

Estudios Genómicos de la Transición Anaerobiosis-Aerobiosis en Procariotas

Lucía Graña

Orientador: Dr. Héctor Romero

Laboratorio de Organización y Evolución del Genoma.

Sección Biomatemática

Facultad de Ciencias

Universidad de la República

ÍNDICE		Págs
Resumen	3
Introducción	4
	<i>Variación en el contenido G+C</i>	5
	<i>Uso de aminoácidos</i>	6
	<i>Oxígeno y metabolismo</i>	8
Objetivo	10
Materiales y métodos	11
	<i>Datos</i>	11
	<i>Relaciones filogenéticas</i>	15
	<i>Reconstrucción de estados ancestrales</i>	17
	<i>Categorización funcional</i>	17
	<i>Identificación de genes específicos de aerobios y anaerobios</i>	19
	<i>Propiedades composicionales</i>	19
Resultados y discusión	21
	<i>Impacto funcional de las transiciones entre los metabolismos anaerobio y aerobio</i>	22
	<i>Propiedades Composicionales</i>	33
Conclusiones	44
Bibliografía	45

RESUMEN

En el año 2002 Naya y colaboradores encontraron una importante correlación entre la aerobiosis y el aumento del contenido G+C genómico, la variabilidad de éste parámetro en particular ha sido ampliamente estudiada.

Para el estudio de la estructura genómica es de gran importancia la disponibilidad de genomas completos pertenecientes a bacterias emparentadas. Con esta información es posible observar el reordenamiento y evolución de los genomas mediante estudios comparativos (Bentley & Parkhill, 2004).

Sabemos que existen diferencias entre genomas de organismos con distintos requerimientos energéticos (Naya et al, 2002) pero dentro de cada genoma buscamos diferencias en la composición génica que se observa en las diferentes enzimas que catalizan las reacciones de las vías metabólicas disímiles entre los organismos con metabolismo aerobio y los que poseen metabolismo anaerobio.

INTRODUCCIÓN

Los microorganismos representan la mayor parte de la diversidad biológica (Pace, 1997), son responsables de la mitad de la producción primaria del planeta, (Field et al. 1998) y llevan a cabo los principales ciclos biogeoquímicos (DeLong & Karl, 2005). Las dinámicas evolutivas y los procesos ecológicos que presentan las comunidades microbianas, tienen un efecto global sobre el flujo de materia y energía, la composición de la atmósfera y el clima global (DeLong & Karl, 2005). Poseen una diversidad metabólica considerada un recurso sumamente valioso para las aplicaciones biotecnológicas (Handelsman et al. 2007, pp.31-32), además, durante los 3.8 mil millones de años de vida en la tierra (2 mil millones más que los organismos eucariotas), los microorganismos se han encargado de mantener el balance biogeoquímico y se han extendido en el tiempo más que todas las otras formas de vida que conocemos (DeLong & Karl, 2005). La extraordinaria diversidad y gama de habitats procariota es el producto de su permanencia (Torsvik et al. 2002).

Existe actualmente una amplia gama de genomas procariotas completamente secuenciados disponibles lo que permite el estudio de sus rasgos y detalles de secuencia, características que revelan los mecanismos mediante los cuales los genomas evolucionan (Bentley & Parkhill, 2004).

Para el estudio de la estructura genómica es de gran importancia la disponibilidad de genomas completos pertenecientes a bacterias emparentadas. Con esta información es posible observar el reordenamiento y evolución de los genomas mediante estudios comparativos (Bentley & Parkhill, 2004).

Junto con la tasa de crecimiento exponencial de genomas secuenciados, se han vuelto cada vez más importantes, los métodos de la genómica comparativa.

Los genomas completamente secuenciados de numerosos organismos ofrecen una gran cantidad de información sobre la biología celular. Es un desafío importante para la bioinformática usar ésta información para descubrir las funciones de las proteínas. Las proteínas que forman parte de un mismo complejo estructural o vía metabólica, se dice que están funcionalmente relacionadas. Las proteínas funcionalmente relacionadas, evolucionan de forma correlacionada, y por lo tanto se encuentran homólogos de ellas en determinados sets de organismos, también relacionados.

La utilidad de los análisis comparativos en las ciencias biológicas es ampliamente reconocida, el entendimiento que se obtiene de la comparación detallada de vías metabólicas usadas por los organismos de muy variados estilos de vida y distancias filogenéticas ofrece un marco de

trabajo que permite caracterizar mecanismos que son aún muy pobremente comprendidos (D'Souza *et al*, 1999).

Los sistemas biológicos muestran frecuentemente modularidad en el sentido en que pueden dividirse en subsistemas casi independientes que realizan funciones diferentes (Parter *et al*, 2007), a cada uno de ellos se pueden asociar proteínas implicadas en éstas funciones. A nivel interespecífico se observa que las proteínas adjudicadas a cada set de reacciones metabólicas comparten homología, el número y tipo de enzimas ubicadas en éstos conjuntos determina cierto perfil para cada organismo.

Los estudios comparativos de las vías metabólicas y enzimas empleadas por los organismos en distintos ambientes o condiciones pueden proveer de importante información sobre la fisiología de muy diversos organismos (D'Souza *et al*, 1999).

Variación en el contenido GC

Los parámetros más usados para describir un genoma bacteriano son: tamaño, geometría, número de replicones y contenido G+C, que describe el contenido de guanina y citosina en una secuencia biológica, la sola variación de éstos parámetros asegura que no existe un genoma bacteriano típico (Bentley & Parkhill, 2004).

Dentro de los procariontes el tamaño del genoma varía mucho y es el resultado de diferentes eventos genéticos como duplicación génica, HGT (Transferencia Horizontal de Genes) y pérdida de genes linaje-específica entre otros fenómenos (Bentley & Parkhill, 2004).

El porcentaje de G+C en los genomas procariontes varía de 25 a 75 (Sueoka, 1962), el porqué de esta variación ha impulsado diversos estudios los cuales han arrojado respuestas de diferente índole. Se ha intentado explicar las fluctuaciones del porcentaje en G+C de los procariontes dentro del marco seleccionista, estos estudios sostienen que el contenido en G+C cambia en respuesta a condiciones ambientales y se lo ha asociado a diferentes variables.

También se ofrece entonces una respuesta neutralista, la cual se manifiesta a favor de un sesgo mutacional que deriva de características particulares en el funcionamiento de la maquinaria de replicación y reparación del ADN, que genera mutaciones, la mayoría de las cuales son eliminadas mediante selección purificadora, pero aquellas no deletéreas son fijadas por deriva génica (Kimura, 1986), ocasionando la variación en el contenido G+C que se observa incluso dentro de un mismo genoma.

Muchas variables ambientales fueron asociadas a la variabilidad del contenido G+C en el marco de los estudios seleccionistas. Algunas de éstas fueron inmediatamente rechazadas, otras no logran

una explicación más que de casos particulares. Las variables más significativas son forma de vida, tamaño del genoma, requerimientos metabólicos y hábitat.

El hecho de que los pares de bases G+C estén unidos por tres puentes de hidrógeno y los pares A+T solo por dos, podría conferir estabilidad en la doble hélice a los organismos con alto contenido en G+C, sobre todo frente a inestabilidad térmica (Yakovchuk *et al.* 2006).

La variación en el contenido G+C observada en los diferentes genomas también se ligó a la temperatura óptima de cada especie, el contenido alto en G+C podría estar asociado a organismos con alta temperatura óptima de crecimiento (Kagawa *et al.* 1984), dado por la estabilidad termodinámica que este rasgo molecular le confiere a las moléculas de ADN, ARN y proteínas (hipótesis termodinámica), un estudio comparativo a nivel intrafamilia permite ver esta correlación (Musto *et al.* 2004).

Las bacterias de vida libre tienen en promedio mayor contenido en G+C que aquellas que son simbiotes o parásitos (Rocha *et al.* 2002), esta diferencia se debería al costo y disponibilidad diferencial de los metabolitos en la célula, ATP y UTP, no sólo son menos costosos energéticamente sino que el ATP por su papel en el metabolismo celular está más disponible dentro de la célula. En un contexto de recursos limitados el sesgo mutacional hacia A+T sería favorecido (Rocha *et al.* 2002).

A su vez, se ha visto asociado el porcentaje de G+C del genoma con su tamaño, una correlación intrínseca es la tendencia de los genomas grandes a tener un mayor contenido en G+C (Musto *et al.* 2006, Moran *et al.* 2002, Rocha *et al.* 2002), a su vez los genomas de mayor tamaño se encuentran en bacterias de vida libre (Rocha *et al.* 2002). Foerstner *et al.* 2005, encontró que existe asociación entre el contenido en G+C y la complejidad de nicho, los genomas de mayor tamaño pertenecen a bacterias de vida libre cuyo hábitat puede ser muy diverso (marino, suelo, etc) mientras que los genomas pequeños se encuentran en simbiotes o parásitos cuyo ambiente es mucho más estable. Se observó también que esto influye en el uso de aminoácidos de cada organismo y de las distintas comunidades (Foerstner *et al.* 2005).

Uso de aminoácidos

El código genético es degenerado, diferentes codones codifican el mismo aminoácido, estos se denominan codones sinónimos. Los cambios en una secuencia codificante que sustituyen un codón sinónimo por otro se consideran silenciosos. Los codones sinónimos se usan con diferente frecuencia, este fenómeno se conoce como sesgo en el uso de codones, los codones que se usan con mayor frecuencia son los codones mayores y estos varían entre organismos e incluso entre genes de un mismo genoma. También el sesgo tiene distinta magnitud entre organismos y entre genes (Hassan *et al.* 2009).

El sesgo en el uso de codones ha sido muy estudiado y se han expuesto diversas explicaciones. Una de éstas propone que se debe al sesgo mutacional y sostiene que la desviación existe porque el patrón mutacional no es al azar, hay codones más susceptibles al cambio y tendrían menor frecuencia de equilibrio (Ermolaeva, 2001). El patrón más significativo que explica los diferentes sesgos de codones entre organismos es el nivel en el contenido G+C. El contenido G+C afecta el uso de codones y la composición de aminoácidos en genes y proteínas (Ermolaeva, 2001).

También hay pruebas que implican a la selección natural en éste fenómeno. Las presiones mutacionales por sí solas no explican el hecho de que los codones sinónimos más frecuentemente usados sean los reconocidos por los ARNt más abundantes. El sesgo mutacional tampoco da cuentas de la variación intragenómica en el sesgo en el uso de codones.

Dada la evidencia de presión mutacional y de selección, el modelo más aceptado hoy en día es el de balance mutación-selección-deriva. Bajo este modelo el sesgo en el uso de codones resulta de selección positiva sobre aquellas mutaciones que aumenten la frecuencia de codones mayores y selección purificadora contra las mutaciones que la disminuyan (Hershberg y Petrov, 2008)

Naya y colaboradores asociaron a los organismos estrictamente aerobios con un mayor contenido en G+C en relación con los anaerobios estrictos y plantearon a la aerobiosis como una factor importante moldeando esta característica (Naya *et al.* 2002). En este trabajo uno de los temas que se tratan es el de cuál es la ventaja selectiva que aporta el alto porcentaje en G+C a un organismo aerobio, este planteo es importante ya que, aunque el oxígeno molecular mata a algunos organismos es esencial para la vida de muchos otros, por esto y por la reactividad química de sus derivados el oxígeno es un gran candidato para estudiar las restricciones en composición de genes y proteínas (Vieira-Silva *et al.* 2008). Se propusieron entonces tres factores que se consideraron ventajosos, que los organismos aerobios presentan frecuencias más bajas de los aminoácidos Cys, Met, Trp, Tyr, Phe y His, los cuales comparten la característica de ser los que se oxidan preferentemente y son codificados por codones ricos en A+T y un aumento en G+C podría disminuir su frecuencia.

Además un porcentaje alto de G+C confiere estabilidad térmica a través de la formación de tres puentes de hidrógeno entre hebras lo que impediría el acceso a radicales de oxígeno que dañan el ADN. Tercero y último, el incremento del contenido G+C aumenta el número de sitios en la tercera posición de codones en las cuales cualquier cambio es sinónimo, esto es favorable en la medida en que las mutaciones en esos sitios no afectan la secuencia de aminoácidos (Naya *et al.* 2002). La variabilidad para cada una de las posiciones de los codones ha sido muy estudiada, se sabe que se comportan de distinta forma. Existen algunas correlaciones independientes del contenido G+C promedio del genoma. El porcentaje de G+C en la primera posición siempre es mayor que el promedio genómico mientras que lo opuesto se da para la segunda posición. El comportamiento en

la tercera posición depende de si el genoma es rico en G+C o no, para el primer caso el porcentaje G+C de la tercera posición es más alto que el promedio genómico mientras que para el segundo es más bajo (con algunas excepciones) (Majumdar et al, 2002).

Oxígeno y Metabolismo

El oxígeno es un elemento muy abundante en la tierra formando parte del agua, de las macromoléculas constituyentes de los seres vivos, y libre en la atmósfera. Esto no fue siempre así, el nivel de oxígeno atmosférico ha sufrido grandes fluctuaciones a través del tiempo, siendo escaso hasta mediados del proterozoico (hace aproximadamente 2000 millones de años) (Buick, 2008). La oxigenación de la tierra y los océanos tiene aún muchas preguntas sin resolver pero es generalmente aceptado que el nivel de oxígeno atmosférico muestra un marcado aumento entre 2.4 a 2.0 mil millones de años atrás, durante lo que se denomina Gran Evento de Oxidación (GOE, por sus siglas en inglés) (Holland, 2006).

La explicación más sencilla para la aparición del oxígeno en la atmósfera en altas concentraciones aproximadamente 2.45Ga atrás es la evolución de las cianobacterias que utilizan el fotosistema II. De todas formas esto es discutido, el oxígeno generado por las cianobacterias se cree era insuficiente para contrarrestar el efecto reductor del H₂ y otros reductores en el sistema atmósfera-océanos (Holland, 2006). Exactamente qué proceso dominó la histórica acumulación de oxígeno en atmósfera y océanos es aún sujeto de debate. Hay dos escuelas de pensamiento importantes, una sostiene que la fotosíntesis oxigénica evolucionó miles de millones de años antes de que la atmósfera estuviera significativamente oxigenada ya que tomó gran cantidad de tiempo alcanzar los niveles actuales de oxígeno, y el otro, mantiene que surgió hace 2.4Ga y causó inmediatamente el GOE (Buick, 2008). Otros cambios importantes en el sistema redox de la atmósfera como la disminución en el flujo hidrotermal de H₂ y H₂S, cambios en la biósfera y en el flujo de nutrientes a los océanos pueden haber sido igualmente importantes (Holland, 2006).

En todo caso el rol de la fotosíntesis es claro y fue indispensable para el desarrollo y mantenimiento de la atmósfera actual (Holland, 2006).

Hoy, el oxígeno, es un elemento fundamental para la vida como constituyente de la biomasa y participando en el metabolismo, pero el cambio generado en la atmósfera por el relativamente súbito aumento en su concentración se considera una transformación ambiental de gran importancia (Segré *et al*, 2005). Este cambio debió impulsar una fuerte presión selectiva sobre los organismos, que favoreció el desarrollo de nuevas vías metabólicas y enzimas, pero también restringió a algunos organismos a microambientes anaerobios (Segré *et al*, 2005, Storz *et al*, 1999). La relación con el oxígeno determina qué ambientes son accesibles para cada microorganismo (Lengeler et al, 1999). Según sus requerimientos oxigénicos los procariotas se clasifican como

aerobios, anaerobios estrictos, anaerobios facultativos y microaerófilos, también existen bacterias aerotolerantes (Lengeler et al, 1999).

La adecuación al uso de oxígeno puede aportar muchos beneficios, pero el metabolismo aerobio presenta también algunos problemas como la toxicidad de los derivados del oxígeno: H_2O_2 , O_2^- y OH^- . La toxicidad resulta en la degradación oxidativa de lípidos, daño oxidativo de las bases de ADN lo que genera mutaciones. En proteínas, el daño está asociado a clivaje de la cadena polipeptídica, modificación de aminoácidos y formación de derivados sensibles a la degradación proteolítica (Stadman et al, 2000).

¿Cómo pueden los microorganismos controlar su exposición a ETOs (especies tóxicas del oxígeno)? Pueden adoptar diversas estrategias, gobernadas muchas veces por factores como la generación endógena o exógena de las especies tóxicas. Los microorganismos pueden neutralizar las ETOs mediante varios mecanismos que incluyen las enzimas SOD (superóxido dismutasas), catalasas, peroxidasas, y variedad de reductasas. También pueden modular la concentración de oxígeno intracelular o el potencial redox minimizando su exposición al daño oxidativo (Fridovich, 1998).

Las células procariotas pueden sobrellevar el daño oxidativo a través de efectivos mecanismos de reparación del ADN (Storz et al, 1999), por lo tanto dentro de las adaptaciones a un metabolismo aeróbico se pueden encontrar aquellas enzimas que median la biotoxicidad del oxígeno.

Las catalasas son responsables de la siguiente reacción:



Las peroxidasas usan diversos agentes reductores para reducir H_2O_2 a $2H_2O$. Estas dos enzimas componen la defensa ante el H_2O_2 .

La defensa frente al O_2^- es llevada a cabo por las superóxido dismutasas (SODs).

