 1996), *Methanococcus maripaludis* (Emery & Sharp, 2011), entre otros.

4.2) Estimación de índices de sesgo en el UCS.

Se analizó comparando el uso observado en cada especie mediante el programa CODONW (Peden, 1999) para obtener el número de codones total y el RSCU (Relative Synonymous Codon Usage) de los genes. Otros parámetros que se calcularon para estudiar el uso de codones son el ENC´, el CAI, el MELP y el valor B; para lo cual se utilizó el programa INCA2.1 (Supek & Vlahovicek, 2004). En todos los casos se tomó como conjunto de referencia a los genes que codifican para proteínas ribosomales dado que son genes conservados de alta expresión.

El Fop es la frecuencia simple de codones óptimos en un gen (Henry & Sharp, 2007).

El RSCU es la frecuencia observada para cada codón en relación a la frecuencia esperada si todos los codones para cualquier aminoácido particular se usara igualmente (Sharp, 1986). El RSCU no depende de los AA en particular o del tamaño de los genes analizados, por lo que permite comparar diferentes genes entre sí (Peden, 1999), sets de datos de diferente tamaño y entre diferentes aminoácidos (Sharp, 1986).

El parámetro ENc es el número efectivo de codones usado por un gen, similar al número de alelos efectivos de una población (Wright, 1990). "Representa el número de codones igualmente usados que generarían el mismo sesgo de UCS observado" (Peden, 1999). Como se define para un gen, es el número de codones usados por cada uno de los AA presentes en el gen. Diferencial ENc es la diferencia entre el número efectivo de codones del genoma y el número efectivo de un subset de proteínas en particular (por ejemplo, los genes que codifican para proteínas ribosomales) y constituye una buena medida del sesgo en el UCS relacionado a la optimización de la traducción (Rocha, 2004).

El método ENc' (ENc prima), permite calcular el número efectivo de codones en relación al estándar que es no uniforme en el UCS (Supek, 2005). Si el valor ENc' de un gen es cercano a 61 hay poca diferencia en el uso de codones.

El índice de adaptación de codones (CAI) mide el grado de sesgo de uso de codones con respecto a un grupo de codones que sabemos que son usados preferentemente (y constituyen el "set de referencia") para una especie determinada (Peden, 1999; Rocha, 2004). Para Sharp & Li (1987) el índice CAI debe ser relativo a un set de genes que fueron determinados experimentalmente como genes de alta expresión. Es una frecuencia similar a Fop pero que pondera los codones sub-óptimos en forma diferente, de acuerdo a cómo se evitan estos codones sub-óptimos en los genes de alta expresión (Henry & Sharp, 2007). El CAI permite comparar entre especies porque es un estadístico que está normalizado (Sharp & Li, 1987).

MELP (MILC-based expression level predictor) es un estadístico que predice cuantitativamente el nivel de expresión de un gen usando datos genómicos y se calcula como la razón entre las distancias de UCS genómicas promedio y las distancias de UCS de un set de genes de referencia;

MELP = MILC genom / MILC ref set

El método MILC es un método de medición independiente del largo y de la composición y cuantifica la distancia existente entre el UCS de un gen y alguna distribución de codones esperada (Supek & Vlahoviček, 2005). La distribución de codones puede derivarse de la composición nucleotídica, de un gen o de un grupo de genes.

El valor B o CUB (Codon Usage Bias) es muy útil cuando se comparan secuencias entre grupos de genes o genomas enteros y que se espera presenten diferente UCS (Supek & Vlahoviček, 2005). Estima la diferencia en el sesgo de codones de un grupo de genes relativo a un segundo grupo de genes de referencia (Karlin *et al*, 1998).

4.3) Identificación de codones óptimos.

Se compararon los resultados de conteo de codones mediante un test χ^2 de contingencia para establecer qué codones son significativamente más usados en genes de alta expresión con respecto al resto de los genes. Estos serán reconocidos como "óptimos" o “mayores” para cada especie.

8. Determinación de ARNt.

Mediante el programa tRNA-scan-SE (Lowe & Eddy, 1997) se identificaron los ARNt de cada genoma y se identificaron los anticodones presentes.

9. Cálculo de coeficientes de selección.

Se calcularon los coeficientes de selección según Sharp y colaboradores (2005), llamado valor S. Este coeficiente permite cuantificar el desvío o sesgo del UCS debido al efecto de la selección. Se estima el desvío del UCS en genes de alta expresión respecto a genes de baja expresión (los últimos reflejarían el sesgo mutacional). Se comparó el valor entre las distintas especies y se discutió en un marco filogenético. El valor S está correlacionado positivamente con el número de operones de rRNA y el número de genes de tRNA en bacterias (Sharp *et al*., 2005).

$P = e^S V / (e^S V + U)$

Con $S = 2Ne\ s$,

$U = 2Ne\ u$

$$V = 2Ne\ v$$

$$S = \ln [(P \times k) / (1-P)]$$

Donde P es la frecuencia en equilibrio esperada del codón y k = U/V que representan las tasas mutacionales actuando entre las variantes de codones en la población efectiva. En genes donde la selección es muy débil como para ser efectiva, la frecuencia de codones está determinada por las tasas de mutación entre ellos: P = V / (V+U) por lo que k = (1-P) / P.

RESULTADOS

1) Determinación de la filogenia.

Las secuencias alineadas para el ARN 16S se usaron para elaborar un árbol filogenético por ML con el modelo HKY85, el resultado del análisis filogenético se muestra en las figuras 1 y 2 y está en concordancia con filogenias publicadas anteriormente (Gribaldo & Brochier-Armanet, 2006; Bintrim *et al.*, 1997; Gao & Gupta, 2007). El árbol se enraizó arbitrariamente en nanoarchaeota. En la figura se pueden ver los nodos basales principales que se corresponden a los *phyla* principales de Crenarchaeota y Euryarchaeota y entre ellas se ubican los organismos *Nanoarchaeum equitans Kin4-M* (agj), *Nitrosopumilus maritimus SCM1* (axj) y *Candidatus Korarchaeum cryptofilum OPF8* (aza) que representan a los potenciales *Phylum* de Nanoarchaeota, Thaumarchaeota y Korarchaeota, cada uno con largas ramas basales.

El *Phylum* Crenarchaota se presenta como un grupo monofilético con un BS del 88%. Está conformado por los órdenes Thermoproteales (BS 77%) y por un nodo polifilético (BS 98%) que incluye a los órdenes Sulfolobales, Desulfurococcales y Acidilobales. El orden Thermoproteales aparece como un nodo monofilético de posición basal con respecto a los otros órdenes con un apoyo estadístico del 77%.

El orden de los Sulfolobales está representado en un nodo monofilético (BS 100%). Los Desulfurococcales aparecen divididos en una rama externa con *Hyperthermus butylicus DSM* 5456 (arf) en posición basal y con otro nodo de BS del 53%. En una de las ramas terminales de este nodo se ubica el organismo *Acidilobus saccharovorans 345-15* (asa) representante del órden de los Acidilobales.

El *Phylum* de Euryarchaeota está representado por un mayor número de organismos que se agrupan en los órdenes Methanopyrales, Thermococcales, Archaeoglobales, Halobacteriales, Methanomicrobiales, Methanocellales, Methanosarcinales, Methanobacteriales, Methanococales y Thermoplasmatales. Los Methanopyrales están representados por *Methanopyrus kandleri* AV19 (adv) y reprenta la rama terminal más basal de los Euryarchaota. A continuación se encuentra a los Thermococcales agrupados con un BS del 100%. El resto de los órdenes se agrupan formando nodos con altos valores de BS también. Debe ser mencionado el orden formado por *Methanocella paludicola SANAE* (bmz) en representación de los Methanocellales que aparece junto a *Uncultured methanogenic archaeon RC-I* (ath) formando un clado con un BS del 100%. Esto es consistente con publicaciones recientes donde ubican a este último organismo ahora llamado *Methanocella arvoryzae* dentro del orden de los Methanocellales (Erkel *et al.*, 2006).

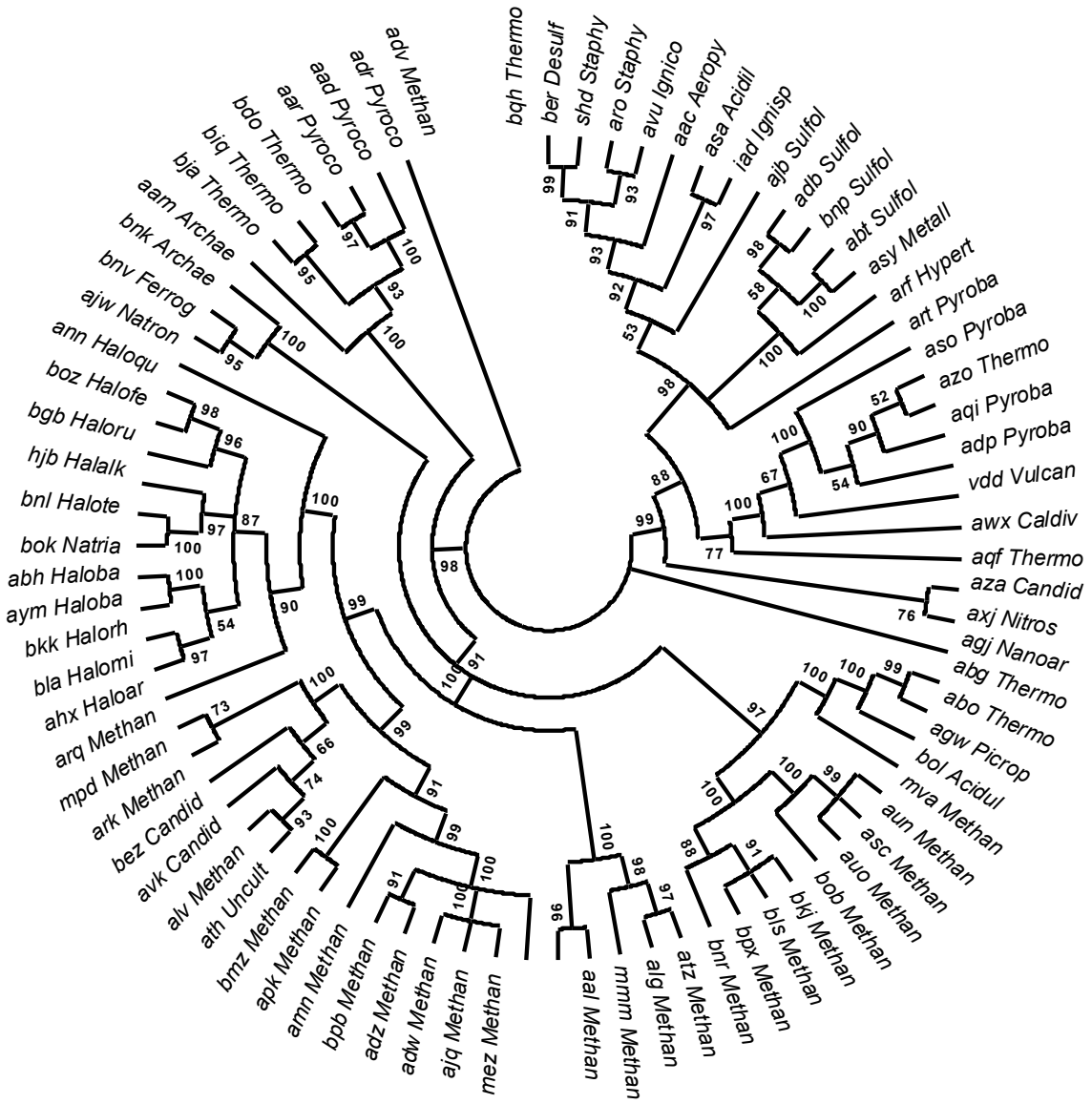

Figura 1. Filogenia Archaea RNA 16S PHYML3.0 con modelo HKY85 y valores de bootstrap (para 100 réplicas).

2) Análisis de correspondencia (COA):

Se analiza el UCS comparando el uso observado en cada especie, mediante el programa CODONW (Peden, 1999) para obtener los valores del número de codones total y el uso relativo de codones sinónimos o RSCU por sus siglas en ingles. El RSCU es la frecuencia observada de cada codón con respecto a la frecuencia esperada si todos los codones para cualquier aminoácido (AA) fueran usados con igual frecuencia (Sharp, 1986).

Con los datos obtenidos, en primer lugar se buscaron patrones de agrupamiento de las proteínas ribosomales en los ejes generados por el análisis de correspondencia (COA) sobre el uso de AAs y para el RSCU (tabla 2 y datos suplementarios). Cuando la selección para la traducción explica la mayor parte de la variabilidad observada en el uso de codones sinónimos (UCS), es decir, es la tendencia principal en el UCS, los genes que codifican para proteínas ribosomales y otros genes de alta expresión se agrupan hacia alguno de los extremos del eje principal en el COA sobre RSCU. Esto sugiere que este tipo de genes tienen un uso de codones particular, compartido, y además que la tendencia que el eje está capturando está asociada a la expresión.

La selección operando sobre la composición de aminoácidos en genes de alta expresión puede generar un patrón similar en los resultados del COA. Esto se observa para todos los grupos dentro de Crenarchaeota. Los organismos analizados muestran una clara tendencia con respecto al primer eje, o eje principal, y la variabilidad interna capturada por el primer eje en promedio es de 20.45% +/- SD 2.75%.

Con respecto al uso diferencial de codones sinónimos, el eje principal de los Desulfurococcales explica en promedio un 9.07% +/- 2.86% y los Sulfolobales 8.79% +/- 2.27%; ambos presentan una tendencia similar baja para un uso diferencial de codones sinónimos. Estos dos grupos, están cercanos filogenéticamente (ver figura 1). La especie *Acidilobus saccharovorans*, único representante en nuestro análisis del Orden de los Acidilobales, también presenta una mayor tendencia a favor de un UCS (19.32%) y se observó agrupamiento en los genes de alta expresión. Sorprende que este orden, se encuentra muy emparentado (ver figura 1) con los grupos Desulfurococcales y Sulfolobales.

En el caso de los Thermoproteales no se encontró un agrupamiento claro de los genes de alta expresión en el COA-RCSU en ninguno de los ejes principales. El análisis mostró que el porcentaje de variabilidad explicada fue de 11.16%, con un desvío estándar de ±4.06%.

Por tanto dentro del *Phylum* de Crenarchaeota no encontramos agrupamiento que evidencie selección para el UCS, aunque si habría una preferencia en el uso de AA por parte de los genes de alta expresión.

Por otro lado, se observa una gran diversidad de patrones en los distintos órdenes del *Phylum* Euryarchaeota.

En algunos grupos encontramos una leve tendencia de selección para el UCS, de acuerdo al criterio de agrupamiento de los genes de alta expresión en los ejes principales y una leve o importante tendencia para el uso de AA; tal sería el caso de los Halobacteriales, Methanobacteriales, Methanosarcinales y Thermococcales.

