

00346

11



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS
DIVISION DE ESTUDIOS DE POSGRADO

EVOLUCION MOLECULAR DE LAS RUTAS
BIOSINTETICAS DE LA LISINA

T E S I S
QUE PARA OBTENER EL GRADO ACADEMICO DE
MAESTRA EN CIENCIAS EN BIOLOGIA CELULAR
P R E S E N T A :
ANA MARIA VELASCO VELASCO

2002

DIRECTOR DE TESIS: DR. ANTONIO EUSEBIO LAZCANO-ARAUJO REYES

TESIS CON
FALLA DE ORIGEN



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

A mis padres, gracias por su apoyo, comprensión y amor.

A mi familia y amigos.

**TESIS CON
FALLA DE ORIGEN**

Agradecimientos:

Especialmente a mi director de tesis Dr. Antonio Lazcano por sus enseñanzas, apoyo y amistad en todo momento. Toño gracias por todo.

Al los miembros de mi comité Tutorial: Dr. Luis Felipe Jiménez García por sus comentarios y observaciones durante el desarrollo de este trabajo, a la Dra. Carmen Gómez Eichelmann por sus valiosos e interesantes comentarios y por el tiempo dedicado a la revisión y agradable discusión del manuscrito.

Al los miembros del jurado Dr. Germinal Cocho, Dr. Pedro Miramontes, Dra. Ana Barahona y al Dr. Arturo Becerra por la revisión de este trabajo.

A José Ignacio Leguina por su apoyo y colaboración al inicio de este trabajo.

A mis compañeros del Laboratorio de Microbiología Sara, Arturito, el Dr. Ulises, Luis, Amanda, Erwin, Héctor, Roberto y los tallarines, por su amistad y apoyo.

A Tere y Lourdes por su ayuda.

Al apoyo recibido por la Beca Robert Cedergren.

Y muy especialmente a mi Universidad, a la UNAM.

TESIS CON
FALLA DE ORIGEN

ÍNDICE

Introducción	
Teorías sobre el origen y evolución del metabolismo	1
Rutas anabólicas de la lisina	5
Artículo	12
Discusión y conclusiones	44
Referencias	46

TESIS CON
FALLA DE ORIGEN

RESUMEN

En el anabolismo de aminoácidos la lisina presenta una naturaleza exclusiva ya que es el único que se sintetiza a partir de dos rutas completamente diferentes. Cada una de estas rutas recibe el nombre del intermediario principal, así, una de ellas es llamada ruta del ácido diaminopimérico (DAP) y la otra del ácido α -aminoadípico (AAA). El hecho de que existan dos rutas anabólicas para la lisina nos ofrece un buen modelo para establecer cual de los modelos propuestos para explicar la evolución del metabolismo fue el que operó en el desarrollo de estas. En este trabajo se reportan los resultados del análisis detallado de las secuencias de cada una de las enzimas que participan en ambas rutas, sin encontrarse ninguna relación evolutiva entre ellas. Los resultados sugieren que la ruta DAP está relacionada con el metabolismo de la arginina, ya que los productos de los genes *lysC*, *asd*, *dapE* y *lysA* pertenecientes a la biosíntesis vía DAP están relacionados con los de los genes *argB*, *argC*, *argE* y *speA*, cuyos productos catalizan diferentes etapas en la biosíntesis de arginina. Por otro lado, los genes *lys20*, *lys4* y *lys12* que pertenecen a la ruta del AAA presentan relación con los genes *leuA*, *leuCD* y *leuB*, respectivamente los cuales están involucrados en la biosíntesis de leucina, al igual que con una serie de genes que pertenecen al ciclo del ácido cítrico y a la síntesis de antibióticos. En base al análisis y comparación de los genes involucrados en DAP y AAA, así como la relación que se encontró entre estas rutas y las de arginina y leucina se concluye que el modelo que operó para ensamblar ambas rutas es el de "patchwork" propuesto por Jensen (1976).

INTRODUCCIÓN

La aparición de las rutas metabólicas biosintéticas básicas fue uno de los eventos principales durante la evolución temprana de la vida en el Arqueano, ya que su aparición llevó a los organismos primitivos a disminuir su dependencia de fuentes exógenas de aminoácidos, purinas y otros compuestos que se acumularon en el medio primitivo como resultado de la síntesis prebiótica de compuestos orgánicos (Fani *et al.*, 1995). Se han propuesto varias teorías para tratar de explicar el origen de las rutas metabólicas y a continuación se analizarán cuatro de éstas teorías.

Teoría sobre el origen y evolución del metabolismo

Horowitz (1945) sugirió que los primeros organismos eran heterótrofos que utilizaban los compuestos químicos sintetizados prebióticamente, y que las enzimas biosintéticas de éstos organismos se originaron por procesos de duplicación genética, pero en un orden inverso al encontrado en las rutas metabólicas actuales. Por ello, ésta explicación se conoce como teoría retrógrada. Así, el último paso en una ruta biosintética fue el primero en ser adquirido en el proceso evolutivo, mientras que el penúltimo fue el segundo, y así sucesivamente hasta completar una ruta metabólica (Figura 1a). Otro supuesto necesario para esta hipótesis es que todos los intermediarios bioquímicos y los productos finales de las rutas biosintéticas básicas estuvieran disponibles en la sopa primitiva. De tal forma que si las primeras células eran heterótrofas para el compuesto **A**, éstas lo tomaba del medio y como resultado de la actividad biológica primordial, éste compuesto empezaba a escasear, a tal punto que limitaba el crecimiento de los organismos, algunos de los cuales poseían una enzima **a** capaz de reaccionar con un compuesto **B** químicamente relacionado y transformarlo en **A**, lo cual daría origen a una ruta de un solo paso. Subsecuentemente el compuesto **B** también se agotaría en el medio y aquellos organismos con una enzima mutante **b** producto de un evento de duplicación de **a** y capaz de mediar la transformación del compuesto **C** en **B**, llevó a los organismos a un incremento en la complejidad de la ruta y continuó de ésta manera hasta que se estableciera una ruta completa. De acuerdo a lo anterior, las enzimas de una ruta metabólica se originaron por una serie de eventos de duplicación genética, por lo que comparten características estructurales y se ordenaron en forma sucesiva uno después de otro semejando el modelo del operón (Horowitz, 1965). En las rutas metabólicas actuales se han encontrado pocos casos en donde enzimas que realizan reacciones consecutivas están codificadas por genes ubicados contiguamente en el genoma y que sean resultado de eventos de duplicación genética. Estos

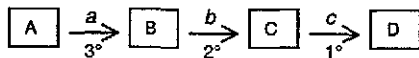
ejemplos incluyen (a) la β -cistationasa y la cistation β -sintasa involucradas en la biosíntesis de leucina; (b) la protoclorofila reductasa y la clorin reductasa, que intervienen en la biosíntesis de bacterioclorofila; y finalmente (c) los productos de *hisA* y *hisF* que participan en la síntesis de histidina (Peretó *et al.*, 1997).

No obstante los ejemplos anteriores, son varios los puntos débiles de esta teoría (Lazcano *et al.*, 1995). Ello incluye el que la mayoría de los intermediarios son químicamente inestables, por lo que su síntesis y acumulación en el medio primitivo sería poco probable; además dichos intermediarios son compuestos aniónicos, fosforilados que no podían pasar por las membranas primitivas debido a la carencia en esas membranas de sistemas de transporte especializado. La teoría retrograda plantea que pasos sucesivos en las rutas metabólicas involucran reacciones químicas similares, lo cual no es frecuente. Por último, como ya se mencionó anteriormente son pocos los ejemplos de genes que codifican enzimas de rutas metabólicas que catalizan reacciones consecutivas que se originaron por eventos de duplicación genética y ubicados de manera contigua en el genoma. Hasta el momento no se ha encontrado ningún caso en el cual las enzimas participantes de una ruta metabólica sean todas producto de eventos de duplicación de un gen ancestral común a todas ellas, y por otro lado la idea de que por ser productos de duplicación los genes de una ruta dada deberían estar ordenado en la forma de un operón tampoco es una hipótesis que encuentre apoyo en lo que se sabe hasta el momento, ya que no se han encontrado ejemplos de enzimas homólogas de una ruta ni dispersos ni formando un operón. Debido a que son solo algunos pares de enzimas que catalizan pasos consecutivos además de estar ubicadas de forma contigua en el genoma, es posible que por lo menos algunas enzimas se originaran de esta forma, mientras que las demás enzimas surgieron por procesos diferentes al propuesto por esta teoría.

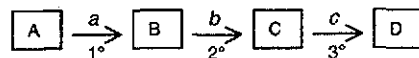
Una propuesta posterior a la de Horowitz fue hecha por Granick (1957, 1965), quien planteó que la estructura de una ruta determinada refleja su historia, basado en la idea de Haeckel de que la ontogenia recapitula la filogenia. Esta hipótesis asume que los compuestos bioquímicos más sencillos son más antiguos que los más complejos, de tal forma que las rutas metabólicas se ensamblaron hacia adelante (Figura 1b) por eventos sucesivos de duplicación y mutación del último gen de la cadena, por lo que las primeras enzimas de una ruta metabólica son más antiguas que las últimas además de que cada nueva etapa que se adquirió fue seleccionada ya que su producto presentaba una función más útil que su precursor. Así, la estructura de varias rutas biosintéticas sugiere que su desarrollo en rutas cada vez más largas involucra el reclutamiento progresivo de etapas enzimáticas, tal es el caso de la biosíntesis de algunos componentes de membrana (Peretó *et al.*, 1997). Uno de ellos corresponde a la presencia del lípido con éter insaturado 2,3-di-*O*-geranylgeranyl-*sn*-glycerol, que es el mayor

componente de la membrana de la arquea hipertermófila *Methanopyrus kandleri*, así como uno de los intermediarios en la biosíntesis del 2,3-di-*O*-fintanil-*sn*-glicerol, el cual es común en otras membranas de arqueas. Esta secuencia de reacciones puede ser interpretada como una ruta progresivamente más compleja, pues se puede suponer que evolucionó por la adición de pasos enzimáticos a una ruta ya existente cuyo producto final es el lípido con éter insaturado. En las filogenias basadas en 16S rRNA, *M. kandleri* se ubica en una de las ramas más antiguas de arqueas por lo que se puede suponer que la estructura de la ruta refleja la filogenia del grupo, de tal forma que los organismos que evolucionaron más recientemente presentan una ruta biosintética de lípidos más compleja (Peretó *et al.*, 1997). Sin embargo, no todas las rutas metabólicas parecen ajustarse a este modelo. Al igual que el modelo anterior, el hecho de que los procesos metabólicos se ensamblaran por el mecanismo hacia adelante implicaría que todas las enzimas de una ruta fueran homólogas; sin embargo, en esta caso no se plantea la necesidad de que los genes de dicha ruta permanezcan de manera contigua como en un operón.

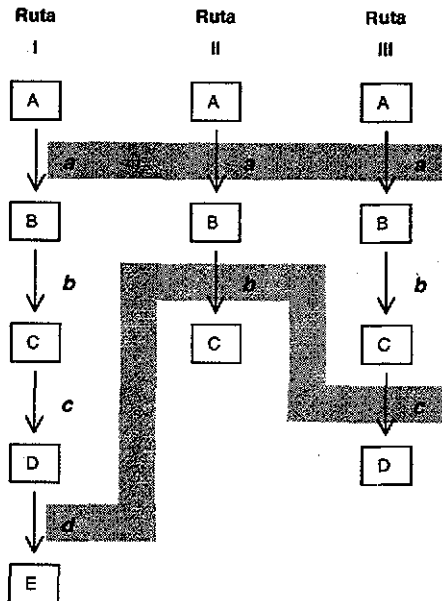
i) Retrógrada



ii) Granick



iii) Patchwork



iv) Semi-enzimática

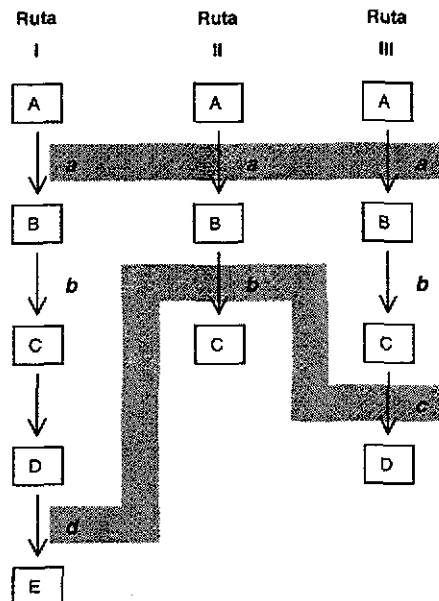


Figura 1. Representación esquemática de las cuatro teorías propuestas para explicar el origen y evolución de las rutas metabólicas. i) Teoría retrógrada, en donde la enzima más antigua de una ruta metabólica es la última, mientras que la primera es la más reciente; ii) teoría de Granick, la enzima más reciente es la última en la ruta, mientras que la más antigua es la primera; iii) teoría de patchwork, en donde las enzimas son producto de eventos de duplicación genética (marcadas con una franja gris) y por lo tanto comparten un origen común, pero intervienen en rutas metabólicas diferentes, llevando a cabo reacciones enzimáticas diferentes; iv) teoría semi-enzimática, las enzimas de origen común están marcadas nuevamente en gris, en tanto que aquellas reacciones que se llevan de manera espontánea no requieren la presencia de enzimas