HO^\bullet tiene como blanco particular el ADN y la causa principal de muerte celular por exposición al oxígeno se debe al daño en el ADN, a su vez las lesiones generadas por oxidantes externos en el ADN son la causa de la mayoría de las mutaciones espontáneas (Fridovich, 1998). La presencia de oxígeno en la atmósfera debió aumentar la tasa de mutación y por lo tanto acelerar la posterior evolución generando materia prima para la selección natural (Fridovich, 1998)

Las fuerzas selectivas que influyen el crecimiento de las redes bioquímicas son sujeto de estudio todavía aunque existen algunas aproximaciones a la respuesta.

Como cualquier otro fenómeno biológico, el metabolismo no puede ser sino un resultado de la evolución, su propia complejidad sugiere la de su historia evolutiva (Cunchillos, 2000). La comparación en el ámbito metabólico entre células actuales pone en evidencia, sobre todo en bacterias pero también en células eucariotas, una gran variabilidad, diferentes formas de heterotrofismo, de autotrofismo fotosintético y quimiosintético, diversas formas de respiración o fermentación, pero bajo esta diversidad se comprueba la existencia de un núcleo metabólico

universal, común a todas las células actuales, formado por unas cincuenta rutas metabólicas (Cunchillos, 2000). Las redes metabólicas representan un set de las reacciones bioquímicas totales que se dan en una célula (Parter, 2007).

Un estudio realizado en proteobacterias con el objetivo de determinar los mecanismos genéticos que contribuyen al crecimiento de las redes metabólicas, y cuál es el rol de la selección en estas modificaciones, develó que los genes que más contribuyeron a la evolución metabólica son ambiente-específicos mientras que los que se mantienen en todo el grupo contribuyen al desarrollo en todos los ambientes. Concluyen que la evolución de las redes metabólicas está dirigida por la adaptación a nuevos ambientes y no por la optimización en los ambientes ya fijados (Pál *et al.* 2005).

Segré *et al.* 2005 realizó una comparación de redes metabólicas posibles bajo condiciones óxicas y anóxicas intentando aclarar el reordenamiento metabólico resultante de este cambio ambiental mayor, concluyó que aunque algunas vías si fueron redirigidas, la mayoría de los cambios se dieron en la evolución de nuevas reacciones y vías (Segré *et al.*, 2005).

El análisis genómico de organismos con diferentes requerimientos de oxígeno debería reflejar estas diferencias.

OBJETIVO

Sabemos que existen diferencias entre genomas de organismos con distintos requerimientos energéticos (Naya *et al.*, 2002) pero dentro de cada genoma buscamos diferencias en la composición génica que se observa en las diferentes enzimas que catalizan las reacciones de las vías metabólicas disímiles entre los organismos con metabolismo aerobio y los que poseen metabolismo anaerobio.

- ❖ Estudiar qué impacto se produce en los genomas cuando ocurren transiciones metabólicas desde la anaerobiosis a la aerobiosis.

Se estudiaran:

- ❖ **Los perfiles funcionales de organismos aerobios y anaerobios para intentar descubrir diferencias de composición que pueden reflejarse debido a las diferentes vías metabólicas entre aerobios y anaerobios.**

- ❖ El aporte de los genes específicos de cada metabolismo energético a las diferencias de composición nucleotídica y aminoacídica comentadas anteriormente.

MATERIALES Y MÉTODOS

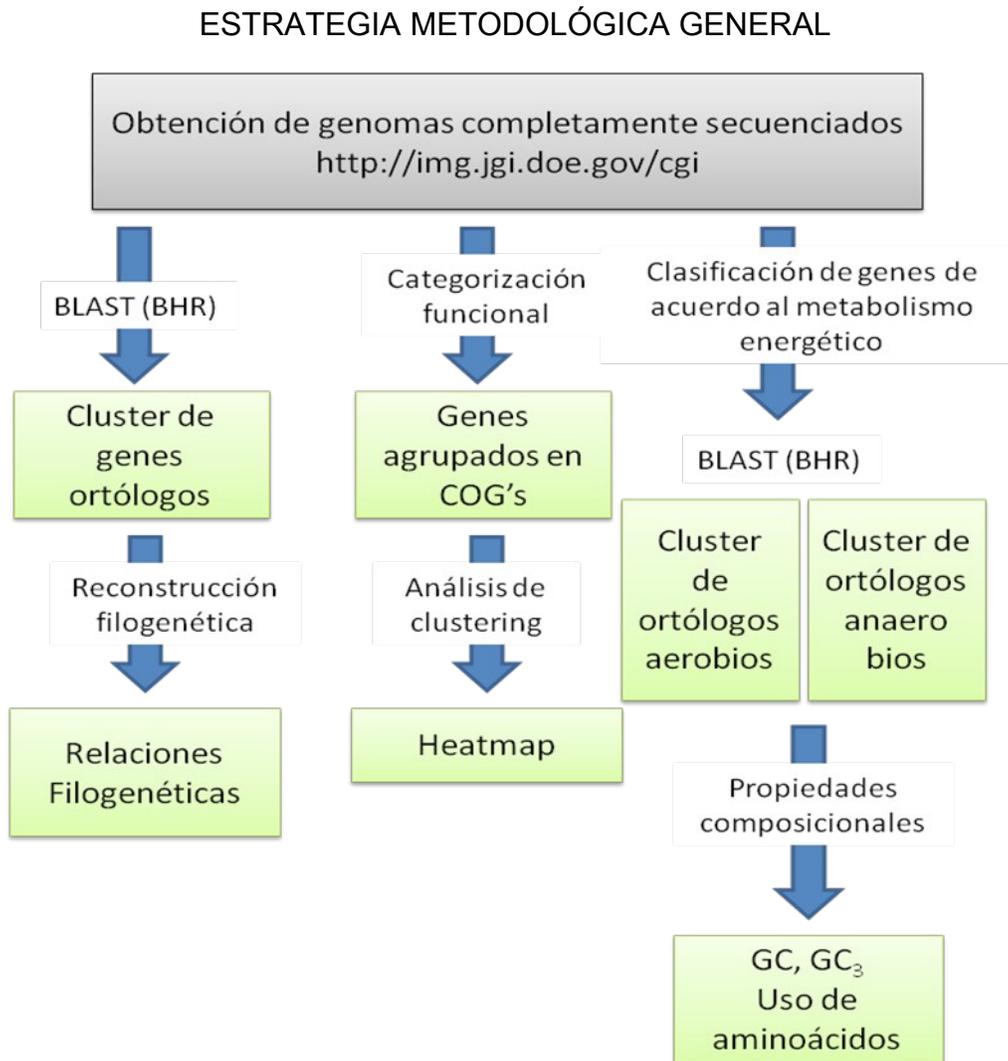


Figura 1. Diagrama de flujo mostrando las tres líneas principales de la estrategia metodológica general. I. Reconstrucción filogenética. II. Análisis de clustering III. Identificación de genes específicos de aerobios y anaerobios para el estudio de sus propiedades composicionales.

DATOS

Se usaron genomas completamente secuenciados pertenecientes a las familias Spirochaetaceae, Alcaligenaceae, Halobacteriaceae, Corynebacteriaceae, Neisseriaceae, obtenidos de la base de datos IMG (<http://img.jgi.doe.gov/cgi>).

Las familias en estudio fueron seleccionadas entre muchas otras por la disponibilidad de información (genomas completamente secuenciados) y por tener diferentes requerimientos oxigénicos dentro del mismo taxa, lo que aporta un contexto de trabajo particular ya que se ha encontrado que los organismos más relacionados filogenéticamente son más concordantes en sus requerimientos oxigénicos (Rocha et al. 2008). Se trabaja a nivel de familia porque en niveles de parentesco más cercano no se encuentran diferencias en cuanto al metabolismo energético y a nivel interfamilia las diferencias aumentan y dificultan atribuir los cambios simplemente uso de oxígeno.

Las especies utilizadas se detallan a continuación y en la Tabla 1.

Spirochaetaceae

Dominio: *Bacteria*

Filo: *Spirochaetes*

Clase: *Spirochaetes*

Orden: *Spirochaetales*

Familia: ***Spirochaetaceae***

Las spirochetas son bacterias helicoidales de vida libre o asociadas a un hospedero, algunas de ellas son patógenos del hombre y otros animales. El estudio de secuencias de ARNr 16S demuestra que son un grupo monofilético dentro de las bacterias. La familia Spirochaetaceae se separa en seis géneros, Borrelia, Brevinema, Cristispira, Spirochaeta, "Spironema" y Treponema.

Alcaligenaceae

Dominio: *Bacteria*

Filo: *Proteobacteria*

Clase: *Betaproteobacteria*

Orden: *Burkholderiales*

Familia: ***Alcaligenaceae***

Los análisis de secuencia demuestran que la familia Alcaligenaceae pertenece al grupo de las *Betaproteobacterias*. Son bacilos o coco bacilos Gram-negativos. Aerobios estrictos aunque algunas especies pueden utilizar nitrato o nitrito como aceptor de electrones, permitiendo así el crecimiento en anaerobiosis (Don *et al.* 2005). Las especies analizadas pertenecen al género *Bordetella*, cuya mayoría son aerobios estrictos y están adaptadas a la vida en estrecha asociación con organismos superiores, son patógenos importantes, comensales o patógenos oportunistas (The Prokaryotes: a handbook on the biology of bacteria. Proteobacteria edited by Martin Dworkin, Stanley Falkow, Eugene Rosenberg, Karl-Heinz Schleifer, Erko Stackebrandt). Dentro de este grupo se encuentra también *Bordetella petrii*, que es la excepción dentro del género ya que es anaerobia y de vida libre (Gross *et al.*, 2008).

Halobacteriaceae

Dominio: *Archaea*

Filo: *Euryarchaeota*

Clase: *Halobacteria*

Orden: *Halobacteriales*

Familia: ***Halobacteriaceae***

Todos los miembros de la familia son halófilos y se los encuentra en gran variedad de ambientes hipersalinos. Presentan una gran variedad morfológica que se debe a la gran diversidad de esta familia que se divide en 14 géneros.

Corynebacteriaceae

Dominio: *Bacteria*

Filo: *Actinobacteria*

Clase: *Actinobacteria*

Subclase: *Actinobacteridae*

Orden: *Actinomycetales*

Suborden: *Corynebacterineae*

Familia: ***Corynebacteriaceae***

Esta familia pertenece al orden Eubacteria, *Corynebacterium* es uno de los géneros más numerosos de las actinobacterias con más de 50 especies, son bacilos gram positivos, inmóviles (Don *et al.* 2005).

Las corinebacterias son habitantes comunes de la mucosa y piel del hombre. Algunas especies son conocidas por sus efectos patógenos en humanos y otros animales (Yassin *et al.*, 2003).

Tabla1. Familias usadas en el estudio y datos relevantes de cada una

Familia	Organism Name	Oxygen Req	Genome Size(Mb)	GC Content(%)	Habitat	Temp. range
Alcaligenaceae						
	<i>Bordetella avium</i> 197N	Aerobio	3,7	61,6	En asociación	Mesofílico
	<i>Bordetella bronchiseptica</i> RB50	Aerobio	5,3	68,1	En asociación	Mesofílico
	<i>Bordetella parapertussis</i> 12822	Aerobio	4,8	68,1	En asociación	Mesofílico
	<i>Bordetella pertussis</i> Tohama I	Aerobio	4,1	67,7	En asociación	Mesofílico
	<i>Bordetella petrii</i> DSM 12804	Anaerobio	5,3	65,5	Acuatico	Mesofílico
grupo externo	<i>Cupriavidus taiwanensis</i>	Facultativo	3,4	68	En asociación	Mesofílico
Spirochaetaceae						
	<i>Borrelia burgdorferi</i> 64b	Microaerofílico	1,6	28,4	En asociación	Mesifílico
	<i>Borrelia garinii</i> PBr	Microaerofílico	1,2	27,9	En asociación	Mesifílico
	<i>Borrelia</i> sp. SV1	Microaerofílico	1,2	28,3	En asociación	Mesifílico
	<i>Borrelia spielmanii</i> A14S	Microaerofílico	1,2	27,7	En asociación	Mesifílico
	<i>Borrelia hermsii</i> DAH	Aerobio	0,9	29	En asociación	Mesifílico
	<i>Borrelia turicatae</i> 91E135	Aerobio	0,9	29,1	En asociación	Mesifílico
	<i>Treponema denticola</i> ATCC 35405	Anaerobio	2,8	37,9	En asociación	Mesifílico
	<i>Treponema pallidum</i>	Anaerobio	1,1	52,8	En asociación	Mesifílico
grupo externo	<i>Brachyspira hyodysenteriae</i>	Facultativo	3	27	En asociación	Mesifílico
Halobacteriaceae						
	<i>Natronomonas pharaonis</i> DSM 2160	Aerobic	2,8	63,1	Acuatico	
	<i>Halogeometricum borinquense</i> DSM 11551	Aerobic	3,9	60	Especializado	
	<i>Haloarcula marismortui</i> ATCC 43049	Aerobic	3,4	61,1	Marino	Mesofílico
	<i>Halobacterium salinarum</i> R1	Anaerobic	2,7	65,7	Especializado	Termofílico
	<i>Halobacterium</i> sp. NRC-1	Facultative	2,6	65,9	Especializado	Mesofílico
	<i>Halomicrobium mukohataei</i> DSM 12286	Facultative	3,1	65	Especializado	Mesofílico
grupo externo	<i>Halorhabdus utahensis</i> DSM 12940	Aerobic	3,1	62,9	Terrestre	Mesofílico
Corynebacteriaceae						
	<i>Corynebacterium glucuronolyticum</i> ATCC 5186	Facultative	2,7	59,2	En asociación	Mesofílico
	<i>Corynebacterium glutamicum</i> R	Facultative	3,3	54,1	Multiple	Mesofílico
	<i>Corynebacterium efficiens</i> YS-314	Facultative	3,1	63,1	Multiple	Mesofílico
	<i>Corynebacterium diphtheriae</i> NCTC 13129	Aerobic	2,4	53,5	Multiple	Mesofílico
	<i>Corynebacterium matruchotii</i> ATCC 33806	Aerobic	2,9	57	En asociación	Mesofílico
	<i>Corynebacterium striatum</i> ATCC 6940	Anaerobic	2,7	59,4	En asociación	Mesofílico
	<i>Corynebacterium pseudogenitalium</i> ATCC 3303	Facultative	2,5	59,6	En asociación	Mesofílico
	<i>Corynebacterium accolens</i> ATCC 49725	Anaerobic	2,4	59,7	En asociación	Mesofílico
	<i>Corynebacterium aurimucosum</i> ATCC 700975	Facultative	2,7	60,7	En asociación	Mesofílico
	<i>Corynebacterium lipophiloflavum</i> DSM 44291	Facultative	2,3	64,8	En asociación	Mesofílico
	<i>Corynebacterium kroppenstedtii</i> DSM 44385	Facultative	2,4	57,5	En asociación	Mesofílico
	<i>Corynebacterium amycolatum</i> SK46	Aerobic	2,6	58,6	En asociación	Mesofílico
	<i>Corynebacterium urealyticum</i> DSM 7109	Aerobic	2,4	64,2	En asociación	Mesofílico
	<i>Corynebacterium jeikeium</i> K411	Facultative	2,5	61,4	Multiple	Mesofílico
grupo externo	<i>Cellulomonas flavigena</i> 134DSM 20109	Facultative	4,1	74	Vida libre	Mesofílico
Neisseriaceae						
	<i>Neisseria cinerea</i> ATCC 14685	Aerobic	1,8	50,8	En asociación	Mesofílico
	<i>Neisseria flavescens</i> SK114	Aerobic	2,2	49,2	En asociación	Mesofílico
	<i>Neisseria gonorrhoeae</i> FA 1090	Aerobic	2,2	52,7	En asociación	Mesofílico
	<i>Neisseria lactamica</i> ATCC 23970	Aerobic	2,1	52,3	En asociación	Mesofílico
	<i>Neisseria meningitidis</i> 053442	Aerobic	2,2	51,7	En asociación	Mesofílico
	<i>Neisseria mucosa</i> ATCC 25996	Aerobic	2,6	51,2	En asociación	Mesofílico
	<i>Neisseria subflava</i> NJ9703	Aerobic	2,3	49	En asociación	Mesofílico
	<i>Eikenella corrodens</i> ATCC 23834	Anaerobic	2,1	55,8	En asociación	Mesofílico
grupo externo	<i>Nitrosomonas europaea</i> ATCC 19718	Aerobic	2,8	50	Multiple	Mesofílico

ANÁLISIS

Relaciones filogenéticas

BLAST

La construcción de las relaciones filogenéticas intrafamilias se realizó con genes ortólogos que se obtuvieron a través de una búsqueda de homología a través del programa BLAST (Altschul et al, 1990) con el criterio de *mejor hit recíproco* (Figura 1), se usaron las secuencias traducidas generadas a partir de las secuencias nucleotídicas mediante el uso de la herramienta *transeq*.

Ortólogos y parálogos son dos conceptos clave dentro de la genómica evolutiva. Mientras que los ortólogos se relacionan a través de la descendencia vertical de un ancestro común, los parálogos se relacionan a través de eventos de duplicación y eventualmente posterior especiación. Los ortólogos típicamente ocupan el mismo nicho funcional en diferentes especies, mientras que los parálogos tienden a diversificarse funcionalmente (Bratlie et al, 2010).