El grupo de los Archaeoglobales y de los Methanococcales muestra un claro agrupamiento de genes de alta expresión, lo que sugiere un UCS y un uso de AA diferencial en los genes de alta expresión capturado por los ejes principales. Los valores de variabilidad interna asociados a los ejes principales en los Archaeoglobales son en promedio de 7.74% (SD ±1.42%) para el eje principal del RSCU y del 18.18% (SD ±0.65%) en el eje principal de uso de AA. En el caso de los Methanococcales, los valores son de 8.68% (SD ±2.00%) para el RSCU y de 18.62% (SD ±1.45%) en el caso del COA para el primer eje principal del análisis del RSCU. En contraposición, los Methanomicrobiales muestran un agrupamiento de genes de alta expresión en el segundo eje del análisis COA sobre el RSCU, lo que sugiere un UCS diferencial en los genes de alta expresión, con un valor de variabilidad explicada promedio de 5.28% (SD ±1.96), contra un 14.2% (SD ±2.20) del primer eje del RSCU. Esta variabilidad interna está correlacionada con el contenido en GC, evidenciando la influencia del sesgo mutacional. Se observa la misma tendencia para el uso de AA en genes de baja expresión (promedio 18.50%, SD ±2.39).

El grupo de los Thermoplasmatales sólo presenta clustering de genes de alta expresión en el primer eje de AA (promedio 26.33%, SD ±1.32).

El caso de los Methanocellales sólo tenemos dos organismos (*Methanocella paludicola* y *Uncultured methanocenica archaeon RC.I*) que presentan el mismo patrón que observamos en los Methanomicrobiales y Methanosarcinales. Es de mencionar que estos grupos aparecen muy cercanos en la filogenia presentada en la figura 1. Los Methanocellales mostraron valores de porcentaje de variabilidad explicada promedio de 18.59%, SD ±0.39 para el primer eje del análisis de correspondencia sobre el RSCU y de 18.22%, SD ±0.93, en el caso de el primer eje del análisis de correspondencia para el uso de AA.

Nombre	RSCU1	RSCU2	GC/RSCU 1	GC/RSCU 2	RSCU1|MELP	S (Sharp 2005)
Acidilobus saccharovorans 345-15 chromosome	SI/NO	SI/NO	0.734	0.0025	87,51%	-0,438954304
Aeropyrum pernix K1	NO	NO	0.6179	0.006	58,96%	0,121851129
Hyperthermus butylicus DSM 5456	NO	NO	0.4403	0.2079	10,95%	0,319964049
Staphylothermus marinus F1	SI	NO	0.0018	0.0035	19,21%	0,088544339
Ignicoccus hospitalis KIN4/I	NO	NO	0.3984	0.0918	39,89%	0,372650371
Desulfurococcus kamchatkensis 1221n	NO	NO	0.0008	0.0028	27,25%	-0,67098298
Thermosphaera aggregans DSM 11486	SI	NO	0.63	0.0004	72,95%	0,463249459
Ignisphaera aggregans DSM 17230	NO	NO	0.3404	0.1302	37,91%	-0,612633136
Staphylothermus hellenicus DSM 12710	SI	NO	0.0002	0.1141	23,32%	0,094144471
Sulfolobus solfataricus P2	SI	NO	0.5651	0.0033	74,63%	-0,438760552
Sulfolobus tokodaii str. 7	SI	NO	0.3566	0.0251	56,31%	-0,353529329
Sulfolobus acidocaldarius DSM 639	NO	NO	0.3984	0.00003	57,33%	-0,495323211
Metallosphaera sedula DSM 5348	SI/NO	NO	0.6537	0.003	78,12%	-0,313736213
Sulfolobus islandicus L.D.8.5	NO	NO	0.2666	0.0065	33,41%	-0,282336862
Pyrobaculum aerophilum str. IM2	NO	NO	0.542	0.0094	36,21%	-0,136658471
Thermofilum pendens Hrk 5	SI	NO	0.7018	0.0154	89,26%	-0,545506491
Pyrobaculum islandicum DSM 4184	NO	NO	0.8198	0.0129	86,71%	-0,373085103
Pyrobaculum calidifontis JCM 11548	NO	NO	0.274	0.4026	76,08%	-0,342108957
Pyrobaculum arsenaticum DSM 13514	NO	SI/NO	0.6877	0.0224	66,64%	-0,155036923
Caldivirga maquilingensis IC-167	SI	NO	0.0181	0.2109	68,80%	-0,219766866
Thermoproteus neutrophilus V24Sta	NO	NO	0.4948	0.0779	28,65%	-0,053576952
Vulcanisaeta distributa DSM 14429	NO	NO	0.0161	0.6116	39,15%	0,021875945
Archaeoglobus fulgidus DSM 4304	SI	NO	0.5661	0.0233	48,01%	0,176111692
Archaeoglobus profundus DSM 5631	SI	NO	0.2722	0.003	46,96%	0,391590413
Ferroglobus placidus DSM 10642	SI/NO	NO	0.3467	0.1349	45,02%	0,449263878
Halobacterium sp. NRC-1	NO	NO	0.5591	0.015	1,37%	0,529255873
Haloarcula marismortui ATCC 43049 (II)	SI	NO	0.5181	0.0009	47,91%	1,239192243
Natronomonas pharaonis DSM 2160	SI	SI	0.2818	0.0679	29,20%	0,967804213
Haloquadratum walsbyi DSM 16790	NO	SI	0.541	0.0485	4,49%	0,202086147
Halobacterium salinarum R1	NO	NO	0.5676	0.0166	1,93%	0,502820489
Halorubrum lacusprofundi ATCC 49239 (II)	SI/NO	SI	0.6855	0.0031	39,00%	1,149556819
Halorhabdus utahensis DSM 12940	SI	NO	0.6055	0.0004	57,26%	0,84329683
Halomicrobium mukohataei DSM 12286	NO	SI	0.6093	0.0136	6,25%	1,201983658
Haloterrigena turkmenica DSM 5511	NO	SI	0.3611	0.0368	0,77%	0,901534831
Natrialba magadii ATCC 43099	SI/NO	NO	0.6217	0.0072	41,48%	0,988479572
Haloferax volcanii DS2 chromosome	NO	SI	0.5395	0.0321	1,56%	1,382052053
Halalkalicoccus jeotgali B3 chromosome	NO	NO	0.5268	0.0041	5,21%	0,816827371
Methanothermobacter thermautotrophicus str. Delta	NO	SI	0.5557	0.0455	0,35%	0,201141362
Methanosphaera stadtmanae DSM 3091	NO	SI	0.0419	0.0849	13,98%	0,986451581
Methanobrevibacter smithii ATCC 35061	SI	NO	0.00001	0.0442	0,10%	0,855450367
Methanobrevibacter ruminantium M1 chromosome	SI	NO	0.0959	0.0483	69,76%	0,779746832
Methanothermobacter marburgensis str. Marburg	NO	SI	0.6136	0.0341	0,11%	0,408362665

Tabla 2 Análisis de clustering para RSCU1 y RSCU2 en CODON W según ejes 1 y 2 respectivamente. **SI** se observó agrupamiento; **NO**, no se observó agrupamiento; **SI/NO** agrupamiento poco claro. **GC/RSCU1** y **CG /RSCU2** valor R^2 del contenido GC en comparación con RSCU en los ejes 1 y 2 respectivamente. **RSCU1|MELP** valor R^2 entre RSCU1 y el índice MELP obtenido en INCA. **S (Sharp 2005)** coeficiente de selección de uso de codones sgún Sharp 2005. MELP (MILC-based expression level predictor) es un estadístico que predice cuantitativamente el nivel de expresión de un gen .

Nombre	RSCU1	RSCU2	GC/RSCU 1	GC/RSCU 2	RSCU1|MELP	S (Sharp 2005)
Methanococcus maripaludis C5	SI	NO	0.2641	0.0005	64,38%	1,532936806
Methanococcus vannielii SB	SI	NO	0.1999	0.0008	59,37%	1,599864509
Methanococcus aeolicus Nankai-3	SI	SI/NO	0.0461	0.0481	8,61%	1,129478295
Methanocaldococcus fervens AG86	SI	NO	0.0172	0.009	1,49%	1,139794563
Methanocaldococcus vulcanius M7	SI	SI/NO	0.0339	0.0071	16,57%	0,674840363
Methanocaldococcus sp. FS406-22 chromosome	SI	SI	0.0366	0.0646	2,24%	1,051530518
Methanocaldococcus infernus ME chromosome	NO	NO	0.0048	0.0554	0,16%	0,450480258
Methanococcus voltae A3 chromosome	SI	NO	0.2332	0.0067	67,70%	1,606947968
Methanocella paludicola SANAE	NO	SI	0.7194	0.0002	24,58%	1,022148593
Uncultured methanogenic archaeon RC-I	NO	SI/NO	0.6415	0.0211	30,79%	0,83636032
Methanospirillum hungatei JF-1	NO	SI	0.7395	0.0052	51,13%	0,582917023
Methanocorpusculum labreanum Z	NO	SI	0.7177	0.0164	2,09%	0,586984196
Methanoculleus marisnigri JR1	NO	SI	0.6623	0.0047	21,16%	-0,001762373
Candidatus Methanoregula boonei 6A8	NO	SI	0.6042	0.0014	19,05%	0,531254215
Candidatus Methanosphaerula palustris E1-9c	NO	NO	0.7702	0.0002	34,44%	-0,090509599
Methanoplanus petrolearius DSM 11571	NO	SI/NO	0.762	0.0006	10,64%	0,75505809
Methanosarcina acetivorans C2A	SI/NO	NO	0.7453	0.0071	46,34%	0,702185547
Methanosarcina mazei Go1	SI	NO	0.6806	0.0069	46,46%	0,773208253
Methanosarcina barkeri str. fusaro	NO	NO	0.6584	0.0006	49,62%	0,580522636
Methanococcoides burtonii DSM 6242	NO	SI	0.6694	0.0204	13,71%	0,644330911
Methanosaeta thermophila PT	SI/NO	NO	0.4515	0.0035	29,02%	0,148784069
Methanohalophilus mahii DSM 5219	NO	SI/NO	0.5781	0.005	0,32%	0,305288017
Methanohalobium evestigatum Z-7303 chromosome	NO	NO	0.3778	0.18	0,34%	0,134373828
Methanopyrus kandleri AV19	NO	NO	0.6885	0.00001	30,80%	0,638135967
Pyrococcus abyssi GE5	NO	NO	0.5443	0.00008	37,89%	0,152273956
Pyrococcus horikoshii OT3	NO	NO	0.2803	0.1061	0,50%	-0,051336158
Pyrococcus furiosus DSM 3638	NO	SI	0.4507	0.0507	4,08%	0,404389746
Thermococcus onnurineus NA1	SI	NO	0.5956	0.0217	28,74%	0,781222347
Thermococcus gammatolerans EJ3	SI	NO	0.6221	0.0124	54,90%	0,223298551
Thermococcus sibiricus MM 739	NO	SI	0.4845	0.0172	4,53%	0,142505369
Thermoplasma acidophilum DSM 1728	NO	NO	0.6368	0.0136	1,79%	0,131473403
Thermoplasma volcanium GSS1	NO	NO	0.3943	0.0714	4,49%	0,009914785
Picrophilus torridus DSM 9790	SI/NO	NO	0.2257	0.1566	29,20%	0,434202906
Aciduliprofundum boonei T469 chromosome	NO	NO	0.3342	0.0268	17,89%	0,032124084
Candidatus Korarchaeum cryptofilum OPF8	NO	NO	0.464	0.0164	35,87%	0,299622526
Nanoarchaeum equitans Kin4-M	NO	NO	0.0005	0.0508	1,70%	0,210013613
Nitrosopumilus maritimus SCM1	SI/NO	NO	0.0015	0.0021	18,95%	0,703511044

CONTINUACIÓN. Tabla 2 Análisis de clustering para RSCU1 y RSCU2 en CODON W según ejes 1 y 2 respectivamente. **SI** se observó agrupamiento; **NO**, no se observó agrupamiento; **SI/NO** agrupamiento poco claro. **GC/RSCU1** y **GC /RSCU2** valor R^2 del contenido GC en comparación con RSCU en los ejes 1 y 2 respectivamente. **RSCU1|MELP** valor R^2 entre RSCU1 y el índice MELP obtenido en INCA. **S** (Sharp, 2005).

3) Determinación de codones óptimos y relación existente con tRNAs presentes.