En las teorías anteriores, la duplicación genética juega un papel relevante en el origen y evolución de las rutas metabólicas básicas, sin embargo una interpretación alternativa de éste proceso es desarrollada en la Teoría de Patchwork (Waley, 1969; Ycas, 1974; Jensen 1976) en la que se propone que las enzimas específicas presentes en las rutas biosintéticas actuales evolucionaron por un proceso de duplicación y reclutamiento de un pequeño conjunto de enzimas ancestrales con baja especificidad que podía reaccionar con un amplio rango de sustratos relacionados químicamente (Figura 1c). La baja especificidad de las primeras enzimas puede representar el mecanismo por medio del cual las células primitivas, que debieron presentar un número pequeño de genes, resolvieron sus limitadas habilidades codificantes. De acuerdo a ésta hipótesis, las enzimas nuevas fueron el resultado de eventos de amplificación seguidos por la divergencia de los genes duplicados, lo cual llevó a limitar y modificar la especificidad de las enzimas ya que las copias de los genes presentaban una serie de mutaciones que ocasionaron cambios conformacionales que podían originar una nueva enzima que llevara a cabo una nueva reacción pero de tipo similar al que le dio origen. Así, la duplicación de genes presentaba la oportunidad de incrementar el tamaño del genoma, así como la especialización de las enzimas divergentes, la especialización en el reconocimiento del sustrato además fue reforzado por el desarrollo de mecanismos de regulación. El modelo de patchwork también supone que las enzimas que catalizan reacciones bioquímicas similares en rutas metabólicas diferentes deben ser homólogas. La existencia de enzimas homólogas que participan en diferentes rutas metabólicas proveen evidencia que apoya esta teoría. Este es el caso de la ancestría común de la piruvato oxidasa y la acetohidroxiácido sintasa, la primera involucrada en la descarboxilación de piruvato a acetato y la segunda en la biosíntesis de aminoácidos ramificados. Otra evidencia que apoya esta teoría es el hecho de que en experimentos de sustitución sitio-específicos de aminoácidos se logró cambiar la actividad de la lactato deshidrogenasa NAD dependiente, por la de malato deshidrogenasa con el cambio de un solo aminoácido. Sin embargo, también existen ejemplos que contradicen esta teoría por la redundancia bioquímica de carácter homólogo y a la falta de homología (Peretó *et al.*, 1997) entre (a) la triptofanasa y la subunidad β de la triptofano sintasa, ambas enzimas piridoxal 5'-dependientes que catalizan reacciones de β -eliminación y (b) las glutamino amido transferasas tipo F y G, las cuales no muestran un origen común. Sin embargo, el mecanismo de patchwork sólo pudo haber operado después de la aparición de la síntesis de proteínas y del desarrollo de enzimas, así como después de la aparición del mundo basado en DNA y proteínas, ya que muy pocas estaban disponibles antes de estos eventos (Lazcano y Miller, 1999). Actualmente se sabe que existen pasos en algunas rutas metabólicas en los que no se requiere la acción de enzimas, llevándose de manera espontánea tales reacciones. Estos casos no son explicados por ninguna de las teorías mencionadas anteriormente ya que estos modelos no contemplaba el hecho de que al menos en algunas rutas hay reacciones que se llevan a cabo de manera espontánea sin la necesidad de una enzima, por lo que Lazcano y Miller (1999), proponen un

modelo más detallado en el que las primeras rutas biosintéticas fueron parcial o totalmente no-enzimáticas (Figura 1d), y que se desarrollaron por la utilización de compuestos prebióticos disponibles en el medio, así como por el uso de compuestos producto de enzimas poco específicas resultado de eventos de duplicación genética que posteriormente se especializaron y fueron usadas en rutas metabólicas diferentes, a esta explicación le dieron el nombre de teoría semi-enzimática.

La mayoría de los pasos en las rutas biosintéticas son mediados por enzimas, mientras que en algunos casos las reacciones son llevadas a cabo de manera espontánea. En otros casos, el paso químico correspondiente a una ruta metabólica puede llevarse a cabo en ausencia de enzimas cambiando los reactivos y las condiciones de la reacción. Un ejemplo de esto es el producto del gen *hisH* que codifica para la glutamino amidotransferasa, que toma parte en la biosíntesis de histidina. En un medio con una elevada concentración de amonía la reacción de adición de NH_3 puede realizarse en ausencia de la enzima codificada por *hisH*. Se propone que esta reacción se llevaba a cabo de manera espontánea en la tierra primitiva utilizando el NH_3 disponible, después se dio el desarrollo de la enzima codificada por el gen *hisH*, seguida por la utilización de glutamina en lugar de NH_3 , cuando este compuesto desapareció de la tierra primitiva (Lazcano y Miller, 1999).

Rutas anabólicas de la lisina

Del conjunto de rutas biosintéticas, la lisina representa un caso peculiar, ya que se conocen dos rutas anabólicas de este aminoácido sin ninguna relación entre sí, lo cual es un caso único entre las rutas metabólicas de síntesis de aminoácidos (Vogel, 1965). Una es la llamada ruta del ácido diaminopimélico, mientras que la otra es la ruta del ácido α -amino adípico, o DAP y AAA respectivamente, por sus siglas en inglés (Figura 2). La primera pertenece a la familia del aspartato junto con las rutas anabólicas de la metionina (Met), la treonina (Thr) y la isoleucina (Ile) (Umbarger, 1978; Scapin y Blanchard, 1996). La ruta AAA es parte de la familia del glutamato, que incluye a la de la glutamina (Gln), la prolina (Pro) y la arginina (Arg) (Umbarger, 1978).

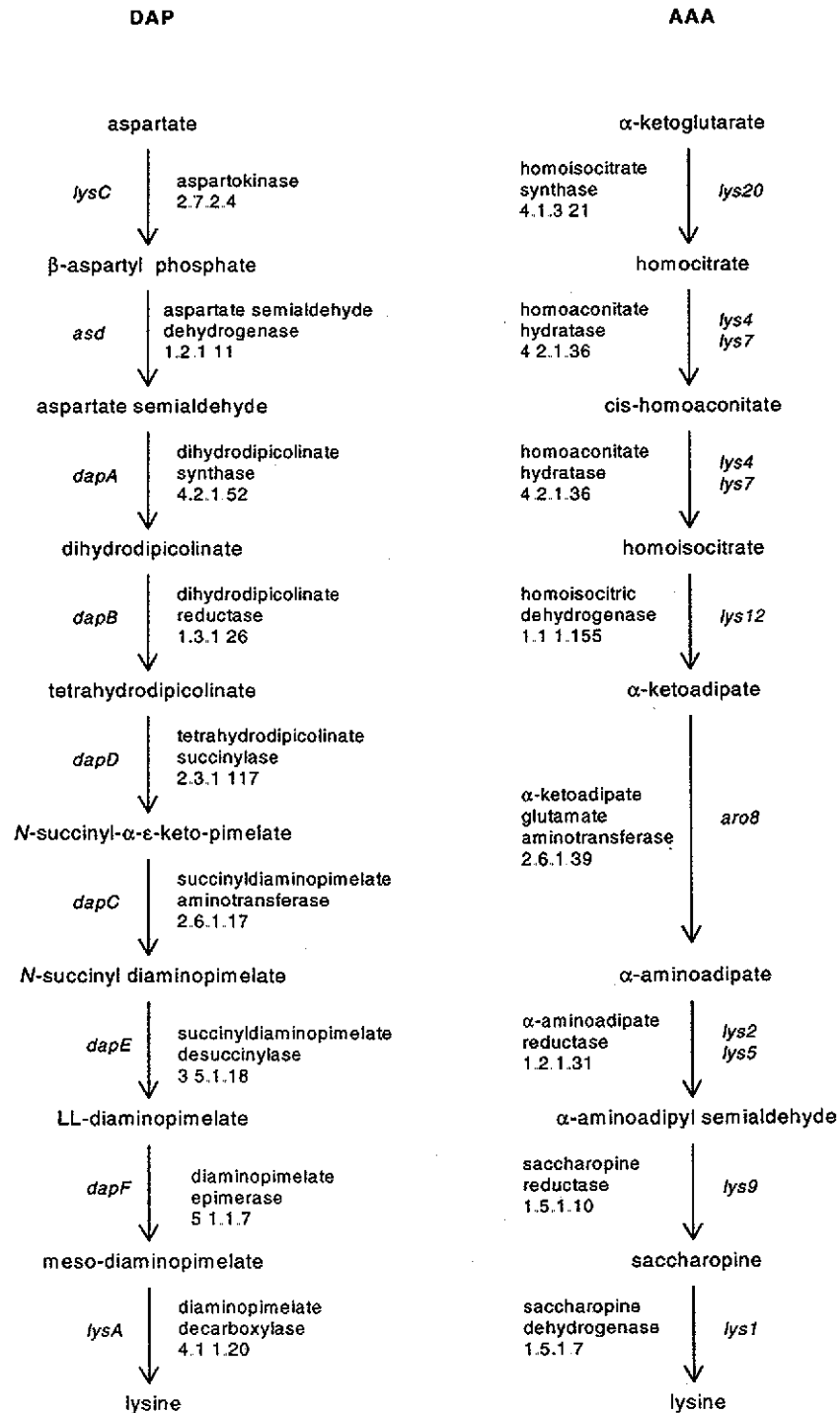
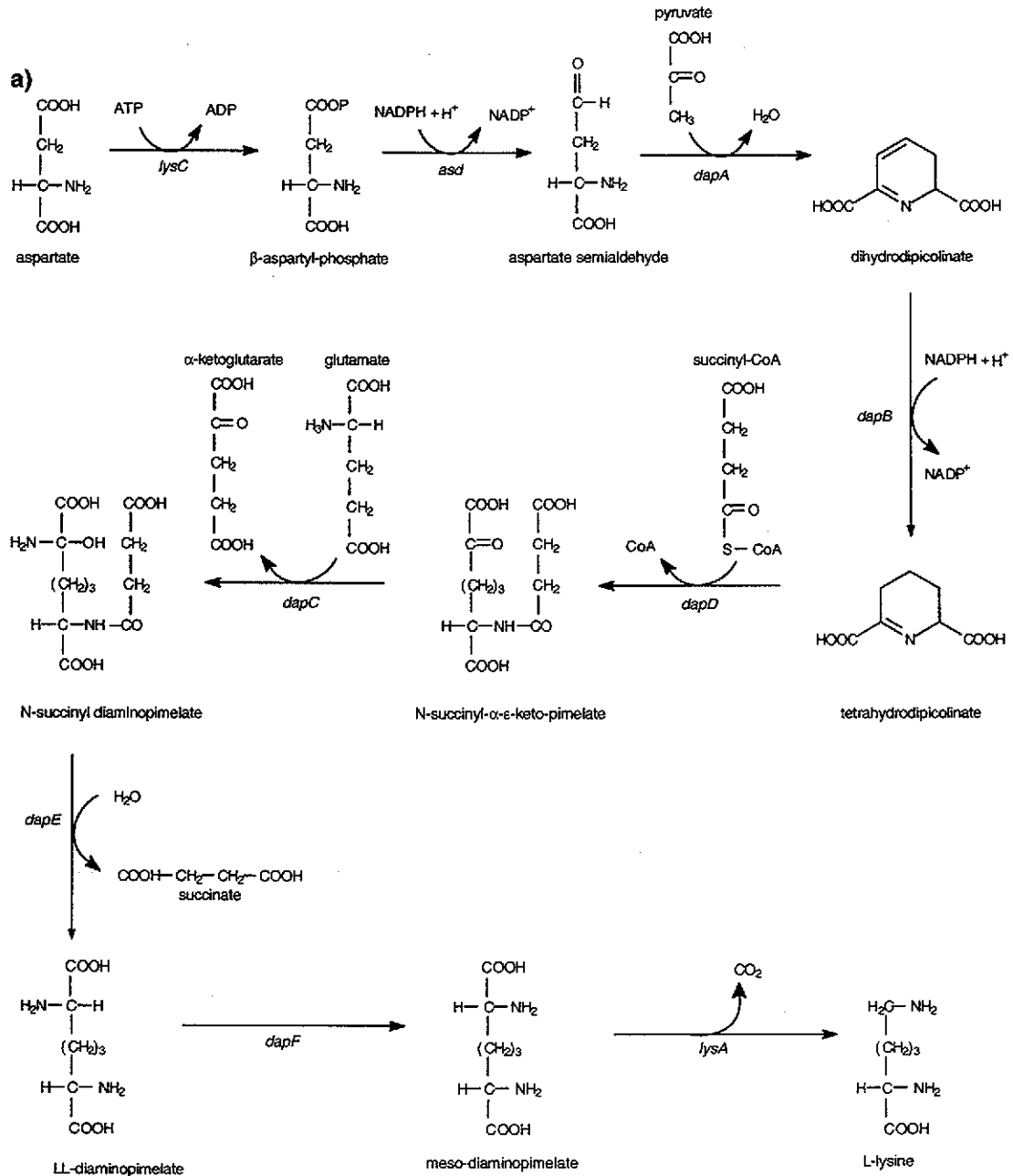
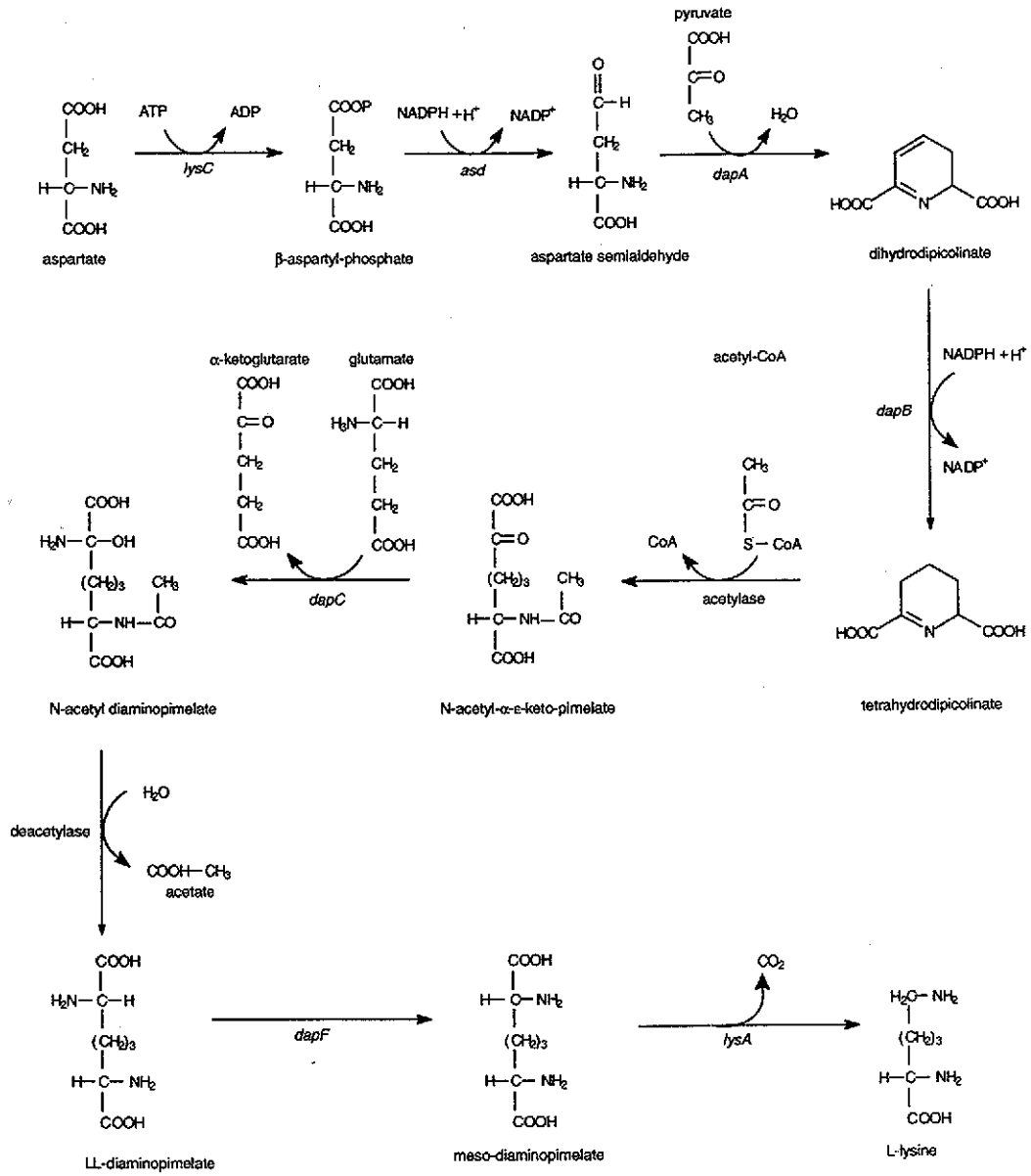


Figura 2. Representación esquemática de las dos rutas biosintéticas de lisina, sus intermediarios, y las enzimas involucradas en cada una de ellas.