El mejor hit es una proteína en el genoma usado como base, que es la más similar a una proteína dada en el genoma "query". El fundamento subyacente es que los ortólogos son más parecidos entre ellos que con cualquier otra proteína de los respectivos genomas. En comparaciones de múltiples genomas, los pares de ortólogos potenciales identificados se pueden unir para formar clusters de ortólogos que se encuentran representados en todos o en un set de los genomas analizados (Tatusov et al, 2003). En este caso y como se muestra en la figura 2 los ortólogos son aquellos genes que presentan hit recíproco para cada organismo.

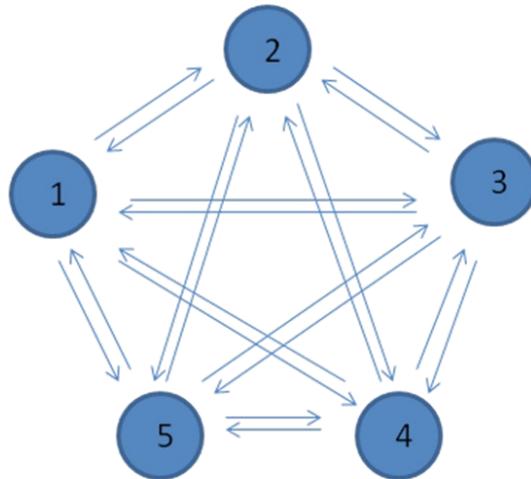


Figura 2. BLAST Best Reciprocal Hit. Cada círculo representa un gen de un genoma. Las flechas indican las relaciones establecidas por el blast, el hit recíproco es necesario para que el gen forme parte del grupo de ortólogos.

Alineamiento

El resultado se procesó y los genes ortólogos obtenidos se alinearon mediante el programa MUSCLE (Edgar, R.C, 2004), el cual lleva a cabo alineamientos múltiples de secuencias. Las secuencias traducidas fueron usadas para la reconstrucción filogenética. MUSCLE usa dos medidas de distancia para la comparación entre dos secuencias: la distancia *kmer* para secuencias sin alinear y la distancia de Kimura para secuencias alineadas. Las secuencias relacionadas tienden a compartir más *kmers*, las cuales consisten de subsecuencias o “palabras”, que lo esperado por azar. Dado un par de secuencias alineadas se calcula su identidad de secuencia lo que se convierte en un estimativo de distancia y se aplica la corrección de Kimura para múltiples sustituciones en un mismo sitio. Las matrices de distancia se agrupan usando UPGMA (*The Unweighted Pair Group Method with Arithmetic Mean*) lo que genera un árbol. UPGMA utiliza un algoritmo que une las dos secuencias mas similares, o las que tienen menor distancia entre sí, estas dos secuencias son tratadas como una unidad y se unen a la siguiente más similar, el algoritmo de clustering es sucesivo (Salemi & Vandamme, 2003). UPGMA es el método de reconstrucción filogenética más sencillo de todos y actualmente no se usa con ese fin.

A partir del orden de ramificación del árbol obtenido por UPGMA, se construye un alineamiento progresivo del cual surge el primer alineamiento múltiple. Para eliminar o minimizar errores, a este primer alineamiento se aplica la medida de distancias de Kimura para cada par de secuencias, lo que genera una nueva matriz de distancias que permite repetir el proceso explicado y generar un nuevo árbol y un nuevo alineamiento. El procedimiento finaliza con el proceso de refinamiento que divide y analiza el árbol en sub-árboles, esto permite afinar las relaciones de distancia entre las

secuencias (Edge R C, 2004). La calidad del alineamiento determina la del árbol filogenético, el alineamiento revela qué sitios se conservan y cuales divergieron a partir de un ancestro común, además el alineamiento puede poseer gaps, lo que indica deleciones o inserciones, y estos datos son también tomados en cuenta para construir las relaciones filogenéticas

Reconstrucción filogenética

El árbol filogenético se obtuvo por el método de máxima verosimilitud utilizando Phylml. Este método usa cálculos probabilísticos para explicar la variación en un set de secuencias. Comienza por establecer un modelo evolutivo de cambio en las secuencias lo que permite estimar las tasas de sustitución. Se estudian todos los árboles posibles, la probabilidad de cada árbol es el producto de las tasas de mutación de cada rama. Para cada columna en el alineamiento se calcula la probabilidad de cada cambio que luego se suman para encontrar la probabilidad de que un árbol dado produzca esa columna en el alineamiento. El procedimiento se repite para todos los sitios del alineamiento, debido a que la tasa de aparición de mutaciones es muy baja, cuanto menos cambios se necesiten para ajustar el árbol a los datos, más probable es (Mount D, 2004).

Reconstrucción de estados ancestrales

El problema de base a resolver en este caso es si la distribución de estados de caracteres en especies actuales provee información sobre estados ancestrales.

En reconstrucción filogenética la asignación del mismo estado a dos o más especies constituye una hipótesis de homología, o sea, los estados de los caracteres se hipotetiza que son similares debido a una continua cadena de herencia desde un ancestro común que también posee ése estado.

La reconstrucción de estados ancestrales se realizó mediante el método de parsimonia, el cual infiere los estados ancestrales minimizando el número de pasos requeridos para explicar la distribución de estados en las especies actuales.

Categorización funcional

Los genes de cada especie fueron agrupados dentro de las “COG categories” previamente establecidas en la base de datos COG (<http://www.ncbi.nlm.nih.gov/COG>). Los COG son clusters de grupos ortólogos de proteínas que en la base de datos KEGG fueron agrupados de acuerdo a la función que cumplen en las categorías que se detallan en la tabla 2.

El procedimiento de construcción de los COGs se basa en la noción de que cualquier grupo de proteínas de genomas distantes que son más similares entre sí que con cualquier otra proteína de sus propios genomas es probable que pertenezcan a una familia de ortólogos, dentro de este sistema se incluye la remoción de parálogos (Tatusov et al, 2000). Los COGs son luego clasificados

en diferentes categorías funcionales. Los COGs tienen varias aplicaciones directas como la predicción de función de proteínas individuales, delinear sets de proteínas candidatas a determinada característica funcional, en general, es una plataforma adecuada para llevar a cabo una variedad de análisis evolutivos con familias de proteínas (Tatusov et al, 2000).

Los genes de cada genoma analizado se encontraron ya distribuidos en las diferentes categorías en la base de datos COG (<http://www.ncbi.nlm.nih.gov/COG>). La distribución de los genes en cada grupo funcional se comparó con técnicas de análisis multivariado, en este caso se usó el programa CLUSTER 3.0.

Tabla 2. Categorías de grupos ortólogos

A	Procesamiento y modificación del ARN
B	Estructura y dinámica de la cromatina
C	Producción y conversión de energía
D	Control del ciclo celular y mitosis
E	Metabolismo y transporte aminoacídico
F	Metabolismo y transporte nucleotídico
G	Metabolismo y transporte de carbohidratos
H	Metabolismo coenzimático
I	Metabolismo lipídico
J	Traducción
K	Transcripción
L	Replicación y reparación de ADN
M	Biogénesis de membrana, pared y envoltura celular
N	Movilidad celular
O	Modificación post-traducciona, chaperonas
P	Metabolismo y transporte de iones inorgánicos
Q	Estructura secundaria
T	Transducción de señales
U	Tráfico y secreción intracelulares
Y	Estructura nuclear
Z	Citoesqueleto
R	General Functional Prediction only
S	Función desconocida

El programa Cluster provee varios algoritmos, en este caso se usó el clustering jerárquico que reúne los ítems en un dendrograma. El clustering permite la asociación de diferentes objetos de acuerdo a su similitud, los ítems con el set de datos que comparten mayor similitud son unidos y

éstos a su vez son asociados con otros, se forma así un dendrograma que refleja el grado de semejanza entre los ítems en estudio, la asociación entre los objetos es máxima si pertenecen al mismo grupo y mínima si ocurre lo contrario. Éste análisis puede usarse para descubrir cierta estructura en los datos.

De esta forma el programa permite vincular a los organismos que tienen perfiles funcionales similares, el resultado puede visualizarse a través de un *heatmap*. El *heatmap* muestra con gradiente de colores la sub y sobrerrepresentación de cada categoría para los diferentes perfiles u organismos.

Identificación de genes específicos de aerobios y anaerobios

A través de un BLAST con el criterio de *mejor hit recíproco* entre organismos aerobios, luego entre anaerobios, se pueden obtener los genes ortólogos compartidos entre las especies que poseen los mismos requerimientos metabólicos. Dentro de este set de ortólogos se encuentran tanto genes exclusivos de aerobios o anaerobios como genes pertenecientes al core (set de ortólogos de toda la familia), para remover estos genes comparamos los obtenidos a partir del blast aerobios y anaerobios con los obtenidos en el blast hecho entre todos los organismos de las familias. De esta forma podemos eliminar los genes que se encuentran en ambos grupos y quedarnos con los que realmente son exclusivos. El objetivo de este procedimiento es intentar establecer qué genes son únicos de cada grupo para analizar luego que contribución hacen a la estructura y composición del genoma.

Propiedades composicionales

Mediante el programa Codonw (Peden, 2005) se analizará el uso de aminoácidos y otras propiedades relevantes (contenido G+C genómico, GC₁, GC₂, GC₃), lo que permite calificar el aporte de cada uno de los genes ya identificados como exclusivos de cada metabolismo, a la composición de su genoma. Codonw trabaja con secuencias y puede realizar numerosos análisis con ellas, permite calcular parámetros genéticos como contenido G+C, uso de codones y de aminoácidos, puede realizar análisis multivariados, cálculo de índices de uso de codones, y de aminoácidos.

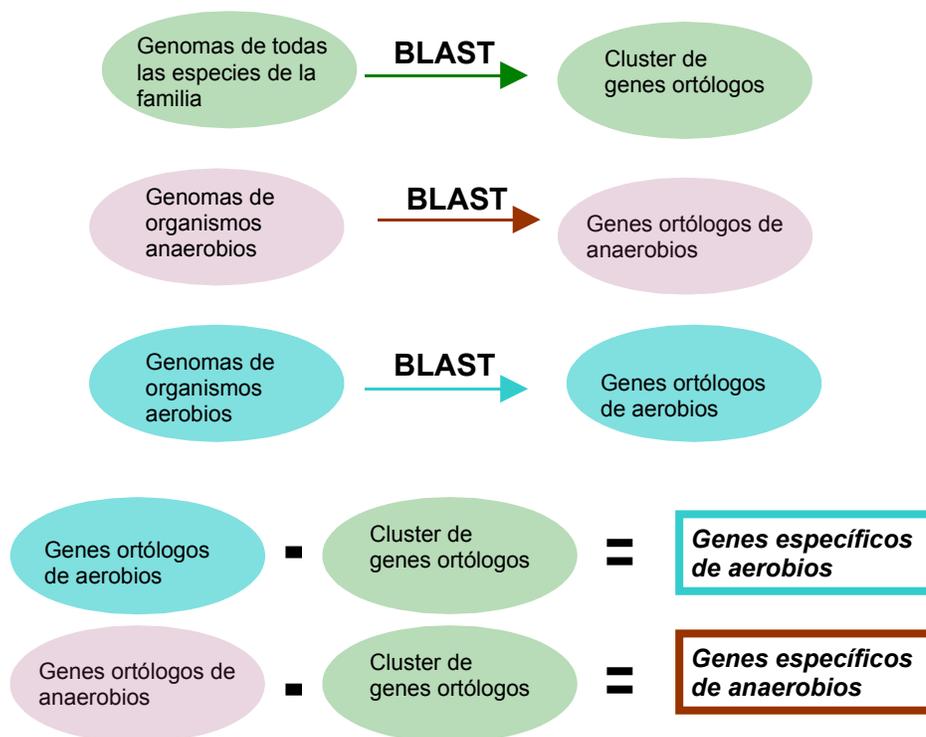


Figura 3. Diagrama mostrando la metodología usada para obtener los genes metabolismo-específicos en aerobios y anaerobios.

El uso de aminoácidos se estudiará para el set de genes ortólogos, y para los genes identificados como metabolismo específicos. Los aminoácidos cisteína (Cys), metionina (Met), triptófano (Trp), tirosina (Tyr), fenilalanina (Phe) y histidina (His) son fuertemente susceptibles a la oxidación (Naya et al, 2002). A su vez, los tripletes que codifican estos residuos están sesgados hacia A+T en los sitios no sinónimos (9 de 14 bases). Específicamente, Tyr y Phe (codificada exclusivamente por A o T en los sitios no sinónimos) son los aminoácidos con la mayor contribución a las diferencias mencionadas antes. Dada la fuerte y generalizada correlación entre el contenido G+C genómico y el contenido en las distintas posiciones de los codones (GC1, GC y GC3), un incremento en el G+C genómico implicaría una disminución de las frecuencias de los residuos más susceptibles a la oxidación (FASO), con la consecuente ventaja selectiva (Naya et al, 2002). FASO se define como la suma de las frecuencias relativas de los aminoácidos antedichos en las secuencias de las bacterias estudiadas, también se va a estudiar cómo varía este carácter.

Los datos generados se tratarán con análisis multivariados, los cuales se usan para simplificar matrices rectangulares en las cuales las columnas representan, en este caso, una medida de uso de aminoácidos, y las filas el organismo al cual corresponden los datos.

El análisis de correspondencia (CA) es una técnica de análisis multivariado que identifica las grandes tendencias en la variación de los datos y distribuye los genes a lo largo de ejes continuos de acuerdo a estas tendencias. El CA mapea los genes del genoma en un espacio multidimensional de acuerdo al uso de los posibles aminoácidos, luego identifica las tendencias más importantes como los ejes del espacio multidimensional que explican la mayor fracción de variación entre los genes (Ermolaeva, 2001).

El contenido de G+C de cada posición de los codones también se obtiene con Codonw y se puede visualizar en histogramas.

RESULTADOS y DISCUSIÓN

Las tres líneas de trabajo diferentes expuestas en la estrategia metodológica general arrojaron diversos resultados, por un lado se consideran los resultados del análisis de clustering y el análisis filogenético, por otro lado analizaremos los estudios realizados sobre los genes metabolismo-específicos obtenidos.

Para cada familia tenemos una filogenia y un heatmap en donde se visualiza cómo está representada cuantitativamente cada categoría funcional en cada perfil funcional, también se pueden ver las asociaciones entre los organismos, resultado del clustering jerárquico.

El análisis filogenético y el funcional se exponen como uno, ya que la filogenia ofrece una referencia para estudiar las relaciones entre las especies, generadas por el clustering. Dentro de éstos se incluye el estudio de las transiciones entre estados metabólicos resultado de la reconstrucción de caracteres ancestrales.

Por último se expondrán los resultados del análisis de propiedades composicionales realizados sobre los genes reconocidos como específicos de aerobios y anaerobios, en los cuales se estudió el contenido G+C, GC₁, GC₂ y GC₃, y uso de aminoácidos.

1-Impacto funcional de las transiciones entre los metabolismos anaerobio y aerobio

Halobacteriaceae

En la figura 4.B se ve el resultado del *clustering* jerárquico. Se puede observar que se agruparon las especies en dos *clusters* bien separados. Uno de ellos está formado por

H.salinarum y dos especies más, una de ellas facultativa, *Halobacterium sp.* y la otra aerobia, *N.pharaonis*. *H. Salinarum* es, de las especies tenidas en cuenta dentro de esta familia, la única que presenta un metabolismo anaerobio. A pesar de que su perfil funcional se diferencia de los perfiles de los microorganismos aerobios, es prácticamente idéntico al perfil de *Halobacterium sp.*

La observación del *heatmap* permite entender porqué *N.pharaonis* forma el *cluster* con *H.salinarum* y *Halobacterium sp.*, pero también se pueden notar grandes diferencias que la separan específicamente tomando en cuenta las siguientes categorías: 'E, H, Q, I, T, U'

Teniendo en cuenta a todos los organismos, tres de las veintidós categorías estudiadas se destacan, 'T' subrepresentada en *H.salinarum* y *Halobacterium sp.*, 'L' y 'D' sobrerrepresentadas en las mismas. La categoría 'T' incluye genes involucrados en la señalización celular, los procariontes monitorean su alrededor a través de sensores intracelulares, sensores unidos a la membrana y sistemas de transporte, vías muy complejas de transducción de señales están ligadas a estos sensores (Lengeler, *Biology of the Prokaryotes*). La categoría 'D' abarca genes implicados en la división celular, control del ciclo celular y partición cromosómica.

La categoría 'L' incluye los genes implicados en replicación, recombinación y reparación del ADN. Las células procariontes pueden sobreponerse a los efectos del daño oxidativo a través de mecanismos eficientes de reparación del ADN, algunas enzimas importantes estudiadas principalmente en *E.coli*, pero presentes en muchos otros procariontes son RecA, RecJ, RecN, RecR, RecG, y las proteínas RuvABC (Stuart L, et al. Evasion of the Toxic Effects of Oxygen, Chapter 15). Esperaríamos que las especies que viven expuestas al oxígeno presentaran la sobrerrepresentación en esta categoría pero los resultados en el *heatmap* muestran lo contrario. Sin embargo esto no es contradictorio ya que a este nivel de granularidad en la clasificación funcional muchas otras enzimas pueden cumplir estos roles.