	Orden	Acidilob.		Desulforo.		Sulfolo.		Thermopr.		Korarch.		Nanoarch.		Thaumar.		Thermopl.		Methanoco.	
AA	**codón/N**	1		8		5		8		1		1		1		4		8	
Ala	GCA	+100/-0	1	+12/-12	7	+40/-0	5	+25/-25	7	+0/-0	1	+0/-0	1	+100/-0	1	+0/-0	4	+37/-25	12
	GCC	+0/-100	1	+0/-12	8	+0/-60	4	+0/-50	10	+0/-0	1	+0/-0	1	+0/-100	1	+0/-25	4	+0/-100	8
	GCG	+0/-0	2	+0/-25	8	+0/-20	5	+0/-37	9	+100/-0	1	+0/-0	1	+0/-0	1	+0/-25	4	+0/-62	0
	GCU	+100/-0	0	+25/-0	0	+20/-0	0	+75/-0	0	+0/-0	0	+0/-0	0	+0/-0	0	+25/-0	0	+62/-0	0
Arg	AGA	+100/-0	1	+25/-25	8	+80/-0	5	+37/-25	5	+0/-100	1	+0/-0	1	+100/-0	1	+25/-25	4	+100/-0	8
	AGG	+0/-100	1	+37/-37	9	+0/-80	5	+25/-25	3	+100/-0	1	+0/-0	1	+0/-100	0	+50/-0	4	+0/-100	0
	CGA	+0/-100	1	+0/-37	8	+0/-20	5	+12/-12	6	+0/-0	1	+0/-100	1	+0/-100	1	+0/-50	2	+0/-87	9
	CGC	+0/-0	1	+37/-12	9	+0/-20	5	+0/-25	8	+0/-0	1	+0/-0	1	+0/-0	1	+50/-25	4	+0/-25	8
	CGG	+0/-0	1	+0/-75	8	+0/-40	5	+0/-75	5	+0/-100	1	+0/-0	0	+0/-0	0	+0/-50	4	+0/-62	0
	CGU	+100/-0	0	+37/-0	0	+0/-20	0	+37/-12	0	+0/-100	0	+0/-0	0	+0/-0	0	+25/-0	0	+0/-75	0
Asn	AAC	+0/-100	1	+50/-12	8	+0/-80	5	+0/-12	3	+100/-0	1	+0/-0	1	+100/-0	1	+25/-0	5	+100/-0	8
	AAU	+100/-0	0	+12/-50	0	+80/-0	0	+12/-0	0	+0/-100	0	+0/-0	0	+0/-100	0	+0/-25	0	+0/-100	0
Asp	GAC	+0/-100	1	+37/-12	7	+0/-100	4	+0/-12	6	+0/-0	1	+0/-0	1	+100/-0	0	+25/-0	4	+87/-0	12
	GAU	+100/-0	0	+12/-37	0	+100/-0	0	+12/-0	0	+0/-0	0	+0/-0	1	+0/-100	0	+0/-25	0	+0/-87	0
Cys	UGC	+0/-0	1	+0/-0	7	+0/-0	2	+0/-12	4	+0/-0	1	+0/-0	1	+0/-0	1	+0/-0	4	+25/-12	8
	UGU	+0/-0	0	+0/-0	0	+0/-0	0	+12/-0	0	+0/-0	0	+0/-0	0	+0/-0	0	+0/-0	0	+12/-25	0
Gln	CAA	+0/-0	1	+12/-37	8	+20/-0	5	+0/-37	5	+0/-100	1	+0/-0	0	+100/-0	1	+0/-75	4	+25/-0	8
	CAG	+0/-0	1	+37/-12	9	+0/-20	5	+37/-0	6	+100/-0	1	+0/-0	1	+0/-100	1	+75/-0	4	+0/-25	0
Glu	GAA	+100/-0	1	+12/-50	8	+60/-0	3	+25/-50	2	+0/-100	1	+0/-0	0	+0/-0	0	+0/-50	4	+75/-0	16
	GAG	+0/-100	1	+50/-12	8	+0/-60	5	+50/-25	5	+100/-0	1	+0/-0	1	+0/-0	1	+50/-0	4	+0/-75	0
Gly	GGA	+100/-0	1	+25/-25	8	+20/-0	5	+12/-0	4	+0/-0	1	+0/-0	1	+0/-100	1	+0/-25	4	+50/-25	9
	GGC	+0/-100	1	+37/-12	8	+0/-0	5	+12/-25	6	+0/-0	1	+0/-0	1	+0/-0	1	+0/-50	4	+0/-62	8
	GGG	+0/-100	1	+0/-87	9	+0/-100	5	+0/-87	8	+0/-100	1	+0/-100	1	+0/-100	1	+0/-75	4	+0/-87	0
	GGU	+100/-0	0	+50/-0	0	+80/-0	0	+87/-0	0	+100/-0	0	+100/-0	0	+100/-0	0	+75/-0	0	+75/-0	0
His	CAC	+0/-0	1	+25/-0	8	+0/-60	5	+0/-25	7	+0/-0	1	+0/-0	1	+100/-0	1	+25/-0	4	+100/-0	8
	CAU	+0/-0	0	+0/-25	0	+60/-0	0	+25/-0	0	+0/-0	0	+0/-0	0	+0/-100	0	+0/-25	0	+0/-100	0
Ile	AUA	+100/-0	0	+37/-0	0	+80/-0	0	+12/-0	0	+100/-0	1	+0/-0	1	+0/-100	0	+50/-0	0	+0/-75	0
	AUC	+0/-100	1	+0/-25	8	+0/-80	5	+0/-62	4	+0/-100	1	+0/-0	1	+100/-0	1	+0/-25	4	+87/-0	8
	AUU	+0/-0	0	+0/-25	0	+0/-0	0	+25/-0	0	+0/-100	0	+0/-0	0	+100/-0	0	+0/-50	0	+0/-12	0
Leu	CUA	+100/-0	1	+25/-25	8	+0/-20	5	+25/-0	10	+0/-0	1	+0/-0	1	+0/-0	1	+0/-0	4	+0/-100	8
	CUC	+0/-100	1	+0/-50	8	+0/-100	3	+0/-50	7	+0/-0	1	+0/-0	1	+0/-0	1	+25/-0	4	+50/-12	8
	CUG	+0/-100	1	+25/-12	8	+0/-80	5	+0/-25	7	+0/-0	1	+0/-0	0	+0/-100	1	+50/-0	4	+0/-50	0
	CUU	+100/-0	0	+0/-0	0	+0/-20	0	+75/-0	0	+0/-0	0	+0/-0	0	+0/-0	0	+0/-0	0	+0/-87	0
	UUA	+100/-0	1	+12/-37	8	+100/-0	5	+25/-12	8	+0/-100	1	+100/-0	1	+0/-0	0	+0/-75	4	+87/-0	8
	UUG	+0/-0	0	+25/-25	8	+20/-0	5	+12/-0	8	+0/-0	1	+0/-100	1	+0/-0	1	+0/-0	4	+25/-0	0
Lys	AAA	+0/-100	1	+0/-62	8	+40/-0	5	+12/-50	4	+0/-100	1	+0/-0	2	+0/-100	1	+0/-100	4	+12/-25	12
	AAG	+100/-0	1	+62/-0	7	+0/-40	5	+50/-12	6	+100/-0	1	+0/-0	0	+100/-0	2	+100/-0	4	+25/-12	0
Phe	UUC	+0/-100	1	+25/-12	8	+0/-20	4	+0/-37	7	+0/-0	1	+100/-0	1	+100/-0	1	+25/-0	4	+100/-0	8
	UUU	+100/-0	0	+12/-25	0	+20/-0	0	+37/-0	1	+0/-0	0	+0/-100	0	+0/-100	0	+0/-25	0	+0/-100	0
Pro	CCA	+100/-0	1	+12/-0	6	+80/-0	5	+25/-0	3	+0/-0	1	+100/-0	1	+100/-0	1	+0/-0	3	+50/-0	8
	CCC	+0/-100	1	+0/-37	8	+0/-80	5	+0/-87	10	+0/-0	1	+0/-0	1	+0/-100	1	+0/-0	4	+0/-100	7
	CCG	+0/-100	0	+25/-0	3	+0/-20	5	+25/-0	2	+0/-0	1	+0/-100	1	+0/-0	0	+0/-0	4	+0/-75	0
	CCU	+100/-0	0	+12/-12	0	+0/-0	0	+62/-0	0	+0/-0	0	+0/-100	0	+0/-100	0	+0/-0	0	+50/-25	0
Ser	AGC	+0/-0	1	+75/-12	8	+0/-0	5	+0/-0	9	+0/-0	1	+0/-0	1	+0/-0	1	+25/-0	4	+37/-0	8
	AGU	+100/-0	0	+25/-12	0	+40/-0	0	+12/-0	0	+100/-0	0	+0/-100	0	+0/-0	0	+75/-0	0	+0/-25	0
	UCA	+0/-0	1	+0/-0	8	+0/-0	5	+12/-12	5	+0/-0	1	+0/-0	1	+0/-0	1	+0/-0	4	+87/-0	8
	UCC	+0/-100	1	+0/-37	8	+0/-80	5	+0/-50	9	+0/-0	1	+0/-0	1	+0/-0	1	+0/-0	4	+0/-12	8
	UCG	+0/-100	1	+0/-37	8	+0/-0	5	+12/-12	8	+0/-0	0	+0/-100	2	+0/-0	1	+0/-25	4	+0/-75	0
	UCU	+100/-0	0	+0/-50	0	+20/-0	0	+37/-0	0	+0/-0	0	+100/-0	0	+0/-100	0	+0/-25	0	+0/-100	0
Thr	ACA	+100/-0	0	+37/-12	9	+40/-0	5	+12/-25	2	+0/-0	1	+0/-0	1	+0/-0	1	+0/-25	4	+100/-0	8
	ACC	+0/-100	1	+12/-0	8	+0/-80	5	+0/-50	8	+0/-0	1	+0/-0	1	+0/-0	1	+0/-0	4	+12/-25	8
	ACG	+0/-100	1	+12/-12	8	+0/-20	5	+0/-12	4	+0/-0	1	+0/-100	1	+0/-0	1	+25/-25	4	+0/-87	1
	ACU	+100/-0	1	+0/-12	0	+60/-0	0	+50/-0	0	+0/-0	0	+0/-0	0	+0/-0	0	+25/-0	0	+0/-100	0
Trp	UGG	+0/-100	0	+0/-62	4	+0/-100	5	+0/-50	7	+0/-100	1	+0/-0	0	+0/-0	0	+0/-100	3	+0/-0	3
Tyr	UAC	+0/-100	1	+0/-0	8	+0/-100	4	+0/-50	7	+0/-0	1	+0/-0	1	+100/-0	1	+25/-0	4	+100/-0	8
	UAU	+100/-0	1	+0/-0	0	+100/-0	0	+50/-0	0	+0/-0	0	+0/-0	0	+0/-100	0	+0/-25	0	+0/-100	0
Val	GUA	+100/-0	0	+62/-12	8	+60/-0	5	+25/-25	7	+0/-0	1	+100/-0	1	+0/-0	1	+0/-0	4	+0/-0	9
	GUC	+0/-100	1	+0/-25	8	+0/-60	5	+0/-37	2	+0/-0	1	+0/-0	1	+0/-0	1	+0/-0	4	+25/-37	8
	GUG	+0/-100	1	+0/-25	10	+0/-40	5	+0/-37	8	+0/-0	1	+0/-100	1	+0/-100	1	+0/-0	4	+0/-100	4
	GUU	+0/-0	1	+12/-0	0	+40/-20	0	+87/-0	0	+0/-0	0	+0/-0	0	+0/-0	0	+25/-0	0	+62/-0	0

Tabla 3 Distribución de codones y número de tRNAs isoaceptores correspondientes a cada codón, por cada Orden filogenético. **N** número de especies pertenecientes al grupo filogenético. **AA** aminoácido. Se indica para cada grupo el promedio simple en forma de porcentaje de uso preferencial de codones sinónimos (**+**) o de exclusión diferencial de codones sinónimos (**-**) para cada codón por AA de cada lado de la barra (*/*) mediante cálculo por tabla de contingencia χ^2 con p<0.01. (+) para +x, +xx y +xxx. (-) para –x, -xx y –xxx.