Una característica importante de la ruta del DAP es la existencia de tres variantes para esta ruta que incluyen: a) una en la cual los intermediarios son succinilados para evitar la formación de una estructura cíclica (Figura 3a) (Kindler y Gilvarg, 1960); b) la segunda consta del mismo número de pasos enzimáticos que la primera, pero en lugar de succinato se utiliza un grupo acetil (Figura 3b) (Sundharadas y Gilvarg, 1967), mientras que c) la última es la variante más corta, ya que en esta ruta se pasa del cuarto intermediario (tetrahidrodipicolinato) hasta el último, que es el meso-diaminopimelato por la acción de la enzima diaminopimelato deshidrogenasa (*ddh*) (Figura 3c) (Misono *et al.*, 1979).



b)



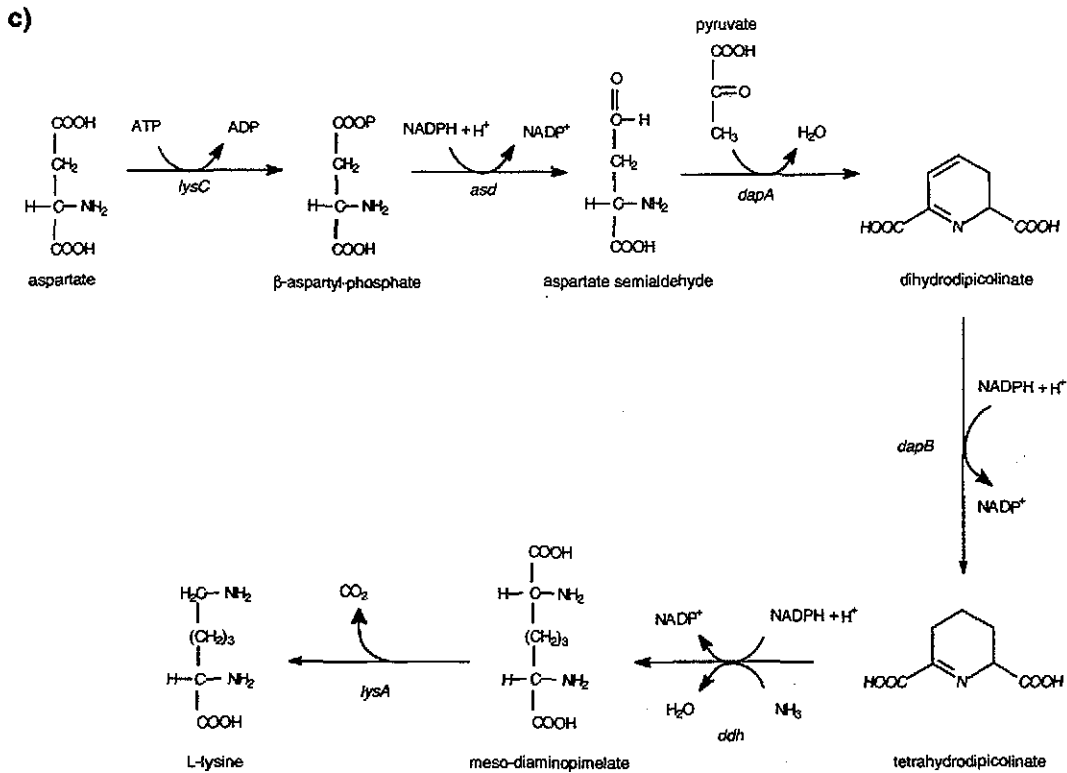


Figura 3. Reacciones llevadas a cabo en dos de las diferentes vías del DAP. a) Variante succinato, b) variante acetato, c) variante deshidrogenasa.

Como se puede apreciar en la Figura 3a y 3b, es interesante hacer notar que las variantes del succinato y del acetato llevan a cabo las mismas reacciones y lo único que las diferencia son los grupos que se utilizan como bloqueadores. La variante más frecuente entre los tres grandes grupos biológicos es la del succinato. Sin embargo, se utilizó como ruta modelo la presentada por *Escherichia coli* (Patte, 1996). Así, a lo largo de este trabajo cuando se haga mención a la ruta DAP, nos estaremos refiriendo la variante del succinato, ya que hasta el momento no se han reportado las secuencias de los genes que intervienen en la variante del acetato, mientras que la variante de la deshidrogenasa se ha localizado en pocos organismos, *Corynebacterium glutamicum* y en algunas especies del género *Bacillus* (Schrumpp et al., 19991; Wehrmann et al., 1998). La disponibilidad actual de genomas completamente secuenciados permite observar que lo que se pensaba era una ruta establecida para todos los organismos que la presentan está muy lejos de ser así. Tal es el caso de la ruta que nos ocupa, ya que mientras en *E. coli* existen hasta tres genes que sintetizan la aspartokinasa (*thrA*, *metL* y *lysC*), la mayoría de los organismos con genoma secuenciado completamente solo presentan un gen para esta enzima, como ocurre en *Rickettsia prowasekii* (Andersson et al., 1998) y *Campylobacter jejuni* (Parkhill et al., 2000), por mencionar solo algunos ejemplos. Por otra parte, el caso contrario lo representan organismos con la ruta del DAP incompleta, como sería el caso de las bacterias

Chlamydia trachomatis (Stephens *et al.*, 1998) y *Aquifex aelicus* (Deckert *et al.* 1998), así como algunas arqueas como *Aeropyrum pernix* (Kawarabayasi *et al.*, 1999), y *Methanobacterium thermoautotrophicum* (Smith *et al.* 1997).

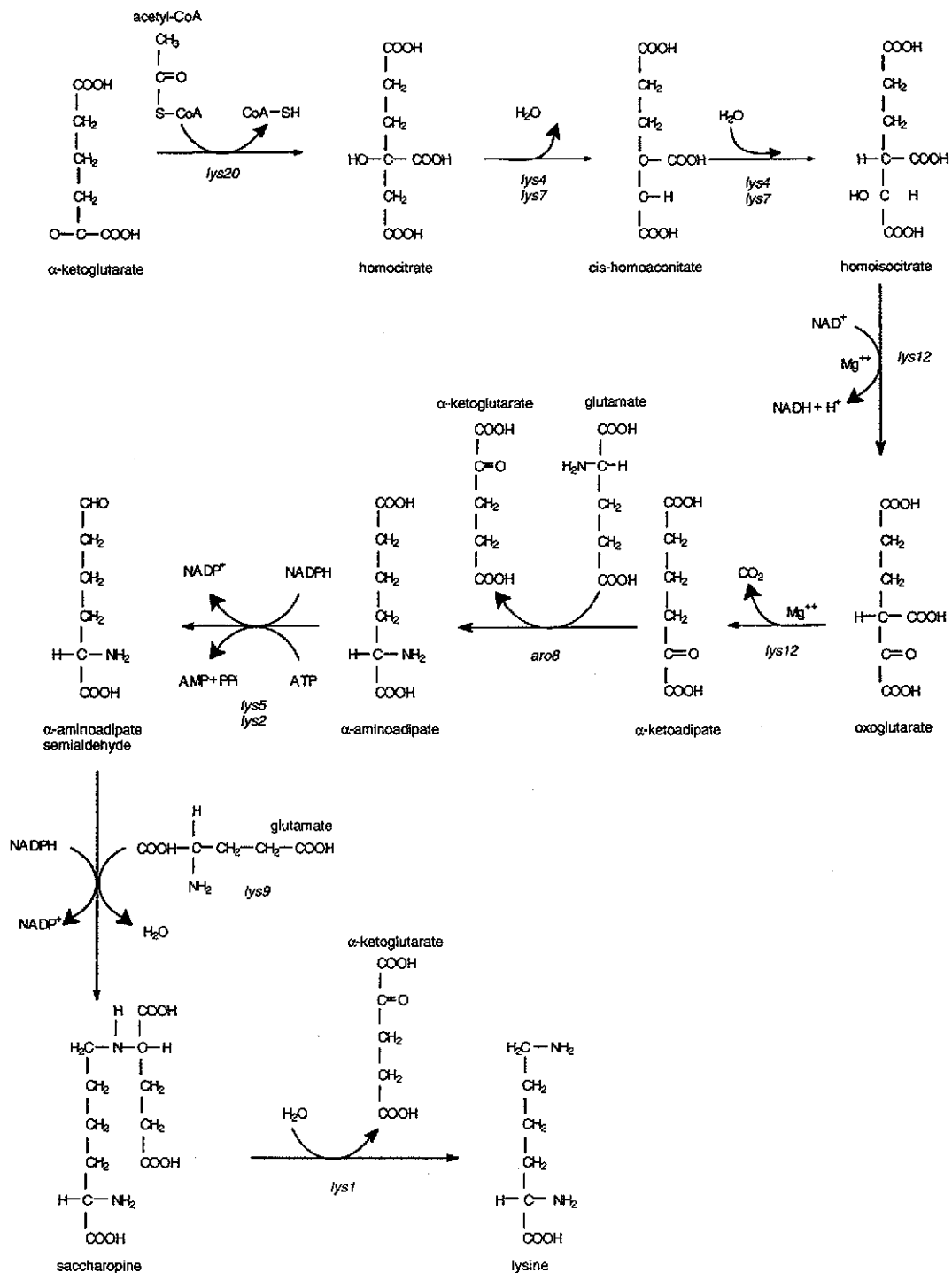


Figura 4. Pasos enzimáticos de la ruta AAA

En lo que se refiere a la ruta AAA, ésta consta de siete pasos enzimáticos (Figura 2) y da inicio con la condensación de α -ketoglutarato y la acetyl-Co, para sintetizar al principal intermediario, que es el ácido α -aminoadípico (Bhattacharjee, 1992). Hasta hace poco se creía que esta ruta era exclusiva de hongos superiores, pero recientemente se descubrió que la bacteria hipertermófila *Thermus thermophilus* sintetiza lisina vía AAA (Kosuge y Hoshino, 1998), por medio de una variante (Figura 4) en la cual al parecer intervienen genes de las rutas biosintéticas de la arginina y de la leucina (Nishida et al., 1999).

En lo que se refiere a la distribución de estas dos rutas entre los organismos de los diferentes grupos biológicos, se sabe, gracias a trabajos hechos por Vogel (1965), que la ruta del DAP es utilizada por eucariontes como hipochitriales, saprolegniales, leptomitales y plantas, eubacterias Gram-positivas y Gram-negativas (Ragan, 1989), y recientemente, se localizó también en metanógenas (Bakhiet *et al.*, 1984). En cambio, desde que se descubrió la ruta del AAA se consideró como un carácter diagnóstico de hongos, ya que solo se había encontrado en chitridiales, zigomicetos, heterobasidiomicetos, euascomicetos, blastocladiales y mucorales. Además de estar presente en hongos, también se ha visto que los euglenoides utilizan la ruta del AAA para sintetizar lisina. La ruta del AAA es tan característica de los hongos, que se considera un rasgo diagnóstico de este grupo. Su presencia en critidios y blastocladiales es congruente con las filogenias basadas en las secuencias de los 18S rRNAs de estos organismos, que demuestran su afinidad con los hongos (Sogin, 1994). Sin embargo, también se ha encontrado en arqueobacterias como *Sulfolobus* y *Thermoproteus* (Ragan, 1989).

El descubrimiento reciente de una vía alternativa para la ruta AAA en *T. thermophilus*, así como la relación que guarda ésta con otras rutas anabólicas nos lleva a preguntarnos si existe alguna relación entre la ruta de *Thermus* y la ruta del DAP. Por otro lado, la existencia de dos rutas metabólicas para la biosíntesis de lisina ofrece un buen modelo para establecer cuál de las teorías propuestas para evolución metabólica se dio en estas. Debido a que los organismos que sintetizan la lisina siguiendo la ruta del AAA son claramente posteriores a los procariontes, se dispone de un modelo que en principio permitirá explicar el surgimiento de una ruta alterna de biosíntesis de la lisina mediante la hipótesis de Waley-Ycas-Jensen. Con el propósito de analizar estas hipótesis se llevó a cabo el análisis de las secuencias que participan en ambas rutas disponibles en los bancos de datos, cuyos resultados se presenta aquí.