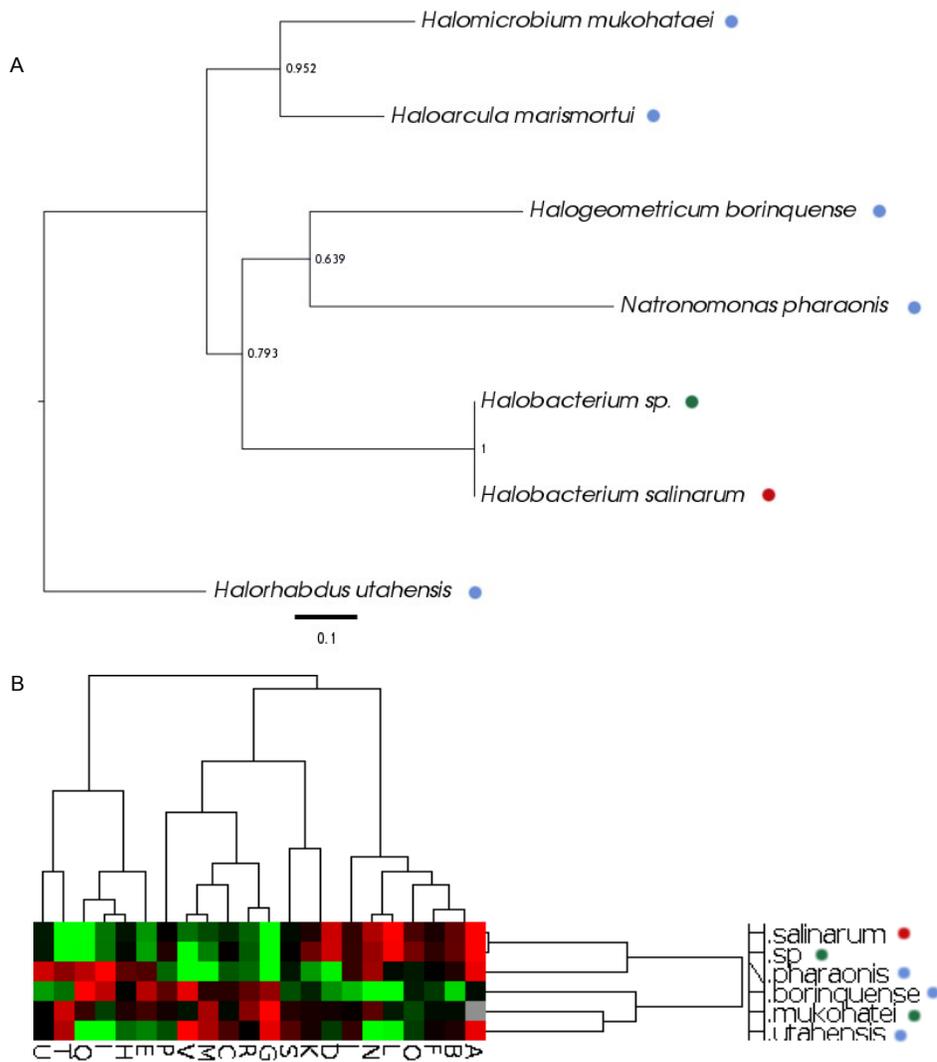


Figura 4. Relaciones filogenéticas (A) y resultados del clustering (B) de la familia Halobacteriaceae.

Aerobios Anaerobios Facultativos

Similares relaciones a las que se desprenden del *clustering* se observan en la filogenia de la familia *Halobacteriaceae* (Fig. 4-A), esto puede deberse a que las agrupaciones de los perfiles estén fuertemente determinadas por las relaciones de parentesco.

Alcaligenaceae

Bordetella petri es la única especie anaerobia considerada en esta familia. En el dendrograma generado por el estudio de *clustering* (Fig. 5-B) la encontramos compartiendo uno, de los dos grupos que claramente se forman, con *B. parapertussis* y *B. bronchiseptica* especies aeróbicas. Aún así, el perfil funcional de *B. petri* se diferencia del resto, siendo destacables los casos de las categorías 'L', 'P', y 'B'. La categoría 'L' incluye los genes implicados en replicación, recombinación y reparación del ADN y se encuentra sobre expresada en *B. petri*. Las categorías 'B' y 'P' se destacan pero por estar menos representadas en *B. petri* que en el resto de las especies, la categoría 'P' engloba a las proteínas encargadas del transporte y metabolismo iónico.

Otro resultado interesante se ve en el *heatmap* al comparar los perfiles funcionales de las especies *B. bronchiseptica* y *B. parapertussis* con los perfiles de *B. avium* y *B. pertussis*, aún cuando sus requerimientos oxigénicos son los mismos los perfiles son completamente opuestos.

En la filogenia (Fig. 5-A) los organismos aerobios forman un grupo monofilético, este es el único caso en que podemos ver los requerimientos oxigénicos asociados a las relaciones de parentesco y las diferencias observadas se pueden estudiar como cambios en la transición de anaerobiosis a la aerobiosis. *B. petrii* es la única especie anaerobia del género *Bordetella* y ha sido aislada de muy diversos nichos ecológicos, los estudios genómicos ya realizados sobre esta especie sugieren que es un eslabón entre las bacterias de vida libre y el resto de las bordetellas, obligatoriamente asociadas a un huésped (Gross et al, 2008). Es posible que la adaptación a un huésped implique la pérdida de genes lo cual puede verse en un principio en la diferencia del tamaño del genoma, más grande para *B. petrii* que para el resto de las especies estudiadas (Gross et al, 2008).

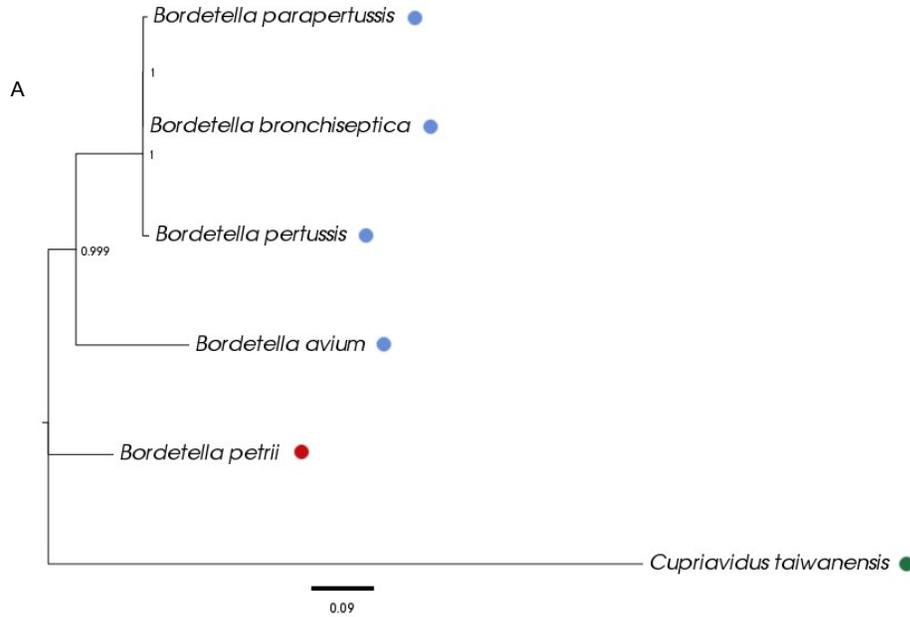
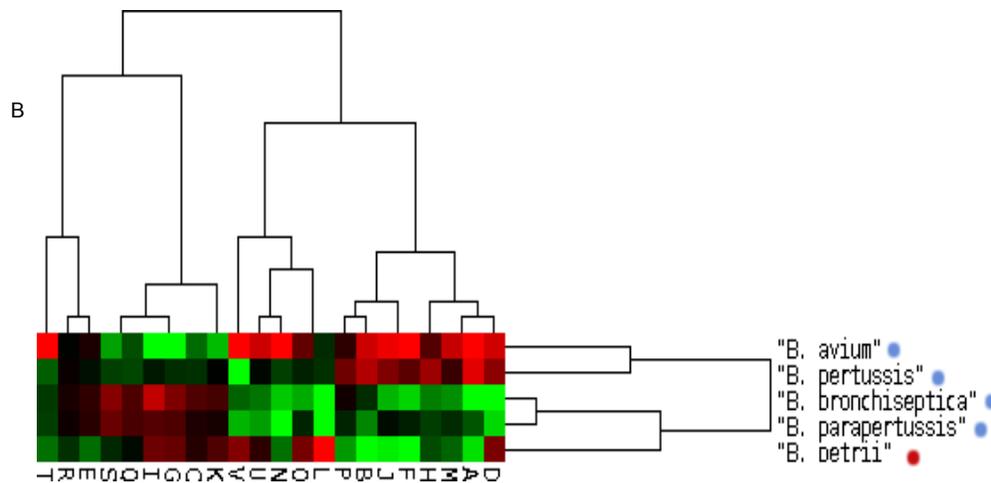


Figura 5. Relaciones filogenéticas (A) y resultados del clustering (B) de la familia Alcaligenaceae.

Aerobios ●

Anaerobios ●

Facultativos ●



Corynebacteriaceae

De todas las familias en estudio ésta es la que presenta mayor diversidad fisiológica en cuanto al uso de oxígeno.

El resultado del clustering no es tan claro como en otras familias pero se pueden ver dos clusters dividiendo a las categorías funcionales. A, B, D, M, H, O, F, J, N, U, forman parte de uno de los clusters y el resto de las categorías se agrupan en el otro. En el heatmap se puede

ver un patrón de expresión casi opuesto para estos clusters, a su vez esto concuerda con los dos grandes clusters que dividen la familia. El cluster formado por A, B, D, M, H, O, F, J, N, U reúne categorías que globalmente se ven sobrerrepresentadas en *C. accolens*, *C. diphtheriae*, *C. glucuronolyticum*, *C. amycolatum* y *C. kroppenstedtii*, el resto de las categorías son nucleadas al mismo nivel observándose sobrerrepresentadas en las especies restantes.

Aunque se puede percibir la separación de dos grupos a nivel de las especies, esta división no muestra asociación con los requerimientos de oxígeno.

El dendrograma no separa organismos con distintos requerimientos oxigénicos sino que tanto aerobios, anaerobios como facultativos se ven dispersos. Esto nos está indicando que la similitud entre perfiles funcionales no se asocia al metabolismo energético, podría estar asociado a las relaciones filogenéticas, para lo cual es necesario analizar la filogenia.

Es posible que las categorías B y N, en las cuales faltan los datos para varias de las especies, no estén aportando información por lo que deberían ser removidas para intentar mejorar el estudio.

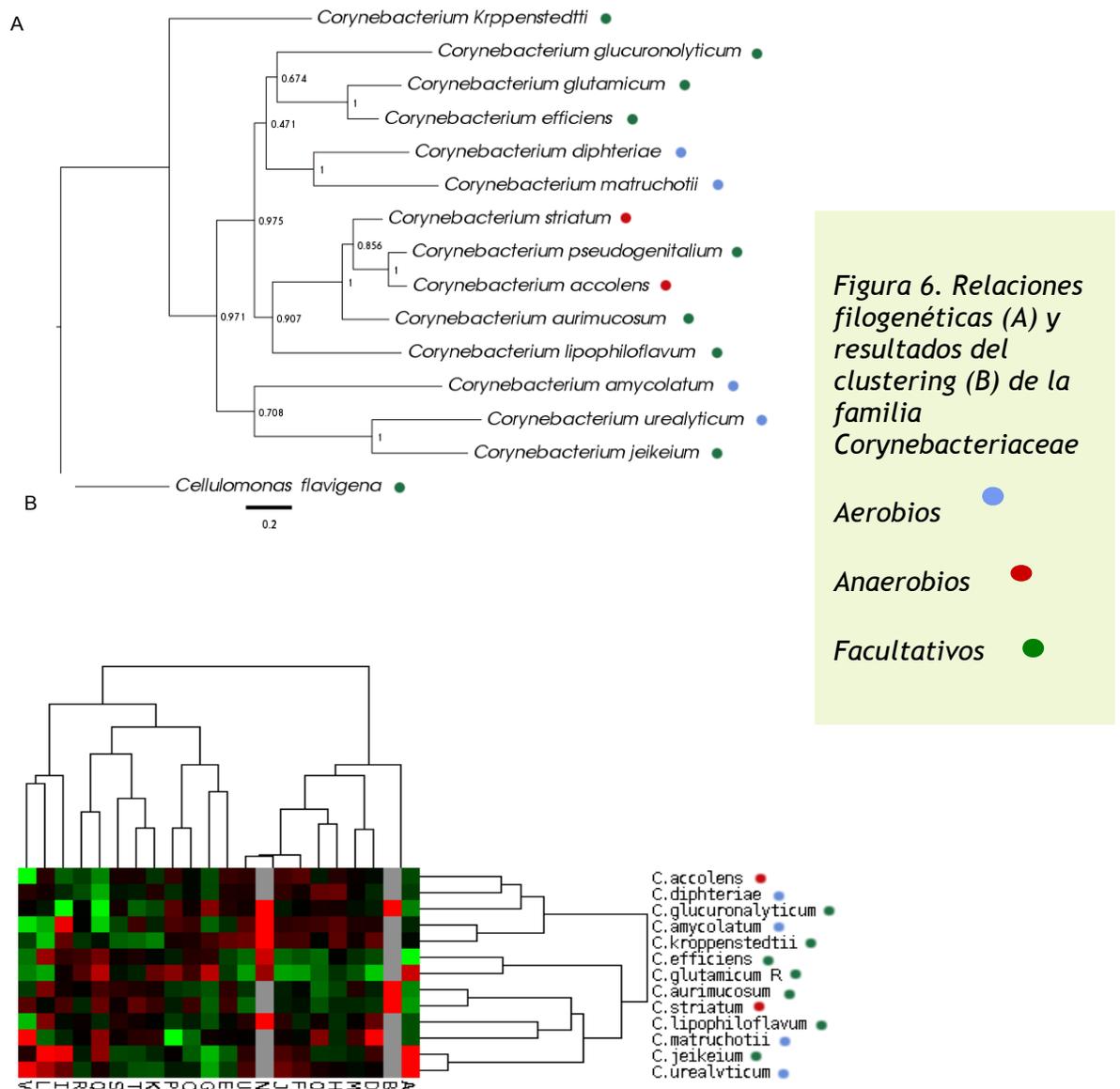
En general la topología del clustering no coincide con la del árbol, son pocos los casos de especies que se asocian en ambos casos, por ejemplo, *C. urealyticum* y *C. jeikeium*, *C. glutamicum* y *C. efficiens*.

En cuanto a los requerimientos de oxígeno dentro de la filogenia, se ve que las especies aerobias no forman un grupo monofilético sino que la aerobiosis estricta aparece en dos eventos diferentes. Esto no sucede en el caso de la anaerobiosis en el que las dos especies consideradas poseen un ancestro común posiblemente anaerobio.

La comparación entre ambos resultados permite ver que especies cercanamente emparentadas en el filogenia poseen patrones opuestos en el *heatmap*, como es el caso de *Corynebacterium striatum*, *Corynebacterium aurimucosum*, *Corynebacterium accolens*. Esto implica que el grado de parentesco de las especies no determina el perfil funcional por lo menos para la familia *Corynebacteriaceae*, el cual es un grupo muy diverso que abarca patógenos e integrantes de la flora humana habitual. Probablemente la especialización de nicho implicó tanto pérdida como ganancia de genes y generó perfiles funcionales en gran medida determinados por esto, podemos decir que otras variables ecofisiológicas están influyendo en el análisis de clustering.

Spirochaetaceae

Esta familia posee la particularidad de contener especies microaerófilas, la microaerofilia implica que el organismo posee enzimas que se inactivan en condiciones de alta disponibilidad de oxígeno, en condiciones de baja presión de oxígeno y ausencia de aceptores de electrones alternativos exhiben crecimiento oxígeno-dependiente (Don *et al.* 2005).



Las treponemas presentan perfiles funcionales idénticos (Fig.7B), ambas son anaerobias pero además se encuentran muy cercanamente emparentadas. *B. turicatae* es una especie aeróbica que se asocia con las treponemas en el dendograma aunque si se observa el *heatmap* se ve que su perfil tiene grandes diferencias con los perfiles de *T.denticola* y *T.pallidum*, de hecho hay una fuerte discordancia en las categorías 'E' y 'V' sobrerrepresentadas en *B.turicatae* y en las categorías 'L, M, O, y U', sobrerrepresentadas en las treponemas. Los genes dentro de la clase 'E' son genes asociados al metabolismo y transporte aminoacídicos, mientras que los incluidos en la clase 'V' están asociados a los mecanismos de defensa. Nuevamente la categoría 'L' se destaca encontrándose subrepresentada en *B. turicatae*.

La categoría funcional C, la cual agrupa a los genes involucrados en producción y conversión de energía, se ve enriquecida en *T.pallidum* y en *T.denticola* con respecto al resto. La cantidad de genes presentes en esta categoría refleja la complejidad bioenergética de cada especie (Castresana, 2001).

B.hermsii también es aeróbica pero se observa asociada a las especies microaerófilicas y su perfil es completamente distinto al de *B.turicatae*. Al contrario de lo esperado la diferencia de composición entre las categorías no consigue nuclear a los organismos aerobios en un mismo grupo pero sí a las especies anaerobias.

La filogenia (Fig.7-A) permite ver que las relaciones generadas en la clusterización son muy similares a las relaciones de parentesco.