	Orden	Methanoba		Methanosa.		M-cellales		Methanomi.		Halobacte.		Archaeog.		Thermoco.		Methanop.	
AA	codón/N	5		7		2		6		11		3		6		1	
Ala	GCA	+40/-20	#	+86/-0	13	+100/-0	4	+100/-0	12	+54/-0	16	+0/-33	3	+0/-33	6	+0/-0	1
	GCC	+0/-100	5	+14/-43	7	+0/-100	2	+0/-67	6	+73/-0	11	+0/-0	3	+33/-50	6	+100/-0	1
	GCG	+0/-40	0	+0/-86	7	+0/-100	2	+0/-100	7	+0/-100	13	+33/-0	3	+0/-50	6	+0/-100	0
	GCU	+100/-0	0	+0/-57	0	+0/-0	0	+0/-17	0	+27/-0	0	+0/-0	0	+67/-0	0	+0/-0	0
Arg	AGA	+100/-0	6	+43/-0	7	+50/-0	2	+50/-17	6	+0/-73	11	+33/-33	3	+50/-17	6	+0/-0	1
	AGG	+0/-60	5	+29/-0	6	+100/-0	2	+0/-50	6	+0/-100	11	+33/-33	3	+33/-0	6	+0/-0	0
	CGA	+20/-40	5	+0/-71	6	+0/-100	2	+0/-67	6	+45/-0	12	+0/-67	4	+0/-100	6	+0/-100	1
	CGC	+0/-40	5	+43/-14	7	+50/-0	2	+100/-0	6	+100/-0	11	+0/-67	3	+0/-33	6	+100/-0	1
	CGG	+0/-80	0	+0/-71	7	+0/-100	2	+0/-100	6	+0/-100	8	+0/-67	3	+0/-67	6	+0/-100	0
	CGU	+20/-20	1	+43/-0	0	+50/-0	0	+100/-0	0	+73/-0	0	+33/-67	0	+0/-17	0	+100/-0	0
Asn	AAC	+100/-0	7	+71/-0	10	+100/-0	2	+67/-0	6	+73/-0	11	+67/-0	3	+50/-0	6	+100/-0	1
	AAU	+0/-100	0	+0/-71	0	+0/-100	0	+0/-67	0	+0/-73	0	+0/-67	0	+0/-50	0	+0/-100	0
Asp	GAC	+100/-0	6	+57/-0	7	+100/-0	2	+67/-0	9	+82/-0	16	+67/-33	3	+67/-0	6	+100/-0	1
	GAU	+0/-100	0	+0/-57	0	+0/-100	0	+0/-67	1	+0/-82	0	+33/-67	0	+0/-67	0	+0/-100	0
Cys	UGC	+0/-20	5	+43/-0	21	+0/-0	3	+33/-0	20	+0/-0	11	+33/-0	3	+17/-0	6	+0/-0	0
	UGU	+20/-0	0	+0/-43	0	+0/-0	0	+0/-33	0	+0/-0	0	+0/-33	0	+0/-17	0	+0/-0	0
Gln	CAA	+60/-0	7	+0/-72	8	+0/-100	2	+0/-100	6	+0/-73	10	+33/-67	3	+17/-67	6	+0/-100	1
	CAG	+0/-60	4	+71/-0	7	+100/-0	2	+100/-0	6	+73/-0	11	+67/-33	4	+67/-17	6	+100/-0	0
Glu	GAA	+100/-0	7	+29/-14	10	+0/-0	2	+0/-17	6	+27/-0	11	+33/-33	3	+0/-67	6	+0/-100	1
	GAG	+0/-100	0	+14/-29	7	+0/-0	2	+17/-0	6	+0/-27	11	+33/-33	3	+67/-0	6	+100/-0	0
Gly	GGA	+40/-40	7	+0/-57	7	+50/-50	2	+33/-17	6	+0/-36	12	+67/-0	3	+0/-17	6	+0/-0	0
	GGC	+0/-40	5	+43/-0	13	+100/-0	3	+50/-0	7	+45/-9	14	+0/-67	3	+17/-33	6	+0/-0	1
	GGG	+0/-80	0	+0/-71	7	+0/-100	2	+0/-100	6	+0/-100	10	+0/-67	3	+0/-100	6	+0/-100	0
	GGU	+60/-0	0	+71/-0	0	+50/-0	0	+100/-0	0	+100/-0	0	+33/-0	0	+100/-0	0	+100/-0	0
His	CAC	+100/-0	6	+86/-0	7	+100/-0	2	+67/-0	6	+64/-0	11	+67/-0	3	+83/-0	6	+100/-0	1
	CAU	+0/-100	0	+0/-86	0	+0/-100	0	+0/-67	0	+0/-64	0	+0/-67	0	+0/-83	0	+0/-100	0
Ile	AUA	+0/-60	0	+14/-86	0	+0/-100	0	+0/-100	0	+0/-73	0	+67/-0	0	+0/-67	0	+0/-100	0
	AUC	+60/-0	6	+71/-14	10	+100/-0	2	+67/-0	6	+73/-0	11	+33/-0	4	+50/-0	6	+100/-0	1
	AUU	+60/-0	0	+14/-0	0	+0/-50	0	+33/-0	0	+0/-54	0	+0/-67	0	+17/-0	0	+0/-0	0
Leu	CUA	+0/-60	5	+0/-71	7	+0/-50	2	+0/-50	7	+0/-36	11	+33/-0	3	+0/-33	6	+100/-0	1
	CUC	+40/-0	5	+71/-14	11	+100/-0	2	+50/-33	6	+82/-0	11	+33/-0	4	+67/-17	5	+0/-0	1
	CUG	+0/-60	0	+0/-29	7	+0/-100	2	+33/-33	6	+0/-18	11	+0/-0	3	+0/-83	5	+100/-0	0
	CUU	+0/-0	0	+71/-0	0	+0/-50	0	+67/-0	0	+18/-9	0	+0/-0	0	+67/-0	0	+0/-100	0
	UUA	+60/-0	7	+0/-71	8	+50/-0	2	+17/-33	6	+0/-27	11	+0/-67	3	+0/-33	6	+0/-100	1
	UUG	+0/-60	1	+0/-43	7	+0/-100	2	+0/-67	7	+0/-100	11	+33/-33	3	+17/-17	5	+0/-0	0
Lys	AAA	+60/-20	7	+0/-100	7	+0/-100	2	+0/-100	6	+0/-82	12	+33/-67	3	+0/-100	6	+0/-100	1
	AAG	+20/-60	1	+100/-0	7	+100/-0	2	+100/-0	8	+82/-0	11	+67/-33	3	+100/-0	5	+100/-0	0
Phe	UUC	+60/-0	7	+86/-0	10	+100/-0	2	+67/-0	6	+100/-0	13	+67/-0	3	+33/-0	6	+100/-0	1
	UUU	+0/-60	0	+0/-86	0	+0/-100	0	+0/-67	0	+0/-100	0	+0/-67	0	+0/-33	0	+0/-100	0
Pro	CCA	+60/-0	6	+43/-14	7	+0/-100	2	+33/-0	6	+18/-18	11	+33/-0	3	+67/-17	5	+0/-0	1
	CCC	+0/-100	0	+0/-29	7	+100/-0	2	+33/-0	6	+36/-18	11	+33/-33	3	+0/-67	5	+0/-100	0
	CCG	+0/-60	0	+43/-29	7	+50/-0	2	+17/-17	6	+0/-36	10	+0/-33	3	+17/-50	6	+100/-0	0
	CCU	+60/-0	1	+0/-29	0	+0/-100	0	+17/-50	0	+18/-9	0	+0/-0	0	+0/-33	0	+0/-100	0
Ser	AGC	+60/-20	7	+29/-0	7	+0/-0	2	+100/-0	6	+18/-18	11	+33/-33	3	+50/-0	6	+100/-0	1
	AGU	+20/-0	0	+0/-14	0	+0/-100	0	+0/-17	0	+9/-9	0	+0/-0	0	+17/-0	0	+0/-0	0
	UCA	+40/-40	7	+0/-0	7	+0/-50	2	+33/-17	6	+0/-73	11	+0/-0	3	+17/-0	6	+0/-0	1
	UCC	+20/-0	9	+57/-0	7	+100/-0	2	+17/-0	6	+82/-9	11	+0/-0	3	+0/-50	6	+0/-0	1
	UCG	+0/-40	0	+0/-86	3	+0/-0	2	+0/-83	6	+0/-36	11	+0/-33	3	+0/-50	6	+0/-100	0
	UCU	+60/-40	0	+0/-43	0	+0/-0	0	+0/-33	0	+9/-9	0	+0/-0	0	+0/-50	0	+0/-0	0
Thr	ACA	+40/-60	5	+14/-14	8	+0/-100	4	+33/-50	10	+0/-45	7	+0/-33	4	+33/-33	6	+0/-0	1
	ACC	+60/-20	5	+57/-0	7	+100/-0	2	+50/-0	6	+45/-0	10	+0/-0	3	+50/-50	6	+0/-0	1
	ACG	+0/-60	4	+0/-71	8	+0/-100	2	+0/-67	6	+9/-9	10	+33/-0	3	+0/-67	6	+0/-0	0
	ACU	+60/-0	0	+0/-29	0	+0/-50	0	+33/-17	0	+0/-27	0	+0/-0	0	+17/-0	0	+0/-0	0
Trp	UGG	+0/-0	5	+0/-100	4	+0/-100	1	+0/-83	2	+0/-91	9	+0/-100	3	+0/-100	5	+0/-0	0
Tyr	UAC	+100/-0	6	+86/-0	7	+100/-0	1	+67/-0	6	+82/-0	12	+33/-0	3	+50/-0	6	+100/-0	1
	UAU	+0/-100	0	+0/-86	0	+0/-100	0	+0/-67	0	+0/-82	0	+0/-33	0	+0/-50	0	+0/-100	0
Val	GUA	+40/-0	6	+29/-0	7	+50/-0	2	+67/-17	6	+27/-0	11	+0/-0	3	+0/-50	6	+0/-0	1
	GUC	+0/-20	6	+57/-0	10	+100/-0	2	+33/-17	6	+91/-0	12	+33/-0	3	+50/-0	6	+0/-0	1
	GUG	+0/-80	2	+0/-57	7	+0/-100	2	+0/-83	6	+0/-91	11	+67/-0	3	+0/-50	6	+0/-0	0
	GUU	+40/-20	0	+0/-14	1	+0/-50	0	+33/-0	0	+18/-9	1	+0/-33	0	+17/-0	0	+0/-0	0

Tabla 3. CONTINUACIÓN. Acidilob.: Acidilobales; Desulfuro.:Desulfurococcales; Sulfulo.: Sulfolobales; Thermopr.: Thermoproteales; Korarch.: Korarchaeota; Nanoarch.: Nanoarchaeota; Thaumar.: Thaumarchaeota; Thermopl.: Thermoplasmatales; Methanoco.: Methanococcales; Methanoba.: Methanobacteriales; Methanosa.: Methanosarcinales; M-celalles.: Methanocellales; Methanomi.: Methanomicrobiales; Halobacte.: Halobacteriales; Archaeog.: Archaeoglobales; Thermoco.: Thermococcales; Methanop.: Methanopyrales.

Mediante el programa RNAt-scan-SE (Lowe & Eddy, 1997) se determinaron los ARNt de cada genoma y se analizó la relación entre las abundancias de ARNt y el UCS. Se determinó mediante test de χ^2 cuáles eran los codones significativamente más utilizados por los genes de alta expresión (codones óptimos) para cada especie (+, en tabla 3), así como también qué codones eran significativamente menos usados (-) en los genes de alta expresión (ver tabla 3). Se encontró que en la mayoría de los grupos hay más codones evitados que aquellos que son seleccionados en forma positiva, la relación entre los codones preferidos referidos como "+" en la tabla 3 y los codones evitados "-" es siempre menor a 1, si dividimos el número total de codones preferidos en el grupo entre el número total de codones evitados (datos suplementarios), excepto para el caso de los Thermoplasmatales que es de 1.07. En el caso de *Acidilobus saccharovorans* la relación es de 1 pero es el único representante analizado para el grupo de los Acidilobales.

El uso de codones óptimos universales (UUC, AUC, UAC y AAC) muestra que en todos los organismos se asoció por lo menos un tRNA isoaceptor. Los codones óptimos universales se encuentran usados en forma preferencial con valores promedio por grupo mayores al 30% en los Methanococcales, Methanobacteriales, Methanosarcinales, Methanocellales, Methanomicrobiales, Halobacteriales, Archaeogloblaes, Thermococcales, Methanopyrales y en el *Phylum* Thaumarchaeota. En otras palabras, en más de el 30% de los organismos de los grupos mencionados se usaron codones sinónimos en forma significativa mayor al resto de los codones sinónimos para ese AA. Los Thermoplasmatales muestran valores del 25% de preferencia de uso para tres de los codones óptimos universales, el triplete AUC no tiene preferencia de UCS estadísticamente significativa. En el caso de los Desulfurococcales, Sulfolobales, Thermoproteales y Acidilobales (tabla 3) se encontró, por el contrario, una tendencia a evitar estos codones universales o bien a no usarlos en forma preferencial. Estos grupos pertenecen al *Phylum* Crenarchaeota y están cercanos filogenéticamente (figura 1). Tampoco hay un uso preferencial de los codones óptimos universales en los casos de *Korarchaeum cryptofilum* o en *Nanoarchaeum equitans*.

La distribución de tRNAs es bastante uniforme a lo largo de todos los individuos, presentando un valor de tRNAs medio de 41, SD ±6. El valor máximo lo presenta *Methanosarcina barkeri* perteneciente a los Methanosarcinales con 56 tRNAs y el valor mínimo lo tiene *Pyrobaculum calidifontis* del orden de los Thermoproteales con 23 tRNAs.

4) Cálculo de coeficientes de selección S.

Se calcularon los coeficientes de selección (S) para cuantificar el grado en el cual la selección ha moldeado el uso de codones sinónimos en genes de alta expresión y se comparó en los distintos grupos filogenéticos (tabla 2 y figura 2). Se estimaron contraponiendo el patrón de UCS en genes codificantes para proteínas ribosomales y los genomas completos de los organismos para los aminoácidos Phe, Tyr, Ile y Asn de acuerdo a lo sugerido por Sharp *et al*. (2005).

En el *Phylum* Crenarchaeota no se observan valores de S mayores a 0.5, (0 -ausencia de selección- y un máximo de 0.46). La tendencia general es de presentar valores de S que sugieren una ausencia de selección (figura 2). Solamente el orden de los Desulfurococcales tiene valores diversos de fuerza de selección para UCS dentro del grupo. Aquí observamos valores que van desde los -0.67 a 0.46 (mediana: 0,11; DS ±0,43). En concordancia al grupo, *A. saccharovorans* muestra un valor de S de -0.44.

Los organismos pertenecientes a Euryarchaeota, así como *Candidatus korarchaeum cryptofilum OPF8*, *Nanorachaeum equitans KIN4-M* y *Nitrosopumilus maritimus SCM1*, como se observa en las figuras 1 y 2, presentan valores de fuerza de selección positivos que van desde 0 hasta S=1.61 en *Methanococcus voltae A3 chromosome* y de S=1.60 en *Methanococcus vannielii SB*. Ambas especies pertenecen al orden de los Methanococcales. Los Halobacteriales también se destacan por tener valores de S altos. En vista de la diversidad encontrada en Euryarchaeota, se clasificaron los grupos como fuertemente seleccionados (FS) cuando S es mayor a 0.5, levemente seleccionados (LS) cuando el coeficiente de selección se encuentra entre los valores de 0.20 y 0.50 y con nula o baja selección cuando este valor es menor a 0.20 (BS), siguiendo la sugerencia de Sharp *et al*. (2005).

- Grupos filogenéticos BS.

Los Thermoplasmatales presentan valores en promedio de 0.19 (med 0.13; DS ±0.22).

- Grupos filogenéticos LS.

Los grupos filogenéticos con valores promedio de coeficiente de selección de Sharp intermedios son los Archaeoglobales, con un valor promedio para el grupo de S=0.34 (med 0.39; DS ±0.14), los Methanosarcinales con valor promedio para el grupo de S=0.47 (med 0.58, DS ±0.27), los Thermococcales con valor promedio para el grupo de S=0.27 (med 0.19; DS ±0.29) y los Methanomicrobiales con valor promedio para el grupo de S=0.39 (med 0.56; DS ±0.35). En este último grupo encontramos a *Candidatus Methanosphaerula palustris* y a *Methanoculleus marisnigri* JR1 que se destacan por tener un comportamiento muy distinto a otros organismos del mismo grupo, sus valores de S son de -0.09 y de 0, respectivamente, mostrando ausencia de selección.

- Grupos filogenéticos con S altos mayores a 0.50.

En este conjunto se encuentra a los Methanococcales con valor promedio para el grupo de S=1.13 (med 1.13; DS ±0.40), los Methanobacteriales con valor promedio para el grupo de S=0.65 (med 0.78; DS ±0.33), los Halobacteriales con valor promedio para el grupo de S=0.89 (med 0.93; DS ±0.34) y los Methanocellales con valor promedio para el grupo de S=0.93 (med 0.93, DS ±0.13).

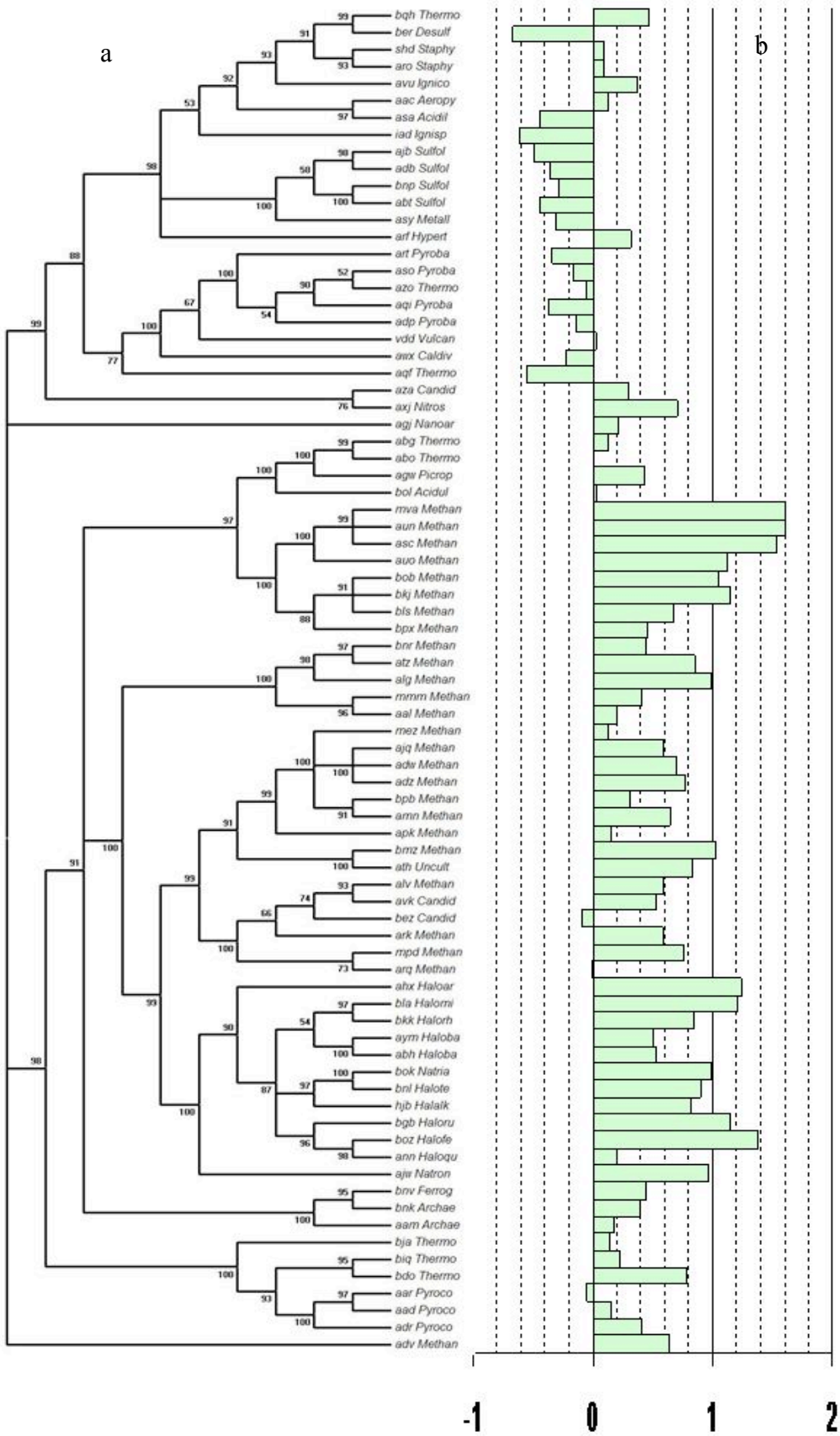

Figura 2. Filogenia de Archaea en base a rRNA 16S en (a) PHYML 3.0 tomando como modelo de sustitución nucleotídica HKY85 (b) gráfico de coeficiente de selección según Sharp (2005) S, donde S = ln [(P x k) / (1-P)]