Molecular evolution of the lysine biosynthetic pathways

AM Velasco, JI Leguina, A Lazcano*

Facultad de Ciencias, UNAM
Apdo. Postal 70-407, Cd. Universitaria,
México D. F. 04510,
MEXICO
e-mail: alar@hp.fciencias.unam.mx

*corresponding author

ABSTRACT

Among the different biosynthetic pathways found in extant organisms, lysine biosynthesis is peculiar because it has two different anabolic routes. One is the diaminopimelic acid pathway (DAP), and the other the α -aminoadipic acid pathway (AAA). A new variant of the AAA route which includes some enzymes involved in arginine and leucine biosyntheses, has been recently reported in *Thermus thermophilus* (Nishida *et al.*, 1999). Here we describe the results of a detailed genomic analysis of each of the sequences involved in the two lysine anabolic routes, as well as of genes from other routes related to them. No apparent evolutionary relationship between the DAP and AAA enzymes was found. Our results suggest that the DAP pathway is related to arginine metabolism, since the *lysC*, *asd*, *dapC*, *dapE* and *lysA* genes from lysine biosynthesis are related to the *argB*, *argC*, *argD*, *argE* and *speAC* genes, respectively, whose products catalyze different steps in arginine metabolism. This work supports previous reports on the relationship between AAA gene products and some enzymes involved in leucine biosynthesis and the tricarboxylic acid cycle (Irvin and Bhattacharjee, 1998; Miyazaki *et al.*, 2001). Here we discuss the significance of the recent finding that several genes involved in the arginine (Arg) and leucine (Leu) biosynthesis participate in a new alternative route of the AAA pathway (Miyazaki *et al.*, 2001). Our results demonstrate a clear relationship between the DAP and Arg routes and between the AAA and Leu pathways.

Keywords: lysine biosynthesis, diaminopimelic acid, α -aminoadipic acid, evolution of metabolic pathways, arginine biosynthesis

INTRODUCTION

Two completely different pathways have evolved separately for lysine (Lys) biosynthesis. One is the so-called diaminopimelic acid route (DAP), and the other the α -amino adipic acid pathway (AAA) (Vogel, 1965). The first route belongs to the aspartate family, and consists of nine enzymatic steps that lead from L-aspartate to Lys (Figure 1). The first two steps catalyze the syntheses of intermediates common to Lys, threonine and methionine, and the next seven ones lead to Lys from the condensation of aspartate semialdehyde with pyruvate to form diaminopimelic acid as the key intermediate (Umbarger, 1978; Pavelka and Jacobs, 1996; Scapin and Blanchard, 1998; Patte, 1996). The DAP pathway also plays a central role in eubacterial cell wall biosynthesis, since meso-diaminopimelate is the immediate precursor for the biosynthesis of peptidoglycan, the main component of the eubacterial cell wall (Cirillo *et al.*, 1994; Wehrmann *et al.*, 1998). Another feature of this pathway is that three different ways to synthesize meso-diaminopimelate and Lys have been described (Figure 1), (a) one in which the intermediates are succinylated to avoid ring formation (Kindler and Gilvarg, 1960; Weinberger and Gilvarg, 1970); (b) a second variant where an acetyl residue is used as a blocking group instead of the succinyl one (Sundharadas and Gilvarg, 1967), and (c) the dehydrogenase pathway, in which tetrahydrodipicolinate is converted to meso-diaminopimelate in a single reaction catalyzed by diaminopimelate dehydrogenase (EC 1.4.1.16) (Misono *et al.*, 1979; White, 1983). Although most bacteria utilize only one of the three variants, experimental evidence has shown that the Gram positive *Corynebacterium glutamicum*, *Bacillus brevis*, *B. cereus*, and *B. licheniformis* can use both the succinyl and dehydrogenase routes (Schrumpf *et al.*, 1991; Wehrmann *et al.*, 1998).

On the other hand, the AAA pathway is part of the glutamate biosynthetic family (Umbarger, 1978). This route involves eight steps for Lys biosynthesis (Figure 1), and requires the condensation of α -ketoglutarate and acetyl-CoA for the synthesis of α -amino adipate and saccharopine as intermediates (Vogel, 1965; Rowley and Tucci, 1970; Bhattacharjee, 1985, 1992). In contrast with the DAP route, it was generally assumed that the AAA route had no variants. However, modified version of this pathway has been recently reported, which involves, in addition to the genes from the AAA pathway sequences from arginine anabolism (Arg), and leucine (Leu) biosynthesis (Nishida *et al.*, 1999).

Of the three known DAP variants, the succinyl-dependent pathway is the most widely distributed (Table 1). It is known to be present in (eu)bacteria (Vogel, 1965; Scapin and Blanchard, 1998; Scapin *et al.*, 1998), the so called lower fungi (LéJohn, 1971, 1974; Ragan and Chapman, 1978, Ragan, 1989), and plants (Vogel, 1965; Bryan, 1980), as well as in Archaea (Bakhiet *et al.*, 1984; Smith *et al.*, 1997; Bult *et al.*, 1996; Klenk *et al.*, 1997). The acetyl-

dependent DAP route has been identified in certain *Bacillus* species (Weinberger and Gilvarg, 1970), although the corresponding genes have not been characterized physically (Liu and Shaw, 1997). Diaminopimelate dehydrogenase activity has been detected only in a small group of bacteria, which include the Gram positive bacteria *C. glutamicum*, *Bacillus* spp., and *Brevibacterium* species (Misono *et al.*, 1986a, 1986b), as well as in some plants such as soybean, maize, and wheat (Wenko *et al.*, 1985).

Until a few years ago it was believed that the AAA route was present only in higher fungi (Matsuda and Ogur, 1969; Rowley and Tucci, 1970; Bhattacharjee, 1992; Weidner *et al.*, 1997) and in euglenoids, where biochemical activity associated with this pathway had been detected (Vogel, 1965; LéJohn, 1971). However, it is now known to be also present in the archaea *Thermoproteus* (Schäfer *et al.*, 1989) and *Sulfolobus* (Ragan, 1989; Schäfer *et al.*, 1989). More recently, an alternate variant of the AAA route was found in the bacteria *T. thermophilus*, which opens the possibility that it may have a wider distribution among prokaryotes (Nishida and Nishiyama, 2000; Nishida, 2001).

The finding of a AAA route variant which involves enzymes from at least three different pathways (AAA, Arg, Leu) raises several questions about its relationship to the other Lys anabolic route. The availability of completely sequenced cellular genomes from the major three biological domains (Bacteria, Archaea and Eukarya) offers the possibility of approaching this issue from a genomic perspective. Here we report the results of a detailed analysis of each of the genes involved in the two Lys anabolic pathways, and show that the DAP genes are related to some genes involved in Arg biosynthesis. Finally, we discuss our results in relationship with the *T. thermophilus* AAA pathway variant.

MATERIAL AND METHODS

The genes involved in the DAP and AAA pathways and their homologs were identified in different public databases (<http://www.ncbi.nlm.nih.gov>; <http://www.tigr.org/tdb/mdb/mdb.html>; <http://genome-www.stanford.edu/Saccharomyces>; <http://www.genome.ad.jp/kegg/kegg2.html>;) through their BLAST submissions using current default settings. The minimum score considered in a match between two sequences was 100. Amino acids sequences were aligned using the default set for multiple alignments from CLUSTALW 1.74 (Thompson *et al.*,1994), and blocks were detected with MACAW 2.0.5 Win 32i, using the Segment Pair Overlap as search method. Secondary structure analysis was performed using the PREDATOR program (http://www.embl-heidelberg.de/cgi/predator_serv.pl).

RESULTS

Exhaustive comparisons of the amino acid sequences of the enzymes catalyzing the different steps in all the known variants of the DAP and AAA routes has confirmed previous contentions of the unrelatedness of the two pathways. The 35 genomes analyzed here include 32 prokaryotic (Bacteria and Archaea) and 3 eucaryal completely sequenced genomes available in the public databases. The genes involved in the DAP pathway in 27 completely sequenced prokaryotic genomes are listed in Table 2. In most organisms the DAP route appears to be incomplete, and in other species it is altogether absent. The lack of Lys biosynthesis in the eubacteria *Borrelia burgdorferi*, *Mycoplasma genitalium*, *M. pneumoniae*, *Ureaplasma urealyticum*, and *Treponema pallidum* is probably a derived character that can be easily explained as a polyphyletic secondary loss that reflects their parasitic lifestyle. These parasitic organisms probably use lysine provided by their host environment. However, as shown by *Haemophilus influenzae*, and the *Helicobacter pylori* and *Neisseria meningitidis* strains, which display all the genes involved in the DAP route, not all parasites have lost this route. On the other hand, in the genomes of the archaea *Pyrococcus horikoshii*, *Pyrococcus abyssi*, and *Aeropyrum pernix* less than the half of the genes required for the DAP lysine biosynthesis have been identified (Table 3).

As a result of the comparative analysis of the DAP and AAA genes in completely sequenced cellular genomes, we report here that the products of the four genes involved in the DAP pathway *lysC*, *asd*, *dapE*, and *lysA*, are related with arginine metabolic enzymes encoded by *argB*, *argC*, *argE*, and *speA*, and *speC*, respectively (Figure 3). These results confirm and extend previous reports of the evolutionary relationship between *dapE* and *argE* (Boyen *et al.*, 1992), and of *lysA* with the *speA* and *speC* pair (Martin *et al.*, 1988; Sandmeier *et al.*, 1994). The similarity values found between the *lysC-argB* and *asd-argC* genes pairs were below the minimum score considered in this work. Thus, a secondary structure analysis was performed, which led to the identification of several structural features common to these two sets of genes. We have found that the products of the *lysC-argB* pair present a β - α - β - α structure formed by 5, 18, 6, and 8 amino acids, respectively. This motif is followed by two regions that have a β - α - β structure composed of 15, 5, and 15 amino acids in the first one, and 5, 10, and 5 amino acids in the second one. The β - α - β - α structure and the two β - α - β motifs are located at the NH₂-end in both *lysC* and *argB* (Figure 2). Secondary structure comparison of the *asd-argC* pair also led to the identification of several structural features shared between them (Figure 3).

The similarity between the AAA sequences *lys20*, *lys21*, *lys4* and *lys12*, and the *leuA*, *leuC*, *leuD* and *leuB* genes respectively, from the leucine biosynthetic route is shown in Figure 5.

Our results confirm and extend previous reports of the homology between *lys4* and the *leuC-leuD* pair (Irvin and Bhattacharjee, 1998), and of the relationship between *lys20* and *leuA* (Nishida *et al.*, 1999). The availability of the *Saccharomyces cerevisiae lys12* sequence has allowed us to demonstrate its relatedness with the *leuB* sequence, which participates in leucine biosynthesis, as well as with genes encoding enzymes that participate in carbohydrate metabolism and the tricarboxylic acid cycle, such as isocitrate dehydrogenase (data not show). The gene *lys2* exhibits a high similarity with a set of genes related with antibiotic biosyntheses, as well as with genes that participate in lipid and fatty acid metabolism. All these enzymes share several blocks with *lys2* (data not show).

The Arg and Leu biosynthetic genes of the 35 completely sequenced genomes studied here were analyzed to establish the relation between these two routes and the two Lys pathways. As noted above, 27 prokaryotes in our sample present at least four DAP genes. The complete set of genes of the AAA pathway was only found in the *S. cerevisiae* genome, while *Deinococcus radiodurans* is the only prokaryote in our sample that has only one copy of the *lys20* gene. The Arg genes are present in 20 prokaryotic genomes, but they are absent in parasitic organisms such as *Rickettsia prowasekii*, *Chlamydia trachomatis*, *M. genitalium*, *U. urealiticum*, the two strains of *H. pylori*, *T. pallidum*, and the three strains of *Chlamydia pneumoniae*. In *M. pneumoniae* only the gene *argI* is present. As shown in Table 3, *S. cerevisiae* presents six Arg genes, while only two of them (*argE*, *argH*) are present in *Drosophila melanogaster*.

At least one gene (*ilvE*) of the Leu anabolic pathway is present in 21 prokaryotic organisms, and the full set of genes of this route is present in sixteen of them, with a distribution that defies a simple explanation. In *P. horikoshii* and *P. abyssi* only the *ilvE* gene is absent. In contrast, in the *H. pylori* strains studied here, the only Leu gene present is precisely *ilvE*. An interesting fact of Leu sequences is that some genes are duplicated or even triplicated, as in the case of the three *leuA* copies found in the archaea *M. jannaschii*, *M. thermoautotrophicum*, and *P. abyssi*, and the two copies of *leuC* and *leuD* genes present in *D. radiodurans*, *T. maritima*, *M. jannaschii*, *M. thermoautotrophicum*, *A. fulgidus* and *P. abyssi*. In the eukaryote *S. cerevisiae* all the enzymes involved in the Leu biosynthesis have been identified, while in *D. melanogaster* only the gene *ilvE* has been identified. None of the Leu genes have been identified in the *C. elegans* genome.

DISCUSSION AND CONCLUSIONS

An exhaustive genomic comparison has demonstrated that a significant number of the DAP genes are related to Arg metabolic genes (Figure 4), as shown by the similarity levels between the gene pairs *lysC-argB*, *asd-argC*, *dapE-argE*, and *lysA-speAC*. This suggests that both routes may have been originally catalyzed by a single set of less-specific enzymes. This possibility is supported by several lines of evidence, including the recent demonstration that *N*-acetylornithine aminotransferase, an enzyme from Arg anabolism (*argD*), also participates in the DAP pathway (Ledwidge and Blanchard, 1999), and by biochemical evidence showing that the *argB*, *argC*, *argD* and *argE* gene products act in the second half of the AAA route recently reported in *T. thermophilus* (Nishida *et al.*, 1999; Miyazaki *et al.*, 2001).

The AAA genes *lys20*, *lys21*, *lys4* and *lys12* are related with practically all the genes from the leucine anabolism, with the sole exception of *ilvE* (Figure 4) indicating the key role played by this gene in the emergence of Leu biosynthesis. These results also suggest that the AAA route was the outcome of an assemblage process that was also produced by a patchwork-like mechanism. These results are in agreement with previous reports by Irvin and Bhattacharjee, (1998) and Nishida *et al.* (1999) on the homology between the *leuC-leuD* pair with *lys4*, and of *lys20* with *leuA* (Nishida *et al.*, 1999; Miyazaki *et al.*, 2001).

Based on the identification of the *T. thermophilus* Lys pathway gene homologs, it has been argued that the same route operates in *P. horikoshii*, *P. abyssi* (Nishida *et al.*, 1999) and *D. radiodurans* (Nishida, 2001). However, the presence of these genes does not necessarily indicate by itself that these organisms use this Lys route variant, since genes related to the DAP pathways are also found in these same genomes (Table 3). Furthermore, the presence of the *lysR* gene in *Thermus*, which is involved in the control of the *E. coli lysA* gene (Kosuge and Hoshino, 1997), suggest at least two alternative explanations: (a) it is possible that in the past *Thermus* had the DAP route, and that its *lysR* gene is a relict of such state. On the other hand, (b) if the *Thermus* pathway was the original Lys biosynthetic route (Miyazaki *et al.*, 2001), then the presence of *lysR* gene in the *T. thermophilus* genome could be the result of a horizontal transfer event.