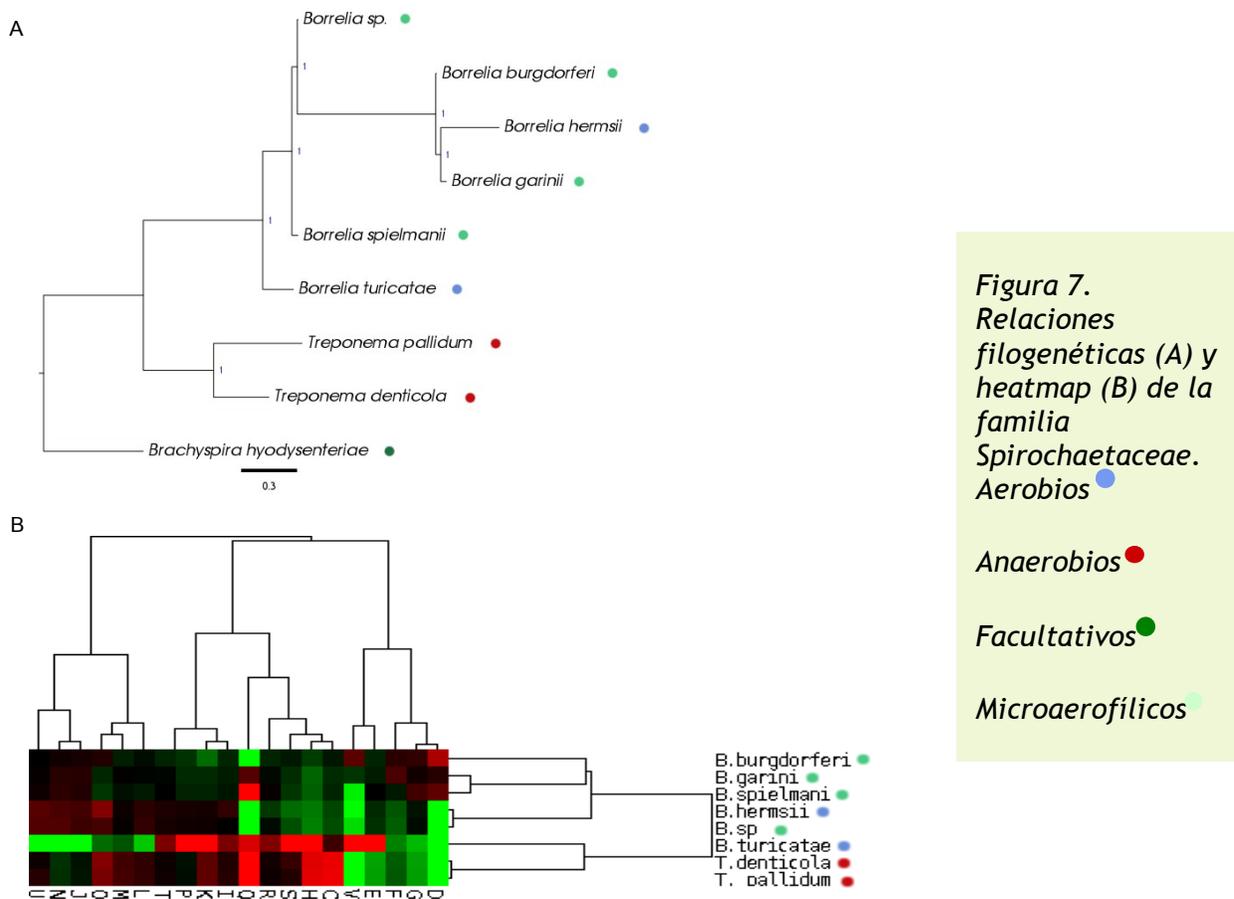


Figura 7. Relaciones filogenéticas (A) y heatmap (B) de la familia Spirochaetaceae. Aerobios ● Anaerobios ● Facultativos ● Microaerófilicos ●

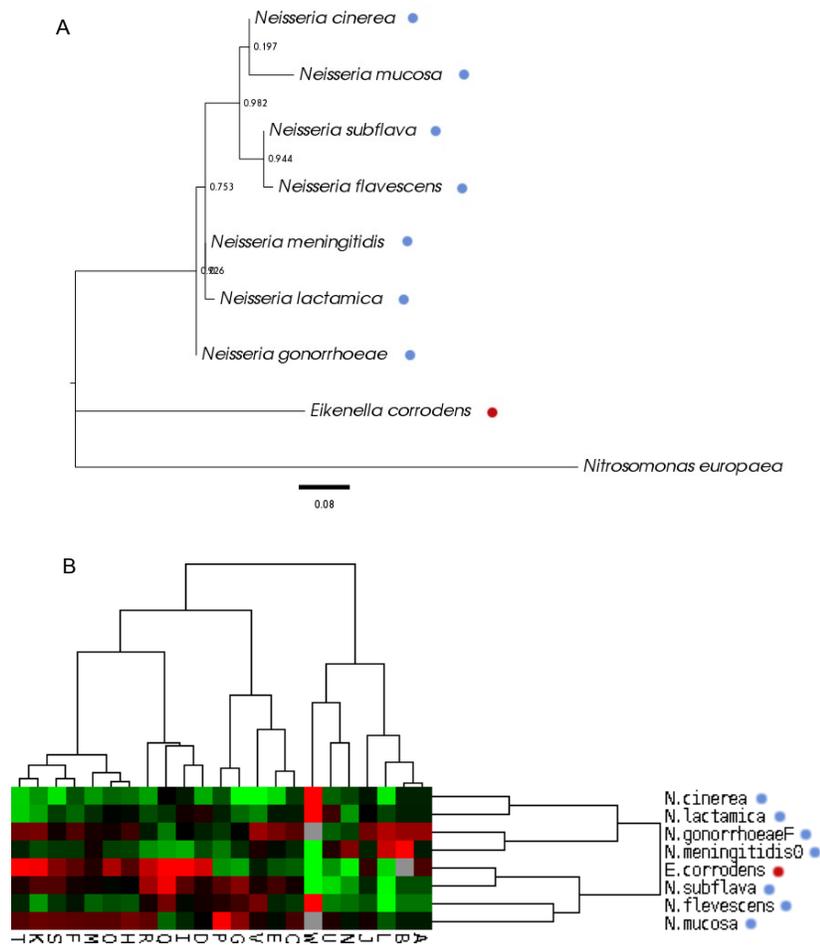


Figura 8. Relaciones filogenéticas (A) y heatmap (B) de la familia Neisseriaceae. Aerobios ● Anaerobios ● Facultativos ● Facultativos ●

Neisseriaceae

Sólo una de las especies consideradas, *Eikenella corrodens*, posee un metabolismo energético anaeróbico, esperaríamos que se separara del resto durante la formación de los clusters pero no es el caso (Fig.8B).

Las categorías que se ven sobrerrepresentadas en *E.corrodens* son, 'D' que agrupa a los genes que controlan el ciclo celular, la división celular y partición cromosómica, la categoría 'I' que representa genes implicados en el metabolismo y transporte lipídicos, 'K' que engloba lo relativo a la transcripción y la 'T' que posee los genes encargados de la transducción de señales.

Las expectativas en este caso eran poder observar en *E.corrodens* un perfil completamente diferente a los perfiles de las neisserias pero el dendrograma de las especies generado de acuerdo

la representación de genes en cada categoría no devela ese patrón, sí la vemos distanciada en la filogenia en la cual las neisserias forman un grupo monofilético.

Aunque muchos organismos de esta familia son comensales de mamíferos o parte de la flora normal, el género *Neisseria* incluye importantes patógenos humanos, todas las *Neisserias* e incluso *Eikenella corrodens* han sido asociados a gran variedad de enfermedades.

En la filogenia (Fig.8A) las neisserias forman un grupo monofilético, esto es coherente con reconstrucciones filogenéticas a partir de ARNr 16s (Don *et al.* 2005), ya que *E. corrodens* pertenece a un género distinto dentro de esta familia. Este es otro caso en que se ve la transición anaerobiosis-aerobiosis.

Transiciones entre estados metabólicos

Tabla 3. Número de transiciones entre los distintos metabolismos energéticos. O+: Metabolismo aerobio, **O-:** Metabolismo anaerobio, **F:** Metabolismo facultativo

De\A	O+	O-	F
O+	0	1	2
O-	0	0	0
F	5	4	0

Los resultados que se observan en la tabla 3 surgen de la reconstrucción de estados de caracteres ancestrales mediante parsimonia. Se observan en total doce transiciones entre los estados metabólicos facultativo, aerobio y anaerobio, para las cinco familias en consideración.

Las transiciones son más frecuentes desde el estado facultativo a el aerobio y el anaerobio, lo que nos permite hipotetizar que es más fácil llevar a cabo este tipo de cambio que el que implica la transición de un metabolismo estricto a otro, esto puede verse como la habilidad de perder o ganar la capacidad de llevar a cabo un metabolismo determinado.

La mayoría de las transiciones se dan desde el metabolismo facultativo, en el cual se posee la capacidad de crecer (o sobrevivir) tanto en condiciones de aerobiosis, como en ausencia de oxígeno, a los metabolismos estrictos aerobio y anaerobio, esto implica la pérdida de capacidades. Esto último supone que es más fácil o probable que sucedan cambios a través de la perdida de una habilidad que a través de la adquisición.

Cada familia aporta en diferente medida a los cambios observados, son más influyentes las que muestran mayor diversidad de estados metabólicos.

Distribución de categorías funcionales de acuerdo a los clusters formados

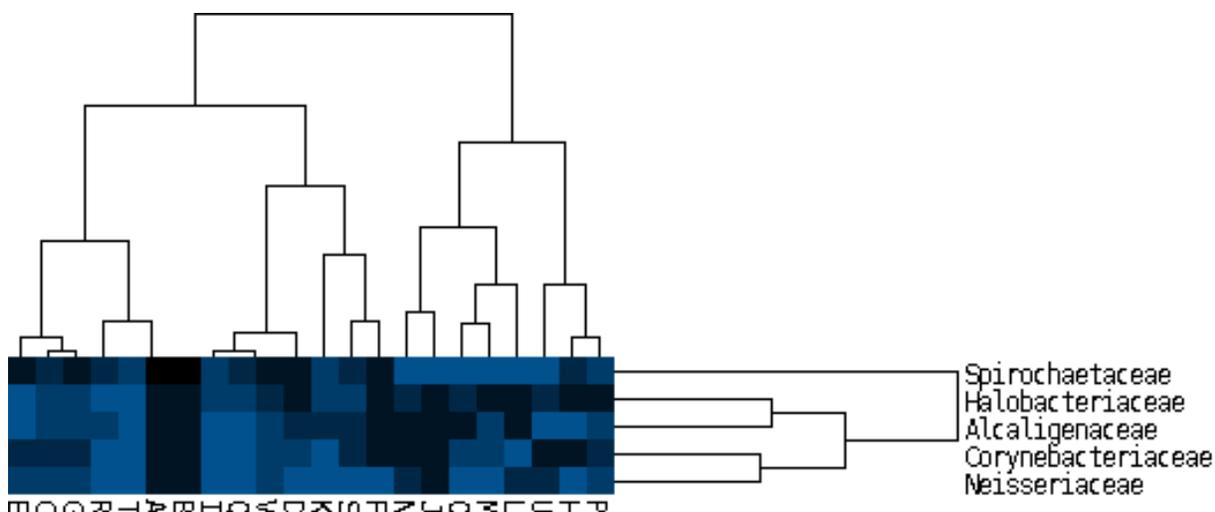


Figura 9. Heatmap del análisis de clustering realizado para los 4 grupos en que se reunieron las categorías funcionales.

Este estudio se realizó para observar si existe alguna asociación de las categorías en cuanto a la distribución en los dendrogramas.

Se definieron cuatro clusters de categorías para cada familia de bacterias, basado en el segundo nivel de disociación que se observa en cada dendrograma, y se asoció a cada categoría el número del cluster (1, 2, 3 o 4) del cual formaba parte en cada una de las familias analizadas. Los 4 grupos en los cuales se colocaron a las diversas categorías se definieron a partir de las primeras dos ramificaciones del clustering para cada caso.

A y B siempre que están presentes forman parte de un mismo grupo, luego encontramos que algunas se asocian frecuentemente (F, T, J), pero no hay categorías que se reúnan en todos los casos.

En algunas de las familias analizadas en cuanto a las distribución de genes dentro de los subsistemas metabólicos se encontraron coincidencias pero no se devela un patrón importante que pueda distinguir a las especies de acuerdo a sus requerimientos oxigénicos, esto estaría indicando que otros aspectos de la ecología y fisiología de los organismos dominan los perfiles funcionales.

Los procariotas poseen un diverso espectro de estrategias bioenergéticas lo cual se ve reflejado en estos resultados e impide la segregación de especies aerobias de anaerobias a través del método usado. Los organismos analizados aquí incluyen además parásitos, endosimbiontes y patógenos que durante el camino recorrido en la adaptación a su nuevo ambiente u hospedero es probable hayan perdido rutas metabólicas importantes (Castresana, 2001). Es importante aclarar que el análisis de clustering está basado en la anotación de los genes en grandes categorías funcionales y que se toma en cuenta cuántos genes posee el organismo en esa categoría no cuáles son esos genes, por lo tanto existen diferencias que no podemos apreciar a través de este estudio.

Si se ven categorías como L, O, T, que son importantes en cada caso y marcan las diferencias, pero la sub o sobrerrepresentación de éstas no se puede asociar a los requerimientos energéticos de las especies.

El trabajo realizado por Segré *et ál.* 2005 mostró que la disponibilidad de oxígeno permite que se lleven a cabo redes metabólicas de mayor complejidad que las alcanzables dentro de un metabolismo anaerobio, el número de ortólogos presentes en la categoría C, la cual agrupa a los genes asociados a los procesos de obtención y conversión de energía, debería reflejar la complejidad metabólica de los organismos (Castresana, 2001), la cual está íntimamente asociada al ambiente en el cuál ese organismo se encuentra, no sólo por las condiciones del medio sino porque en una considerable medida determina los metabolitos disponibles para el organismo.

Es importante observar que en este estudio la categoría C no se destaca dentro del resto de las categorías como una variable importante, si comparamos la cantidad de genes presentes en el grupo funcional C en cada especie vemos que ese número prácticamente no varía dentro de las familias y tampoco se notan diferencias entre organismos aerobios y anaerobios, sí varía cuando se realiza una comparación interfamilia, las especies que poseen menor cantidad de genes en C son las que pertenecen a la familia Spirochaetaceae y las que poseen los mayores valores son las agrupadas en la familia *Alcaligenaceae*. Si se da en algunos casos como el de *Neisseriaceae* que el porcentaje de genes dentro del genoma, que pertenecen al grupo C, es menor en las especies anaerobias, pero esta no es una relación que se mantenga para el resto de las familias.

En general los clusters formados durante el análisis funcional no están determinados por los requerimientos de oxígeno de los organismos, muchas veces imitan la disposición en el árbol filogenético pero con importantes excepciones como el caso de *Corynebacteriaceae* en que son totalmente diferentes, pero en éste caso la diversidad eco-fisiológica es realmente grande, habiendo

organismos con formas de vida que implican un alto grado de especialización, por ejemplo parasitismo, y esto puede estar viéndose reflejado en el estudio de perfiles metabólicos.

Como se mencionó en la introducción la aerobiosis ha sido descrita como una variable de alta inercia filogenética, lo cual hace difícil el estudio de este rasgo entre bacterias cercanamente emparentadas, en este caso la elección de las familias se debió a que dentro de ellas hay organismos con diferentes requerimientos de oxígeno.

La mayoría de las transiciones metabólicas que se dan en las filogenias del análisis son desde el metabolismo facultativo a la aerobiosis o anaerobiosis, solo se da un caso de transición de aerobiosis estricta a anaerobiosis. Los organismos facultativos tienen la capacidad de usar el oxígeno como aceptor final de electrones y además poseen enzimas que les permiten vivir en anaerobiosis. La transición de este estado a la aerobiosis o anaerobiosis estricta no implicaría un cambio tan radical como el de las transiciones aerobiosis-anaerobiosis.

Existen estudios que dan mucha importancia a la transferencia horizontal de genes en la evolución de las vías metabólicas (Pal et al, 2005), esto puede explicar la diversa gama de estados que se observa dentro de una misma familia como es el caso de *Corynebacteriaceae*. También se ha observado una fuerte relación entre las vías metabólicas de organismos del mismo género (Forst y Schulten), esto también se vió en las familias estudiadas aquí, en las que generalmente los requerimientos de oxígeno cambian entre especies de géneros diferentes.

2-Propiedades Composicionales

Distribuciones del contenido G+C

Los antecedentes indican que existe relación entre el contenido G+C y los requerimientos oxigénicos en procariontas (Naya et al, 2002), en la tabla 1 se expone el porcentaje de G+C genómico para cada organismo en estudio, de la simple observación de éstos datos se puede concluir que no existen diferencias importantes entre especies aerobias y anaerobias, a diferencia de lo notado anteriormente.

Los histogramas que se presentan en este apartado comparan la distribución del contenido G+C para la primera, segunda y tercera posición de los codones, entre los genes del core y los genes metabolismo específicos tanto aerobios como anaerobios en cada familia. Las tres posiciones tienen, en general, un comportamiento que se repite en los diferentes genomas. La primera posición siempre tiene un porcentaje mayor que el promedio genómico, en la segunda sucede lo contrario, y la tercera es la más variable en éste sentido ya que es alta o baja si lo es

el porcentaje G+C promedio del genoma. Los histogramas de contenido G+C para las posiciones de los codones presentados aquí muestran, en general, concordancia con estas observaciones

El uso de aminoácidos está directamente influenciado por el contenido en G+C genómico.