DISCUSIÓN

En este trabajo nos planteamos estudiar el UCS en Archaea, evaluando particularmente si hay selección para el uso de codones sinónimos y como es el patrón filogenético de esta característica. Para definir selección traduccional en el UCS deberían ser identificables cuatro condiciones. En primer lugar debe ser observable un agrupamiento de los genes de alta expresión (genes codificantes para proteínas ribosomales y factores de elongación) en uno de los extremos de alguno de los ejes principales generados en el COA (análisis multivariado, ver materiales y métodos y material suplementario). Esto se evidencia mediante inspección visual y se estima mediante la correlación entre los índices de expresión basados en el UCS de los genes (ej. MELP) y el valor en los ejes generados por el COA (Musto *et al.,* 2003; Sharp & Li, 1987; Iriarte *et al.*, 2011). Los valores de correlación significativos sugieren que la mayor parte de la variabilidad observada en el UCS podría ser explicada por la selección para la traducción mediante un uso de codones específico asociada a la expresión en los genes.

La segunda condición hace referencia a que los codones óptimos universales UUC, AUC, UAC y AAC aparezcan como significativamente más utilizados en genes de alta expresión. La selección para una traducción eficiente está relacionada con la concentración de tRNAs disponibles en la célula y de esta forma se establece la tercera condición y es que exista una distribución diferencial de concentración de tRNAs para los distintos codones sinónimos. Por último; cuando se calcula la fuerza de selección S debe ser mayor a cero.

Siguiendo estos criterios cuando estudiamos el comportamiento del UCS en Archaea identificamos dos grupos principales. El primer grupo no presenta selección traduccional para el UCS y está conformado por los *Phyla* Crenarchaeota, Korarchaeota y Nanoarchaeota. El segundo grupo en contraposición, cumple con al menos tres de las cuatro condiciones impuestas para determinar selección traduccional para el UCS. En este último encontramos a los *Phyla* Euryarchaeota y Thaumarchaeota.

Un aspecto que debemos tener en cuenta es si los patrones observados no son producto de sesgos mutacionales. Como se observa en la figura 3 hay una independencia entre el contenido GC genómico y el uso relativo de codones sinónimos calculado por CODONW (figura 3a) y esta independencia también se observa cuando se compara el contenido en GC3. Tampoco se observa una relación entre la fuerza de selección de Sharp y el contenido en GC genómico o el de las terceras posiciones sinónimas (figura 3b). Esta independencia es aún más evidente para los organismos que presentaron S mayores a 0.5 (Figura 3b, triángulos rojos).

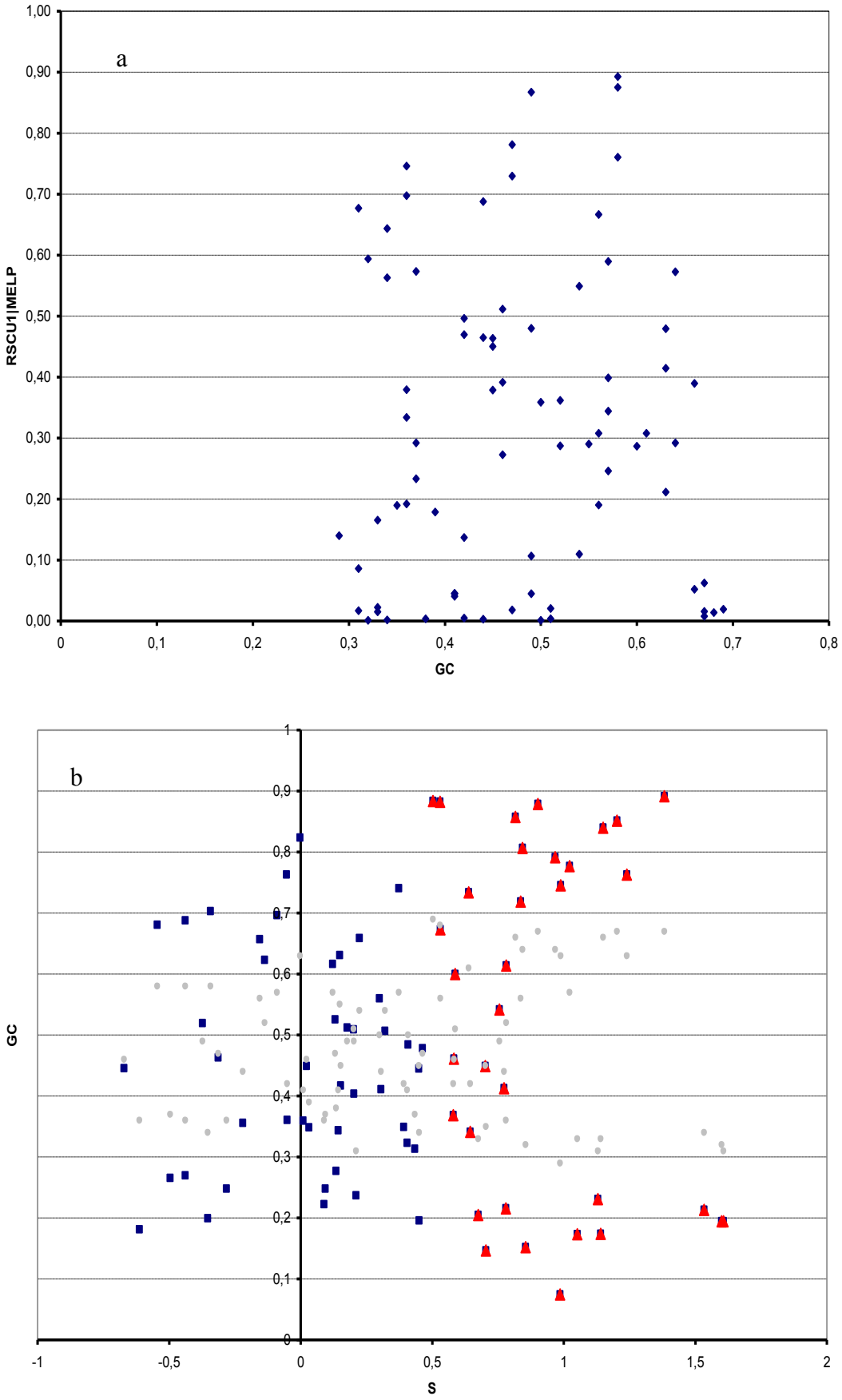

Figura 3.**a)** Independencia del contenido **GC genómico** con respecto a al valor R^2 entre **RSCU1** y el parámetro **MELP b)** Independencia de el contenido GC con respecto de la fuerza de selección de Sharp **S.** Puntos grises indican contenido **GC genómico**, cuadrados azules indican contenido en **GC3** y los triángulos rojos hacen referencia al contenido **GC3** para los organismos con S>0.50.

Como se definió anteriormente los codones óptimos son aquellos que encontramos significativamente con mayor frecuencia de uso en los genes de alta expresión (tabla 3).

En el *Phylum* Crenarchaeota, en todos los grupos analizados se encontró por lo menos algún codón que aparece con mayor frecuencia en los genes de alta expresión. Ninguno de ellos se corresponde a los codones óptimos universales (tabla 3). En este grupo los codones preferidos son: GCA, GCU, AGA, CGU, AAU, GAU, GAA, GGA, GGU, AUA, UUA, CUA, UUG, AAG, UUU, CCA, CCU, AGU, UCU, ACA, ACU, AUA, GUA y GUU. La tendencia es que los codones terminen en U o en A.

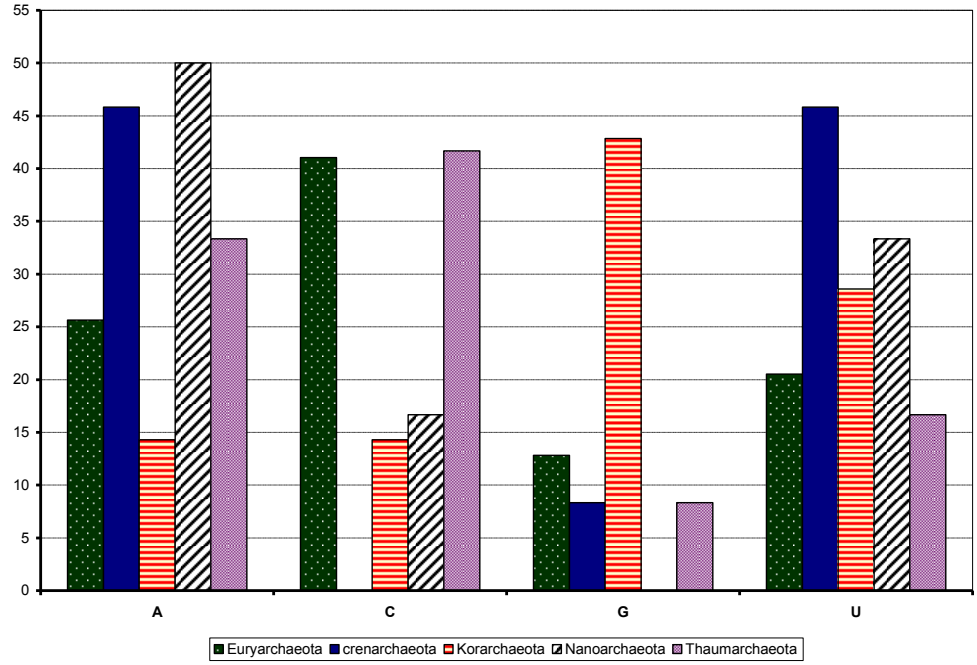

Figura 4. Frecuencias relativas de preferencia de uso de los nucleótidos **A C G U** en la tercera posición para los codones óptimos para cada uno de los Phyllum Euryarchaeota, Crenarchaeota, Korarchaeota, Nanoarchaeota y Thaumarchaeota.

Para los organismos del *Phylum* Euryarchaeota se observó una mayor diversidad de preferencia de UCS (tabla 3 y figura 4). Los codones que aparecen como preferidos en por lo menos el 40% de los casos son: GCU, GCC, GCA, CGU, CGC, AGG, AGA, AAC, GAC, UGC, CAA, CAG, GAA, GAG, GGA, GGC, GGU, CAC, AUC, CUC, CUU, UUA, AAG, UUC, CCA, CCC, CCG, CCU, AGC, AGU, UCA, UCC, ACA, ACC, ACU, UAC, GUA, GUC y GUU. Hay un uso preferencial de los codones óptimos universales en todos los órdenes, pero se constató una excepción en los Thermoplasmatales; donde aparece con mayor frecuencia el codón AUA para Ile en lugar de AUC. Se usaron preferentemente codones terminados en C o en A.

El estudio de UCS en *N. equitans* mostró preferencia por los codones: GGU, UUA, UUC (Codón Universal Optimo, CUO), CCA, UCU y GUA. El organismo *N. maritimus* mostró una preferencia por

GCA, AGA, AAC, GAC, CAA, GGU, CAC, AUU, AAG, UUC, CCA y UAC. En este caso se usaron los cuatro codones universales óptimos. El *Phylum* Thaumarchaeota es un grupo que en nuestra filogenia aparece como hermano de Korarchaeota, con divergencia relativamente reciente. *K. cryptofilum* del *Phylum* Korarchaeota mostró un patrón menor de UCS, con preferencia por los codones GCG, AGG, AAC, GGU, AUA, AAG y AGU; con AAC como CUO. Sería interesante profundizar en esta relación evolutiva con un mayor número de organismos representantes de ambos grupos y establecer si estos grupos muestran patrones de UCS de transición evolutiva intermedios entre los *Phyla* principales Crenarchaeota y Euryarcheaota.

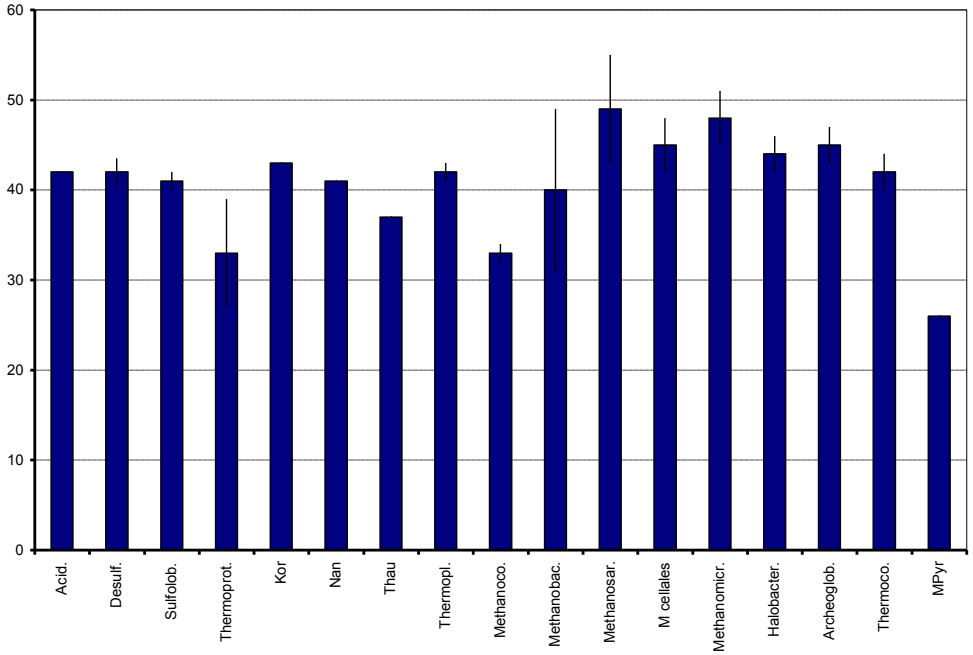

Figura 5. Distribución promedio del conteo de tRNAs por Orden Filogenético.