The fact that all the homologous genes in the DAP-Arg routes and in the AAA-Leu pathways follow the same step order, strongly suggest that they are the outcome of a duplication of an ancestral gene cluster. Although none of these genes are arranged in an operon fashion in the sample studied here, the order in which they operate may resemble the order of genes in the original cluster. This interpretation is equivalent to the explanation advocated by Fani *et al.* (2000) to explain the evolution of the operons *nifDK* and *nifEN*, which are involved in nitrogen

fixation. As of today, however, it is not clear which is the origin of the isolated, non-homologous enzymes, such as those encoded by *dapB*, *dapD* and *dapF* in the DAP route, and those that belong to the AAA pathway (*lys7*, *lys5*, *lys9*, *lys1* and *aro8*).

The fact that several organisms lack complete sets of DAP genes raises the issue of how can they synthesize obtain the final product with an "incomplete" pathway. There are several possible answers to this question. First, it is possible that at the primary structure level such genes have evolved beyond recognition, but may be detectable with conserved secondary and tertiary structure features (Aurora and Rose, 1998). Second, it is also possible that a low specificity enzyme catalyzes similar reactions in different pathways. An excellent example of this phenomenon is provided by the aminotransferases, as shown by several reports of the low specificity these enzymes exhibit and their dual activity (Buchli et al, 1995; Jensen and Gu, 1996; Iraqui et al., 1998). This includes the *E. coli* acetylornithine aminotransferase (*argD*), an enzyme which has shown recently reported to participate both in arginine biosynthesis (Figure 3) and in the DAP transamination reaction (Ledwidge and Blanchard, 1999), which until recently was believed to be catalyzed by the diaminopimelate aminotransferase encoded by the recently detected *dapC* gene. Third, it is possible that some of these reactions for which no catalyst can be detected are spontaneous and require no enzyme.

In summary, the assembly of both the DAP and AAA routes can be explained as outcomes of a series of gene duplication events followed by their later specialization. This is not discordant with the fact that some Arg genes (Figure 4) are involved in the *Thermus* Lys biosynthesis via the AAA pathway. Since if the two routes for lysine biosynthesis were assembled via the patchwork mechanism (Jensen, 1976), the enzyme encoded by the *argD* gene, which participates in the Arg, the DAP and the *Thermus* AAA pathway (Ledwidge and Blanchat, 1999; Miyazaki *et al.*, 2001) provides an example of how enzymes could have been shared in the past among different routes. The multifunctional activity of this enzyme must be the result of the well-known low aminotransferase specificity (Jensen and Gu, 1996).

Finally, the close evolutionary relationship between the DAP and Arg pathways not only suggest that during early biological evolution Lys and Arg were used indiscriminately, but also supports previous contentions of the older antiquity of the DAP route over the AAA biosynthesis. It may also helps to understand at least in part the codon assignments for these two amino acids. Lys and Arg are sterically-related amino acids that share (a) many physico-chemical properties; (b) at least some similar codons (Table 3); and, as shown here, (c) have homologous biosynthetic enzymes that are the outcome of duplication events. It is thus possible that the concerted evolution of an entire set of genes took place, eventually leading to a less-ambiguous code. This interpretation is consistent with the coevolution theory of the genetic code suggested

by Wong (1975), who has argued that the current codon assignments reflect the sequence of appearance of new amino acids within the earliest biochemical systems (Table 4). If this the case, then one can especulate that the Lys codons (AAA and AAG) and two of the Arg codons (AGA and AGG) evolved from a primitive, less-accurate code, and that the additional Arg codons CGU, CGC, CGA, CGG were recruited during a later stage of evolution of the genetic code.

Acknowledgments

We thank Drs. Stanley L. Miller and Henry Vogel for several helpful discussions and comments. This work was supported by project PAPIIT IN213598 (UNAM, México) to A.L., and by the Dr. Robert Cedergren Fellowship (Canada) to A. M. V. This paper was completed during a sabbatical leave in the laboratory of Professor Ricardo Amils (Universidad Autónoma de Madrid, Spain), with support from DGAPA-UNAM (A. L.)

REFERENCES

1. Alm RA, Ling LSL., Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, deJonge BL, Carmel G, Tummino PJ, Caruso A, Uria-Nickelsen M, Mills DM, Ives C, Gibson R, Merberg D, Mills SD, Jiang Q, Taylor DE, Vovis GF, Trust TJ. (1999). Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397**: 176-180.
2. Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Pontén T, Alsmark UCM, Podowski RM, Nässtrand AK, Eriksson AS, Winkler HH, Kurland CG. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**: 133-140
3. Aurora R, Rose GD. (1998). Seeking an ancient enzyme in *Methanococcus janaschii* using ORF, a program based on secondary structure comparisons. *Proc Natl Acad Sci USA* **95**: 2818-2823
4. Bakhtiet N, Forney FW, Sthaly DP, Daniels L. (1984). Lysine biosynthesis in *Methanobacterium thermoautotrophicum* is by diaminopimelic acid pathway. *Curr. Microbiol.* **10**: 195-198.
5. Bhattacharjee JK. (1985). α -aminoadipate pathway for the biosynthesis of lysine in lower eukaryotes. *CRC Crit. Rev. Microbiol.* **12**: 131-151.
6. Bhattacharjee JK. (1992). Evolution of α -aminoadipate pathway for the synthesis of lysine in fungi. *In The Evolution of Metabolic Function*. Mortlock RP (ed), CRC Press, London, pp 47-80.
7. Blattner FR, Plunkett III G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* **277**: 1453-1474
8. Boyen A, Charlier D, Charlier J, Sakanyan V, Mett I, Glansdorff N. (1992). Acetylornithine deacetylase, succinyl-diaminopimelate desuccinylase, and carboxypeptidase G2 are evolutionary related. *Gene* **116**: 1-6.

-
9. Bryan JK. (1980). Synthesis of the aspartate family and branched-chain amino acids. In *The Biochemistry of Plants*. Vol. 5. Miñin BJ (ed), Academic Press, New York, pp 402-452.
 10. Buchli R, Alberati-Giani D, Malherbe P, Köhler C, Broger C, Cesura AM. (1995). Cloning and functional expression of a soluble form of kynurenine/ α -aminoacidipate aminotransferase from rat kidney. *J. Biol. Chem.* **270**: 29330-29335.
 11. Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, Fitzgerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NSM, Weidman JF, Fuhrmann JL, Nguyen D, Utterback TR, Kelley JM, Peterson JD, Sadow PW, Hanna MC, Cotton MD, Roberts KM, Hurst MA, Kaine BP, Borodovsky M, Klenk HP, Fraser CM, Smith HO, Woese CR, Venter JC. (1996). Complete genome sequence of the methanogenic Archaeon, *Methanococcus jannaschii*. *Science* **273**:1058-73
 12. Cirillo JD, Weisbord TR, Banejee A, Bloom BR, Jacobs WR Jr. (1994). Genetic determination of the meso-diaminopimelate biosynthetic pathway of *Mycobacteria*. *J. Bacteriol.* **176**: 4424-4429.
 13. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krog A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**: 537 - 544
 14. Deckert G, Warren PV, Gaasterland T, Young WG, Lenox AL, Graham DE, Overbeek R, Snead MA, Keller M, Aujay M, Huber R, Feldman RA, Short JM, Olson GJ, Swanson RV. (1998). The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392**: 353-358
 15. Fani R, Gallo R, Liò P. (2000). Molecular evolution of nitrogen fixation: the evolutionary history of the *nifD*, *nifK*, and *nifN* genes. *J. Mol. Evol.* **51**: 1-11
 16. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb J-F, Dougherty BA, Merrick JM, McKenney K, Sutton G, FitzHugh W, Fields C,

-
- Gocayne JD, Scott J, Shirley R, Liu L-I, Glodek A, Kelley JM, Weidman JF, Phillips CA, Spriggs T, Hedblom E, Cotton MD, Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Brandon RC, Fine LD, Fritchman JL, Fuhrmann JL, Geoghagen NSM, Gnehm CL, McDonald LA, Small KV, Fraser CM, Smith HO, Venter JC. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496-512.
17. Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton R, Lathigra R, White O, Ketchum KA, Dodson R, Hickey EK, Gwinn M, Dougherty B, Tomb JF, Fleischmann RD, Richardson D, Peterson J, Kerlavage AR, Quackenbush J, Salzberg S, Hanson M, van Vugt R, Palmer N, Adams MD, Gocayne J, Weidman J, Utterback T, Watthey L, McDonald L, Artiach P, Bowman C, Garland S, Fujii C, Cotton MD, Horst K, Roberts K, Hatch B, Smith HO, Venter JC. (1997). Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* **390**: 580-586
 18. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton GG, Kelley JM, Fritchman JL, Weidman JF, Small KV, Sandusky M, Fuhrmann JL, Nguyen D, Utterback TR, Saudek DM, Phillips CA, Merrick JM, Tomb JF, Dougherty BA, Bott KF, Hu PC, Lucier TS, Peterson SN, Smith HO, Hutchinson III CA and Venter JC. (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**: 397-403
 19. Fraser CM, Norris SJ, Weinstock GM, White O, Sutton GG, Dodson R, Gwinn M, Hickey EK, Clayton R, Ketchum KA, Sodergren E, Hardham JM, McLeod MP, Salzberg S, Peterson J, Khalak H, Richardson D, Howell JK, Chidambaram M, Utterback T, McDonald L, Artiach P, Bowman C, Cotton MD, Fujii C, Garland S, Hatch B, Horst K, Roberts K, Sandusky M, Weidman J, Smith HO, Venter JC. (1998). Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* **281**: 375-88
 20. Glass JI, Lefkowitz EJ, Glass JS, Heiner CR, Chen EY, Cassell GH. (2000). The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*. *Nature* **407**: 757-762.
 21. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, Gill SR, Nelson KE, Read TD, Tettelin H, Richardson D, Ermolaeva MD, Vamathevan J, Bass S, Qin H, Dragoi I, Sellers P, McDonald L, Utterback T, Fleischmann RD, Nierman WC, White O, Salzberg SL, Smith HO, Colwell RR, Mekalanos JJ, Venter JC, Fraser CM. (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholera*. *Nature* **406**: 477-483.

-
22. Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li BC, Herrmann R. (1996). Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* **24**: 4420-4449
 23. Iraqui I, Vissers S, Cartiax M, Urrestarazu A. (1998). Characterization of *Saccharomyces cerevisiae* *aro8* and *aro9* genes encoding aromatic aminotransferases I and II reveals a new aminotransferase subfamily. *Mol. Gen. Genet.* **257**: 238-248.
 24. Irvin SD, Bhattacharjee JK. (1998). A unique fungal lysine biosynthesis enzyme shares a common ancestor with tricarboxylic acid cycle and leucine biosynthetic enzymes found in diverse organisms. *J. Mol. Evol.* **46**: 401-408.
 25. Jensen RA (1976). Enzyme recruitment in the evolution of new function. *Annu. Rev. Microbiol.* **30**: 409-425.
 26. Jensen RA, Gu W. (1996). Evolutionary recruitment of biochemically specialized subdivisions of family I within the protein superfamily of aminotransferases. *J. Bacteriol.* **178**: 2161-2171.
 27. Kalman S, Mitchell W, Marathe R, Lammel C, Fan J, Hyman RW, Olinger L, Grimwood J, Davis RW & Richard Stephens R. (1999). Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. *Nature Genetics* **21**: 385 – 389
 28. Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, Tabata S. (1996). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**: 109-136
 29. Kawarabayasi Y, Hino Y, Horikawa H, Yamazaki S, Haikawa Y, Jin-no K, Takahashi M, Sekine M, Baba S, Ankai A, Kosugi H, Hosoyama A, Fukui S, Nagai Y, Nishijima K, Nakazawa H, Takamiya M, Masuda S, Funahashi T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Aoki K, Kubota K, Nakamura Y, Nomura N, Sako Y and Kikuchi H. (1999). Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aeropyrum pernix* K1. *DNA Res.* **6**: 83-101.

-
30. Kawarabayasi Y, Sawada M, Horikawa H, Haikawa Y, Hino Y, Yamamoto S, Sekine M, Baba S, Kosugi H, Hosoyama A, Nagai Y, Sakai M, Ogura K, Otuka R, Nakazawa H, Takamiya M, Ohfuku Y, Funahashi T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Aoki K, Nakamura Y, Robb TF, Horikoshi K, Masuchi Y, Shizuya H and Kikuchi H. (1998). Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3. *DNA Res.* 5: 55-76
31. Kindler SH, Gilvarg C. (1960). *N*-succinyl-L- α,ϵ -diapimelic acid deacylase. *J. Biol. Chem.* 235: 3532-3535.
32. Klenk HP, Clayton RA, Tomb J, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey EK, Peterson JD, Richardson DL, Kerlavage AR, Graham DE, Kyrpides NC, Fleischmann RD, Quackenbush J, Lee NH, Sutton GG, Gill S, Kirkness EF, Dougherty BA, McKenney K, Adams MD, Loftus B, Peterson S, Reich CI, McNeil LK, Badger JH, Glodek A, Zhou L, Overbeek R, Gocayne JD, Weidman JF, McDonald L, Utterback T, Cotton MD, Spriggs T, Artiach P, Kaine BP, Sykes SM, Sadow PW, D'Andrea KP, Bowman C, Fujii C, Garland SA, Mason TM, Olsen GJ, Fraser CM, Smith HO, Woese CR and Venter JC. (1997). The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* 390: 364-370.
33. Kobashi N, Nishiyama M, Tanokura M. (1999). Aspartate kinase-independent lysine synthesis in an extremely thermophilic bacterium, *Thermus thermophilus*: lysine is synthesized via α -amino adipic acid not via diaminopimelic acid. *J. Bacteriol.* 181: 1713-1718.
34. Kosuge T, Hoshino T. (1997). Molecular cloning and sequence analysis of the *lysR* gene from the extremely thermophilic eubacterium, *Thermus thermophilus* HB27. *FEMS* 157: 73-79.
35. Kosuge T, Hoshino T. (1998). Lysine is synthesized through the α -amino adipate pathway in *Thermus thermophilus*. *FEMS* 169: 361-367.
36. Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessieres P, Bolotin A, Borchert S, Borriss R, Boursier L, Brans A, Braun M, Brignell SC, Bron S, Brouillet S, Bruschi CV, Caldwell B, Capuano V, Carter NM, Choi SK, Codani JJ, Connerton IF, Cummings NJ, Daniel RA, Denizot F, Devine KM, Düsterhöft A, Ehrlich SD, Emmerson PT, Entian KD, Errington J, Fabret C, Ferrari E, Foulger D, Fritz C, Fujita M, Fujita Y, Fuma S, Galizzi A, Galleron N, Ghim S-Y, Glaser P, Goffeau A, Golightly EJ, Grandi G, Guiseppi