La disminución en el porcentaje de G+C disminuye la probabilidad de observar alguna de estas bases en la segunda posición de los codones, lo que aumenta la probabilidad de que sea ocupada por A o T, esto permitiría a escala global disminuir la frecuencia de uso de aminoácidos susceptibles a la oxidación (Cys, Met, Trp, Tyr, Phe y His) los cuales son codificados en su mayoría por codones ricos en A+T. A su vez la variación en esta posición es la más estrecha (Muto & Osawa, 1987), la segunda posición de los codones muestra baja variabilidad debido a la selección purificadora sobre los cambios no-sinónimos (Rocha et al, 2010).

En estos resultados el contenido G+C de la segunda posición de los codones no tiene un patrón común para todos los aerobios ni para todos los anaerobios.

En Alcaligenaceae (Figura 11) y Halobacteriaceae (Figura 10) el porcentaje de contenido G+C en la primera y tercera posición se adapta a las generalidades, los genomas de éstas familias son ricos en G+C y por lo tanto la tercera posición es rica también, a su vez, la primera posición posee un porcentaje mayor que la segunda. Las curvas de distribución de GC₂ muestran un mayor porcentaje en los genes específicos de aerobios que en el core de esas especies, en cambio las especies anaerobias de estas familias tienen menor contenido G+C en la segunda posición de los genes específicos. Esto aunque se ajusta a lo comentado anteriormente y podría explicar las diferencias de composición aminoacídicas notadas en la literatura, no se continúa para el resto de las familias en las que las distribuciones no tienen estas coincidencias.

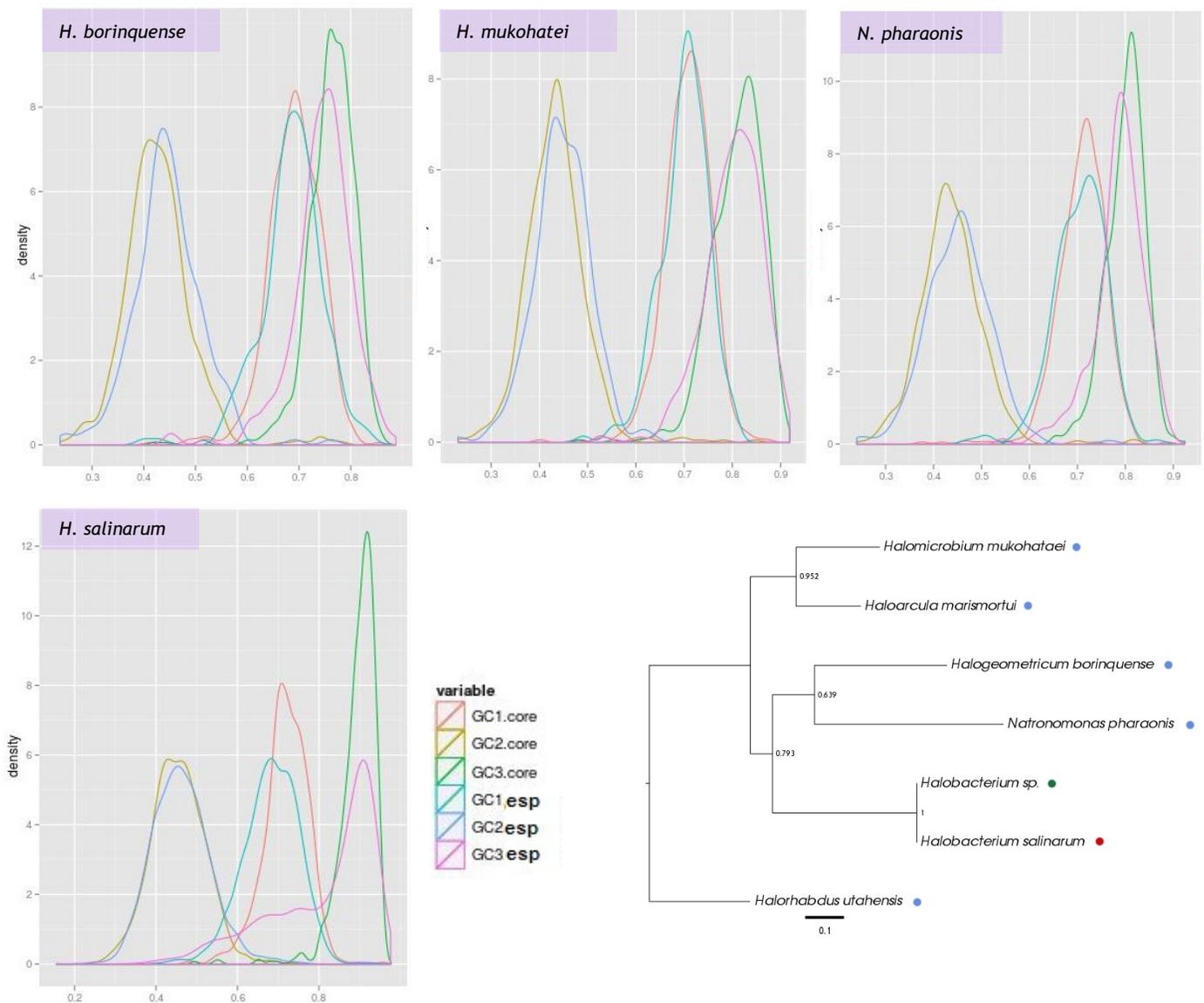


Figura 10. Histogramas mostrando la distribución de G+C para las tres posiciones de los codones tanto para genes específicos como para el core y árbol filogenético de la familia Halobacteriaceae

En las especies aerobias de la familia Spirochaetaceae (*B. hermsii* y *B. turicatae*) (Figura 12) las tres posiciones muestran el comportamiento general comentado, la segunda posición tiene contenido en G+C mayor en los genes del core en la mayoría de los casos estudiados. GC₃ es bajo para todas las especies, esto concuerda con el contenido G+C bajo de toda la familia, la excepción en éste caso es *T. pallidum*, ya que el porcentaje de G+C en la tercera posición es muy amplio, también en esta especie se observan diferencias en la segunda posición, los genes metabolismo específicos poseen un contenido G+C mayor que el core.

En la familia Neisseriaceae (Figura 14) la variación entre posiciones se ajusta al modelo general. Una misma posición, no tiene diferencias entre core y genes específicos en aerobios y

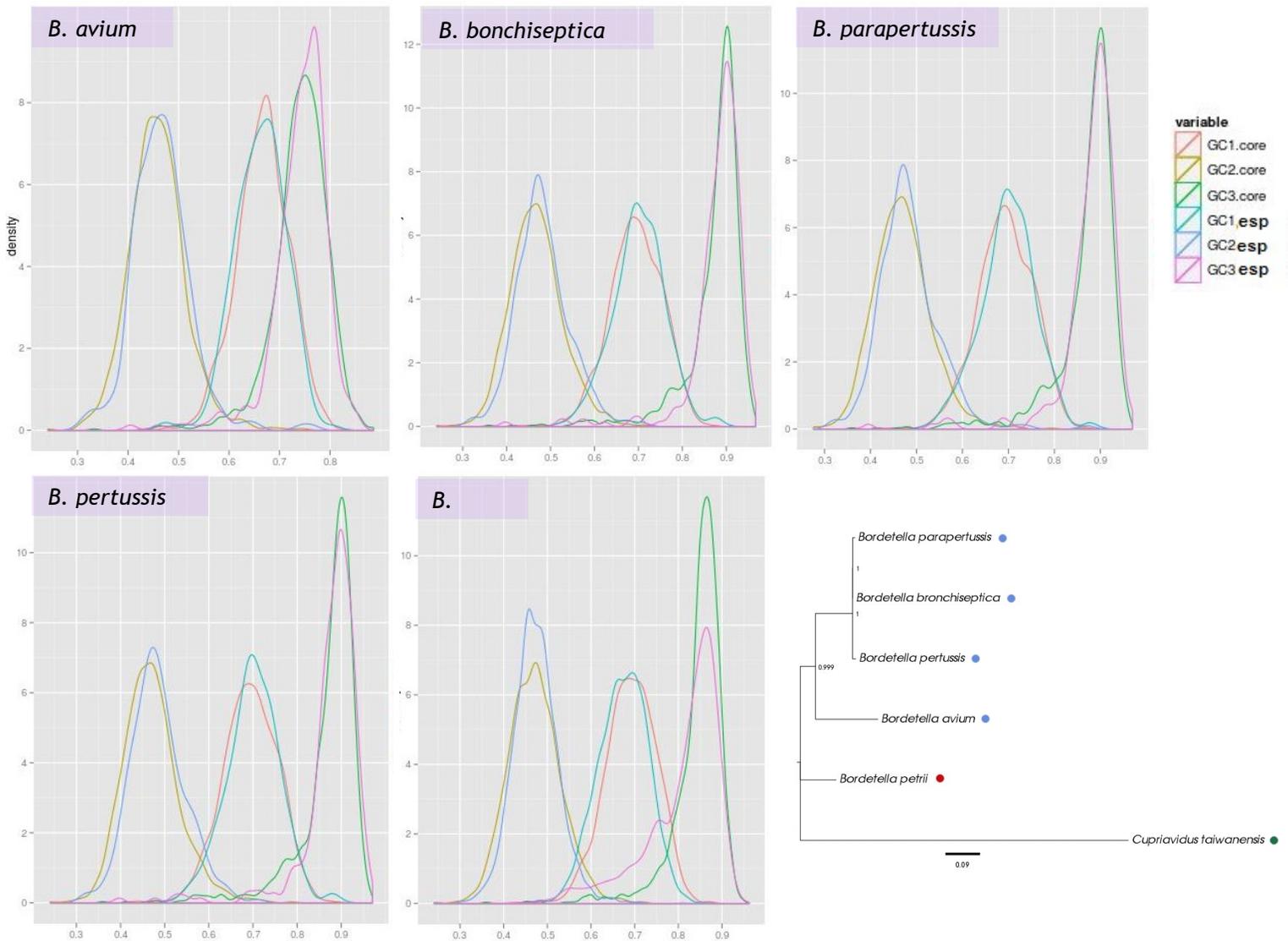


Figura 11. Histogramas mostrando la distribución de G+C para las tres posiciones de los codones tanto para genes específicos como para el core y árbol filogenético de la familia Alcaligenaceae

sí tiene en la especie anaerobia (*E. corrodens*), siendo menor el porcentaje en el core que en los genes específicos para las tres posiciones.

Dentro de la familia Corynebacteriaceae (Figura 13) las posiciones de los codones poseen la forma general excepto por un caso, el de *C. diphtheriae*, en el cual la distribución de GC₃ es anormalmente amplia. En cuanto a GC₂, todas las especies sin distinción de requerimientos oxigénicos, poseen un porcentaje más alto en genes específicos, la única diferencia es que en el caso de los anaerobios es mayor.

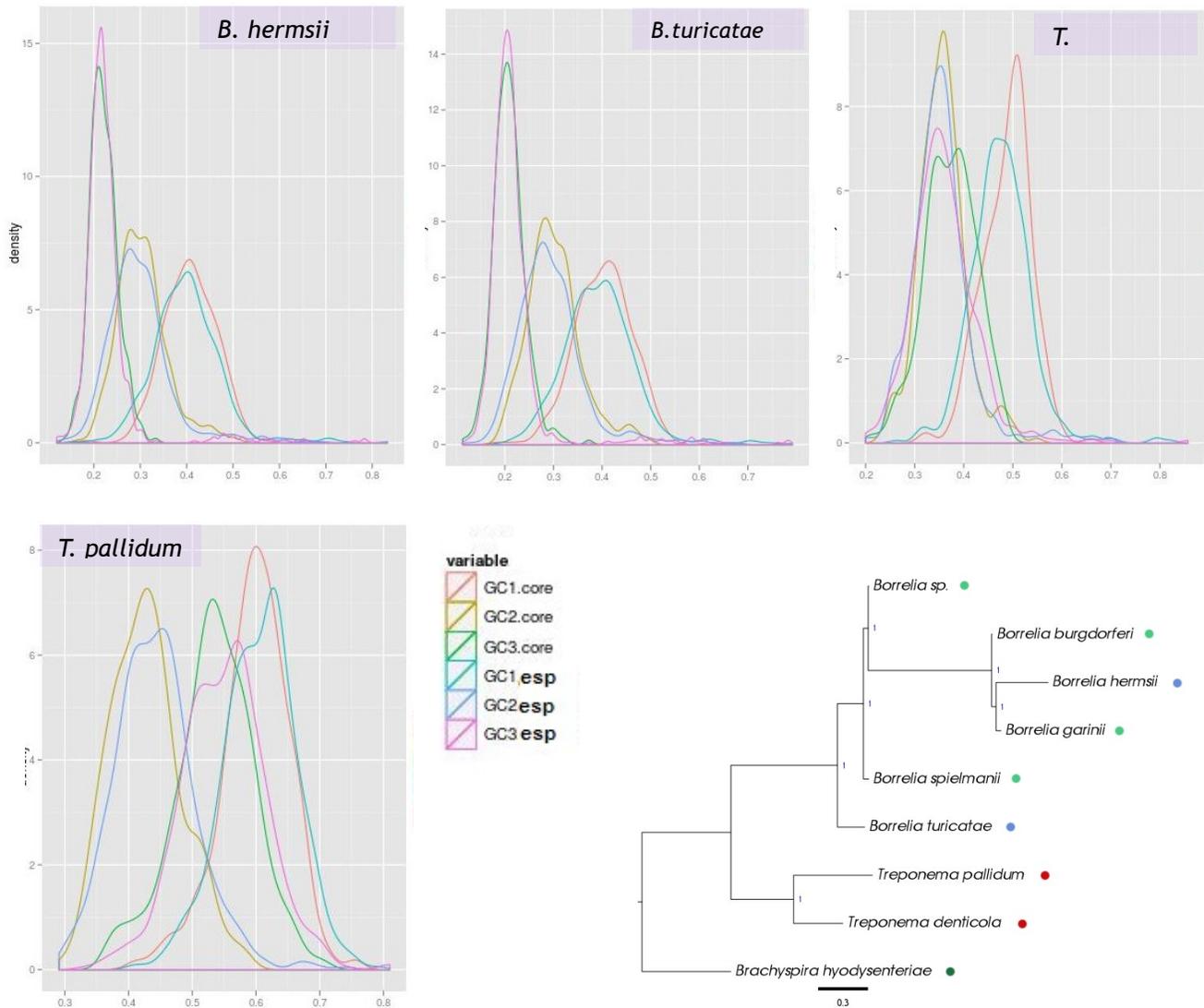


Figura 12. Histogramas mostrando la distribución de G+C para las tres posiciones de los codones tanto para genes específicos como para el core y árbol filogenético de la familia Spirochaetaceae