AA	codón	Crenarchaeota	Korarchaeota	Thaumarchaeota	Nanoarchaeota	Euryarchaeota
	GCU					
	GCG					
	GCC					x
Ala	GCA	x		x		x
	CGU					
	CGG					
	AGG					x
	CGC					x
	CGA					
Arg	AGA	x		x		x
	AAU					
Asn	AAC			x		x
	GAU					
Asp	GAC					x
	UGU					
Cys	UGC					x
	CAG			x		
Gln	CAA					x
	GAG					x
Glu	GAA	x				x
	GGU					
	GGG					
	GGC					x
Gly	GGA	x				x
	CAU					
His	CAC			x		x
	AUU					
	AUC					x
Ile	AUA					
	CUU					
	UUG	x				
	CUG					
	CUC					x
	UUA	x			x	x
Leu	CUA	x				
	AAG	x		x		x
Lys	AAA					
	UUU					
Phe	UUC		x	x	x	x
	CCU					
	CCG					x
	CCC					x
Pro	CCA	x	x	x	x	x
	AGU					
	UCU					
	UCG					
	AGC					x
	UCC					x
Ser	UCA					x
	ACU					
	ACG					
	ACC					x
Thr	ACA	x				x
Trp	UGG					
	UAU					
Tyr	UAC		x	x		x
	GUU					
	GUG					
	GUC					x
Val	GUA	x			x	x

Figura 6. Distribución de codones ótpimos para cada aminoácido por Phyllum. Se indican los codones óptimos universales en sombreado.

La distribución de tRNAs es relativamente homogénea a lo largo de los organismos analizados. La media de tRNAs presentes por organismo es de 41 (SD ± 6), alcanzando su valor mínimo en *P. calidifontis* (Thermoproteales) con 23 tRNAs y sus valores máximos en *M. barkeri* (Methanosarcinales) y en *M. ruminantium* (Methanobacteriales) ambos con 56 tRNAs. Si analizamos las cantidades promedio por grupo filogenético (ver figura 5) encontramos un total de 42 para los Acidilobales, 42 ± 2 en los Desulfurococcales, 41 ± 1 en los Sulfolobales, 33 ± 6 en Thermoproteales, 42 ± 1 en Thermoplasmatales, 33 ± 1 para los Methanococcales, 40 ± 9 en Methanobacterias, 49 ± 6 en Methanosarcinales, 45 ± 3 para los Methanocellales, 48 ± 3 en Methanomicrobiales, 44 ± 2 en Halobacterias, 45 ± 2 para Archaeoglobales, 42 ± 2 Thermococcales, 26 para *M. kandleri,* 43 en *K. cryptofilum, N. equitans* mostró un conteo de 41 y 37 en *N. maritimus.*

Luego de analizar la distribución de tRNAs y determinar cuáles tripletes aparecen como preferidos para cada AA, evaluamos cuál de estos tripletes tienen un tRNA asociado (figura 6). Encontramos que para Crenarchaeota los codones preferidos con tRNA asociado son GCA, AGA, GAA, GGA, UUA, CUA, UUG, AAG, CCA, ACA y GUA. De la misma forma determinamos que en Euryarchaeota los codones preferidos con tRNA asociado son GCC, GCA, CGC, AGG, AGA, AAC, GAC, UGC, CAA, GAA, GAG, GGA, GGC, CAC, AUC, CUC, UUA, AAG, UUC, CCA, CCC, CCG, AGC, UCA, UCC, ACA, ACC, UAC, GUA y GUC. En este *Phylum* vemos que los codones óptimos son preferidos si terminan en C o en A. *N. equitans* muestra como codones más frecuentes a UUA, UUC, CCA y GUA, con preferencia por codones terminados en A. En representación de Thaumarchaota, *N. maritimus* tiene como codones óptimos a GCA, AGA, AAC, CAA, CAC, AAG, UUC, CCA y UAC por

lo que muestra sesgo hacia los tripletes terminados en C o en A. Por último *K. cryptofilum,* tiene como codones preferidos a GCG, AGG, AAC, AUA y AAG; tripletes terminados en G preferentemente. Si se observa la figura 4 se puede ver que se mantuvieron los patrones de preferencia de elección de los nucleótidos en la posición de balanceo que se obtuvieron mediante estudio por tabla de χ^2 de contingencia.

Cuando se analizó la fuerza de selección para UCS según Sharp (2005), se mantuvo el mismo patrón diferenciado entre Euryarchaeota y Crenarcheota (ver figura 2), donde se distingue una ausencia de fuerza de selección para UCS para Crenarchaeota y una diversidad de patrones con coeficientes mayores a cero para los *Phylum* Euryarchaeota, Thaumarchaeota, Nanoarchaeota y Korarchaeota.

Si se relaciona la distribución de los coeficientes S con el número de tRNAs por cada especie se observa una correlación despreciable (R^2 0.005) tomando el total de organismos analizados. Sin embargo, si nos enfocamos en el grupo de organismos con S>0.50 la correlación aumenta a R^2 0.148 (figura 7).

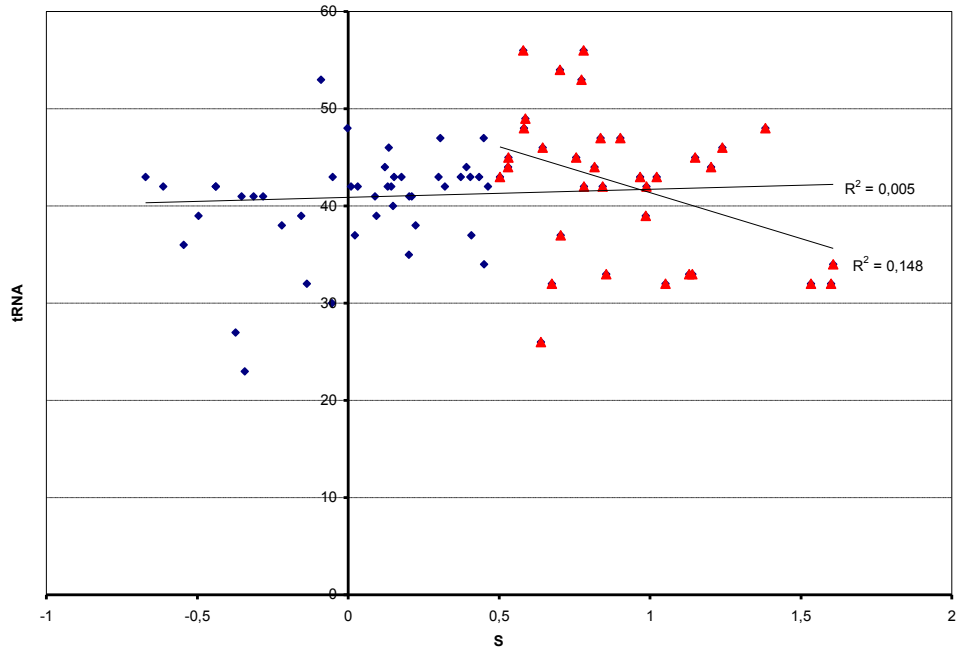

Figura 7. Relación entre la fuerza de Selección de Sharp **S** y el número de tRNAs detectados por tRNA-scan. Los puntos azules muestran la relación entre todos los organismos y los triángulos rojos muestran la relación para aquellos organismos cuyo S>0.50.

Comparación entre grupos filogenéticos.

De acuerdo a las cuatro condiciones fijadas anteriormente se puede afirmar que en el *Phylum* de Crenarchaeota conformado por los Ordenes Acidilobales, Desulfurococcales, Sulfolobales y Thermoproteales no hay selección traduccional en el UCS, porque no presenta clustering diferencial para los genes de alta expresión y no hay un uso incrementado de los codones óptimos universales. Sin embargo se encontraron codones usados en forma preferencial con tRNAs asociados que reconocen mejor los tripletes terminados en A. Cuando analizamos el contenido en GC de estos organismos (Tabla 1) podemos ver que en promedio está sesgado hacia AT, por lo que los patrones de uso de codones están debidos a sesgos mutacionales propios de los grupos.

El *Phylum* de Euryarchaeota mostró una mayor diversidad de patrones de sesgo de uso de codones sinónimos.

El organismo *Aciduliprofundum boonei* muestra un comportamiento similar a los representantes del Orden de Thermoplasmatales (todos acidófilos) y se agrupan muy cercanamente en la filogenia basada en RNA 16S (figuras 1 y 2), por lo que en este trabajo este grupo fue analizado en conjunto. Los Thermoplasmatales se caracterizaron por no mostrar agrupamiento para ninguno de los dos ejes de análisis de expresión en el COA y presentar muy bajos valores de S (en promedio 0.19; figura 2). El análisis del uso relativo de codones obtenido por CODONW muestra valores de correlación muy bajos con el índice MELP (en promedio $R^2 = 0.13$) que es indicativo del nivel de expresión de un gen (Tabla 2). Este grupo mostró 15 codones óptimos (tabla 3, continuación), lo cual sin contar los codones STOP ni la Met y el Trp constituye un 27%. Se usan tres de los cuatro COU, el triplete AUC para Ile figura con tRNA asociado pero tiene una frecuencia de uso por debajo de la que uno esperaría al azar (-25%); en cambio figura AUA como el codón más frecuente en la secuencia génica en los genes de alta expresión pero no se detectó el tRNA correspondiente. El tRNA con anticodón GAT reconoce AUC, y por balanceo no puede reconocer al codón AUA porque resultaría en un contraproducente apareamiento entre dos purinas. Los codones óptimos tendieron a terminar en G o en C. Si se tiene en cuenta que el contenido en GC promedio de este grupo es de sólo 41%, llama la atención la tendencia a que los tripletes terminen en G o C. Sin embargo hay trabajos publicados en los que se observó una correlación entre los nucleótidos GC y la temperatura (Musto *et al.*, 2004). Es posible que la pequeña desviación encontrada en el UCS y en el mínimo valor de fuerza de selección S, pueda ser explicado por la historia de vida de estos organismos que además de acidófilos, son termófilos (temperatura de crecimiento óptimo de 56°C en *Thermoplasma acidophilum*, por ejemplo).

El grupo de los Methanococcales se caracteriza por un bajo contenido en GC (33% promedio del grupo) y es uno de los Órdenes que mostraron un coeficiente de fuerza selectiva S más altos (figura 2). El análisis de agrupamiento de genes de alta expresión en alguno de los ejes principales del COA y la relación con MELP (R^2 promedio = 0.28; tabla 2) mostraron evidencia positiva a favor de que exista selección traduccional para el UCS. Se identificaron 21 codones óptimos con tRNA correspondiente (tabla 3) y este grupo mostró preferencia también por el uso de los cuatro COU. Los tripletes aparecen terminados en C o en A. En este grupo se cumplen las cuatro condiciones que establecimos para determinar la existencia de selección traduccional mediante sesgo en el UCS. Dentro de este grupo podemos destacar a los organismos *M. maripaludis, M. vannielii* y *M. voltae* como organismos con influencia fuerte de la selección natural (tabla 2, figura 2 y tabla 3). Queda demostrado en este grupo que el patrón observado de UCS no responde a los sesgos mutacionales propios del genoma y es altamente selectivo.

Los Methanobacteriales, como grupo, muestran selección traduccional en el UCS. Las distintas especies muestran diferente grado de agrupamiento por COA y relación MELP y en promedio el coeficiente S es de 0.65 (tabla 2). Se contaron 25 codones óptimos con tRNA asociado y tienden a terminar en A o C y se usaron en forma preferencial los cuatro COU. El contenido en GC promedio es del 40%, por lo que el bajo sesgo observado aparecería como independiente de la composición nucleotídica. En este grupo se destaca el organismo *M. ruminantium* que mostró más evidencia de selección que el resto, con valores de S=0.78, y R^2 de 0.70 entre el uso relativo de codones sinónimos y el MELP.

Dentro del Orden de los Methanosarcinales se observa que como grupo presenta selección traduccional intermedia con un S promedio de 0.47, pero en este caso el contenido en GC (44% en promedio) parece ser muy influyente como se observa mediante la relación R^2 entre los usos relativos de codones para el primer eje principal del COA y el contenido en GC (ver tabla 2). El análisis de clustering de COA no mostró un claro agrupamiento para los genes de alta expresión en relación a los ejes de expresión. La relación con el índice MELP confirma que hay un sesgo relativo al uso de codones y que el mismo se relaciona con la expresión. En el análisis de distribución y patrones de codones óptimos y tRNAs asociados se muestra que hay 26 codones óptimos con C en la tercera posición. Este patrón de preferencia por codones terminados en C reafirma que este sesgo es debido al contenido nucleotídico del genoma. Los cuatro COU también en este caso mostraron un uso preferencial, pero no queda claro que su uso se deba a que son codones óptimos universales o porque son codones terminados en C.

En cuanto a los Methanocelalles con alto valor S=0.93 y relativamente alto contenido en GC (promerdio 56%, ver tabla 1), parecen repetir el comportamiento de los Methanosarcinales. Del mismo

modo a lo que sucedía en ese grupo, en los 22 codones óptimos la preferencia por el nucleótido C en la posición de balanceo es clara y también se usaron los cuatro COU. El sesgo observado en el uso de codones sinónimos está relacionado al contenido en GC.

En el clado Methanomicrobiales, se observaron valores intermedios de S (0.39) y su contenido en GC es de 54%. El análisis de COA no indicó agrupación para los genes de alta expresión en el primer eje del análisis pero sí para el segundo eje, y este patrón no estaría relacionado con el contenido en GC, ya que el valor R^2 promedio del grupo es cercano a cero (tabla 2, columna GC/RSCU2). Los codones óptimos universales aparecen con uso preferencial, junto a otros 28 con tRNAs asociados y terminan preferentemente en C. Se cumplen tres de los cuatro criterios establecidos para determinar la existencia de selección traduccional por UCS.

Es importante destacar que estos últimos tres grupos son cercanos filogenéticamente (figura 1); en particular los Ordenes de Methanocellales y Methanosarcinales. Este gran clado filogenético tiene composición nucleotídica similar y comparten similaridades en los codones óptimos elegidos.