-
- G, Guy BJ, Haga K, Haiech J, Harwood CR, Hénaut A, Hilbert H, Holsappel S, Hosono S, Hullo M-F, Itaya M, Jones L, Joris B, Karamata D, Kasahara Y, Klaerr-Blanchard M, Kieln C, Kobayashi Y, Koetter P, Keningstein G, Krogh S, Kumano M, Kurita K, Lapidus A, Lardinois S, Lauber J, Lzarevic V, Lee S-M, Levine A, Liu H, Masuda S, Maüel C, Médigue C, Medina N, Mellado RP, Mizuno M, Moesti D, Nakai S, Noback M, Noone D, O'Reilly M, Ogawa K, Ogowara A, Oudega B, Park S-H, Parro V, Phol TM, Portetelle D, Porwollik S, Prescott AM, Presecan E, Pujic P, Purnelle B, Rapaport G, Rey M, Reynolds S, Rieger M, Rivolta C, Rocha E, Roche B, Rose M, Sadaie Y, Sato T, Scanlan E, Schleich S, Schroeter R, Scoffone F, Sekiguchi J, Sekowska A, Seror SJ, Serror P, Shini B-S, Soldo B, Sorokin A, Tacconi E, Takagi T, Takahashi H, Takemura K, Takeuchi M, Tamakoshi A, Tanaka T, Terpstra P, Tognoni A, Tosato V, Uchiyama S, Vandebol M, Vannier F, Vassarotti A, Viari A, Wambutt R, Wedler E, Wedler H, Weitzenegger T, Winters P, Wipat A, Yamamoto H, Yamane K, Yasumoto K, Yata K, Yoshida K, Yoshikawa H-F, Zumstein E, Yoshikawa H, Danchin A. (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**: 249-256
37. Ledwidge R, Blanchard JS. (1999). The dual biosynthetic capability of *N*-acetylornithine aminotransferase in arginine and lysine biosynthesis. *Biochemistry* **38**: 3019-
38. LéJohn HB. (1971). Enzyme regulation, lysine pathways and cell wall structures as indicators of major lines of evolution in fungi. *Nature* **231**:164-168.
39. LéJohn HB. (1974). Biochemical parameters of fungal phylogenetics. *Evol. Biol.* **7**: 79-125.
40. Liu L, Shaw PD. (1997). A possible role for acetylated intermediates in diaminopimelate and tabtoxinine-b-lactam biosynthesis in *Pseudomonas syringae* pv. *tabaci* BR2.024. *J. Bacteriol.* **179**: 5922-5927.
41. Martin C, Cami B, Yeh P, Stragier P, Parsot C, Patte JC. (1988). *Pseudomonas aeruginosa* diaminopimelate decarboxylase: evolutionary relationship with other amino acid decarboxylases. *Mol. Biol. Evol.* **5**: 549-559.
42. Matsuda M, Ogur M. (1969). Separation and specificity of the yeast glutamate- α -keto adipate transaminase. *J. Biol. Chem.* **244**: 3352-3358.
43. Misono H, Ogasawara M, Nagasaki S. (1986a). Purification and properties of meso-diaminopimelate dehydrogenase from *Brevibacterium* sp. *Agric. Biol. Chem.* **50**: 1329-1330

-
44. Misono H, Ogasawara M, Nagasaki S. (1986b). Characterization of meso-diaminopimelate dehydrogenase from *Corynebacterium glutamicum* and its distribution in bacteria. *Agric. Biol. Chem.* **50**: 2729-2734.
 45. Misono H, Togawa H, Yamamoto T, Soda K. (1979). Meso- α,ϵ -diaminopimelate D-dehydrogenase: distribution and the reaction product. *J. Bacteriol.* **137**: 22-27.
 46. Miyazaki J, Kobashi N, Nishiyama M, Yamane H. (2001). Functional and evolutionary relationship between arginine biosynthesis and prokaryotic lysine biosynthesis through α -aminoadipate. *J. Bacteriol.* **183**:5067-5073
 47. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, White O, Salzberg SL, Smith HO, Venter JC, Fraser CM. (1999). Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**: 323-329 .
 48. Nishida H. (2001). Distribution of genes for lysine biosynthesis through the aminoadipate pathway among prokaryotic genomes. *Bioinformatics* **17**: 189-191
 49. Nishida H, Nishiyama M. (2000). What is characteristic of fungal lysine synthesis through the α -aminoadipate pathway?. *J Mol Evol.* **51**: 299-302
 50. Nishida H, Nishiyama M, Kobashi N, Kosuge T, Hoshino T, Yamane H. (1999). A prokaryotic gene cluster in synthesis of lysine through the amino adipate pathway: a key to the evolution of amino acid biosynthesis. *Genome Res.* **9**: 1175-1185
 51. Parkhill J, Achtman M, James KD, Bentley SD, Churcher C, Klee SR, Morelli G, Basham D, Brown D, Chillingworth T, Davies RM, Davis P, Devlin K, Feltwell T, Hamlin N, Holroyd S, Jagels K, Leather S, Moule S, Mungall K, Quail MA, Rajandream MA, Rutherford KM, Simmonds M, Skelton J, Whitehead S, Spratt BG, Barrell BG. (2000a). Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* **404**: 502-506
 52. Parkhill J, Wren BW, Mungall K, Ketley JM, Churcher C, Basham D, Chillingworth T, Davies RM, Feltwell T, Holroyd S, Jagels K, Karlyshev AV, Moule S, Pallen MJ, Penn CW, Quail MA, Rajandream MA, Rutherford KM, van Vliet AHM, Whitehead S, Barrell BG. (2000b). The

genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* **403**: 665 - 668.

53. Patte JC. (1996). Biosynthesis of threonine and lysine. In *Escherichia coli and Salmonella*. Neidhardt FC, Curtis R III, Ingraham JL, Lin EC, Low KB, Magasanik, Magasanik B, Resnikoff WS, Riley M, Schaechter M and Umbarger HE. ASM Press, Washington DC., pp 528-541
54. Pavelka M Jr, Jacobs W Jr. (1996). Biosynthesis of diaminopimelate, the precursor of lysine and a component of peptidoglycan, is an essential function of *Mycobacterium smegmatis*. (1996). *J. Bacteriol.* **178**: 6496-6507.
55. Ragan MA. (1989). Biochemical pathways and the phylogeny of the eukaryotes. In *The Hierarchy of Life*. Fernholm B, Bremer K and Jornvall H. (eds), Elsevier, pp 145-160.
56. Ragan MA, Chapman DJ. (1978). *A Biochemical Phylogeny of the Protists*. Academic Press, New York. pp 105-111
57. Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, Hickey EK, Peterson J, Utterback T, Berry K, Bass S, Linher K, Weidman J, Khouri H, Craven B, Bowman C, Dodson R, Gwinn M, Nelson W, DeBoy R, Kolonay J, McClarty G, Salzberg SL, Eisen J and Fraser CM. (2000). Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res.* **28**: 1397-1406
58. Rowley B, Tucci AF. (1970). Homoisocitric dehydrogenase from yeast. *Arch. Biochem. Biophys.* **141**: 499-510.
59. Sandmeier E, Hale TI, Christen P. (1994). Multiple evolutionary origin of pyridoxal-5'-phosphate-dependent amino acid decarboxylase. *Eur. J. Biochem.* **221**: 997-1002.
60. Scapin G, Blanchard J. (1996). Enzymology of bacterial lysine biosynthesis. In *Escherichia coli and Salmonella*. Neidhardt FC, Curtis R III, Ingraham JL, Lin EC, Low KB, Magasanik, Magasanik B, Resnikoff WS, Riley M, Schaechter M and Umbarger HE. ASM Press, Washington DC. Vol II pp 279-324.

-
61. Scapin G, Cirilli M, Reddy SG, Gao Y, Vederas JC, Blanchard JS. (1998). Substrate and inhibitor binding sites in *Corynebacterium glutamicum* diaminopimelate dehydrogenase. *Biochemistry* **37**: 3278-3285.
62. Schäfer S, Paalme T, Vilu R, Fuchs G. (1989). ¹³C-NMR study of acetate assimilation in *Thermoproteus neutrophilus*. *Eur. J. Biochem.* **186**: 695-700.
63. Schrupf B, Schwarzer A, Kalinowski J, Pühler A, Eggeling L, Sahl H. (1991). A functionally split pathway synthesis in *Corynebacterium glutamicum*. *J. Bacteriol.* **173**: 4510-4516.
64. Shirai M, Hirakawa H, Kimoto M, Tabuchi M, Kishi F, Ouchi K, Shiba T, Ishii K, Hattori M, Kuhara S, Nakazawa T. (2000). Comparison of whole genome sequences of *Chlamydia pneumoniae* J138 from Japan and CWL029 from USA. *Nucleic Acids Res.* **28**: 2311-2314
65. Simpson AJG, Reinach FC, Arruda P, Abreu FA, Acencio M, Alvarenga R, Alves LMC, Araya JE, Baia GS, Baptista CS, Barros MH, Bonaccorsi ED, Bordin S, JM Bove, Briones MRS, Bueno MRP, Camargo AA, Camargo LEA, Carraro DM, Carrer H, Colauto NB, Colombo C, Costa FF, Costa MCR, Costa-Neto CM, Coutinho LL, Cristofani M, Dias-Neto E, Docena C, El-Dorry H, Facincani AP, Ferreira AJS, Ferreira VCA, Ferro JA, Fraga JS, França SC, Franco MC, Frohme M, Furlan LR, Garnier M, Goldman GH, Goldman MHS, Gomes SL, Gruber A, Ho PL, Hoheisel JD, Junqueira ML, Kemper EL, Kitajima JP, Krieger JE, Kuramae EE, Laigret F, Lambais MR, Leite LCC, Lemos EGM, Lemos MVF, Lopes SA, Lopes CR, Machado JA, Machado MA; Madeira AMBN, Madeira HMF, Marino CL, Marques MV, Martins EAL, Martins EMF, Matsukuma AY, Menck CFM, Miracca EC, Miyaki CY, Monteiro-Vitorello CB, Moon DH, Nagai MA, Nascimento ALTO, Netto LES, Nhani A Jr, Nobrega FG, Nunes LR, Oliveira MA, de Oliveira MC, de Oliveira RC, Palmieri DA, Paris A, Peixoto BR, Pereira GAG, Pereira HA Jr, Pesquero JB, Quaggio RB, Roberto PG, Rodrigues V, Rosa AJ de M, de Rosa VE Jr, de Sá RG, Santelli RV, Sawasaki HE, Silva ACR, da Silva AM, da Silva FR, Silva WA Jr, da Silveira JF, Silvestri MLZ, Siqueira Wj, de Souza AA, de Souza AP, Terenzi MF, Truffi D, Tsai SM, Tsuhako MH, Vallada H, Van Sluys MA, Verjovski-Almeida S, Vettore AL, Zago MA, Zatz M, Meidanis J, Setubal JC. (2000). The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* **406**: 151-157.
66. Smith DR, Doucette-Stamm LA, Deloughery C, Lee H-M, Dubois J, Aldredge T, Bashirzadeh R, Blakely D, Cook R, Gilbert K, Harrison D, Hoang L, Keagle P, Lumm W, Pothier B, Qiu D, Spadafora R, Vicare R, Wang Y, Wierzbowski J, Gibson R, Jiwani N, Caruso A, Bush D, Safer H, Patwell D, Prabhakar S, McDougall S, Shimer G, Goyal A, Pietrovski S, Church GM,

-
- Daniels CJ, Mao J-I, Rice P, Nolling J and Reeve JN. (1997). Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J. Bacteriol.* **179**: 7135-7155
67. Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, Mitchell W, Olinger L, Tatusov RL, Zhao Q, Koonin EV, Davis RW. (1998). Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* **282**: 754-759
68. Sundharadas G, Gilvarg C. (1967). Biosynthesis of 2,6-diaminopimelate acid in *Bacillus megaterium*. *J. Biol. Chem.* **242**: 3983-3988.
69. Tettelin H, Saundres NJ, Heidelberg J, Jeffries AC, Nelson KE, Eisen JA, Ketchum KA, Hood DW, Peden JF, Dodson RJ, Nelson WC, Gwinn ML, DeBoy R, Peterson JD, Hickey EK, Haft DH, Salzberg SL, White O, Fleischmann RD, Dougherty BA, Mason T, Ciecko A, Parksey DS, Blair E, Cittone H, Clark EB, Cotton MD, Utterback TR, Khouri H, Qin H, Vamathevan J, Gill J, Scarlato V, Masignani V, Pizza M, Grandi G, Sun L, Smith HO, Fraser CM, Moxon ER, Rappuoli R, Venter JC. (2000). Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* **287**: 1809-1815.
70. Thompson JD, Higgins DG, Gibson TJ. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673-4680
71. Tomb J-F, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA, Nelson K, Quackenbush J, Zhou L, Kirkness EF, Peterson S, Loftus B, Richardson D, Dodson R, Khalak HG, Glodek A, Ckenney K, Fitzgerald LM, Lee N, Adams MD, Hickey EK, Berg DE, Gocayne JD, Utterback TR, Peterson JD, Kelley JM, Cotton MD, Weidman JM, Fujii C, Bowman C, Watthey L, Wallin E, Hayes W, Borodovsky M, Karp PD, Smith HO, Fraser CM, Venter JC. (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**: 539 - 547.
72. Umbarger HE. (1978). Amino acid biosynthesis and its regulation. *Annu. Rev. Biochem.* **47**: 533-606
73. Vogel HJ. (1960). Two modes of lysine biosynthesis among the lower fungi: evolutionary significance. *Biochem. Biophys. Acta* **41**: 172-173.