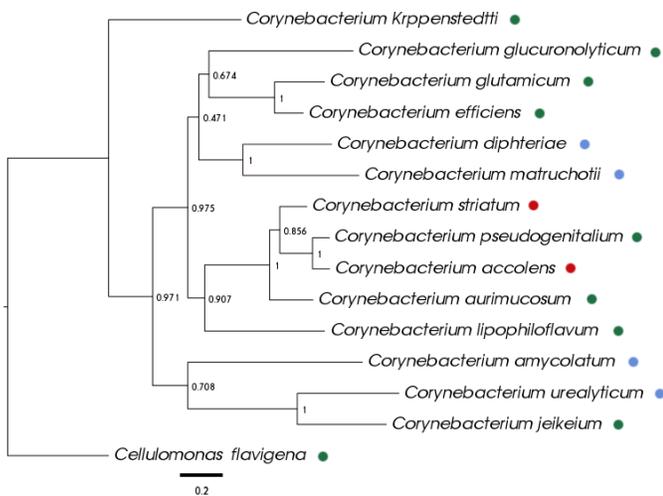
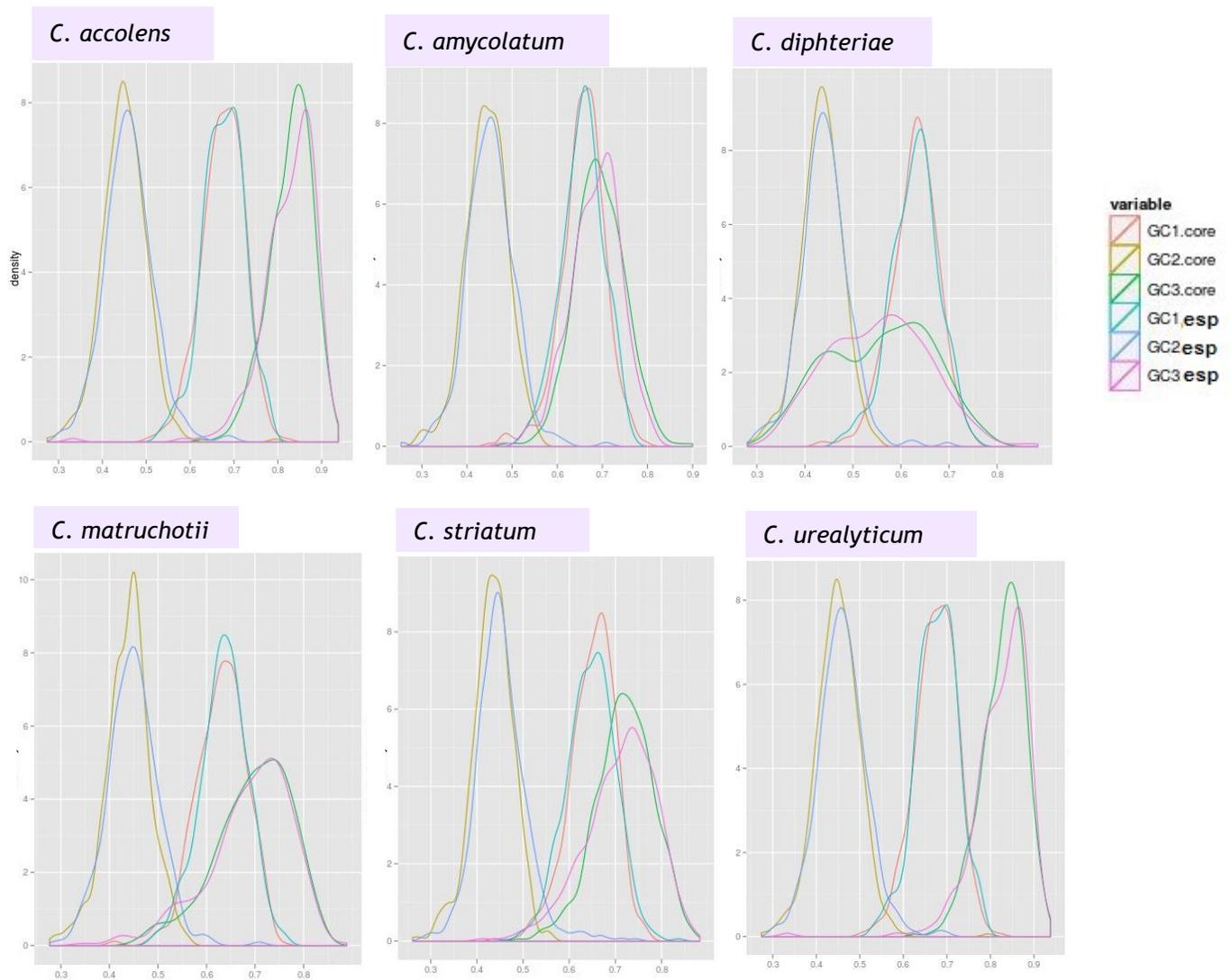


Figura 13. Histogramas mostrando la distribución de G+C para las tres posiciones de los codones tanto para genes específicos como para el core y árbol filogenético de la

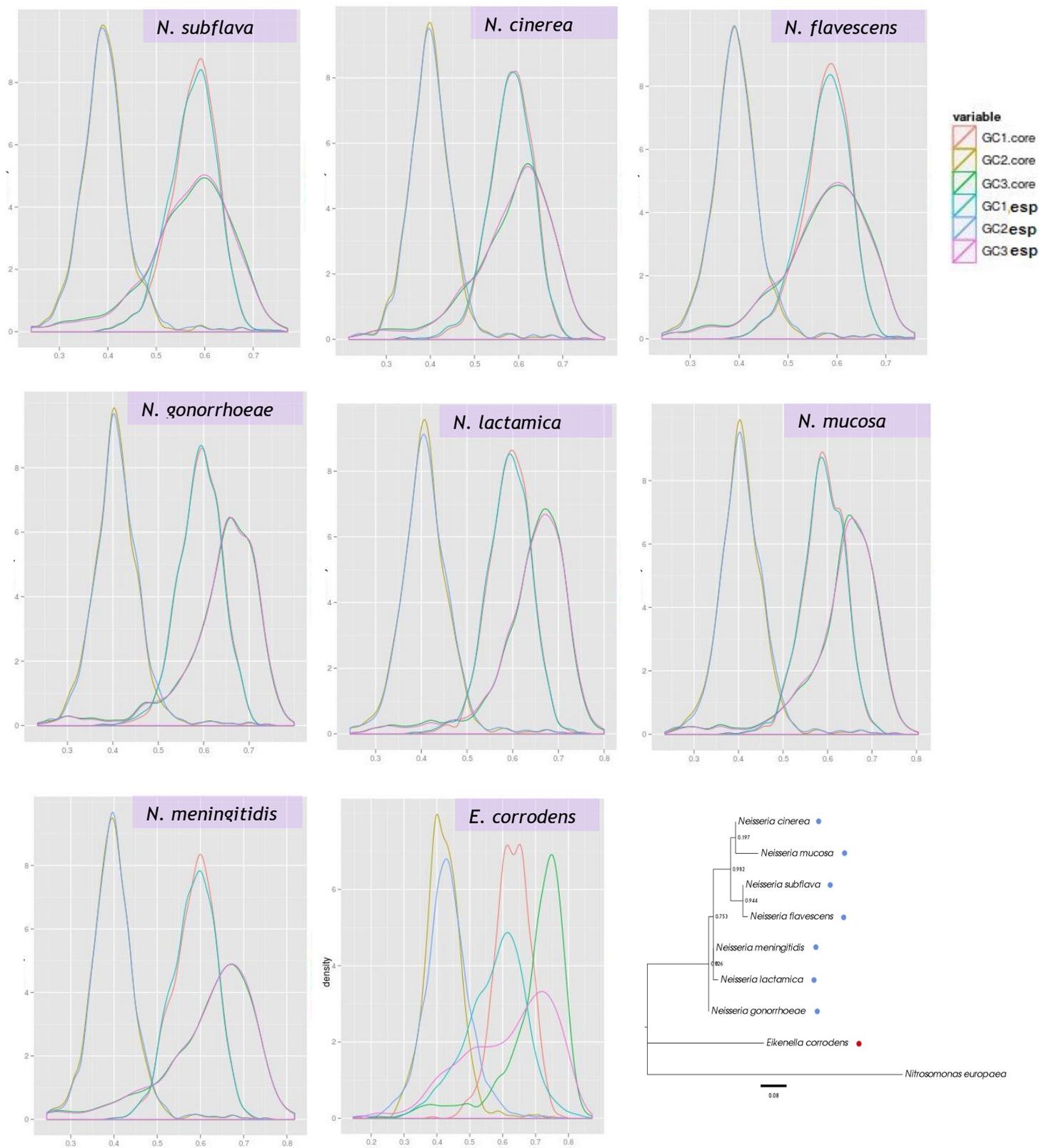


Figura 14. Histogramas mostrando la distribución de G+C para las tres posiciones de los codones tanto para genes específicos como para el core y árbol filogenético de la familia Neisseriaceae

Uso de aminoácidos

La figura 15 muestra el resultado del análisis de correspondencia realizado con los datos de uso de aminoácidos de los grupos de genes ortólogos de cada familia.

Se observa que las especies se distribuyen principalmente agrupándose de acuerdo a las relaciones de parentesco, vemos cómo en las pequeñas aglomeraciones que se forman, el factor común es la familia a la que pertenecen los organismos. Aún así hay algunas especies que sobresalen, por ejemplo en el caso de la familia *Spirochaetaceae*, las dos especies anaerobias *T.dentícola* y *T.pallidum* no se agrupan con el resto y en el caso de la familia *Neisseriaceae*, *E.corrodens* se ubica por fuera del grupo de las neisserias.

En ambas familias las especies que se alejan del grupo son anaerobias mientras que las que se agrupan son aerobias en el caso de *Neisseriaceae* y aerobias y microaerofílicas en el caso de *Spirochaetaceae*, a su vez éstas especies están distanciadas filogenéticamente del resto, son de distinto género que el resto del grupo, por lo tanto no se puede asociar la distancia simplemente con las diferencias metabólicas.

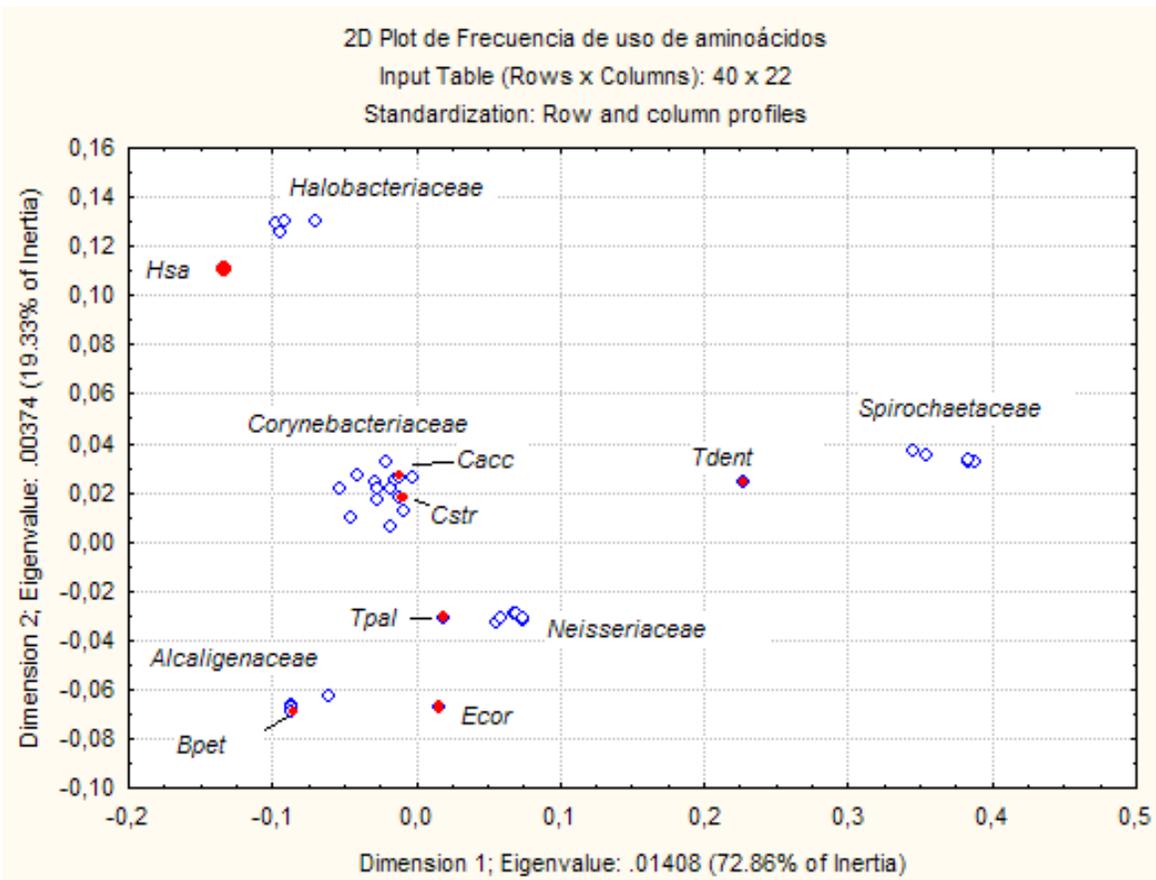


Figura 15. Resultado del análisis de correspondencia realizado con las frecuencias de uso de aminoácidos de genes del core de cada especie. Aerobios ○ Anaerobios ●

Se estudiaron también las frecuencias de uso de aminoácidos susceptibles a la oxidación (FASO).

Los resultados del análisis de FASO se pueden separar en dos partes, por un lado se hizo una comparación entre organismos aerobios y anaerobios, tanto para el uso de estos aminoácidos en genes metabolismo específicos (Fig.16) como para el cluster de genes ortólogos de las familias (Fig.17). Por otra parte se comparó dentro de cada familia el FASO, contrastando entre aerobios y anaerobios en genes metabolismo específicos y genes del core.

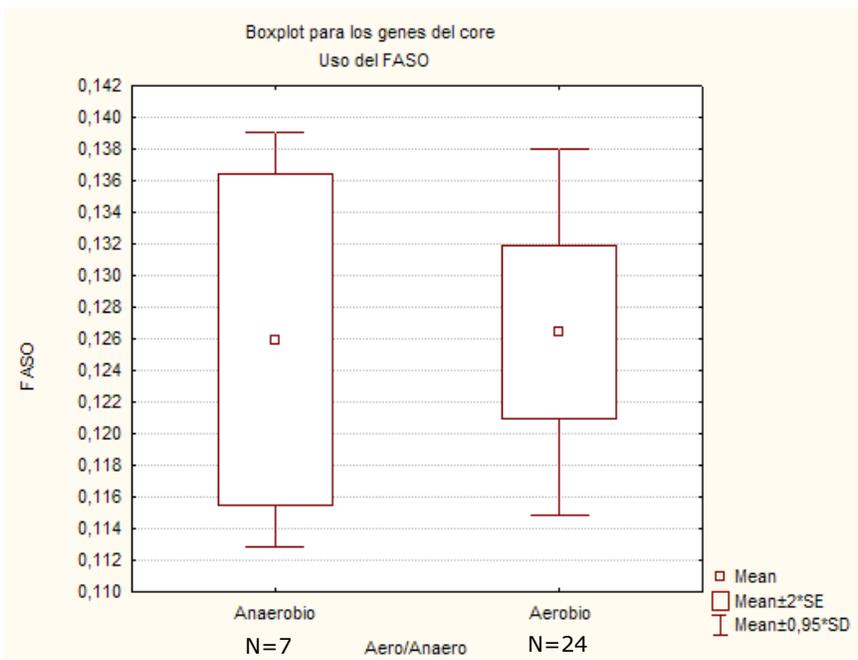


Figura 16.
Comparación de FASO entre aerobios y anaerobios sólo para los genes del core.

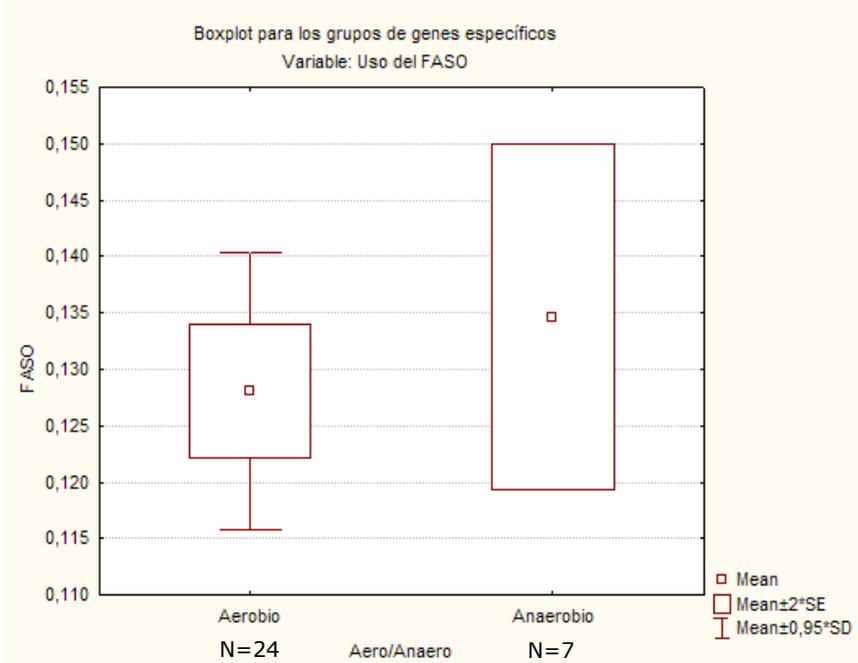


Figura 17.
Comparación de FASO entre aerobios y anaerobios sólo para los genes específicos.

En la primera instancia, la media para el uso del FASO de los genes del core es similar entre aerobios y anaerobios, la distribución es un poco más amplia para los anaerobios, esto puede verse en la figura 16.

Las frecuencias de uso de FASO calculadas para los genes metabolismo-específicos de anaerobios muestra un comportamiento diferente (Fig.17), su media se encuentra por encima de la calculada para el core y también es mayor que la de los aerobios.