Un grupo en el que la selección natural para el UCS es muy fuerte es el de los Halobacteriales, con un S=0.89. Se caracterizan además por su alto contenido en GC (promedio de 64%). Este grupo presentó clustering positivo según el COA en alguno de los ejes principales (tabla 2), uso de COU y preferencia por codones óptimos terminados en C con tRNAs asociados. Podemos destacar los organismos *H. marismortui, H. lacusprofundi*, *H. utahensis* y *N. magadii.* En las cuatro especies se evidenció clustering positivo según el COA, y cuando se analizó la relación de los patrones de uso relativo de codones sinónimos con el contenido en GC genómico se encontró una correlación R^2 cercana a 0.61. También es alta la relación con el índice MELP, indicador directo de niveles de expresión por uso relativo de codones sinónimos (tabla 2), así como los valores de S. En vista de que se cumplen las cuatro condiciones determinantes, se puede afirmar que para este grupo hay selección traduccional para el UCS.

Los Archaeoglobales, con contenido en GC de 45%, coeficiente de selección Sharp de 0.34 y preferencia por codones terminados en C, G o A se presenta como un grupo medianamente selectivo porque cumple con las condiciones impuestas previamente (tabla 2 y tabla 3).

El grupo de los Thermococcales es un grupo que muestra una gran diversidad interna. Como grupo presenta un bajo nivel de selección traduccional porque si bien no todos los organismos analizados cumplieron con el primer criterio, sí mostraron preferencia por codones óptimos con frecuencias diferenciales y tRNAs asociados (incluyendo los cuatro CUO) y un valor intermedio de S (0.27). Los codones óptimos terminan en C preferentemente.

Por último, en los Methanopyrales, *M. kandleri* como único representante analizado; presentó seleción débil para el UCS. No fue posible determinar un agrupamiento positivo según el COA. Sin embargo, se identificaron codones óptimos con tRNA asociado (GCC, CGC, AAC*, GAC*, CAC, AUC*, CUA, UUC*, AGC y UAC*) (figura 3 continuación) de los cuales algunos son además COU (*). Otro punto a favor de la selección para el uso preferencial de codones sinónimos es el coeficiente de Sharp mayor a cero (S=0.64). La tendencia que presentan los codones a terminar en C puede ser atribuíble al alto contenido GC del 61%, esta correlación con el contenido genómico se confirma por la relación R^2 entre GC y el RSCU1 de 0.68 (tabla2) pero a pesar del fuerte sesgo mutacional actuante se observa una relación R^2 entre RSCU1 y MELP de 0.31 indicativo de selección para el UCS.

En los *Phyla* con menos representantes (Korarchaeota, Nanoarchaeota y Thaumarchaeota), sólo *N. maritimus* en representación de Thaumarchaeota presenta selección para el uso de codones sinónimos.

CONCLUSIONES.

El uso de codones sinónimos varía entre las distintas especies en Archaea. En algunos organismos se cumplen las condiciones necesarias para afirmar que hay selección natural traduccional para eficiencia y precisión, mediante sesgo en el uso de codones sinónimos. Para los distintos AA hay múltiples especies de tRNA con distintos anticodones que aparecen en abundancias diferenciales; los codones traducidos por las especies de tRNA más abundantes son los preferidos en los genes altamente expresados pero también los estilos de vida en las especies influyen en los patrones encontrados. Los organismos que mostraron o no este comportamiento están relacionados filogenéticamente y forman clusters filogenéticos con estrategias comunes de UCS. Aparentemente una vez adquirido determinado patrón de uso de codones, éste se mantiene en el tiempo. Los grupos filogenéticos analizados que presentaron mayor grado de selección para uso de codones sinónimos, coinciden con una mayor radiación y abundancia de especies.

BIBLIOGRAFÍA.

Akashi, H. (1995). Inferring weak selection from patterns of polymorphism and divergence at" silent" sites in Drosophila DNA. Genetics, 139(2), 1067-1076.

Akashi, H. (2001). Gene expression and molecular evolution. Curr. Opin. Genet. Dev. 11; 660-666.

Allen, M. A., Lauro, F. M., Williams, T. J., Burg, D., Siddiqui, K. S., De Francisci, D., WY Chong, K., Pilak, O., Chew, H. H., De Maere, M. Z., Ting, L., Katrib, M., Ng, C., Sowers, K. R., Galperin, M. Y., Anderson, I. J., Ivanova, N., Dalin, E., Martinez, M., Lapidus, A., Hauser, L., Land, M., Thomas, T., Cavicchioli, R. (2009). The genome sequence of the psychrophilic archaeon, Methanococcoides burtonii: the role of genome evolution in cold adaptation. The ISME J. 3: 1012-1035.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of molecular biology, 215(3), 403-410.

Ann-Christin Lindås and Rolf Bernander. (2013). The cell cycle of archaea. Nature Reviews Microbiology. doi:10.1038/nrmicro3077

Behura, S. K., & Severson, D. W. (2012). Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. PloS one, 7(8), e43111.

Behura, S. K., & Severson, D. W. (2013). Codon usage bias: causative factors, quantification methods and genome‐wide patterns: with emphasis on insect genomes. Biological Reviews, 88(1), 49-61.

Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., & Blüthgen, N. (2013). Efficient translation initiation dictates codon usage at gene start. Molecular systems biology, 9(1).

Bernander, R. (2003). The archaeal cell cycle: current issues. Molecular microbiology, 48(3), 599-604.

Bernardi, G. (1993). The vertebrate genome: isochores and evolution. Mol. Biol. Evol. 10: 186-204.

Bintrim, S. B., Donohue, T. J., Handelsman, J., Roberts, G. P., Goodman, R. (1997). Molecular phylogeny of Archaea from soil. Proc. Natl. Acad. Sci. 94, 277-282.

Botzman, M., & Margalit, H. (2011). Variation in global codon usage bias among prokaryotic organisms is associated with their lifestyles. Genome Biol,12(10), R109.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., & Forterre, P. (2008). Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. Nature Reviews Microbiology, 6(3), 245-252.

Bulmer, M. (1987). Coevolution of codon usage and transfer RNA abundance.Nature, 325(6106), 728-730.

Bulmer, M. (1988). Are codon usage patterns in unicellular organisms determined by selection‑mutation balance?. Journal of Evolutionary Biology,1(1), 15-26.

Bulmer, M. (1991). The selection-mutation-drift Theory of synonymous codon usage. Genetics 129: 897-907.

Cannarozzi, G., Schraudolph, N.N., Faty, M., Rohr, P., Friberg, M., Roth, A., Gonnet, P., Gonnet, G., Barral, Y., (2010). A role for codon order in translation dynamics. Cell 141, 355-367.

Carlini, D.B. (2004). Experimental reduction of codon bias in the drosophila alcohol dehydrogenase gene results in decreased ethanol tolerance of adult flies. J. Evol. Biol. 17, 779-785.

Cavalier-Smith, T. (2002). The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification. Int. J. Syst. Evol. Microbiol. 52: 7-76.

Charlesworth, B. (2013). Stabilizing Selection, Purifying Selection, and Mutational Bias in Finite Populations. Genetics, 194(4), 955-971.

Chen, S. L., Lee, W., Hottes, A. K., Shapiro, L., & McAdams, H. H. (2004). Codon usage between genomes is constrained by genome-wide mutational processes. Proc. Natl. Acad. Sci., 101(10), 3480-3485.

Comeron, J. M., & Aguadé, M. (1998). An evaluation of measures of synonymous codon usage bias. Journal of molecular evolution, 47(3), 268-274.

Comeron, J.M. & Kreitman, M. (1998). The correlation between synonymous and nonsynonymous substitutions in Drosophila: mutation, selection or relaxed constraints? Genetics 150: 767-775.

De Koning, B., Blombach, F., Bronus, S. J. J., Van Der Oost, J. (2010). Fidelity in archaeal information processing. Hindawi Publishing Corp. doi:10.1155/2010/960298

Doolittle, R. F. (2002). Microbial genomes multiply. Nature 416: 697-700.

Dressaire, C., Picard, F., Redon, E., Loubière, P., Queinnec, I., Girbal, L., & Cocaign-Bousquet, M. (2013). Role of mRNA stability during bacterial adaptation. PloS one, 8(3), e59059.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32: 1792-1797.

Elkins J.G., Podar, M., Graham, D. E., Makarova, K.S., Wolf, Y., Randau, L., … Stetter, K.O. (2008). A korarchaeal genome reveals insights into the evolution of the Archaea. Proc. Natl. Acad. Sci., 105(23), 8102-8107. doi : 10.1073/pnas.0801980105.

Emery, L. R. (2011). Codon usage bias in Archaea.

Emery, L.R. & Sharp, P.M. (2011). Impact of translational selection on codon usage bias in the archaeon Methanococcus maripaludis. Biol. Lett. 7: 131-135. doi:10.1098/rsbl.2010.0620

Erkel, C., Kube, M., Reinhardt, R., & Liesack, W. (2006). Genome of Rice Cluster I archaea—the key methane producers in the rice rhizosphere. Science, 313(5785), 370-372.

Eyre-Walker, A. (1996). Synonymous codon bias is related to gene length in Escherichia coli: selection for translational accuracy?. Molecular biology and evolution, 13(6), 864-872.

Fang, H., Oates, M. E., Pethica, R. B., Greenwood, J. M., Sardar, A. J., Rackham, O. J., ... & Gough, J. (2013). A daily-updated tree of (sequenced) life as a reference for genome research. Scientific reports, 3.

Gao, B. & Gupta, R.S. (2007). Phylogenomic analysis of proteins that are distinctive of Archaea and its main subgroups and the origin of methanogenesis. BMC Genomics 8:86.

Gingold, H. & Pilpel, Y. (2011). Determinants of translational efficiency and accuracy. Molecular Systems Biology 7: 481. doi:10.1038/msb.2011.14

Gingold, H., Dahan, O., & Pilpel, Y. (2012). Dynamic changes in translational efficiency are deduced from codon usage of the transcriptome. Nucleic acids research, 40(20), 10053-10063.

Gouy, M. & Gautier, C. (1982). Codon usage in bacteria: correlation with gene expressivity. Nucleic acids research 10: 7055-7074.

Grantham, R., Gautier, C. Gouy, M., Jacobzone, M., Mercier, R. (1981). Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic acids research 9, R43-R74.

Grantham, R., Gautier, C., Gouy, M., Mercier, R., Pavé, A. (1980). Codon catalog usage and the genome hypothesis. Nucleic acids research 8: r49-r62.

Gribaldo, S. & Brochier-Armanet, C. (2006). The origin and evolution of Archaea: a state of the art. Phil. Trans. R. Soc. B 361; 1007-1022.

Guo, F. B., Ye, Y. N., Zhao, H. L., Lin, D., & Wei, W. (2012). Universal pattern and diverse strengths of successive synonymous codon bias in three domains of life, particularly among prokaryotic genomes. DNA research, 19(6), 477-485.

Gupta, R. S. (1998a). Life´s third domain (Archaea): an established fact or an endangered paradigm? : A new proposal for classification of organisms based on protein sequences and cell structure. Theor. Popul. Biol. 54:91-104.

Gupta, R. S. (1998b). Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among Archaebacteria, Eubacteria and Eukaryotes. Microbiol. Mol. Biol. Rev. 62:1435-1491.

Gupta, R. S. (1998c). What are archaeabacteria: life´s third domain or monoderm prokariotes related to Gram-positive bacteria? A new proposal for the classification of the prokaryotes organisms. Mol. Microbiol. 29: 695-707.

Gupta, R. S. (2010). Application of conserved indels for understanding microbial phylogeny. Chapter 7, Book: molecular phylogeny of microbial organism. Edited by A. Oren & R. T. Papke. Caister AcademicPress.

http://books.google.es/books?hl=es&lr=&id=a5t9DYZwccC&oi=fnd&pg=PA135&dq=gupta+archaea&ots=ejy5NOxOO&sig=4yCFj8ZaGqPaFweyQeI93R48h3w#v=onepage&q=gupta%20archaea&f=true

H K Stenøien. (2004). Adaptive basis of codon usage in the haploid moss Physcomitrella patens. Heredity. doi:10.1038/sj.hdy.6800547

Henderson, E., Oakes, M., Clark, M. W., Lake, J. A., Matheson, A. T. & Zillig, W. (1984). A new ribosome structure. Science 225, 510-512.

Henry, I. & Sharp, P. (2007). Predicting gene expression level from codon usage bias. Mol. Biol. Evol. 24: 10-12.

Hershberg, R. & Petrov, D.A. (2009). General rules for optimal codon choice. PLoS Genet. 5(7): e1000556 doi:10.1371/journal.pgen.1000556

Hershberg, R., & Petrov, D. A. (2008). Selection on codon bias. Annual review of genetics, 42, 287-299.

Hilterbrand, A., Saelens, J., & Putonti, C. (2012). CBDB: the codon bias database. BMC bioinformatics, 13(1), 62.

http://es.wikipedia.org/wiki/Archaea

http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi

Huber, H., Hohn, M. J., Stetter, K.O., & Rachel, R. (2003). The phylum Nanoarchaeota: present knowledge and future perspectives of a unique form of life. Research in microbiology, 154(3), 165-171.

Ikemura, T. (1981 a). Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. J. Mol. Biol. 151 (3): 389-409.

Ikemura, T. (1981 b). Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons of its protein genes. J. Mol. Biol. 146: 1-21.

Ikemura, T. (1982). Correlation between the abundance of yeast tRNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice pattern of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs. J. Mol. Biol. 158: 573-597.

Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2 (1): 13-34.

Ikemura, T., & Ozeki, H. (1983, January). Codon usage and transfer RNA contents: organism-specific codon-choice patterns in reference to the isoacceptor contents. In Cold Spring Harbor symposia on quantitative biology(Vol. 47, pp. 1087-1097). Cold Spring Harbor Laboratory Press.

Ingvarsson, P.K. (2008) Molecular evolution of synonymous codon usage in Populus. BMC Evolutionary Biology 8:307.

Iriarte, A, Baraibar, J.D.; Romero, H.; Musto, H. (2011). Selected codon usage bias in members of the class Mollicutes. Gene 473, 110-118.

Karlin, S., Campbell, A. M., & Mrázek, J. (1998). Comparative DNA analysis across diverse genomes. Annual review of genetics, 32(1), 185-225.

Karlin, S., Mrázek, J., & Campbell, A. M. (1998). Codon usages in different gene classes of the Escherichia coli genome. Molecular microbiology, 29(6), 1341-1355.

Karr, E. A., Ng, J. M., Belchik, S. M., Sattley, W. M., Madigan, M. T., Achenbach, L.A. (2006). Biodiversity of methanogenic and other Archaea in the permanently frozen Lake Fryxell, Antartica. Appl. Environ. Microbiol. 72, 1663-1666.

Kawahara-Kobayashi, A., Masuda, A., Araiso, Y., Sakai, Y., Kohda, A., Uchiyama, M., ... & Kiga, D. (2012). Simplification of the genetic code: restricted diversity of genetically encoded amino acids. Nucleic acids research,40(20), 10576-10584.

Kimura, M. (1977). Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 267, 275-276.

Klumpp, S., Dong, J., & Hwa, T. (2012). On ribosome load, codon bias and protein abundance. PloS one, 7(11), e48542.