-
74. Vogel HJ. (1965). Lysine biosynthesis and evolution. In *Evolving Genes and Proteins*. V. Bryson and HJ Vogel (eds), Academic Press, New York, pp 25-40.
75. Waley SG. (1969). Some aspects of the evolution of metabolic pathways. *Comp. Biochem. Physiol.* **30**: 409-425.
76. Wehrmann A, Phillipp B, Sahm H, Eggeling L. (1998). Different modes of diaminopimelate synthesis and their role in cell wall integrity: a study with *Corynebacterium glutamicum*. *J. Bacteriol.* **180**: 3159-3165.
77. Weidner G, Steffan B, Brakhage AA. (1997). The *Aspergillus nidulans lysF* gene encodes homoaconitase, an enzyme involved in the fungus-specific lysine biosynthesis pathway. *Mol. Gen. Genet.* **255**: 237-247.
78. Weinberger S, Gilvarg C. (1970). Bacterial distribution of the use of succinyl and acetyl blocking groups in diaminopimelate acid biosynthesis. *J. Bacteriol.* **101**: 323-324.
79. Wenko LK, Treick RW, Wilson KG. (1985). Isolation and characterization of a gene encoding meso-diaminopimelate dehydrogenase from *Glycine max*. *Plant Mol. Biol.* **4**: 197-204.
80. White O, Eisen JA, Heidelberg JF, Hickey EK, Peterson JD, Dodson RJ, Haft DH, Gwinn ML, Nelson WC, Richardson DL, Moffat KS, Qin H, Jiang L, Pamphile W, Crosby M, Shen M, Vamathevan JJ, Lam P, McDonald L, Utterback T, Zalewski C, Makarova KS, Aravind L, Daly MJ, Minton KW, Fleischmann RD, Ketchum KA, Nelson KE, Salzberg S, Smith HO, Venter JC, Fraser CM. (1999). Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science* **286**:1571-1577
81. White PJ. (1983). The essential role of diaminopimelate dehydrogenase in the biosynthesis of lysine by *Bacillus sphaericus*. *J. Gen. Microbiol.* **129**: 739-749.
82. Wong JTF. (1975). A co-evolution theory of the genetic code. *Proc. Natl. Acad. Sci. USA* **72**:1909-1912.
83. Ycas M. (1974). On the earlier states of biochemical systems. *J. Theor. Biol.* **44**: 145-160

FIGURE CAPTIONS

Figure 1.

Schematic representation of the two lysine biosynthetic pathways. a) Three DAP variants for the synthesis of meso-diaminopimelate and lysine, and their relationship with other members of the aspartic acid family biosynthetic routes for threonine, methionine and isoleucine. *thrA* and *metL* are aspartokinases involved in the biosynthetic pathways of threonine and methionine respectively. b) The eight enzymatic steps of the AAA route.

Figure 2.

Schematic representation of secondary structure consensus in alignments between several sequences of a) aspartokinases, *lysC*; b) *N*-acetylglutamate kinase, *argB* and c) alignment between *lysC* and *argB*. Ovals symbolize α -helix structure, and squares β -strands.

Figure 3.

Alignment and secondary structure of *asd* and *argC* genes. Red color indicate α -helix structure and blue β -strands.

Figure 4.

Lysine biosynthetic genes and their evolutionary relationship to other sequences. Genes in boxes located at the same level of the Leu, AAA variants, citric acid cycle, DAP and Arg metabolism are homologs. *argD* and *dapC* are in dotted boxes since the *argD* product can carry out the function of the *dapC* product but no similarity can be detected between their primary structures.

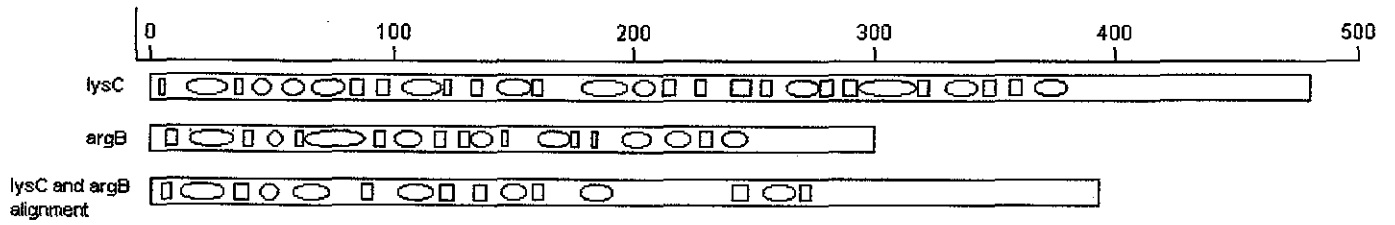


Figure 2

TESIS CON
FALLA DE ORIGEN

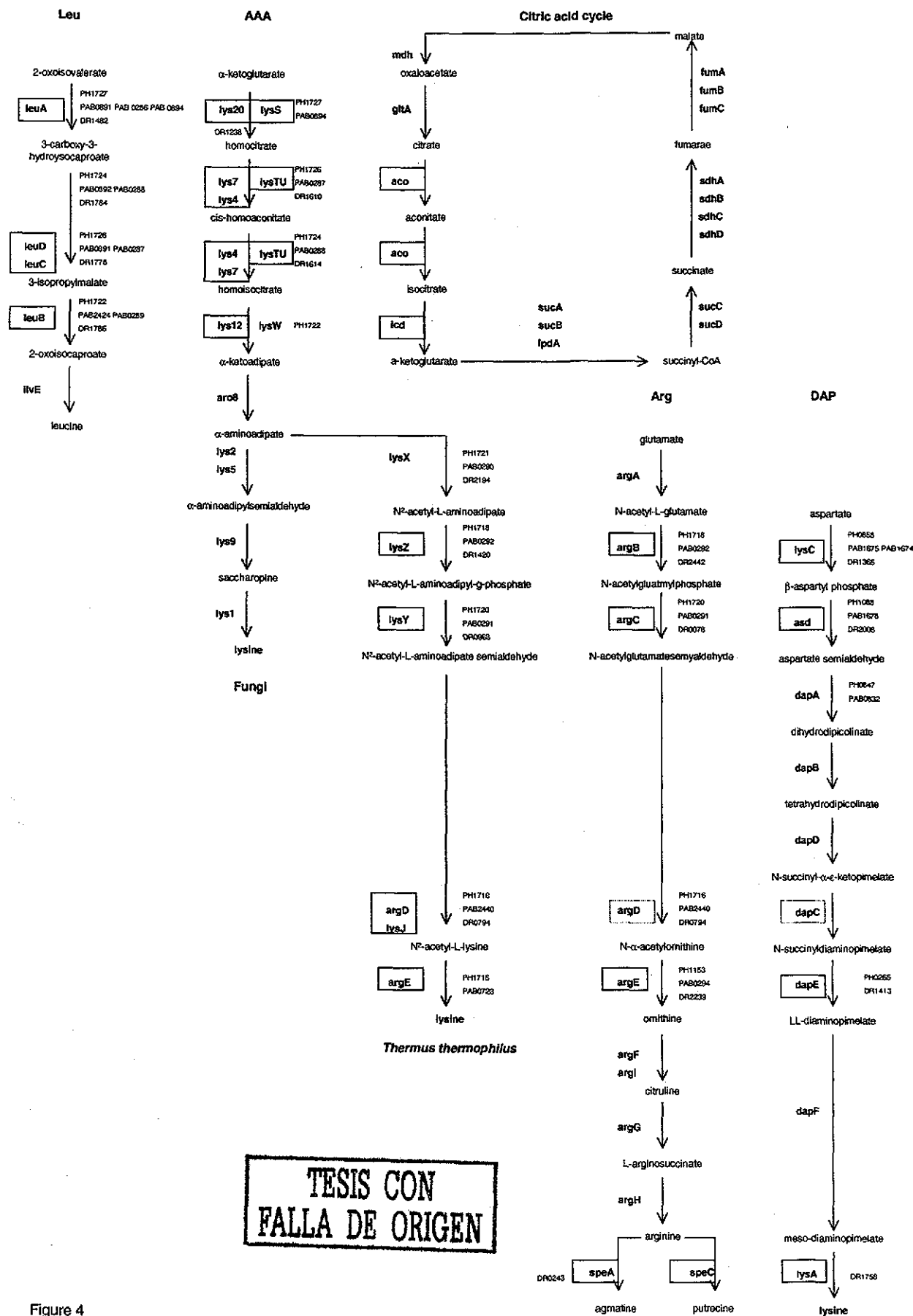


Figure 4

Table2.-Lysine biosynthetic enzymes and related sequences

Enzyme	EC	Gene	ID	Enzyme	EC	Gene	ID
aspartokinase	2.7.2.4	<i>lysC, ask</i>	aq1152, AF0700, APE1140, <i>lysC</i> -Bs, <i>dapG</i> , Cj0582, TC0641, CP0803, <i>lysC</i> -Cpj, DR1365, HP1229, <i>jhp1150</i> , MJ0571, MTH802, Rv3709c, NMA1701, NMB1498, PAB1675, PH0858, RP753, YER052C, sir0657, TM1518, TM0547, VC0547, VC0391, XF2225, XF1116	acetylglutamate kinase	2.7.2.8	<i>argB</i>	aq2068, AF1280, APE1461, <i>argB</i> , b3959, MJ0069, MTH183, Rv1654, PAB0292, sir1898, TM1784
aspartate-semialdehyde dehydrogenase	1.2.1.11	<i>asd</i>	aq1866 AF1506 APE1143 <i>asd</i> Cj1023c CP0804 <i>cpj:asd</i> CPn1048 CT363 DR2008 HP1189 <i>jhp1114</i> MJ0205 MTH799 Rv3708c NMA0351 NMB2079 PAB1678 PH1088 RP316 YER069W sir0549 TM1523 VC2107 VC2036 XF1371	N-acetyl-gamma-glutamyl-phosphate reductase	1.2.1.38	<i>argC</i>	aq_1879 AF2071 <i>argC</i> -Ec MJ1096 MTH846 PH1720 Rv1652 YER069W
succinyl-diaminopimelate desuccinylase	3.5.1.18	<i>dapE</i>	aq_547, AF0051, AF0904, APE0632, Cj1048c, DR1413, <i>dapE</i> -Ec, HP0212, <i>jhp0198</i> , MJ0457, Rv1202, Rv2141c, NMA1730, NMB1530, PH0265, RP874, TM1666, VC2152, XF0116	acetylornithine deacetylase	3.5.1.16	<i>argE</i>	<i>argE</i> -Bs, b3957, PH1153
diaminopimelate decarboxylase	4.1.1.20	<i>lysA</i>	aq_1208 AF0800 <i>lysA</i> Cj0314 DR1758 HP0290 <i>jhp0275</i> MJ1097 MTH1335 Rv1293 NMA0468 NMB1976 sir0504 TM1517 VC0125 XF1116	arginine decarboxylase	4.1.1.19	<i>speA</i>	sir0662 sir1312 b2938 HP0422
homocitrate synthase	4.1.3.21	LYS20, LYS21	YDL182W YDL131W DR1238	ornithine decarboxylase	4.1.1.17	<i>speC</i> , <i>SPE1</i>	YKL187W
cis-homoaconitate	4.2.1.36	LYS4	YDR234W	2-isopropylmalate synthase	4.1.3.12	<i>leuA</i>	aq_356 aq_2090 AF0957 AF0219 <i>leuA</i> -Bs Cj1719c b0074 HI0986 MJ1195 MJ0503 MJ1392 MTH1630 MTH1481 MTH0723 Rv3710 NMA1270 NMB1070 PAB0890 PAB0286 PAB0894 PH1727 sir0186 sir1564 TM0552 TM0553 XF1818
				3-isopropylmalate dehydratase large subunit	4.2.1.33	<i>leuC</i>	aq_940 AF2199 <i>leuC</i> Cj1717c DR1610 DR1778 HI0988 MJ1003 MJ0499 MTH1631 MTH1386 Rv2988c NMA1450 NMB1036 PAB0891 PAB0287 PH1726 YGL009C sir1470 TM0554 TM0281 XF2375

3-isopropylmalate dehydratase small subunit	4.2.1.33	leuD	aq_1398 AF0629 AF1761 leuD Cj1716c DR1784 DR1614 MJ1277 MJ1271 MTH1387 MTH829 Fv1475c NMA1452 NMB1034 PAB0892 PAB0288 PH1724 sll1444 TM0555 TM0292 XF2374
aconitate hydratase	4.2.1.3	aco	aq_1784 AF1963 APE1618 citB Cj0835c DR1720 Fv2987c NMA2052 NMA1761 NMB0433 NMB1572 Fv799 YLFR304C XF0290 XF0292
3-isopropylmalate dehydrogenase	1.1.1.85	leuB	aq_244 AF0628 leuB Cj1718c DR1785 boo73 HI0987 MJ0720 MTH1388 Fv2995c NMA1456 NMB1031 PAB0289 PAB2424 PH1722 YOR136W sir1517 TM0556 XF2372
isocitrate dehydrogenase	1.1.1.42	icd	aq_1512 AF0647 APE0689 citC DR1674 DR1540 b1136 HP0027 ihp0023 MJ1596 MTH184 RP265 YNL037C YCL018W sir1289 TM1148 XF2596

homoisocitric dehydrogenase 1.1.1.155 LYS12 YIL094C

Table 4. Lys and Arg codons

Lys codons	Arg codons
<u>AAA</u>	<u>AGA</u>
<u>AAG</u>	<u>AGG</u>
	CGU
	CGC
	CGA
	CGG

DISCUSIÓN Y CONCLUSIONES

La comparación entre enzimas de las rutas DAP y AAA demuestra que no existe ninguna relación entre ellas, es decir, que surgieron de manera independiente. La homología entre diferentes enzimas participantes en la ruta DAP y el metabolismo de arginina ya habían sido reportadas (Martin *et al.*, 1988; Boyen *et al.*, 1992; Sandmeier *et al.*, 1994), así como entre las ruta AAA y de biosíntesis de Leu y el ciclo de ácido cítrico (Irvin y Bhattacharjee, 1998; Miyazaki *et al.*, 2001). Nosotros encontramos, además de los datos ya reportados, que los genes *argB* y *lysC*, aunque no presentan similitud a nivel de estructura primaria, si lo presentan a nivel de estructura secundaria. Se detectó también la homología entre *argC* y *asd*, y entre *lys12* y *argB*. Todo ello indica fuertemente que tanto la ruta DAP, como AAA se ensamblaron por medio de un mecanismo de "patchwork". Sin embargo es interesante hacer notar que en la ruta catabólica de lisina vía ácido pipercolico la reacción que va de piperidin-6-carboxilato a amino adipato semialdehído parece ser una reacción espontánea (http://www.expasy.ch/cgi-bin/show_image?J3) , lo cual indicaría que ésta ruta catabólica posiblemente se ensambló mediante el mecanismo propuesto por la teoría semi-enzimática.