Spirochaetaceae

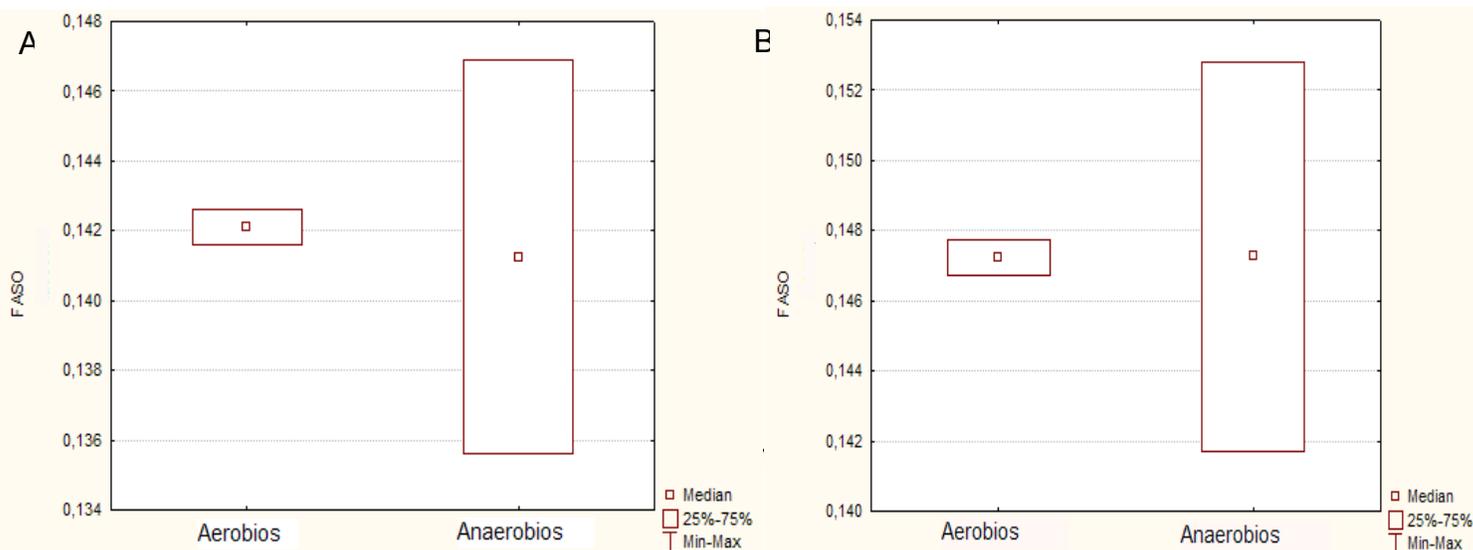


Figura 18. Comparación de el FASO entre aerobios y anaerobios A. Genes del core. B. Genes específicos

Corynebacteriaceae

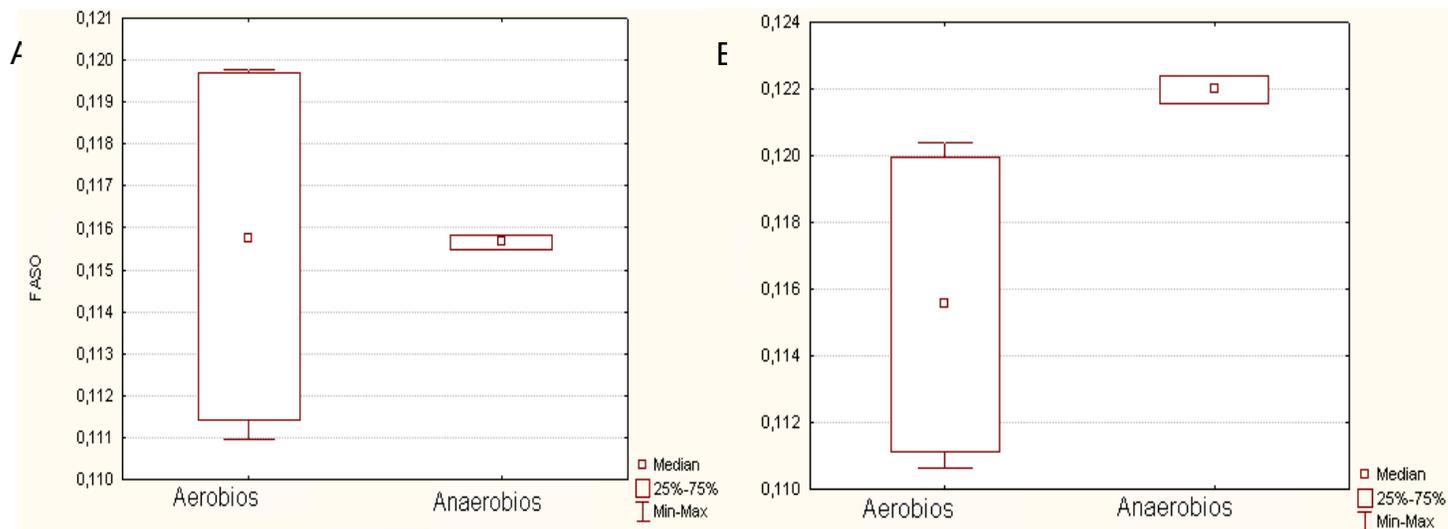


Figura 19. Comparación de el FASO entre aerobios y anaerobios A. Genes del core. B. Genes específicos

La frecuencia media obtenida para aerobios a partir de los genes metabolismo específicos no muestra diferencias con la frecuencia media de los genes del core de los mismos.

Sólo en anaerobios hay diferencias entre el FASO de genes del core y metabolismo-específicos, mostrando los últimos una media más alta.

La comparación intrafamilia de FASO se muestra sólo para las familias que tienen más de una especie en cada categoría. Los resultados (incluidos los que no se muestran) dieron en general un aumento del FASO para los genes específicos de anaerobios con respecto al core. Las especies aerobias no tienen diferencias entre genes específicos y core.

Como se vio anteriormente, se analizó mediante CA (Análisis de Correspondencia) como varían las frecuencias de uso de aminoácidos para todos los organismos en estudio, esto mostró con algunas excepciones que las especies pertenecientes a la misma familia se agrupan lo que indica que la variabilidad dentro de las familias es baja pero no entre familias. Excepto el caso de un par de familias (*Corynebacteriaceae* y *Alcaligenaceae*), los organismos anaerobios se separan del resto de las especies. Esto último indica diferencias en el uso de aminoácidos entre especies aerobias y anaerobias de una misma familia, pero no se agrupan especies de acuerdo al requerimiento de oxígeno independientemente de su familia. Cabe destacar que las especies que se separan del resto de su familia son también las más distantes del grupo filogenéticamente.

El CA realizado con el uso de aminoácidos de genes específicos (no se muestra) no tuvo diferencias con el realizado para los genes ortólogos de las familias.

La comparación de FASO entre aerobios y anaerobios realizada para todas las especies sin distinción entre familias mostró que hay diferencias entre los genes metabolismo-específico y los genes del core.

Las principales diferencias aparecen cuando se analizan los genes metabolismo-específicos de anaerobios, la frecuencia media de uso de aminoácidos susceptibles a la oxidación, no se modifica entre core y genes específicos de aerobios, sí cambia en anaerobios siendo mayor la media para el conjunto de genes específicos que para el core.

Se estudió también el aporte de cada familia a las diferencias observadas en el plot general, y se pudo ver que generalmente los organismos anaerobios tienen mayor frecuencia de uso de aminoácidos susceptibles a la oxidación siendo la diferencia más pronunciada cuando se contrasta el FASO de genes específicos de aerobios y anaerobios que cuando se analiza el core. El FASO de los genes específicos de anaerobios es más alto que el de los genes del core, mientras que en aerobios no difiere tanto entre genes del core y genes específicos.

El uso de aminoácidos está directamente influenciado por el contenido en G+C genómico. Se estudió como varía el contenido en G+C para las tres posiciones de los codones (GC_1 , GC_2 , GC_3) en genes del cluster de ortólogos y en genes identificados como específicos.

De acuerdo a lo propuesto en los antecedentes la diferencia de uso de los aminoácidos susceptibles a la oxidación se refleja en el contenido en G+C de los genomas de organismos aerobios y anaerobios, siendo las distintas posiciones de los codones influyentes en distinta medida.

En los resultados obtenidos no se puede establecer correlación entre el contenido en G+C de las posiciones de los codones y las diferencias en el uso de aminoácidos. Esperábamos notar cambios entre el aporte al %GC del genoma de genes ortólogos y de genes metabolismo específicos. Aunque si se observan diferencias en varios casos no responden, en general, a lo esperado.

Conclusiones

En conclusión, el análisis de *clustering* deja ver diferencias entre perfiles metabólicos pero éstas no son las mismas para los organismos que comparten requerimientos de oxígeno sino que en cada familia los perfiles varían de forma distinta y podría atribuirse a otros caracteres eco-fisiológicos, por ejemplo, al estilo de vida y tamaño del genoma.

También parecen ser influyentes en algunos casos las relaciones filogenéticas.

No se pudo detectar un gran impacto debido a la transición anaerobiosis-aerobiosis a nivel de perfiles funcionales, por lo menos al nivel de granularidad que se llevó a cabo en éste estudio.

Se pudieron ver diferencias de uso del FASO entre genes específicos del metabolismo anaerobio y genes del core de especies anaerobias, también entre los primeros y genes específicos para el grupo aerobio. El resultado más notorio en este caso es que el FASO en anaerobios no parece tener ninguna restricción, siempre se observa un aumento en el FASO en genes específicos respecto de los ortólogos de la familia, mientras que en los aerobios no se ven variaciones entre genes del core y genes específicos.

Perspectivas

El trabajo realizado indica que no existe un gran impacto en los genomas al producirse la transición anaerobiosis-aerobiosis pero se necesitaría de todas formas afinar el estudio para poder detectar genes necesarios para llevar a cabo un cambio de metabolismo energético.

El análisis de los organismos facultativos en cuanto a sus genes específicos podría explicar porque la mayoría de transiciones observadas se dan de un metabolismo facultativo a uno estricto y sólo algunas son estricto-estricto.

Los genes identificados como específicos deberían de ser contrastados con las enzimas específicas de metabolismos oxigénicos y no oxigénicos determinadas por el trabajo de Segré (Segré et al, 2005) para desglosar las diferencias entre la aerobiosis y la anaerobiosis que quizás no sean tan importantes a nivel de genes.

BIBLIOGRAFÍA

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.(1990) "*Basic local alignment search tool.*" J Mol Biol. Oct 5; 215(3):403-10.
- Bentley S D, Parkhill J. (2004). "*Comparative genomic structure of prokaryotes.*" Annu.Rev.Genet.38:771-91.
- Bratlie M. S, Johansen J, Sherma B, Huang D W, Lempicki R A y Drabløs F. 2010 "*Gene duplications in prokaryotes can be associated with environmental adaptation*" BMC Genomics, 11:588
- Buick, R. (2008) "*When did oxygenic photosynthesis evolved?*"Phil. Trans. R. Soc. B 27 August 2008 vol. 363 no. 1504 2731-2743
- Camiscol E, Tamarit J, Ros J. (2000). "*Oxidative stress in bacteria and protein damage by reactive oxygen species.*" INTERNATL MICROBIOL (2000) 3:3-8
- Casjens S. 2000. "Borrelia genomes in the year 2000". J Mol Microbiol Biotechnol Oct;2(4):401-10
- Castresana J. (2001) "Comparative Genomics and Bioenergetics" Biochimica et Biophysica Acta 1506 (2001) 147-162

- Cunchillos Chomin (2000) "Las principales etapas de la evolución del metabolismo celular. Una aproximación evolucionista al estudio del metabolismo." Asclepio-Vol. LII-2
- DeLong EF, Karl DM (2005) "Genomic perspectives in microbial oceanography" *Nature*. 437:336–342. 3.
- Don J. Brenner, Noel R. Krieg, George M. Garrity, James T. Staley (2005) *Bergey's manual of systematic bacteriology*. 2nd ed, Springer Press.
- D'Souza MG, Huan J, Sutton S, Romine M, N Maltsev. (1999) "An Environment for Comparative Analysis of Metabolic Subsystems and Automated Reconstruction of Metabolic of Microbial Consortia and Individual Organisms from Sequence Data" *Mathematics and computer science division*. Agronne National Laboratory.
- Ermolaeva D. María. (2001) "*Synonymous Codon Usage in Bacteria*" *Curr. Issues Mol. Biol.* 3(4): 91-97.
- Field B C, Behrenfeld M J, Randerson J T, Falkowski P. (1998) "*Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components*" *Science Vol. 281 no. 5374 pp. 237-240 DOI: 10.1126/science.281.5374.237*
- Foerstner K, von Mering C, Hooper S, Bork P. (2005). "*Enviroments shape the nucleotide composition of genomes.*" *EMBOreports* Dec; 6(12):1208-13.
- Gross R, Guzman C, Sebahia M, Martins dos Santos V, Pieper D, Koebnik R, Lechner M, Bartels D, Buhrmester J, Choudhuri J, Ebsensen T, Gaigalat L, Herrmann S, Khachane A, Larisch C, Link S, Linke B, Meyer F, 276(5313):734-40. Mormann S, Nakunst D, Rückert C, Schneiker-Bekel S, Schulze K, Vorhölter J, Yevsa T, Engle J, Goldman W, Pühler A, Göbel U, Goesmann A, Blöcker H, Kaiser O and Martinez-Arias R. 2008. "*The missing link: Bordetella petrii is endowed with both the metabolic versatility of environmental bacteria and virulence traits of pathogenic Bordetellae*" *BMC Genomics*, 9:449.
- Handelsman, J. 2007. *Metagenomics and microbial communities*. In: *Encyclopedia of the Life Sciences*. John Wiley and Sons, Ltd. Chichester, UK.
- Hassan S, Mahalingam V, Kumar V. (2009) "*Synonymous Codon Usage Analysis of Thirty Two Mycobacteriophage Genomes*" *Advances in Bioinformatics*, Volume 2009, Article ID 316936.
- Ruth Hershberg & Dmitri A. (2008) "*Selection on codon bias*" *Petrov Annu. Rev. Genet.* 42:287–99.
- Hrinrich D. Holland. (2006) "*The oxygenation of the atmosphere and oceans*" *Phil. Trans. R. Soc.* 361, 903-915.
- Imlay J. (2008). "*How obligatory is anaerobiosis?*" *Molecular Microbiology*. 68(4), 801–804.

- Imlay J. (2002). *"How Oxygen Damages Microbes: Oxygen Tolerance and Obligate Anaerobiosis."* ADVANCES IN MICROBIAL PHYSIOLOGY .Vol 46:111-53.
- Irwin Fridovich. (1998) *"Oxygen toxicity: a radical explanation"* The Journal of Experimental Biology 201, 1203-1209.
- Y Kagawa, H Nojima, N Nukiwa, M Ishizuka, T Nakajima, T Yasuhara, T Tanaka & T Oshima. (1984) *"High guanine plus cytosine content in the third letter of codons of an extreme thermophile. DNA sequence of the isopropylmalate dehydrogenase of Thermus thermophilus."* The Journal of Biological Chemistry, 259, 2956-2960.
- Kimura, M., (1985). *"The Neutral Theory of Molecular Evolution Reprint."*, Cambridge
- Lengeler J W, Drews G, Schlegel H G. *"Biology of the prokaryotes"* Georg Thieme Verlag. 984pp (1999)
- Majumdar S, Gupta K S, Sundararajan V S & T. C. Ghosh T C. (1999) *"Compositional Correlation Studies among the Three Different Codon Positions in 12 Bacterial Genomes"*.Biochemical and Biophysical Research Communications.Volume 266, Issue 1, Pages 66-71
- Mount, David W. 2004 *"Bioinformatics. Sequence and genome analysis"* Cold Spring Harbor Laboratory Press. Second Edition
- Musto, H. et al., 2004. *"Correlations between genomic GC levels and optimal growth temperatures in prokaryotes."* FEBS Letters, 573(1-3), 73-77
- A Muto & Osawa S. 1987 *"The guanine and cytosine content of genomic DNA and bacterial evolution."* Evolution. 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. Sep;55(3):260-4.
- Pace N R. (1997), *"A molecular view of microbial diversity and the biosphere"* Science 276(5313):734-40.
- Pál C, Papp B, Lercher MJ. (2005) *"Adaptive evolution of bacterial metabolic networks by horizontal gene transfer"* Nature Genetics; 37(12):1372-75.
- Parter M, Kashtan N, Alon U. (2007) *"Environmental variability and modularity of bacterial metabolic networks"* BMC Evolutionary Biology. 7:169.
- Peden J. 2005. Codonw. v.1.4.2. Available from: <http://sourceforge.net/projects/codonw/>
- Robert C. Edgar (2004) *"MUSCLE: multiple sequence alignment with high accuracy and high throughput"* Nucleic Acids Research. 32(5) Pp. 1792-1797.

- Rocha EPC, Danchin A (2002) "Base composition bias might result from competition for metabolic resources" Trends in Genetics 18:291-4.
- Rocha EPC, Feil EJ (2010) "Mutational Patterns Cannot Explain Genome Composition: Are There Any Neutral Sites in the Genomes of Bacteria?" PLoS Genet 6(9): e1001104. doi:10.1371/journal.pgen.1001104
- Segré D, Raymond J. (2006). "The Effect of Oxygen on Biochemical Networks and the Evolution of Complex Life". Science. Vol. 311. no. 5768, pp. 1764 - 1767
- STADTMAN, E. R. and LEVINE, R. L. (2000), "Protein Oxidation." Annals of the New York Academy of Sciences, 899: 191–208.
- Sueoka N (1962). "On the genetic bases of variation and heterogeneity of DNA base composition." PNAS. 48:582-92.
- Tatusov R L, Fedorova N D, Jackson J D, Jacobs A R, Kiryutin B, Koonin E V, Krylov D M, Mazumder R, Mekhedov S L, Nikolskaya A N, Rao B S, Smirnov S, Sverdlov A V, Vasudevan S, Wolf Y, Yin J J, y A N Darren. 2003. "The COG database: an updated version includes eukaryotes" BMC Bioinformatics; 4: 41.
- Torsvik, V. Ovreas, L. & T. F. Thingstad. (2002) "Prokaryotic Diversity--Magnitude, Dynamics, and Controlling Factors" Science.Vol. 296 no. 5570 pp. 1064-1066
- Vieira-Silva S, Rocha EPC (2008) "An Assessment of the Impacts of Molecular Oxygen on the Evolution of Proteomes." Mol Biol Evol 25:1931-42.
- Yassin A F, Kroppenstedt R. M and W. Ludwig. (2003) *Corynebacterium glaucum* sp. nov. Int J Syst Evol Microbiol 53, 705-709.