Knight, R.D., Freeland, S.J., Landweber, L.F. (2001). A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. Genome Biol., 2 (4): RESEARCH0010.

Kurland, C. G., (1991). Codon bias and gene-expression. FEBS Letters 285: 165-169.

Lake, J. A., Henderson, E. Oakes, M. & Clark, M. W. (1984). Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. Proc. Natl. Acad. Sci., USA 81, 3786-3790.

Li, G. W., Oh, E., & Weissman, J. S. (2012). The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature,484(7395), 538-541.

Li, W. H. (1987). Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. Journal of molecular evolution,24(4), 337-345.

Lind, P. A., & Andersson, D. I. (2013). Fitness costs of synonymous mutations in the rpsT gene can be compensated by restoring mRNA base pairing. PloS one, 8(5), e63373.

Liu, Q. (2012). Mutational Bias and Translational Selection Shaping the Codon Usage Pattern of Tissue-Specific Genes in Rice. PloS one, 7(10), e48295.

Lobry, J.R. & Chessel D. (2003). Internal correspondence analysis of codon and amino-acid usage in thermophilic bacteria. J. Appl. Genet. 44: 235-261.

Lowe, T. M. & Eddy, S. R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic acids research 25: 955-964.

Lynn, D.J., Singer, G.A.C., Hickey, D.A. (2002). Synonymous codon usage is subject to selection in thermophilic bacteria. Nucleic acids research 30; 4272-4277.

Madigan, M.T. (2000). Extremophilic bacteria and microbial diversity. Ann. Missouri Bot. Gard. 87, 3-12.

Madigan, M.T. & Marrs, B.L. (1997). Extremophiles. Sci. Amer. 276, 82-87.

Madigan, M.T. & Orent, A. (1999). Termophilic and halophilic extremophiles. Curr. Opin. Microbiol. 2, 265-269.

Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2013). Comparative genomics of defense systems in archaea and bacteria. Nucleic acids research, 41(8), 4360-4377.

Marquez, R., Smit, S., Knight, R. (2005). Do universal codon-usage patterns minimize the effects of mutation and translation error? Genome Biol. 6:R91.

Matte-Tailliez, O., Brochier, C., Forterre, P., & Philippe, H. (2002). Archaeal phylogeny based on ribosomal proteins. Molecular Biology and Evolution, 19(5), 631-639.

Matte-Tailliez, O., Brochier, C., Forterre, P., Philippe, H. (2002). Archaeal phylogeny based on ribosomal proteins. Mol. Biol. Evol. 19, 631-639.

Medrano Soto, L.A. (2005). Uso de codones, traducibilidad, niveles de expresión y transferencia horizontal: ¿hemos sobreinterpretado nuestros organismos modelo? Tesis doctorado Universidad Nacional Autónoma de México.

Mian Zhou, Jinhu Guo, Joonseok Cha, Michael Chae, She Chen, Jose M. Barral, Matthew S. Sachs, et al. (2013). Non-optimal codon usage affects expression, structure and function of clock protein FRQ. Nature. doi:10.1038/nature11833

Michail M. Yakimov, Violetta La Cono, Vladlen Z. Slepak, Gina La Spada, Erika Arcadi, Enzo Messina, Mireno Borghini, et al. (n.d.). Microbial life in the Lake Medee, the largest deep-sea salt-saturated formation. Scientific Reports. doi:10.1038/srep03554

Musto, H. (2001). Translational selection on codon usage in Xenopus laevis. Mol. Biol. Evol. 18: 1703-1707.

Musto, H., Naya, H., Zavala, A., Romero, H., Alvarez-Valín, F., Bernardi, G. (2004). Correlations between genomic GC levels and optimal growth temperatures in prokaryotes. FEBS Letters 573: 73-77.

Musto, H., Naya, H., Zavala, A., Romero, H., Alvarez-Valín, F., Bernardi, G. (2006). Genomic GC level, optimal growth temperature, and genome size in prokaryotes. Biochem. Biophys. Res. Commun. 347: 1-3.

Musto, H., Romero, H. & Zavala, A. (2003). Translational selection is operative for synonymous codon usage in Clostridium perfringens and Clostridium acetobutylicum. Microbiology 149, 855-863.

Muto, A. & Osawa, S. (1987). The guanine and cytosine content of genomic DNA and bacterial evolution. Proc. Natl. Acad. Sci., USA 84: 166-169.

Naya, H., Romero, H., Zavala, A., Alvarez, B., Musto, H. (2002). Aerobiosis increases the genomic guanine plus cytosine content (GC%) in Prokaryotes. J. Mol. Evol. 55: 260-264.

Nei, M. (2005). Selectionism and neutralism in molecular evolution. Molecular biology and evolution, 22(12), 2318-2342.

Nishida, H. (2012). Comparative analyses of base compositions, DNA sizes, and dinucleotide frequency profiles in archaeal and bacterial chromosomes and plasmids. International journal of evolutionary biology, 2012.

Nishida, H. (2013). Genome DNA sequence variation, evolution, and function in bacteria and archaea. Curr. Issues Mol. Biol, 15, 19-24.

Novoa, E. M., & Ribas de Pouplana, L. (2012). Speeding with control: codon usage, tRNAs, and ribosomes. Trends in Genetics, 28(11), 574-581.

Otto X. Cordero and Martin F. Polz. (n.d.). Explaining microbial genomic diversity in light of evolutionary ecology. Nature Reviews Microbiology. doi:10.1038/nrmicro3218

Pace, N. R. (1997) A molecular view of microbial diversity and the biosphere. Science 276, 734-740.

Patrick M Erwin, Mari Carmen Pineda, Nicole Webster, Xavier Turon, and Susanna López-Legentil. (2013). Down under the tunic: bacterial biodiversity hotspots and widespread ammonia-oxidizing archaea in coral reef ascidians. The ISME Journal. doi:10.1038/ismej.2013.188

Peden,J.F. (1999) Analysis of codon usage. PhD Thesis, University of Nottingham, UK.

Ping Yu, Yuan Yan, Qing Gu, and Xiangyang Wang. (n.d.). Codon optimisation improves the expression of Trichoderma viride sp. endochitinase in Pichia pastoris. Scientific Reports. doi:10.1038/srep03043

Piovani, Rosina. (2007). Uso de codones sinónimos. Una revision acerca de las fuerzas que gobiernan este fenómeno en la evolución. Trabajo especial I. Licenciatura en bioquímica.

Plotkin, J.B. & Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. Nature Rev. Genet. 12 doi:10.1038/nrg2899.

Precup, J. & Parker, J. (1987). Missense misreading of asparagine codons as a function of codon identity and context. J. Biol. Chem. 262, 11351-11355.

Quax, T. E., Wolf, Y. I., Koehorst, J. J., Wurtzel, O., van der Oost, R., Ran, W., ... & van der Oost, J. (2013). Differential Translation Tunes Uneven Production of Operon-Encoded Proteins. Cell reports, 4(5), 938-944.

Ran, W., & Higgs, P. G. (2012). Contributions of speed and accuracy to translational selection in bacteria. PloS one, 7(12), e51652.

Rest, J.S. & Mindell, D.P. (2003). Retroids in Archaea: phylogeny and lateral origins. Mol. Biol. Evol. 20 (7): 1134-1142.

Retchless, A.C & Lawrence, J.G. (2011). Quantification of codon selection for comparative bacterial genomics. BMC Genomics 12: 374 http://www.biomedcentral.com/1471-2164/12/374.

Rocha, E.P.C. (2004 a). Codon usage bias from tRNA´s point of view: redundacy, specialization, and efficient decoding for translation optimization. Genome Res. 14: 2279-2286.

Rocha, E.P.C. (2004 b). The replication-related organization of bacterial genomes. Microbiology 150, 1609-1627.

Rocha, E.P.C. & Antoine Danchin. (2002). Base composition bias might result from competition for metabolic resources. Trends Genet.18, 291-294.

Roche, E.D. & Sauer, R.T. (1999). SsrA-mediated tagging caused by rare codons and tRNA scarcity. The EMBO J. 18: 4579-4589.

Roth, A. C. (2012). Decoding properties of tRNA leave a detectable signal in codon usage bias. Bioinformatics, 28(18), i340-i348.

Roth, A., Anisimova, M., & Cannarozzi, G. M. (2012). Measuring codon usage bias. Codon evolution: mechanisms and models. New York: Oxford University Press Inc, 189-217.

Sapp, J. (2006). Two faces of the prokaryote concept. Int. Microbiol. 9, 163-172.

Sarmiento, F., Mrázek, J., & Whitman, W. B. (2013). Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon Methanococcus maripaludis. Proc. Natl. Acad. Sci., 110(12), 4726-4731.

Schmidt, A., Rzanny, M., Schmidt, A., Hagen, M., Schütze, E., & Kothe, E. (2012). GC content‐independent amino acid patterns in Bacteria and Archaea. Journal of basic microbiology, 52(2), 195-205.

Shabalina, S. A., Spiridonov, N. A., & Kashina, A. (2013). Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity. Nucleic acids research, 41(4), 2073-2094.

Sharp, P.M. & Li, W.H. (1986). Codon usage in regulatory genes in Escherichia coli does not reflect selection for “rare” codons. Nucleic acids research 14, 7737-7749.

Sharp, P.M. & Li, W.H. (1986). Codon usage in regulatory genes in Escherichia coli does not reflect selection for “rare” codons. Nucleic acids research 14, 7737-7749.

Sharp, P.M. & Wen-Hsiung Li (1987). The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. Nucleic acids research 15: 1281-1295.

Sharp, P.M. & Wen-Hsiung, L. (1987). The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. 4(3): 222-230.

Sharp, P.M., Bailes, E., Grocock, R.J., Peden, J.F, Sockett, R.E. (2005). Variation in the strenght of selected codon usage bias among bacteria. Nucleic acids research 33: 1141-1153.

Sharp, P.M., Emery, L.R., Zeng, K., (2010). Forces that influence the evolution of codon bias. Phil. Trans. R. Soc. B. 365, 1203, 1212.

She, Q., Singh, R.K., Confalonieri, F., Zivanovik, I., Allard, G.; et al. (2001). The complete genome of the crenarchaeon Sulfolobus solfataricus P2. Proc. Natl. Acad. Sci., 98, 7835-7840.

Singer, G. A., & Hickey, D. A. (2003). Thermophilic prokaryotes have characteristic patterns of codon usage, amino acid composition and nucleotide content. Gene, 317, 39-47.

Spang, A., Hatzenpichler, R., Brochier-Armanet, C., Rattei, T., Tischler P., Spieck, E., Streit, W., Stahl D.A., Wagner, M., Schleper, C. (2010). Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. Trends Microbiol. 18, 331-340.

Subramaniam, A. R., Pan, T., & Cluzel, P. (2013). Environmental perturbations lift the degeneracy of the genetic code to regulate protein levels in bacteria. Proc. Natl. Acad. Sci., 110(6), 2419-2424.

Sueoka, N. (1962). On the genetic basis of variation and heterogeneity of DNA base composition. Proceedings of the National Academy of Sciences of the United States of America, 48(4), 582.

Supek, F., & Vlahoviček, K. (2005). Comparison of codon usage measures and their applicability in prediction of microbial gene expressivity. BMC bioinformatics, 6(1), 182.

Supek, F., Vlahovicek, K. (2004). INCA: synonymous codon usage analysis and clustering by means of self-organizing map. Bioinformatics 20, 2329-2330.

Suzuki, H., Saito, R., & Tomita, M. (2009). Measure of synonymous codon usage diversity among genes in bacteria. BMC bioinformatics, 10(1), 167.

Thanaraj, T.A. & Argos, P. (1996). Protein secondary structural types are differentially coded on messenger RNA. Protein Science 5, 1973-1983.

Thiele, I., Fleming, R. M., Que, R., Bordbar, A., Diep, D., & Palsson, B. O. (2012). Multiscale modeling of metabolism and macromolecular synthesis in E. coli and its application to the evolution of codon usage. PLoS One, 7(9), e45635.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic acids research, 22(22), 4673-4680.

Timothy J Williams, Michelle A Allen, Matthew Z DeMaere, Nikos C Kyrpides, Susannah G Tringe, Tanja Woyke, and Ricardo Cavicchioli. (2014). Microbial ecology of an Antarctic hypersaline lake: genomic assessment of ecophysiology among dominant haloarchaea. The ISME Journal. doi:10.1038/ismej.2014.18

Tom A. Williams, Peter G. Foster, Cymon J. Cox, and T. Martin Embley. (2013). An archaeal origin of eukaryotes supports only two primary domains of life. Nature. doi:10.1038/nature12779

Tonghai Yu, Jinsong Li, Yang Yang, Liu Qi, Biaobang Chen, Fangqing Zhao, Qiyu Bao, Jinyu Wu (2011). Codon usage patterns and adaptive evolution of marine unicellular cyanobacteria Synechococcus and Prochlorococcus. Mol. Phylogenet. Evol. doi:10.1016/j.ympev.2011.09.013.

Wald, N., Alroy, M., Botzman, M., & Margalit, H. (2012). Codon usage bias in prokaryotic pyrimidine-ending codons is associated with the degeneracy of the encoded amino acids. Nucleic acids research, 40(15), 7074-7083.

Waldman, Y. Y., Tuller, T., Keinan, A., & Ruppin, E. (2011). Selection for translation efficiency on synonymous polymorphisms in recent human evolution.Genome biology and evolution, 3, 749.

Wan, X. F., Xu, D., Kleinhofs, A., & Zhou, J. (2004). Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. BMC Evolutionary Biology, 4(1), 19.

Warnecke, T. & Hurst, L.D. (2011). Error prevention and mitigation as forces in the evolution of genes and genomes. Nature Reviews, Genetics 12; 875-881.

Wen Wei & Feng-Biao Guo. (2010). Strong strand composition bias in the genome of Ehrlichia canis revealed by multiple methods. Open Microbiol J. 4, 98-102.

Woese, C.R. (1987). Bacterial evolution. Microbiol. Rev. 51, 221-271.

Woese, C.R. (1998). The universal ancestor. Proc. Natl. Acad. Sci. 95, 6854-6859.

Woese, C.R., Kandler, O., Wheelis, M.L. (1990). Towards a natural system of organism: proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. 87, 4576-4579.

Wohlgemuth, S. E., Gorochowski, T. E., & Roubos, J. A. (2013). Translational sensitivity of the Escherichia coli genome to fluctuating tRNA availability. Nucleic acids research, 41(17), 8021-8033.

Wu, D., et al. (2009). A phylogeny-driven genomic encycolpaedia of Bacteria and Archaea. Nature, 462, 1056-1060.

Zhang, Z., Li, J., Cui, P., Ding, F., Li, A., Townsend, J. P., & Yu, J. (2012). Codon Deviation Coefficient: a novel measure for estimating codon usage bias and its statistical significance. BMC bioinformatics, 13(1), 43.