Recientemente se ha sugerido que la ruta biosintética de lisina más antigua debe ser la nueva variante de la AAA descubierta en la bacteria *T. thermophilus* (Nishida *et al.*, 1999). La mayoría de los genes de esta variante están organizados en un cúmulo entre los que se encuentran las enzimas homocitrato sintasa (*lys20*) y la homoaconitato hidratasa (*lys4*) de la ruta del AAA característica de hongos, así como el gen *argB* y un homólogos de *argC*. También interviene en la síntesis los productos de los genes *argD* y *argE*, los cuales no se localizan en el cluster. La participación de estos genes en la variante de AAA de *T. thermophilus* ha sido demostrada experimentalmente (Nishida *et al.*, 1999). Además de *Thermus* éste cúmulo se ha localizado mediante el análisis de genomas en *P. horikoshii* (Nishida *et al.*, 1999), *P. abyssi* (Nishida y Nishiyama, 2000) y en *D. radiodurans* (Nishida , 2001). En el cluster de *P. horikoshii* no están presentes la homocitrato sintasa ni la homoaconitato hidratasa; sin embargo, se localizan sus homólogos *leuA*, *leuC*, *leuD* y además *leuB*. Como solo hay una copia de estos genes en el genoma de *P. horikoshii* se piensa que podrían intervenir tanto en la síntesis de leucina como en la de la variante de AAA (Nishida *et al.*, 1999). Si esto se comprobara experimentalmente se estaría documentando un ejemplo claro de como evolucionan las rutas por medio del mecanismo de patchwork.

Los organismos en los que se ha localizado el cluster de *T. thermophilus* se encuentran en los árboles filogenéticos ubicado en ramas muy antiguas, cercanas a la divergencia entre los tres grandes dominios Arquea, Bacteria y Eucaria, por lo que se plantea que debe ser la primera

ruta que se ensambló para la biosíntesis de lisina (Nishida *et al.*, 1999). Sin embargo, se debe mencionar que en *T. thermophilus* también se ha encontrado el gen *lysR*, el cual está relacionado con la síntesis de lisina vía DAP ya que es el gen regulador de *lysA*. También es interesante el hecho de que *P. horikoshii*, *P. abyssi* y *D. radiodurans* presenten algunos genes de la ruta del DAP como *lysC*, *asd*, *dapA*, *dapC*, *dapE* y *lysA* como se muestra en la Tabla 3 del artículo, lo cual sugeriría que en estos organismos se contó en algún momento de su historia evolutiva con la presencia de dos rutas sintéticas de la lisina, la ruta de *Thermus* y la del DAP. Sin embargo, en los dos géneros de *Pyrococcus* y en *D. radiodurans* aún falta demostrar experimentalmente por medio de cual ruta es sintetizada la lisina. La presencia de genes de la ruta DAP en estos organismos nos lleva a preguntarnos si es realmente la variante de *Thermus* la más antigua o si es la más reciente.

Los resultados presentados en este trabajo contrastan con los obtenidos por Nishida *et al.* (1999), ya que como resultado de la comparación de las secuencias de aminoácidos de las enzimas que interviene en la síntesis de lisina vía DAP y AAA encontramos una relación evidente entre varios genes de la ruta DAP y Arg y entre los de AAA y Leu, pero no encontramos relación entre las rutas de DAP y Leu ni entre AAA y Arg. Sin embargo a nivel funcional en el trabajo de Nishida *et al.* (1999) se reporta que en la nueva variante de AAA presente en *T. thermophilus* se ven involucrados genes tanto de la ruta de AAA como de Arg y de Leu, sin que por el momento se haya detectado actividad alguna de los genes relacionados con la ruta DAP.

Por otro lado, la ruta del DAP y el metabolismo de la Arg presentan enzimas homólogas producto de eventos de duplicación genética, las cuales en un principio probablemente eran poco específicas, pero después de un tiempo presentaron una diferenciación en su especificidad, dando origen así a las enzimas involucradas en ambas rutas. Debido que la Lys y la Arg presentan propiedades fisicoquímicas muy parecidas, es posible que en un principio estos dos aminoácidos se hayan utilizado de manera indistinta en organismos ancestrales. La Lys y la Arg presentan un conjunto de codones similares (AAA, AAG para lisina y AGA, AGG para arginina), lo cual es consistente con la teoría co-evolutiva del código genético (Wong, 1975), la que propone que la estructura de los codones refleja las relaciones entre aminoácidos, de tal forma que cuando un aminoácido es derivado de otro, la composición de sus codones es muy similar, variando únicamente en una sola base. Si esto es cierto, entonces, los otros cuatro codones que codifican para Arg (CGU, CGC, CGA, CGG), son el producto de un reclutamiento en una etapa posterior en la evolución del código genético. Por lo que se podría suponer en base a estos argumentos que la ruta del DAP es la más antigua.

Referencias

1. Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Pontén T, Alsmark UCM, Podowski RM, Näslund AK, Eriksson AS, Winkler HH, Kurland CG. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**: 133-140
2. Bakhiet N, Forney FW, Sthaly DP, Daniels L. (1984). Lysine biosynthesis in *Methanobacterium thermoautotrophicum* is by diaminopimelic acid pathway. *Curr. Microbiol.* **10**: 195-198.
3. Bhattacharjee JK. (1992). Evolution of α -aminoadipate pathway for the synthesis of lysine in fungi. In *The Evolution of Metabolic Function*. Mortlock RP (ed), CRC Press, London, pp 47-80.
4. Boyen A, Charlier D, Charlier J, Sakanyan V, Mett I, Glansdorff N. (1992). Acetylmithine deacetylase, succinyldiaminopimelate desuccinylase and carboxypeptidase G2 are evolutionary related. *Gene* **116**: 1-6
5. Deckert G, Warren PV, Gaasterland T, Young WG, Lenox AL, Graham DE, Overbeek R, Snead MA, Keller M, Aujay M, Huber R, Feldman RA, Short JM, Olson GJ, Swanson RV. (1998). The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392**: 353-358
6. Fani R, Lió P, Lazcano A. (1995). Molecular evolution of the histidine biosynthetic pathway. *J.Mol. Evol.* **41**: 760-774
7. Granick S. (1957). Speculations on the origin and evolution of photosynthesis. *Ann. N. Y. Acad. Sci.* **69**: 292-308.
8. Granick S. (1965). The evolution of heme and chlorophyll. In *Evolving genes and proteins*. Bryson V and Vogel HJ (eds). Academic Press New York, pp 67-88
9. Horowitz NH. (1945). On the evolution of biochemical synthesis. *Proc. Natl. Acad. Sci. USA.* **31**: 153-157

-
10. Horowitz NH. (1965). The evolution of biochemical synthesis retrospect and prospect. In *Evolving genes and proteins*, Bryston V and Vogel HJ (eds), Academic Press, New York. pp 15-23.
 11. Irvin SD, Bhattacharjee JK. (1998). A unique fungal lysine biosynthesis enzyme shares a common ancestor with tricarboxylic acid cycle and leucine biosynthetic enzymes found in diverse organisms. *J. Mol. Evol.* **46**: 401-408
 12. Jensen RA (1976). Enzyme recruitment in the evolution of new function. *Annu. Rev. Microbiol.* **30**: 409-425
 13. Kawarabayasi Y, Hino Y, Horikawa H , Yamazaki S, Haikawa Y, Jin-no K, Takahashi M, Sekine M, Baba S, Ankai A, Kosugi H, Hosoyama A, Fukui S, Nagai Y, Nishijima K, Nakazawa H, Takamiya M, Masuda S, Funahashi T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Aoki K, Kubota K, Nakamura Y, Nomura N, Sako Y and Kikuchi H. (1999). Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aeropyrum permix* K1. *DNA Res.* **6**: 83-101
 14. Kindler SH, Gilvarg C. (1960). *N*-succinyl-L- α,ϵ -diaminopimelic acid deacylase. *J. Biol. Chem.* **235**: 3532-3535
 15. Kosuge T, Hoshino T. (1998). Lysine is synthesized through the α -amino adipate pathway in *Thermus thermophilus*. *FEMS* **169**: 361-367
 16. Lazcano A, Díaz-Villagómez E, Mills T, Oro J. (1995). On the levels of enzymatic substrate specificity: implications for the early evolution of metabolic pathways. *Adv. Space Res.* **15**: 345-356
 17. Lazcano A, Miller SL. (1999). On the origin of metabolic pathways. *J. Mol. Evol.* **49**: 424-431
 18. Martin C, Cami B, Yeh P, Stragier P, Parsot C, Patte JC. (1988). *Pseudomonas aeruginosa* diaminopimelate decarboxylase: evolutionary relationship with other amino acid decarboxylases. *Mol. Biol. Evol.* **5**: 549-559
 19. Misono H, Togawa H, Yamamoto T, Soda K. (1979). Meso- α,ϵ -diaminopimelate D-dehydrogenase: distribution and the reaction product. *J. Bacteriol.* **137**: 22-27

-
20. Miyazaki J, Kobashi N, Nishiyama M, Yamane H. (2001). Functional and evolutionary relationship between arginine biosynthesis and prokaryotic lysine biosynthesis through α -amino adipate. *J. Bacteriol.* **183**:5067-5073
 21. Nishida H. (2001). Distribution of genes for lysine biosynthesis through the amino adipate pathway among prokaryotic genomes. *Bioinformatics* **17**: 189-191
 22. Nishida H, Nishiyama M. (2000). What is characteristic of fungal lysine synthesis through the α -amino adipate pathway?. *J Mol Evol.* **51**: 299-302
 23. Nishida H, Nishiyama M, Kobashi N, Kosuge T, Hoshino T, Yamane H. (1999). A prokaryotic gene cluster in synthesis of lysine through the amino adipate pathway: a key to the evolution of amino acid biosynthesis. *Genome Res.* **9**: 1175-1185
 24. Parkhill J, Wren BW, Mungall K, Ketley JM, Churcher C, Basham D, Chillingworth T, Davies RM, Feltwell T, Holroyd S, Jagels K, Karlyshev AV, Moule S, Pallen MJ, Penn CW, Quail MA, Rajandream MA, Rutherford KM, van Vliet AHM, Whitehead S, Barrell BG. (2000). The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* **403**: 665 – 668
 25. Patte JC. (1996). Biosynthesis of threonine and lysine. In *Escherichia coli and Salmonella*. Neidhardt FC, Curtis R III, Ingraham JL, Lin EC, Low KB, Magasanik, Magasanik B, Resnikoff WS, Riley M, Schaechter M and Umberger HE (eds). ASM Press, Washington DC., pp 528-541
 26. Peretó J, Fani R, Leguina JI, Lazcano A. (1997). Enzyme evolution and the development of metabolic pathways. In *New beer in an old bottle: Eduard Buchner and the growth of biochemical knowledge*, Cornish-Bowden A (ed). Universitat de valencia, Valencia , Spain. pp173-198
 27. Ragan MA. (1989). Biochemical pathways and the phylogeny of the eukaryotes. In *The Hierarchy of Life*. Fernholm B, Bremer K and Jornvall H. (eds), Elsevier, pp 145-160
 28. Sandmeier E, Hale TI, Christen P. (1994). Multiple evolutionary origin of pyridoxal-5'-phosphate-dependent amino acid decarboxylase. *Eur. J. Biochem.* **221**: 997-1002.

-
29. Scapin G, Blanchard J. (1996). Enzymology of bacterial lysine biosynthesis. In *Escherichia coli and Salmonella*. Neidhardt FC, Curtis R III, Ingraham JL, Lin EC, Low KB, Magasanik, Magasanik B, Resnikoff WS, Riley M, Schaechter M and Umbarger HE (eds). ASM Press, Washington DC. Vol II, pp 279-324
 30. Schrupf B, Schwarzer A, Kalinowski J, Pühler A, Eggeling L, Sahm H. (1991). A functionally split pathway synthesis in *Corynebacterium glutamicum*. *J. Bacteriol.* **173**: 4510-4516
 31. Smith DR, Doucette-Stamm LA, Deloughery C, Lee H-M, Dubois J, Aldredge T, Bashirzadeh R, Blakely D, Cook R, Gilbert K, Harrison D, Hoang L, Keagle P, Lumm W, Pothier B, Qiu D, Spadafora R, Vicare R, Wang Y, Wierzbowski J, Gibson R, Jiwani N, Caruso A, Bush D, Safer H, Patwell D, Prabhakar S, McDougall S, Shimer G, Goyal A, Pietrovski S, Church GM, Daniels CJ, Mao J-I, Rice P, Nolling J and Reeve JN. (1997). Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J. Bacteriol.* **179**: 7135-7155
 32. Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, Mitchell W, Olinger L, Tatusov RL, Zhao Q, Koonin EV, Davis RW. (1998). Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* **282**: 754-759
 33. Sogin ML. (1994). The origin of eukaryotes and evolution into major kingdoms. In *Early life on earth*. S. Bengtson (ed). Nobel symposium No. 84 . Columbia University Press, New York
 34. Sundharadas G, Gilvarg C. (1967). Biosynthesis of 2,6-diaminopimelate acid in *Bacillus megaterium*. *J. Biol. Chem.* **242**: 3983-3988
 35. Umbarger HE. (1978). Amino acid biosynthesis and its regulation. *Annu. Rev. Biochem.* **47**: 533-606
 36. Vogel HJ. (1965). Lysine biosynthesis and evolution. In *Evolving Genes and Proteins*. V. Bryson and HJ Vogel (eds), Academic Press, New York, pp 25-40
 37. Waley SG. (1969). Some aspects of the evolution of metabolic pathways. *Comp. Biochem. Physiol.* **30**: 1-11.

-
38. Wehrmann A, Phillipp B, Sahm H, Eggeling L. (1998). Different modes of diaminopimelate synthesis and their role in cell wall integrity: a study with *Corynebacterium glutamicum*. *J. Bacteriol.* **180**: 3159-3165
 39. Wong JTF. (1975). A co-evolution theory of the genetic code. *Proc. Natl. Acad. Sci. USA* **72**:1909-1912
 40. Ycas M. (1974). On the earlier states of biochemical systems. *J. Theor. Biol.* **44**: 145-160